Wrist-based emotion recognition for human-computer interaction

Niklas Strengell

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 22.11.2019

Thesis supervisor:

Prof. Stephan Sigg

Thesis advisor:

M.Sc. Timo Idänheimo

Aalto University
School of Electrical Engineering

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF THE
MASTER'S THESIS

Author: Niklas Strengell

Title: Wrist-based emotion recognition for human-computer interaction

Date: 22.11.2019 Language: English Number of pages: 7+110

Department of Communications and Networking (Comnet)

Professorship: Ambient Intelligence

Supervisor: Prof. Stephan Sigg

Advisor: M.Sc. Timo Idänheimo

The role of emotion in human-computer interaction (HCI) has seen an increase in interest during the last decades. Technological advancements have made studying them much more viable for example because of the availability of affordable and accurate wrist-based sensors. However, this subfield of HCI still lacks theory and it has many unsolved engineering problems, especially considering naturalistic and automated emotion recognition. This thesis provides an overview of wrist-based emotion recognition in human-computer interaction by tying in the views and theoretical background of emotion from philosophy, psychology, neuroscience and economics. The thesis also includes an experimental set-up in naturalistic settings. The experiment uses an Empatica E4 device that can be worn on the wrist and which can be used to measure electrodermal activity (EDA) and heartrate variability (HRV). Both EDA and HRV are known biomarkers for various emotional reactions, such as emotional arousal or mental stress. The study explores the possibilities of EDA and HRV to measure emotional arousal and valence. Furthermore, the correlations between psychological surveys and emotional biosignal markers are explored. We used the Affect Intensity Measure (AIM) -survey, which measures the intensity of experienced and shown emotion, and Rational-Experiential Inventory (REI) -survey, which measures an individual preferred style of information processing. Five custom experiments and a data analysis method with custom analyzer code were designed for this thesis. Our findings suggest that EDA is a good marker for arousal, but that HRV is a problematic measure. Furthermore, we found evidence that there would be correlations between psychological traits and biosignals. However, there were limitations within our experiments. In conclusions, we provide suggestions for futher research and a new theoretical framework that could be used to understand emotions better in HCI.

Keywords: emotions, human-computer interaction, affective computing, AIM, REI-40, skin conductance, heart rate variability, Empatica E4

AALTO-YLIOPISTO
SÄHKÖTEKNIIKAN KORKEAKOULU

DIPLOMITYÖN TIIVISTELMÄ

Tekijä: Niklas Strengell

Työn nimi: Ranteesta mitattavat tunteet tietokoneen ja ihmisen vuorovaikutuksessa

Päivämäärä: 22.11.2019 Kieli: Englanti Sivumäärä: 7+110

Tietoliikenne- ja tietoverkkotekniikan laitos (Comnet)

Professuuri: Ambient Intelligence

Työn valvoja: Prof. Stephan Sigg

Työn ohjaaja: DI Timo Idänheimo

Kiinnostus tunteiden merkityksestä ihmisen ja tietokoneen vuorovaikutuksessa on kasvanut. Teknologian kehityksen myötä tunteisiin liittyviä biosignaaleja voidaan mitata hyvinkin huomaamattomasti esimerkiksi rannetietokoneilla. Alan teoria on kuitenkin vähäistä ja erityisesti naturalistiseen ja automatisoituun tunteiden tunnistamiseen liittyy monia ratkaisemattomia teknologisia ongelmia. Tämän diplomityön tarkoituksena on tarjota lukijalleen kattava teoreettinen näkemys monilta tieteen aloilta, jotka tutkivat tunteita. Työ yhdistää tunteisiin liittyvää teoriaa filosofiasta, psykologiasta, neurotieteestä sekä ihmis-tietokone-vuorovaikutkusen tutkimuksesta rakentaakseen yhtenäisen teoreettisen viitekehyksen ongelman ymmärtämiseksi. Työhön kuuluu myös kokeellinen osuus, jossa mitataan tunteita oikeassa ympäristössä. Kokeessa käytetään Empatica E4-rannetietokonetta, jolla voidaan mitata ihon sähkönjohtavuutta (EDA) ja sydämen sykevälivaihtelua (HRV). Sekä EDA että HRV ovat molemmat tunnettuja biosignaaleja erilaisissa tunnetiloissa. Kokeen tarkoitus on tutkia EDA:n ja HRV:n kykyä mitata tunteellista virittäytyneisyyttä ja tunnearvoa. Tämän lisäksi koe tutkii erilaisten psykologisten kyselylomakkeiden korrelaatioita mitattujen biosignaalejen välillä. Kokeessa käytetään Affect Intensity Measure (AIM) -kyselykaavaketta, joka mittaa koettujen ja näytettyjen tunteiden vahvuutta, sekä Rational Experiential Inventory (REI) -kyselykaavaketta, joka mittaa yksilön suosimaa sisäisen tiedonkäsittelyn menetelmää. Koetta varten kehitettiin viisi koeasetelmaa ja metodi, jolla voitiin analysoida mitattua dataa. Tulokset vahvistavat käsityksen, että EDA on hyvä virittäytyneisyyden mittari, mutta HRV:n käytössä löydettiin vain ongelmia. Tuloksissa on myös todisteita psykologisten luonteenpiirteiden ja biosignaalien korrelaatiolle. Lopussa annamme suosituksia seuraaville tutkimuksille ja esittelemme kehittämämme uuden teoreettisen viitekehyksen, jolla tunteita voisi ymmärtää paremmin ihmisen ja tietokoneen vuorovaikutuksessa.

Avainsanat: tunteet, tunneälykäs teknologia, ihmisen ja tietokoneen vuorovaikutus, AIM, REI-40, ihon sähkönjohtavuus, sydämen sykkeen välinvaihtelu, Empatica E4

Wonder is the first of all passions.

Rene Descartes

This work was the most challenging and the most excruciating work I ever did during my studies at Aalto University. Yet, it was the most fun I ever had during my studies. For over half a year, I was allowed to study a topic I was interested in: the human experience and specifically, its emotions. In addition, I could combine that with all what I had learned at school: signal processing, neuroscience, design, programming, data science, and experimental research. Also, I give my respect to Bratislawa Yoghurt and Entropy, as true learning and ideas always form outside the lecture halls.

Thank you Stephan, for giving me the feedback I needed. Thank you Timo, for coaching me through the process of writing a thesis. And I can not thank RELEX Solutions enough for funding such a stupidly ambitious idea that me and Svante had for theses. A special place is reserved for those who took their time to be participants in our experiments. Thank you parents for forc— strongly suggesting for me to go engineering school. Signal processing, after all those years, is actually really cool. Being a Master of Science is even cooler. And thank you Svante, for all those endless discussions about the human experience we had during this time. There will be hopefully more yet to come - and I sense that this is only the beginning!

Finally, thank you Anna, for being in my life every day. You keep the wonder alive in me.

Otaniemi, 22.11.2019

Niklas L. P. Strengell

Contents

Abbreviations

Abbreviations

AI	artificial intelligence
AIM	Affect Intensity Measure (a survey)
ANS	autonomous nervous system
BRS	Baroreflex sensitivity
BVP	blood volume pulse
CL	cognitive load
EA	experiential ability
ECG	electrocardiography
EDA	electrodermal activity
EE	experiential engagement
ENS	enteric nervous system
EQ	emotional intelligence
FFT	fast fourier transform
HANV	high arousal, negative valence
HAPV	high arousal, positive valence
HCI	human-computer interaction
HRV	heart rate variability
IBI	interbeat interval
NN	norma-to-normal
LANV	low arousal, negative valence
LAPV	low arousal, positive valence
PNS	para-sympathetic nervous system
PPG	photoplethysmography
REI	Rational-Experiential Inventory (a survey)
RSA	Respiratory sinus arythmia
RA	rational ability
RE	rational engagement
SCR	skin conductance reaction
SNS	sympathetic nervous system
UI	user interface
UX	user experience

1 Introduction

This thesis started as an idea on a normal work day at the UX team at RELEX Solutions. We pondedered: what if we could measure emotions? What could we achieve then? How could we use that for our user research process - or the future of human-computer interaction?

The curious idea started turning into a reality as a form of two theses projects, but the idea quickly met its first stumbling block: there really aren't any do-it-yourself manuals for emotion recognition systems. Even though the topic has been studied for over a century in various scientific fields such as psychology, economics, and more recently neuro- and computer science, there is no formative work that would tie in the problem of emotion together from all of these views. For example, most of human-computer interaction literature deals with emotion as something very singular and concrete that could be produced by certain stimuli and could then be measured. Economics is only really interested on the decision-making component and other simplifications. Psychology deals with constructs, but lacks neuroscientific evidence; neuroscience has lots of data but lacks theory. And yet everybody says the same: emotions are important.

Thus, we felt that such a multidisciplinary work should be produced. A work that goes through the theory of multiple scientific disciplines which have studied emotion: philosophy, psychology, economics, neuroscience and human-computer interaction. In addition, that work should be supplemented by real experimental data. Although the major goal for that experiment is to provide data for the case company's emotion recognition project, we will nonetheless utilize scientific discipline and methods in it. We think that there really isn't any greater teacher of scientific knowledge then nature itself. The researcher should expose themselves to the problems of real life, not just theory. And emotion recognition has many practical engineering problems as well. The first is a naturalistic setting: we have studied emotions for a long time in a lab, but no so much in real-life environments where noise and confounds are not easily controllable. Second, we know that there are multiple biosignals that we can measure, but we still don't know well how they actually tie into the actual experienced emotion. Third is the problem of personality: the science likes to deal with regressed means and averages, but something so personal as emotion should take into account the variability of personality as well, because "the average man" may not exist at all.

And this is the purpose of this thesis: gather what we already know about emotions, design an experiment based on that knowledge, validate our metrics and couple them with personality types - and finally, show how to move forward from what we already know.

1.1 Statement of the problem

The statement of the problem is wide: "how to measure emotions from the wrist during human-computer interaction?". To understand it better, it could be divided in to three parts: the theoretical part, the engineering part and the experimental part.

The first problem, the theoretical part, is to understand and formalise what emotion could be and why emotion is so important. This should be approached in a multidisciplinary way, because each before mentioned field has their unique insights but also their own blind spots. It is not enough to just say that something causes an emotion. Emotion is a much more complex phenomenon then just a simple stimuli-reaction reduction. Thus, it is much more important to understand how an emotion forms and what components it has and how it affects our cognition and behaviour. This has been done in neuroscience and economics, but not really in human-computer interaction. This of course, is not the first time when this problem has been realised and laid out. A similar problem statement was already laid out in 1997 by Rosalind Picard in her book *Affective Computing* [131]: Picard argues that the future of computing should be affective. She points out that, the scientific findings from the last decades indicate that emotion plays an essential role in decision making, perception, learning and even more - they influence the very mechanisms of our rational thinking. Too much emotion or too little emotion impairs your decision making. Thus if we want computers to be genuinely intelligent and interact with us in a natural way, we need to give computers the ability to recognise and understand our emotions and even to have and express them. And over 20 years after the publishing of Picard's book [131] that problem is still relevant [25][132][179]. It has not been solved, but it's importance has been noted and many more questions have arisen.

That same problem is also of interest on a more practical level: the software industry is becoming more and more interested in understanding the totality of what happens when their customer's use their products [68][127]. The drivers that lead to retained usage and customer satisfaction are much more complex than just usability or other task related measures. In other words, we are moving from just studying usability towards studying the user experience. But according to Marc Hassenzahl [67], approaches to user experience in HCI lack theory and empirical investigation. It seems important to better understand user experience itself, its determinants and situational/personal mediation and to validate this understanding. But this is already a philosophical problem: how can you study and measure something so vague as an experience? One possibility would be to measure an important part of it that is at least somewhat measurable, such as emotional arousal or emotional valence, which seem to form a key part of any user experience [69]. Psychophysiological methods, which can measure emotional processes through measurement of physiological changes, are thus becoming of interest for HCI researchers, because they can bring valuable complement to qualitative and quantitative subjective reports and observational analyses [51]. However, this technology is still in it's infancy and further research has to be done to improve these methods and develop non-invasive technologies.

Which leads us to the second problem: the engineering part. The signals from which we can derive emotional cues are multiple. We could measure non-verbal cues such as eye gaze, pupil size, tone of voice, prosody, posture, body movements and facial expressions for emotional information, just as human do. Or we could tap into signals that are unattainable for humans, such as skin conductance, muscle potential, cardiovascular measures and brain waves. From a computational point of view, all these sources are physiological signals with common difficulties of analysis and interpretation [179].

The third problem is the experimental problem. As said, the emotional measurement should happen a in naturalistic setting. This requires the measuring devices to be robust and reliable, yet small and unobtrusive. Technologically, this is becoming more and more viable with the advent of mood rings and other personal measuring devices, and we can finally move out of the lab. However, confounds and noise are then of course increased and experimental design should change accordingly. In addition, when we move out of highly controlled lab experiments, one key element of human condition starts to affect our experiments - the personality. When the variability of the environment increases, so does the variety of strategies how people interact with that environment increase as well. We want to tie this variable into our measurements as well, and see whether personality constructs could be used as a priori to understand the baselines and correlations between psychophysiological methods.

Thus we suggest that a successful combination of experimental research with multidisciplinary theoretical framework would help the further development of psychophysiological research methods and increase our understanding of user experience. We do not claim that this work would provide groundbreaking results, but rather we want to show that with the recent technological advancements, HCI research with psychophysiology is viable in a naturalistic setting even with small samples and low resources. Furthermore, this kind of work would most certainly be of value to other study fields who are interested in emotional biosignals and human experience, such as psychology or economics.

1.2 Background and Need

Our motivation for this is that emotions are important for every interaction in our lives - also the ones with computers. The traditional narrative, still widely upheld during the last century, considered emotions too fuzzy for any decision making and intolerable to be shown outwardly [131, p. 1-2]. We should reason and be rational when making decisions. Emotional experiences - they have traditionally been left to poets and songwriters. Show them around at your home to your wife and kids or enjoy them at the movies - but don't bring them to work. On the same note, emotions were surely outside of any credible scientific research. Not so much anymore. Emotional intelligence's (EQ) importance is emphasized in workplaces [54]. Research is challenging the view of of purely rationalistic decision making [8][33]. Marketing slogans and political campaigns are touting us the power of both reason and emotion. But apart from the hype, still too little wisdom from our emotional experiences spread into the structures and processes that shape our human societies. And this view is still very much alive and prevails even in the most advanced and cutting-edge fields such as the software industry. Emotion is often overlooked both in the development of the software [73] and in the organisation itself [57][156]. But it shouldn't.

Emotion is also an important of design. Every product or service will elicit certain emotions in their users [125]. But a product that was designed by a designer, is ultimately not used by that designer, but by some other end-user. A product can only have intended functionalities or characteristics, but ultimately it is the end user whose perceptions, actions and consequences are the judge. And some of those consequences are emotions, such as dissatisfaction or pleasure. That is why understanding emotion in the context of design is important [67]. This is why, we talk about user-centered design because we want to design for the user. And to design for the user we need to think and feel like the user. Surely, empathy is a crucial part of design in general. It is the first step in the much raved "design thinking, a framework on learning how to think like a designer [21]. And some designers have truly lived up to their task "of putting yourself into someone else's shoes" such as Griffin [58] in *Black like me*, where he tried to live in the southern states as a black man during the 50's (note that Griffin was white). Or the industrial designer Patricia Moore, who assumed the role of a 85-year-old-woman, who very clearly explains her motive:

> "How could I expect to understand their [elderly people's] needs, to help design better products, develop new concepts, without a sharper sense of their daily experience?" [119]

Fellow researchers from Aalto University have also described a method of *bodystorming*, a way of acting out scenarios in their real environment and context to find out novel solutions to the presented design problems [129]. These methods are viable and their effectiveness proven. Experiential empathy is relevant now and will be in the future. But there is a further challenge: how can we be sure that we pay attention to the actually important things? How can we be certain that we are really empathizing, tapping into other people's experiences and not just confirming our own biases?

Although the importance of empathy in design is acknowledged in the field, much of that emotional design remains intuitive: designers base their choices on how *they* think and feel, and not on any scientifically based objective measurements [146]. Your experience is your experience and my experience is my experience. Even though they might be alike, they will never be the same. While acting out the daily lives and experiences of the users, the designers might still miss important things or make false judgments and observations because of their own internal biases. Even in the shoes of the other, the designer will value and notice different things - if only even subconsciously. And to some degree this might be even detrimental, as the "acting out" might further create a confirmation bias because now the designers think that they have empathized with the user. But it is still better than nothing or a purely utilitarian approach. Empathy and reading people's emotions is something that comes naturally to us humans. We don't need to be trained for it nor do we have to create any technologies for that. There most likely isn't any grand conspiracy against scientifically validating other people's experiences: most likely it is not done because it is an incredibly hard thing to do. Thus the emotional understanding has been left to the designers and their subjective judgments alone. Objective evaluation was left for the places easily measurable.

From the beginning of human-computer-interaction research, the clear emphasis was on pragmatic matters such the utility and usability of the software [123], which were measured through methods such as the usability test and the acceptance criteria list. Pleasantness of the users' experience or affection towards the thing was not what would be measured - it would be a result if all the other criteria were met. Now, a further interest in user experience [68] and an increased viability of psychophysiological measuring [51] is changing this thinking. Automated emotion recognition has seen a new boom due to speed increases and lower costs of computing in general, as well as the thorough digitization of our daily lives, where almost everybody is holding an internet-connected supercomputer in their pockets for most of the day [30]. Furthermore, modern machine learning methods are giving computers unseen abilities and emotion recognition is moving forward with this development at a fast pace [142]. A further sign of this trend is the fact that this thesis is done in collaboration with a company, which wants to integrate a kind of psycho-physiology based method into their user research process to further improve the understanding of their users' needs and experiences during their real work in the application.

As digital technologies are pervading our daily lives more and more, our relationship to them changes as well. The first computers were quite utilitarian in nature, working machines meant to increase the performance of calculation intensive tasks or documentation of important papers. People were trained to use computers - not the other way around. This, of course, could not last forever.

> "Computing is not about computers anymore, it is about living."

phrased Nicholas Negroponte [120], the creator of MIT's Media Lab, 25 years ago and in today's smartphone-fueled era his statement is the living reality. You can not pay a bill or buy a bus ticket without the help of a computer - it is a necessity nowadays, not a luxury. The computerisation of our lives is a trend that does not seem to stop and it has already evolved from the desktop to the mobile. This is also a reason why this thesis is part of the research done at the Ambient Intelligence group of Aalto University's Department of Communications and Networking. Ambient intelligence refers to a vision of future where computing starts to vanish from specific computers to everywhere in our environment. In an ambient intelligence world the technological devices have completely integrated into our surroundings with only the user interface visible. Those computers would help us in our daily tasks and doings in an effective and natural way. For those interactions to be and feel natural, any real ambient intelligence or humanized computing must thus understand us humans as we are - emotional beings. For the future of computing to take the next step, our emotional behaviour and communicating should be understood by computers to make our interactions with them feel more familiar and fluid. To achieve this vision, we should be able to measure emotions during those moments unobtrusively, yet with high accuracy and fidelity. Even though facial recognition software is a common day practice for every teenager with a phone, there is still a long way ahead to reach the real, genuine human-like interaction with computers like we did with the sympathetic R2-D2 of *Star Wars* or the menacing HAL in *2001: A Space Odyssey.*

The benefits of recognizing emotions in natural context could also spread to the psychological and physiological health of individuals. This trend can already be seen in the rising amount of heart-rate tracking watches and "mood rings" [31][79], which measure skin conductivity or heart rate variability. Not to mention about new possibilities for naturalistic research in humanistic sciences. Furthermore in science, one might hear of qualitative research and quantitative research. Often the emotional side is only left on the side of qualitative research, where people are interviewed about their subjective feelings about the matter at hand. This is also the case with software development and user research, as experiences and emotions are usually only done through the application of interviews and observation. The quantitative side then has the spotlight of the natural sciences, such as physics - or in the case of software development, usability studies. Of course, there are quantitative and automated methods outside natural sciences, but often in there also we tend to focus on things easily measurable such as time or money spent. In software companies the marketing department might emphasize the importance of feeling [64][145] and how they can be measured through surveys, but therein lies also the trap of subjectivity [51]. This is also a point raised out by this thesis. The research in this thesis in not conducted in terms of quantitative or qualitative but in terms of subjective and objective - and how these can be combined. The subjective experience might always be out of reach for objective research methods. Emotion is not a signal that can be measured directly. But what can be done, is to measure multiple signals, to combine them to patterns and then understand the meaning through the context where they arise. We do not need to measure the whole of the experience and the emotions that a person has at any given time. Neither would this be practical. But we could measure some components of the emotions. Especially interesting are the physiological signals that we can not consciously control and would be out of the normal repertoire of human perception, such as skin conductance and heart rate variability.

A quick note about the case company The case company, to whom this development project was done, is a finnish software company. The company sells various planning and optimisation software solutions for retail companies. Their main product is a cloud-hosted planning software. It integrates to customer's ERP software and enables calculations and workflows with their own data, such as sales forecasting or inventory managament. Furthermore, many end-users handle mission critical work tasks such as order proposals in the product, making smooth and error-free operation a top priority.

The software handles complex tasks and is thus itself also, complex. Although complex problems usually require complex solutions [164], complexity does not need to be complicated, even though it so often is [126, p. 4-8]. Complexity can be made simple, through the usage of recurring design patterns, constant interactions and by architectural choices of layering the complexity. This complexity also proposes more practical problems: not a single designer or developer can handle all of the actions and complexities in the product, making objective analysis and clear communication of design decisions and research important within and between the teams.

Thus there is a growing interest for not just designing the product, but rather understanding the totality of the experience our users have while using the software.

1.3 Purpose of the Study

The purpose of this study is to understand better the theory of emotion in the context of human-computer interaction, and the data gathering and data analysis problems that come with emotion recognition in a naturalistic setting. We specifically work with heart rate variability and electrodermal activity. For the case company, it would help with the user research process. This thesis is part of a project, which aims to integrate emotion recognition to the company's user research process. The results from this thesis will be used in the development of that process. This thesis gives answers to the viability of the different metrics, which can be derived from HRV and EDA data, and suggestions to how the results from those EDA and HRV data should be interpreted.

The first key component of this vision of emotion recognition aided user research is to have the ability to record emotional cues in a naturalistic settings - that is, as they would happen during a normal work day. This requires a reliable device, capable of medical grade accuracy, yet one that would cause no disturbance to its wearers. For this research, we decided to use a device called E4 by the company Empatica. It seemed to fill all the needs. As wrist-worn and light-weight, it shouldn't be any more disturbing then a normal wrist-watch. Empatica E4 can record electrodermal activity (EDA) and blood volume pulse (BVP), from which we can calculate the heart-beats and the variance between those successive heart-beats is called the heart-rate variability (HRV). E4 can capture very high quality data, with EDA signal having around 90% of the data detected as clean signal and BVP signal having 80% of the data is in the acceptable range [45]. These are promising and interesting metrics also for user research, as they are biosignals that are related to emotional phenomena [37][158]. In addition, they have also been used in research to understand decision-making [33][8], problem-solving [157][141] and human-computer-interaction (HCI) [51].

As already pointed out, biosignals are not simple to analyse. They are not simple stimuli-to-reaction input-output-systems, but rather part of very complex biological machinery that keep up the organism's current biological state. However, we do know that these metrics relate to certain neural activations that themselves relate to emotional processing. For example, EDA has been shown to be correlated with sympathetic nervous activity [37], which is usually activated during the so-called "fight-or-flight" moments. HRV then has been shown to be correlated with both sympathetic nervous system activity as well as the parasympathetic nervous activity [158], which is responsible for "rest-and-digest" or "feed-and-breed" beahaviours.

At our initial recordings with our wrist sensor during usability studies at the case company we noticed something. We found out that human-computer interaction indeed elicits both electrodermal activity and heart rate variability - a fact already well-noted in literature as well. However, the data was extremely hard to analyze. The data was very noisy and inconsistent. To further complicate the situation, even during an exactly similar usability test, the reactions from person to person were different - even if they would have described their experience with similar words. Thus we understood that the first step for us would need to be to understand what happens in the readings: how they relate to emotional processing and what are the most important metrics, and how personality affects them.

Psychology and our human intuition knows already that personality affects emotional behaviour and vice versa [80][93][121][133][138][173][180]. Although probably good for mental health and human culture in general, this is bad news for measuring affective phenomena: personal differences of emotional behaviour and regulation make purely signal-based and analytical models hard or impossible to use for recognizing the emotions at hand. Thus introducing personality as a priori for data analysis might be a good idea, to enable the normalisation of data and think about the possiblity of grouping people together based on their data.

We also want to point out that the purpose of this technology is by no means to remove the need for usability studies, practical empathy, bodystorming or any other methods. This is a mean to enhance it further - a new tool in the box. Same goes for the ability to measure our own emotions. Your own opinion should still be the guiding principle. Technology is there to only assist you to see something that is hidden from our bare eyes.

1.4 Research Questions

This study works with the following research questions:

- RQ1: What is an emotion in the context of human-computer interaction?
- RQ2: Why is emotion important for human-computer interaction?
- RQ3: How to measure EDA and HRV non-obtrusively in naturalistic settings?
- RQ4: How to measure arousal from EDA or HRV signals?
- RQ5: How to measure valence from EDA or HRV signals?
- RQ6: How to take personality into account for psychophysiological measurements?

In addition, we do have a multiple hypothesis based on our previous work as well as research literature:

- H1: EDA is an indicator of emotional arousal
- H2: HRV is an indicator of emotional valence
- H3: the self-assessed affect intensity (AIM) is negatively correlated with the intensity of psychophysiological metrics

2 Background

We are now in a territory that should be familiar to any reader: the human experience, it's emotions and the technology that affects us no matter where we go. This is an interesting field to be in because everybody will always be the ultimate master and expert of their own subjective experience. Here, you always have something to which you can relate to. However, there usually is much more underneath then we what we can see at the first glimpse - and this certainly holds true for the daily condition we call being a human. The division of thinking into emotions and reason is not a clear one - if even true at all [34]. It took our body and emotions millions of years to evolve into the complex highly functioning biological machine which we now find ourselves. With very little conscious effort are we able to feed, breed, repair ourselves and to observe: see and hear all our surroundings, to taste our food and feel the breath of fresh air - what joy is this, to behold a body! Doubt that if you will, as certainly can we not give you any better proof of your existence.

"I think, therefore I am."

The much used (and abused) quote by Rene Descartes [39] has a certain ring of truth to it. Descartes was ahead of his time in many ways. His critique on the fallibility of *passions* and how they erroneously bias our judgments was a true one [40]. He was also probably the first one to methodologically categorize emotions, classifying them into six basic passions: wonder, love and hatred, desire, and joy and sadness. He also distinguished between pure passions, which are bodily-based perceptions, and intellectual emotions, in his words a kind of *volitions* which arise "from the soul" - a view later honed by others [150]. But Descartes had errors: our cognition is not easily dividable [34][36][35]. The co-operation between emotion and reason is a complex matter on which some light has been shed during the last century. We need to use both in order to navigate in this world and to live a meaningful human life. Emotion does not always cloud our judgment: in fact, it is necessary for good judgment. Thus, in this chapter, the reader will be introduced into the finer details of our body, brain and the emotional processing going on, as well as the technology that shapes it every day - the same technology that ultimately is the enabling factor behind our research as well.

2.1 Emotions

"What is an emotion?"

asked the early psychologist and philosopher William James [81], quite rightly, as to date there hasn't been any definitive answers. Emotions are elusive in nature. Everybody has them, everybody feels them, yet no-one can really quite put into words what they are. Poets and artists have known their importance in the human experience and not even objective science can dispute their importance in the human experience. Studies have shown that emotion plays a crucial role not only in our mental health well-being [28] but decision-making as well [8]. However, knowing of their existence or prove of their importance does not yet give us any clear view on what they exactly are, where or how they arise. If *why* we feel is hard a question, then *how* we feel is not that much easier either.

The complexity and ambivalence of the mechanics of affection can be seen easily in the vagueness of terminology and the multitude of different emotional theories we have. Maybe the right one is the 19th century James-Lange theory of emotion that suggest emotions occur as a result of physiological reactions to events [82]. Or maybe it is the the two-factor theory of emotion which says that emotion is based on two factors: physiological arousal and cognitive label, and we must always identify the reason for the arousal to label it and then only do we experience it as an emotion [148]. Appraisal theory takes this idea even further, suggesting that emotional interpretation can and will happen even without any prior physiological arousal and that emotions are extracted from our evaluations of the environment [104]. (Un)fortunately, the amount of different theories doesn't stop there nor do we have a clear consensus on how to quantitatively measure them: there is a long, on-going dispute between discrete models with a few basic categories, or continuous models, where all different emotions are rated on some (orthogonal) axes. However, the field is now moving away from these sterile debates of absolute rights and wrongs towards a more productive work on the details of cognition–emotion interactions [61].

2.1.1 The Qualia Problem

> "The really hard problem of consciousness is the problem of experience. When we think and perceive there is a whir of information processing, but there is also a subjective aspect."

This famous quote by the philosopher David Chalmers [27] is called the hard problem of consciousness or *the qualia problem* and it has buggered scientist and layman perhaps millenia if not for an eternity. It also poses a very pragmatic problem of subjectivism: our interactions with other people are inevitable and constant. We thus need to have some kind of common understanding and language on almost each specific topic. But how do we lay out those common grounds and principles? How do we reach out to each other's experience, if even briefly and partly? Language sure is a great tool, but how often than not have you experienced a misunderstanding due to using a certain word or manner, which you only meant well but was understood differently? This is what we would call, if you allow, the *pragmatic problem of consciousness*: how do we establish a certain common ground when talking of subjective experiences?

The *qualia* problem will not go away in the foreseeable future. If anything modern science shows us, it is that experiences can vary drastically. We experience things differently. For example, our vision is not only affected by the physical characteristics of our eyes, but by cultural differences as well [153]. And curiously, the same goes for perception of emotion: it varies from culture to culture [169][113][112], from individual to individual [62][116]. We might not see, hear, taste or even feel things in the same way. And it is almost impossible to prove or disprove that two experiences would be the same. But we don't need to delve too deep in the mind-body problem or the nature of reality in order to study feeling. Quite the opposite, we can even merge them too here, just like the Enlightenment-era philosopher and an avid admirer and critic of Descartes, Baruc Spinoza said:

> "The object of the idea constituting the human mind is the body." [167].

What Spinoza suggests here is that, our conscious thoughts do not arise nor exist in a vacuum. They arise from stimuli of the outer world, and they are modified from the signals from our body. Modern researchers have taken this idea further, most famously the neuroscientist Antonio Damasio with his somatic marker hypothesis, and they have found some evidence to support this idea [34][36][35]. In his hypothesis Damasio suggests a mechanism how the somatic markers caused by the body are processed in our brain and thus affect our reasoning [33]. Body *affects* our thoughts and vice versa. And isn't it curious that the same word, affect, means "to have an effect on something" and "to touch feelings" [1]?

2.1.2 Emotion, mood, feeling and personality

The first important step towards our understanding of emotions is to define terminology. Traditionally in poetry and fiction literature many words such as *feeling*, *emotion* or *affection* are interchangeable and their definitions loose. Academic literature hasn't fared much better and the word *emotion* is used to refer to a variety of phenomena, from mild to huge and from experiential to behavioral [60]. Ross Buck put it quite nicely three decades ago, saying that it is in state of "conceptual and definitional chaos" [23]. Luckily, research has moved forward and conceptions have been more rigorously defined. In this thesis, we follow the terminology laid out in the paper *The future is so bright I gotta wear shades* by James J. Gross [61] (also, see figure 1):

- *affect* or *affections* is a wide umbrella term to cover all *affective* phenomena and processes.
- *attitudes* are relatively fixed (but can and do change) believes about goodness or badness of things. They bias one's thinking and feelings toward objects and other persons.
- *moods* are not as fixed as attitudes, but still somewhat stable and long-lasting. They don't need to have any specific object, but can rather exist on their own. They then bias other affections for their duration.
- *emotions* are reactions to different situations that are relevant to one's current goals and have a short duration, lasting from seconds to a few minutes. These reactions consist of *appraisals* that then give rise to different physiological and psychological changes in the individual.
- *feeling* is the conscious perception of an emotion.

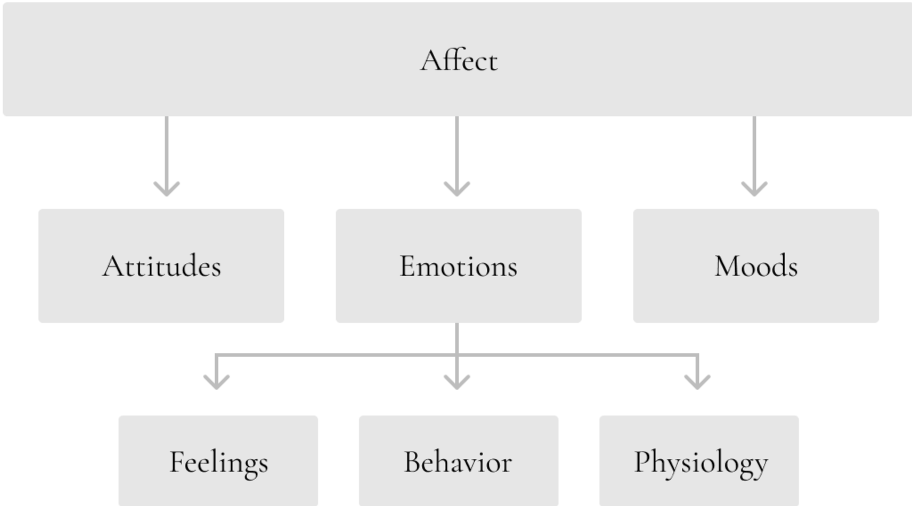

Figure 1: How the terminology of different affective phenomena can be grouped. Adopted from [61].

This semantical categorization already helps with the abstraction of emotions and gives some needed order when writing about them. And surely everyone can reflect from their own being that these are not purely constructed hierarchies, but actually reflect some level of reality. Anyone who has ever had a bad morning - woken up with the wrong foot so to say - knows how their emotions end up very much on the negative side during the rest of the day. We aren't just guided by our emotions, but rather different aspects of many affective modulations. Our personality might give us a certain baseline from which we experience and show emotions for our whole lives, and our current mood biases our thoughts and feelings for days or weeks. Lastly, the reactive emotions might spark us to action or inaction but fade away in the matter of minutes (for author's visualisation of this mental hierarchy, see figure 2). But this still leaves many doors open: how do the emotions arise to our consciousness? What is their innate hierarchy in ourselves: why, how and where do emotions spring to life?

2.1.3 Emotions in the body

According to neuroscientist Antonio Damasio [35], emotions seem to precede cognition, intelligence and brains. Every living thing, from amoeba to primates, can automatically solve the basic problems of life: finding sources of energy, transforming that energy, maintaining proper internal and external living conditions, repairing injuries and fighting off external threats as well as reproducing. One single word that refers to these life regulating processes is *homeostasis*. What is incredible, is

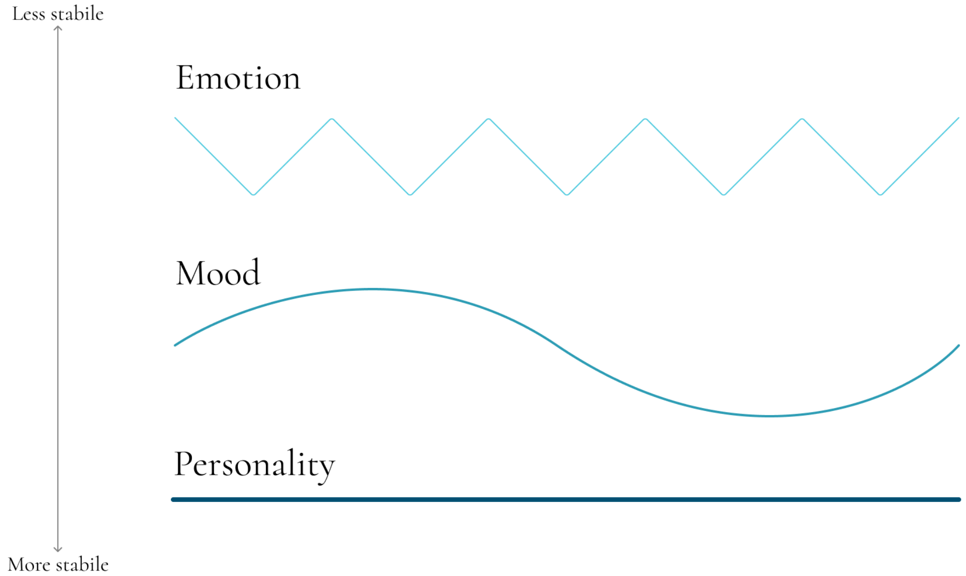

Figure 2: How the relationship between emotion, mood and the personality can be imagined. Personality is rather stable and fixed through lifetime, wheres as moods change constantly albeit not as fast as emotions.

that every single living organism is an incredibly skillful homeostasis maintaining machine. Then depending on the complexity of the organism, the maintenance of the homeostasis varies from the simple to the complex. In the bottom, there are simple reactions or reflexes, such as *withdrawing* or *approaching* relation to an object. Higher up in the processes, you can have increases or decreases in activity (*arousal*) and even higher up, we can already find cognitive behaviours such as *competitive* or *cooperative* responses.

Damasio has a great metaphor of imagining this homeostasis regulating machine as a branching tree, where complexity increases as we move towards the top (see figure 3). On the lowest level, we have basic reflexes, metabolic regulations and the immune system. Moving one level higher up, the tree can be thought to branch for the first time as we have pain and pleasure behaviors: pain causes the organism to move away from an object, where as pleasure makes it move towards it. Then on a later stage, we move up to drives and motivations. With these Damasio means the innate, intrinsic biological drives needed to support organism's survival, such as hunger or sex-drive. They are already related to emotions, but are more primal and simplistic. Lastly, we have emotions-proper (which can then be experienced as conscious feelings). The tree can also be thought to be inverted in ratio of conscious-unconscious action. The lower in the tree you are, the less control an individual has over those workings and moving up on the tree it becomes more and more a matter of conscious choice.[34][35]

2.1.4 The cognitive aspects of emotion

Damasio further divides the emotions-proper into *primary* emotions and *secondary* emotions, which are somewhat comparable to what Descartes described in his philosophy. By this, he creates a helpful distinction between the cognitively-created emotions and the non-cognitively-created emotions. What he suggests is that some of our emotions would be innate: we react to them first (primarily) with emotion, and then only later with cognition [34]. Examples would include the startling sensation after hearing a loud bang or screaming if someone scares you unexpectedly. These are the kind of stimuli, that are mainly processed in the lower and older limbic brain structures. Secondary emotions, on the other hand are first cognitive, also activating our prefrontal cortex and somatosensory brain areas, and only later emotional, activating the limbic structures. These arise in the individual after a sufficient development phase, when the relations and connections of emotions to objects are understood [131]. For example, the feeling of anxiety when you realise you have a deadline by which to complete a Master's thesis. The point is, the primary ones are innate and "hard-wired" biological responses to stimuli around us. They would thus be more or less universal across all human beings (or even some other sentient beings). But the latter ones, the secondary emotions, would be learned through experience leading to a great variety and complexity of associated responses to experiences. Depending on the outcome of our personal experience, we would learn to associate experiences into negative, positive or neutral outcomes. This process continues throughout our lives, giving us a unique combination of responses varying

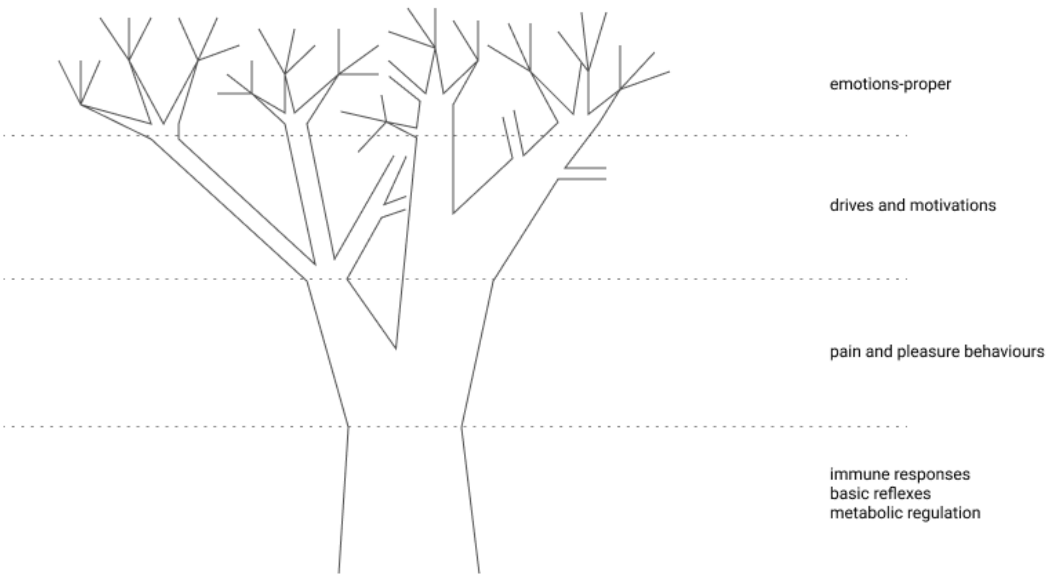

Figure 3: The different levels and complexity of automated homeostatic regulations. This figure has been adopted from Damasio's book *Looking for Spinoza* [35, p. 32]

from person to person. And according to Damasio, it is this specific distinction in emotions that give us our special ability to reason about the world [34].

2.2 Emotion and cognition

Others have also noted the necessity of emotions and how they relate not only to our mental well-being but our decision making and motivation as well. Already in the 1960s the Nobelist Herbert Simon argued well in his essay *Motivational and emotional controls of cognition* [163], that any real model of cognition must include emotions. This paper was one of the earliest ones to try and integrate affective phenomena into the information processing view of human cognition. His view is that emotions act as an *interrupt system* which can temporarily override other programs if real-time needs of higher priority are presented. Thus he viewed emotion as an response to the sudden changes of our environment which demand our immediate attention and thus arouse our autonomic nervous system.

Herbert Simon later refined his theories into the concept of *bounded rationality* [71]. Bounded rationality means that humans can not take into account all possibilities of action for purely rational optimization. Our rational abilities are limited due to the computing limitations of our brain, the tractability of the decision problem, and the time available to make the decision. We thus need something that can limit the possible search space into a limited amount of possibilities [164]. In *Models of man* [162], Simon states that most people are only partly rational and rest of their actions are irrational. Humans always need to make decision based on heuristics, using the regularities found in our environments and experiences as shortcuts to limit the types of utility functions, to recognize the costs of gathering and processing information as well as to understand that there can be a "multivalued" utility function with multiple goals that need to be satisfied.

2.2.1 Somatic marker hypothesis

Antonio Damasio has proposed in the paper *The Somatic Marker Hypothesis and the Possible Functions of the Prefrontal Cortex* [33] a possible mechanism for the workings of the heuristics described by Simon which give us the unique ability of human reasoning.

Reasoning is the ability to consider a set of alternatives, the imagination of a series of possible outcomes, to which we can then apply logic to determine which alternative would yield the best possible outcome. The problem with pure reasoning is that it is incredibly taxing. It requires constant attention, memory and most of all, time. None of which the human brain has in abundance. Identifying alternatives, generating outcomes and weighing their worth can take a long time, even for the simplest of decisions. We would thus need a system, that would limit the set of possibilities to consider to a (literally) *reasonable* amount. Damasio suggests, that it is exactly the secondary (and to some degree primary) emotions that help us bias the outcomes automatically, getting quickly rid of the ones with negative or non-useful outcomes. Through secondary emotions and their learned associations between objects and outcomes, we can "jump into conclusions".

The mechanics of this hypothesis lies in the so-called "marker" signals that influence our unconscious and conscious responses to a certain stimuli. These signals arise in the bio-regulatory processes and express themselves later in emotions and feelings. The name "somatic" (the root word for somatic is soma, old Greek for "body") implies that these signals are related to the body-state representation and regulation in the brain. The direct and in-direct signals coming from our body to our brain, the information about the world, are played out as “as-if” scenarios in our head and simulated in our brain about the possible consequences of our decisions. A certain imagined or perceived situation will then arouse a respective body-state based on the previous experience that it was similar to, leading us to believe that a similar outcome should be expected this time as well.

Somatic markers in the somatosensory parts of the brain are also needed to limit the amount of possible actions in order for rational decision making to work. As decisions are often related to maintaining the homeostasis of the organism, it makes sense that much of these decision-making markers are (or are at least related to) those relative processes that balance the organism’s homeostasis. Hence the “goodness” and “badness” of our decisions are often negated by these somatosensing markers. We can think of emotions as a form of prediction. They must not always be right to be useful. We can use our emotional system to quickly generate a much more limited set of likely alternatives, which can then be given to more sophisticated but resource-consuming brain areas responsible for reasoning. By then employing the logical principles on our previously generated and limited set, we are even more likely to have a better and more realistic forecasts about future.[34]

Like Spinoza and his idea of the essence, a divine body, from which both reason and emotion arise [167], there is no opposition to emotion or logic in Damasio’s theory. It is not reason nor feeling that triumphs but their correct balance used in a fruitful away.

2.2.2 Dual system theory of cognition

The idea of dual system theory in cognition has received widespread attention lately especially through the popular science book, *Thinking fast and slow* by Daniel Kahneman [87]. Kahneman’s (and his late colleague’s Amos Tversky’s) statement, which is based on many decades of both his own and other’s research, is that human cognition can be thought to be dichotomous. We have a System 1, the fast and intuitive system, which operates always on the background and tries to make sense of the vast information we are always exposed to. However, as this system relies on previous experience and uses cognitive shortcuts to do faster decisions, it can very often lead us into misjudgments and biases. System 2, however, is slow and rational one, and acts as an error-correcting mechanism to this system. These two systems co-operate (and sometimes don’t) and give us the unique properties of the human cognition.

Kahneman and Tversky’s work lies on a lot of previous research and the idea between two parallel system has been around at least since Aristotles time [47, p. 712]. One of the most influential and pioneering theories has been the cognitive-experiential self-theory (or CEST) by Seymour Epstein [48][46][47]. Seymour refined a lot previous work in the field of psychology and psychoanalysis with a more rigorous and scientific framework. He replaced Freud’s repressed sexual unconscious [50] and Jung's mystical collective unconscious [85] with a more adaptive system, which introduced the distinction between "the Experiential Self" and "the Rational Self". This experiential self is a system that learns from direct experience through association. It is motivated by emotion. The rational self on the opposite is mostly verbal and it operates on the principles of logical analysis. This system receives a lot of it's information from culture and other people by the means of verbal communication.

The balance of using these both systems relates to mental well-being and health; an imbalance leads to unhappiness and miscommunication, cognitive dissonance or even mental disorders. And the co-operation between these systems is not always a harmonious or an easy task, given that the experiential system has evolved over millions of years and the rational self has not but is rather a very new system. A basic principle of CEST is that people construct their own theory of reality that is divided in two: a world theory, created by the experiential system, and the self theory, created by the rational self, and these two are connected by various propositions.

The experiential system deals with *schemata*, which are gathered from emotionally significant past experiences. These schemata are all tied together, where everything affects everything else, to form a very coherent and adaptable belief-system leading to surprisingly functional and complex behaviours. An unassimilable, emotionally significant experience leads to the disorganisation of the system. At its lower levels of operation, it is a crude system that processes information, but does this automatically, rapidly, effortlessly, and efficiently . At the higher levels, and in interaction with the rational system, the experiential system can be a source of intuitive wisdom and creativity. Although it primarily deals with concrete events and images, it can form generalisations and abstractions through the usage of narratives, metaphors and prototypes.

The rational self is the opposite of this system. It is deliberate, requires a lot of conscious effort, operates on language, and constructs itself with *beliefs*. These beliefs can be highly abstract and have a long-term delay in gratification. Much of the current technological advancement can likely be granted to the workings of this system's work, but in everyday situations it is not very practical and it's long term adaptability is still to be tested.

Rational System	**Experiential System**
Analytic	Holistic
Intentional	Automatic
Rational	Emotional
Mediates behaviour by conscious appraisal	Mediates behaviour through "feel"
Slow for delayed action	Fast for immediate action
Easily changed through reason	Resistant to change
Conscious	Preconscious

Table 1: The charasteristics of the different systems. Applied from [46][47].

Furthermore, it is important to understand that in CEST all human behaviour is assumed to be a joint cause of these two systems operating in coherence (or not). The balance in the usage between these two systems shifts all the time. Some of it is determined by individual differencesnsuch as personality, upbringing and education, but situational variables are also important: some situations require more intuitive experience, some more rational reasoning. Emotional arousal and relevant experience are considered to shift the balance of influence in the direction of the experiential system. A simplification of what happens during an emotionally significant moment is that the experiential self searches our memories for a fitting past experience including the emotions that were active at that time. These recalled feelings influence our futher processing and reactions, which in animals are reactions and in humans conscious and unconscious thoughts and actions. If the reactivated feelings were pleasant, they motivate further action to reproduce those feelings. If they were not, avoidance is preferred.

CEST also argues that most of the time we are relying on our experiential system or are at least greatly affected by it, because it is so effortless and deals very well with daily problems in life. Furthermore, as the experiential self is so tied together with emotion as it can be seen more lucrative option than the cold, dispassionate rational system.

A futher construct in CEST that we still need to go through and understand, is *the need principle*. Epstein gathered previous psychoanalysis theories with their respective needs and instead of picking one, brought them together. These needs are:

1. maximising pleasure and minimizing pain (Freud)
2. maintaining a stable, coherent conceptual system (Rogers, Lecky)
3. need for relatedness (Fairbairn)
4. overcoming feelings of uncertainty and enhancing self-esteem (Adler)

But CEST doesn't lift any of them above the other. In contrary, Epstein argues that these basic needs are equally important and serve to check and balance each other out. Similarly, all human behaviour is thought to be a compromise between all of these four basic needs. If one need is prioritised at the expense of the others, the need to fulfil the others increases subsequently. This can lead to maladaptive behaviour where we sacrifice everything else for the fulfilment of only one thing or even to the creation and life-long pursuit of "false goals". False goals are such goals that even if they are achieved, they are felt as disappointments, because they failed to fulfil the underlying, true basic needs.

False goals are a product of rationalization. As we've pointed out, as a rapid system the experiential self is constantly affecting our thoughts with subtle biases in the form of moods and feelings. As this process is automatic and unconscious, it can appear to form out of nowhere, consequently prompting people to search for an explanation in their rational system - where it is of course not to be found. In normal, healthy, functioning life, this is of course adaptive - it makes our cognition quick, yet flexible. But if this rationalisation is combined with a sacrificed basic need, it can create an extreme case of "false goal". If a person fails constantly to pay attention to their actual needs, a false goal with rationalised belief system is created; "I would be happy if it only were for this...". Epstein observed this deep, problematic frustration with life in many of his patients and argued that this was also the case with the patients of Freud etc. Interestingly, newer research has found evidence that emotional intelligence and happiness with life, are correlated with high usage levels of *both* systems - the experiential and rational [151].

The rational system can also influence the experiential system. As we pointed out it naturally acts as correcting and checking system for the automatic responses - this is for example what we refer to as controlling ourselves to achieve delayed gratification ("I will only eat the chocolate after the food as it is healthier"). It can thus fine-tune our goals and needs into a much more productive behaviours. An experiential system that fails to listen to the reason of the rational system can lead to impulsive and addictive behaviours [74]. In addition, a repetition of conscious behaviour can make the behaviour proceduralized and it can move into the experiental system. This is thought to be crucial in the forming of new habbits: for example, doctors who prefer rationalistic thought to experiential thinking, are less likely to wash their hands before procedures [165].

2.2.3 Dual cognition in HCI

CEST was influential for cognitive science and psychotherapy, because it integrated these two aspects and showed how they act together, wisely bringing out the pros and cons of each system. It is also interesting to see how theories from different fields such as Simon's bounded rationality, Epstein's CEST and Damasio's somatic marker hypothesis are so similar.

Our statement in this thesis is that this kind of thinking will be as important for the field of human-computer interaction and design of interactive technology systems as well. Good design, a great product, not only solves a problem for us but creates an emotional connection to use. Similarly, a good user interface should be able to tap into the usage of both of these systems. Much of "bad design" can be attributed to lack of this. If an UI or product was expected to work in a certain way, but it didn't, it will most likely lead to a less extreme case of "a false goal": cognitive dissociation for the user. Things didn't work as you expected: there was a gulf between expectation and the real outcome and now you are frustrated. Furthermore, the bias towards utilitarian computing had dominated computer science, with UIs that require excessive usage of your rational processing. Now, this has changed, but probably only for the worse: in today's of ubiquitous digitalization we are facing the superficial and mediocre design everywhere, which only taps into the fallacies of our experiential system: things are only created to be beautiful, addictive, marketable.

We argue, that we should aim for the middle road and create software and technology that is humane and understandable and that helps us to achieve our goals - without sacrificing any of our basic needs or abusing our cognition's weak points. Now and in the future, when AI systems with a lot of prediction power are becoming more and more common place, we will give many of our decisions in to the hands of computers. To keep humans in the loop, we should create our technical systems so that our ability to error check them and understanding of their complex behaviour is not compromised and is made as easy as possible. In addition, because software isn't anymore about just UIs and utility, but rather daily living, services and self-identification, we should look to their development processes in a much more wider perspective and understand how they affect our needs on a much wider scale.

And ultimately, if we want any change of the machines to truly understand us, we will need to give them the ability to understand us as we are - emotional beings.

2.3 Emotion and the user experience

As we have already laid out in this thesis, emotions form a crucial part in almost every aspect of human life - and this includes our interactions with computers. We want our experience with technology to be pleasant and aligned with our goals. Where as academia usually speaks only of human-computer interaction (HCI) the industry has coined a term much wider to cover more than mere interaction: the user experience (UX). By definition of Don Norman, who coined the term when working at Apple [127], "user experience" encompasses all aspects of the end-user's interaction with the company, its services, and its products" [171]. For a good user experience, a product must satisfy the needs of the user. This should also happen in an elegant and simple way, so that those products should be joyful to use. Hence, UX covers the user interface (UI) and usability but is not limited to them. A good, usable UI is an essential part of the UX. However, usability studies and metrics have long dominated the research part of user experience [172]. Part of this lies most likely not on any conscious discrimination, but on the fact that usability is easy to define. With terms such as effectiveness, efficiency and satisfaction that can be measured with task error rate, completion time and customer satisfaction surveys, usability studies were the natural inclination to lean towards to. Furthermore, they are still definitely part and parcel of any software development process. Trying to "measure" any experience is by definition a non-trivial task. Being able to measure any part of it is a huge success. Being able to measure a crucial part of it is an immediate breakthrough. And emotions are exactly that, even by the words of Don Norman, who used to preach about the functionality of the product being the ultimate goal [124].

> "The total experience of a product covers much more than its usability: aesthetics, pleasure, and fun play critically important roles. There was no discussion of pleasure, enjoyment and emotion. Emotion is so important that i wrote an entire book, Emotional Design, about the role it plays in design. [126]"

Professor Marc Hassenzahl well formalises this thinking further in his essay *The Thing and I: Understanding the Relationship Between User and Product* [67]. First of all, user experience is *always* subjective. Consequently, any intended experience created by the designer of the product will most likely not be the *real* experience. These actual experiences will vary from person to person, because of their different underlying standard and needs. Also, situations and time may change these experiences - even between the same product and the same user. Second of all, all products have a certain *character* that display the product's capabilities. These capabilities can either be pragmatic, meaning helping the user to achieve their goals, or hedonic, meaning either providing stimulation for the user or acting as a tool of self-expression or provoking certain memories in them. A good product can of course have multiple of these characteristics. To give an example: imagine you had to drive a nail into the wall and for this you would need to get a hammer. As a very pragmatic option, you could just go to the shop and buy a hammer. It does the job. Or, you could buy a more expensive, professionally made toolkit with a hammer. This could spark a new hobby in you (stimulation) and your visiting friends would see your handy-man capabilities (self-identification). Or you could use that old childhood hammer you got as a gift from your grandmother (memory provoking). The final choice will then be made depending on your real needs (and available resources) during that situation. But whatever you value and thus choose, the characteristics of those different products had a certain *appealingness* and caused emotional reactions. These emotional reactions and their desirability also differentiate between what was needed and what was had. Satisfaction may be related to the fulfilment of goals, whereas pleasure may be related to the unexpected.

In the same essay, Hassenzahl also brought one other concern: approaches to user experience in HCI lack theory and empirical investigation [67]. It would be important to understand user experience itself, its determinants and situational and personal mediation and to validate this understanding. The benefits of this would be twofold: first, designers would better understand how user's perceive and act with products. Second, it would allow for the operalisation and measurement of key elements. Both will inform design and lead to better, more satisfying and more pleasurable products. So, if emotions guide our behaviour and decision-making with products and thus need to be involved already in the design phase as well as the validation phase, then how do we it? To understand that better, let us look into how those emotional reactions to design could form.

2.3.1 The three levels of emotional design

In the book *Emotional Design - Why we love (or hate) everyday things?* [125] the cognitive scientist and usability expert Don Norman agrees that our emotional system is extremely important. It is not just a relic of our animal ancestry, but a crucial part of our decison making system that gives value judgements on what is important and what is unimportant. He proposes that this system could be split into three distinctive, yet interconnected levels: the visceral, the behavioural and the reflective (see figure 5). The neuroscientific framework that he uses as inspiration for this proposal is the so-called "Triune Brain theory" [110](see also figure 4). It's basic idea is that our brains have been shaped by different stages of evolution, forming three distinctive layers: the reptilian brain, which gives rise to our instincts, the mammalian brain, which is the house of our emotions, and the human brain, which is the seat of rational and verbal thought. It should be noted, that this theory is now disputed by modern evidence which suggests that processing is happening all over the brain structures, not just on specific areas [91]. But it is nonetheless a very helpful conceptualisation that helps us to give structure and hierarchy for the different peculiarities of our cognition. Norman likewise agrees, and calls his framework a purely conceptual one. However, a conceptual framework that is been backed by experimental evidence [125, p. 21].

So from this theory, Norman devised his own framework of different processing levels. All of these influence our experiences with the world and thus design affects our emotions similarly on three different levels (visualisation shown in figure 5). A good product would correctly activate all those levels, and as he concludes in the

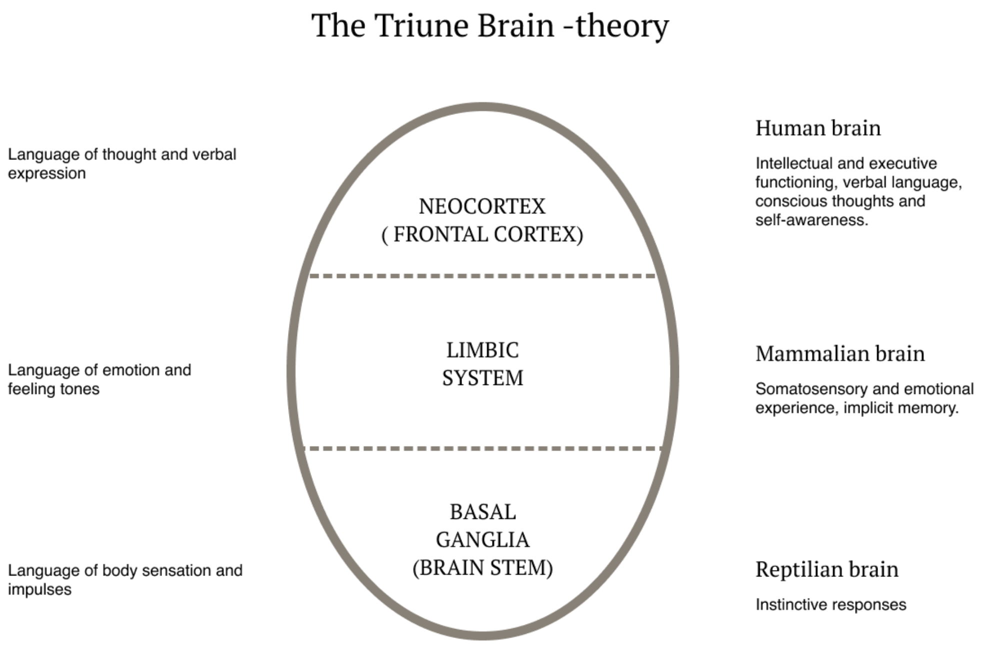

Figure 4: The triune brain theory by Paul Mclean [110]. This is not anymore scientifically valid [91], but is nonetheless a helpful and clarifying concept.

book: "technology should bring more to our lives than the improved performance of tasks: it should be richness and enjoyment [125, p. 101]."

The first level, **visceral**, is the innate level. These are the emotions we are born with. Visceral design deals very much with the appearance of the product. How it looks, how it feels and what sets it apart from other product at a first glance. It could be said that this level deals with the "superficial", tapping into the user's attitudes, beliefs and feelings. A lot of the current design goes into this level, for example in the form of branding.

The second level is **behavioural**. These are the emotions that affect and rise from our behaviour and our expectations. This is how we expect our product to work and help us to perform the tasks in the world we want. This is very much related to usability (if not the same), and can be tested in the same way. Did I correctly do this task? How long it took me to complete this task? Behavioural level doesn't explicitly deal with emotions *per se*, but with the consequences: if a product doesn't work as I expected it to work, I'm most likely to experience a negative feeling.

The third level is the **reflective**. This is the intellectual and rationalising level. This is the level where we ask ourselves: does this product tell a story? Does this thing have any meaning for me? If the visceral level dealt with the superficial - "I love that watch, because it is so beautiful!" - the reflective level takes it further: "what do my friends think about me wearing this watch?" Or "is this expensive watch something that I could later give as a gift to my son later in life?" In a sense, this level too doesn't deal with emotions in itself, but it can affect them especially by

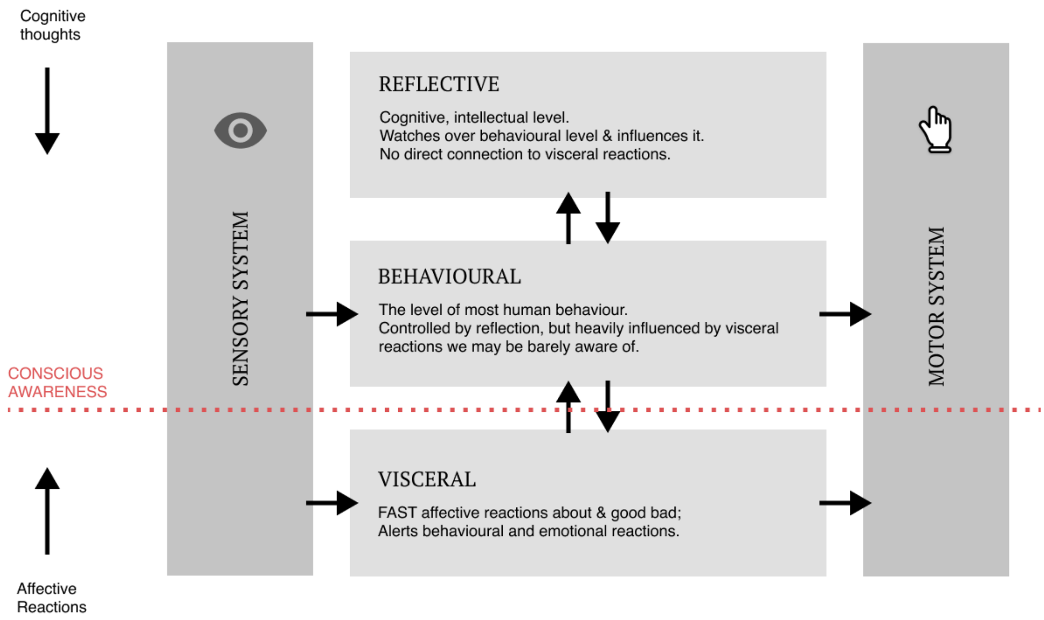

Figure 5: The three levels of emotional processing. Adopted from [41], which in turn was re-adopted from [125].

refining the primary reactions: "well, that watch is beautiful, but would I really enjoy using it for a longer time?"

2.3.2 Different modes

However, the same emotional reaction does not always mean the same perceived feeling. It is as important to understand not only the user, but their activity. Hence it is important to also understand different user modes, that Hassenzahl *et al.* [67][70] have proposed.

Depending on the context, the user has different kinds of needs and aspirations that need to be fulfilled: maybe a certain goal or a need for mental stimulation. As these different situations can be very diverse, they pose a problem for our prediction. For this, Hassenzahl & et al. suggest to focus to defining the *usage modes* [67]. Specifically, they distinguish that software usually has two qualities: a "pragmatic" one and a "hedonistic" one [70]. Respectively, these two qualities are used in **goal mode** and **action mode**. The names can be little bit misleading as usage always has goals. However, in the goal mode the goal is in the forefront. The software has a utility, a function, that it needs to fulfill. The software is used to accomplish a certain goal. An example would be a spreadsheet or a calendar software. In action mode it is the opposite, where using the product can be an end to itself and goals are shifting constantly. An example would be a computer game.

What is important for our thesis is that in these two modes users experience emotions differently. In action mode, high arousal could be experienced as pleasant, and the users would describe their state as "excited". In goal mode, the opposite could be true and high arousal would be experienced as mounting anxiety. In goal mode, low arousal would be preferred and experienced as relaxation - in action mode, low arousal would lead to boredom [67].

2.3.3 How to measure the experience?

Measuring the actual experience is not out of our hands. Humans can empathize with each other by the means of observation and verbal communication. Thus, traditional methods of measuring the experience of the user include for example subjective interviews and observational studies [140]. Some modern methods such as bodystorming put the designer to the user's place [129] by playing out the scenarios. However, all these still have the problem of subjectivity, and the intersubjectivity and validity of these approaches is hard to guarantee [51]. Also, even if the designer would make the final interpretation from this research material, information might be lost through faulty interpretations and erroneous readings. In worst case data could be used to enforce subjective prejudices. There are more objective and repeatable observational techniques such as video analysis, but the evaluation of the material is a very long and laborious process. As part of a larger trend in understanding the user experience [67] more tools are becoming available to evaluate it, and some of them utilize psychophysiological methods [51].

An added challenge with measuring the experience is the personality of it. Users are different and react differently. Personality has been shown to modulate and affect at least preferences between applications [4], learning habbits [90], immersion [181], social network site behaviour [5], satisfaction with software usage [38], posture and gestures displayed during computer usage [7] and website choices [94].

To conclude, we can say that the measurement of experience in HCI research is a challenging and complex issue. Theoretical frameworks have been laid out and practical methods for measurements presented. Yet, a systematic discussion of their effectiveness and applicability in different contexts remains lacking [67]. Affect (or emotion) would be an interesting and promising field of research for UX, as it forms such a crucial part of any experience. However, the same issues still remain here: although a number of techniques for measuring affect have been developed, a systematic discussion of their effectiveness and applicability in different contexts is missing. As computing shifts to increasingly collaborative and ubiquitous models, it is important to discuss affect measurement that happens beyond the individual level [160] and in naturalistic settings [51]. To give suggestions on how to work with these problems, we will present in the next chapter the possibilities of affect measurement with EDA and HRV and show how they have already been applied in UX research.

2.4 Measuring emotions from the wrist

As we have laid out, emotions and the body are very much intertwined, with emotions causing changes in our body and our bodily changes arousing emotions. What this also means is that we can measure those changes. And with the new advancements of technology, we can do this quite comfortably and unobtrusively from the wrist. In this thesis, we focus on two different metrics that can be measured from a single wrist-worn device. Our first point of interest is the skin and its continuous variation in its electrical properties and how it changes with time and context, called the electrodermal activity (EDA). The second point of interest is the heart, the traditional symbol and home of all emotions, whose beats and their changes in time and frequency we are interested in, also called the heart rate variability (HRV). Both of these phenomena relate to different activations of autonomic nervous system (ANS) and the relative balance of those activations in the two subsystems of ANS: the sympathetic nervous system (SNS), which is responsible for fight-or-flight behaviours, and parasympathetic nervous system (PNS), which is responsible for rest-and-digest behaviours [109].

ANS is a largely unconscious controlling and monitoring system for multiple bodily functions [83]. As mentioned, ANS consists of SNS and PNS, but it also has a third component, so called enteric nervous system (ENS), which controls the gastrointestinal tract and can act independently (thus sometimes dubbed "the second brain" [53]). Sensory nervous system is also often included in any discussion about ANS, because its inputs can change the autonomic tone [109]. The best (and very helpful) characterization of the two subsystems is that the SNS acts as a quick response mobilizing system and the PNS as a more slowly activated dampening system. However, even these concepts do not hold up in all cases since there are clear examples of where the two systems work together to carry out physiological functions [109]. As said, we are interested in ANS, because it is thought to be linked to emotional activity [95]. Research has suggested that ANS activity can distinguish between emotions [29][44][107][161] and some go so far as to suggest that emotion is nothing but an evolutionary by-product of the neural regulation happening in the ANS [134]. However, conclusions are still far from clear: it can be said with some certainty that emotions are closely tied to ANS activity and vice versa, but further evidence is needed and methodological obstacles need to be overcome before empirical studies can adequately test theories and resolve controversies [95][106].

2.4.1 Electrodermal activity

In one simple sentence it could be said, that electrodermal activity (EDA) is a phenomenon where skin's conductivity increases due to external or internal stimuli that is physiologically or emotionally arousing [37, p. 201]. But the actual mechanism is a little bit more complicated.

The skin's electrical resistance varies with the state of sweat glands in the skin. The skin has two kinds of sweat glands: apocrine and ecrine, which are classified mainly by their location and function. Apocrine sweat glands open into hair follicles and they are thus located mainly in armpits and genital areas, where as eccrine sweat glands are found almost all over the body. Both are linked to emotions and stress, but the eccrine sweat glands are the ones that are relevant for EDA. The main function of eccrine sweat glands is regulating body's temperature and maintaining a good grip in the soles and palms. It has been suggested that these palmar and plantar sweat glands are more responsive to emotional stimuli [37, p. 202]. As the palmar glands are innervated by SNS, their sweating is controlled by the SNS. Thus if SNS is highly aroused, the followed liberation of acetylcholin causes rapid fluctuations in the eccrine sweat gland, which in turn increases skin conductance [14]. Thus, in this way, we can say that skin conductance can be an indicator of emotional and sympathetic responses [37, p. 203].

The actual neural mechanisms and pathways involved in the control of EDA are numerous and complex [37, p. 203], but the sympathetic activation can originate at least from the following brain areas: the limbic system (which supports a variety of functions including emotion, behavior and motivation) [154], hypothalamus (which regulates many bodily processes by controlling hormone release) [154], the premotor cortex (which helps in planning and guiding movement) [154] and the brain stem (which connects the brain to the rest of the nervous system and helps in a wide variety of roles, such as the regulation of cardiac and respiratory function) [143]. Given these brain areas, the following functional roles of EDA have been suggested:

1. amygdala activated affectional processes [154], such as fear [77]
2. thermoregulatory processes [154]
3. situations requiring fine motor control [43]
4. situations requiring orienting and attention [43]

However, in practical applications it is usually enough to simplify and say that EDA relates to psychophysiological arousal. As discussed in the earlier chapter,

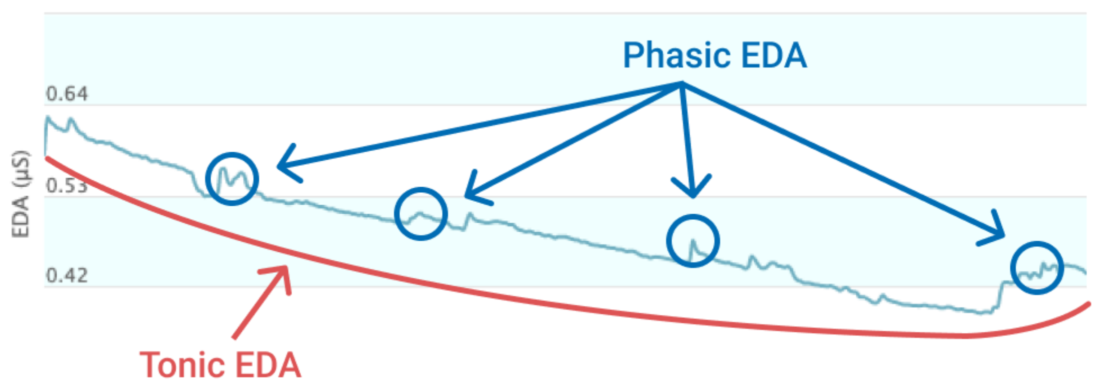

Figure 6: This graph depicts an EDA signal with a slowly ascending tonic component and a few phasic components.

arousal is thought to be one main axis in dimensional emotion theories. Arousal does not equal emotion though, but is a component of emotion which can be experienced in different ways as discussed in the usage modes. Research confirms that EDA is highly correlated with self-reported emotional arousal [100]. But what makes EDA truly interesting is that majority of people can not control these reactions consciously [16, p. 3].

The EDA can be divided into two main components, the **tonic component** and the **phasic component**. Generally speaking, the tonic component refers to the slow changes in the signal. It is the baselevel of the signal or DC-component in signal processing terms. Phasic component refers to the rapid changes or peaks in the signal. For a visual clarification, please refer to figure 6. One peak in the phasic component is usually referred to as a skin conductance reaction or an SCR (for a visual definition, see figure 7). Both tonic and phasic levels are important and related to arousal [16, p. 4], but tonic signal might be harder to analyse because it varies within and between individuals thus making any comparison or normalisation of values hard [16, p. 5]. Furthermore, EDA is phenomenon that can be caused by many different activations in our nervous system, thus any measured SCR can also be event-related (ER-SCR) or non-specific (NS-SCR) and this needs to be taken into account and balanced in any experiments [16]. There are personal differences that affect EDA readings, such as the thickness of the skin, which affects the maximum and minimun SCL and SCR levels [37]. As EDA has a quite consistent re-testability, some have suggested that EDA would be a relatively stable subject trait related to behavioural and psychological individual differences [37, p. 210].

One suggested measure of skin conductance in *A Guide for Analysing Electroder-*

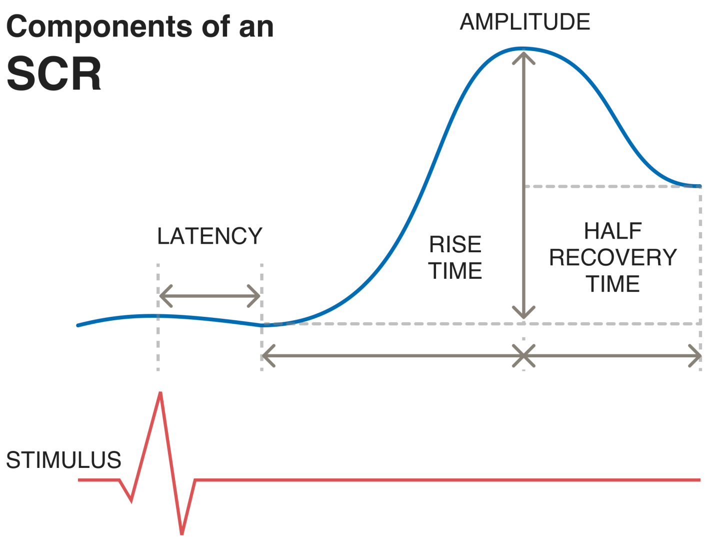

Figure 7: The components of an event-related SCR. Adopted from [37].

mal Activity (EDA) Skin Conductance Responses (SCRs) for Psychological Experiments [16] is the frequency of NS-SCRs (typically 1-3 per/min during rest and over 20 per/min in high arousal situations). These components are more straightforward to compute - though then it is important to use a fixed time period of measurement when concentrating on the frequency of the peaks. This was also the chosen measurement in Antonio Damasio's somatic marker hypothesis experiment [33].

2.4.2 Heart rate variability

Heart rate variability (HRV) is the physiological phenomenon of the variation in time between heartbeats. Heart rate variability is not the same as heart rate even though they can be measured with similar sensors. An average heartbeat, say 80 beats-per-minute, can have a hugely different heart rate variability and a healthy heart displays these complex patterns [159]. HRV as a phenomena is an emergent property of various interdependent bodily regulatory systems, which work on different time scales and help us to adapt to our environmental and physiological challenges [115]. It is generated by heart-brain interactions and the activation cycles in ANS [158]. For example, activation in the sympathetic nervous system (SNS) raises the heart rate and activation in the parasympathetic nervous system (PNS) lowers the heart rate [65]. High variability in non-linear biological systems provides the flexibility to deal with uncertain and constantly changing environments [177]. Hence, a high HRV is also usually considered to be a good thing, a sign of a healthy heart [9]. HRV has been linked to many physiological reactions [152], health conditions [115] and it is also interesting psychologically as it has been found to be related to emotional processing and other cognitive states [3][13][115][135].

HRV data is often also called interbeat intervals (IBI) and its measurement units are milliseconds. These interbeat intervals are calculated as the time distance between each R-peak in the QRS complex of the ECG wave. Hence, the term RR interval is also found in literature. In this thesis, we will most often refer to normal-to-normal (NN) intervals, which also suggests that the measured heartbeats are normal. For a visualisation of a normal heartbeat and the variability, please refer to figure 8. In this thesis, the time-domain HRV data is analyzed to give the following variables:

SDNN The standard deviation of NN intervals.

$$SDNN := \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(NN_i - \overline{NN})^2} \tag{1}$$

Both SNS and PNS activity contribute to the SDNN and it has been find to be highly correlated with ULF, VFL and LF band power (see table 2 for their definitions), whose respective relationships then depend on the measurement conditions. It is usually calculated between a longer time period such as 24 hours, with the conventional short-term recording standard being 5 minutes [26]. Ulta-short measurements such as 60 s to 240 s are not unheard of, but they change what contributes to this metric [158]. Longer recordings gather a wider variety of phenomena of reactions to environmental changes, such as changed workloads, anticipatory nervous system activity and circadian rhythms, in addition to the cardiorespiratory regulation [56]. In short term measurements, the main contribution comes from PNS-meditated respiratory sinus arythmia (RSA) [159]: simply put, it means that the heart rate increases when breathing in and decreases when breathing out.

SDSD The standard deviation of successive differences.

$$SDSD := \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}((NN_{i+1} - NN_i) - \overline{NN_{diff}})^2} \tag{2}$$

This metric is similar to SDNN, but only represents short-term variability [97]. One study with ultra short term measurements found it correlated with HF and RMSSD [147].

RMSSD The root mean square of successive differences.

$$RMSSD := \sqrt{\frac{1}{n-1}\sum_{i=1}^{n-1}(NN_{i+1} - NN_i)^2} \tag{3}$$

RMSSD reflects the variance from beat-to-beat and is primarily used to measure the changes in HRV mediated by the vagus nerve [159]. Vagus nerve is the main component of PNS and it mediates the actions of heart, lungs and digestive tract and

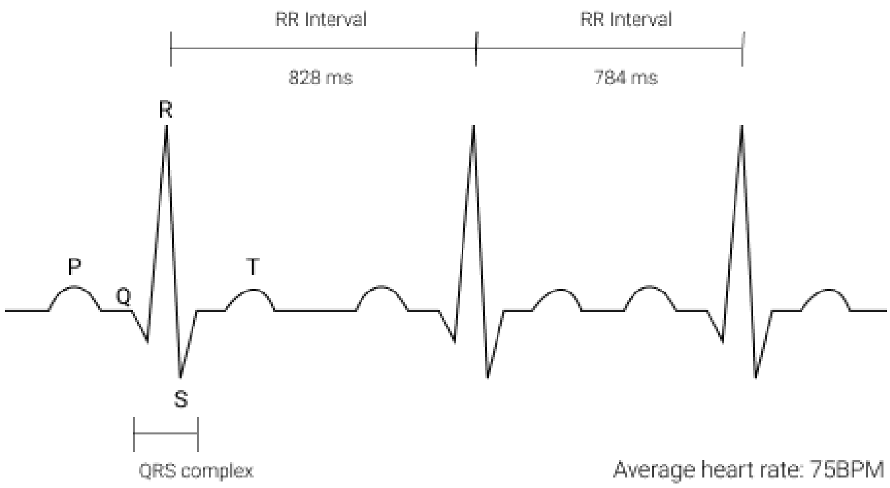

Figure 8: Schematic representation of heart rate variability. The same average heartbeat can have huge variance in HRV.

provides sensory information about them [11]. Vagus nerve is partially responsible for the PNS activation of the heart muscle, which controls the lowering of the heart rate. The vagal tone is correlated with capacity to regulate stress responses and it is linked to mental and physical health [17].

RMSSD as a metric is correlated with HF power and it is more influeced by PNS activity than SDNN.

NN50 and pNN50 The number of successive differences that differ by more than 50 ms and their proportion to all intervals.

$$pNN50 \coloneqq P(|RR_{i+1} - RR_i| > 50\,\mathrm{ms}) \tag{4}$$

The pNN50 is closely correlated with PNS activity [175]. It is also highly correlated with HF power and RMSSD [92].

NN20 and pNN20 The number of successive differences that differ by more than 20 ms and their proportion to all intervals.

$$pNN20 \coloneqq P(|RR_{i+1} - RR_i| > 20\,\mathrm{ms}) \tag{5}$$

The pNN20 is a relatively new marker, that is closely related to pNN50. Some newer research suggests it would be more sensitive to vagal function [6] and better at distinguishing certain pathological conditions [118] than pNN50, but some disagree and say that it would be equivalent to pNN50 [78].

Other time domain metrics There are other time-domain measurements such as SDANN (the standard deviation of the average normal-to-normal interbeat interval) and SDRR (standard deviation of all beats) [158], but these are not used in this thesis.

Frequency domain HRV can also be analyzed in frequency domain. Frequency domain data is divided into different components (or bins). Commonly the spectrum is divided into four components: the four most commonly found are high-frequency (HF) band (0.15-0.40 Hz), low frequency (LF) band (0.04-0.07 Hz) and very low frequency (VLF) band (0-0.04 Hz) [158]. A fifth component is a novel approach, where the HF band is split into another component called the mid frequency (MF) band (0.07-0.15 Hz). Furthermore, and ultra low frequency (ULF) band is sometimes distinguished from VLF [159]. However, even though the naming is quite standard across literature, different divisions of the components do exist between researchers due to different signal processing methods, methodological choices and advancements in technology. The definitions of the different frequency components and their linked phenomena can be seen in table 2.

Especially HF and MF components seem to be related to emotional arousal and cognitive load, probably due to them being an indicator of PNS activity. HF component's amplitude has been linked to state anxiety [84]. MF component's (and

Frequency	Related phenomena
ULF - Ultra low (0.003 Hz)	Uncertain [158]
VLF - Very low (0-0.04 Hz)	Uncertain [158]
LF - Low (0.04-0.15 Hz)	Baroreceptor activity [115], thermoregulatory fluctuations of blood vessels [98], SNS and PNS activity [158], deep breathing, sighs [20]
MF - Mid (0.07-0.15 Hz)	Mental load [72], postural changes [72], emotional strain [122]
HF - High (0.15-0.45 Hz)	Respiration [72], PNS activity [12], anxiety [84], RSA [63]

Table 2: The different spectral components or bins of HRV and their related phenomena.

specifically 0.1 Hz frequency's) decrease has been found to decrease under time pressure [122] and emotional strain [122]. However, the same authors advice against using HRV as a measure of mental and especially cognitive workload, particularly where system safety or occupational risks are at stake [122].

In contrary, Károly Hercegfi from the University of Budapest has later published studies that support the usefulness of MF component to measure mental load [72]. This is interesting as the research group has also developed this methodology into a whole system called the INTERFACE, which they claim is capable of identifying the relatively weak points of the HCI in naturalistic settings. This is also a purpose of our research and we thus find the investigation of HRV and MF components interesting although the conclusions from studies are definitely not clear. Lin *et al.* [108] have also found significant correlations between HRV and mental effort. They did spectral analysis on the HRV data, dividing the frequencies to LF (0.04-0.15 Hz) and HF (0.15-0.45 Hz) and found that the low frequency correlated with mental effort.

Non-linear measurements As discussed earlier, HRV is a non-linear phenomenon. Hence, it would also make sense to measure HRV with non-linear metrics. Non-linearity means that the relationship between variables can not be plotted as a straight line. Non-linear measurements measure the complexity and unpredictability of a time-series, which arises from the interactions of highly non-linear and dynamic systems regulating the HRV. In this thesis, we are interested in the following non-linear metrics: Poincaré plot, S, SD1, SD2, SD1/SD2-ratio and sample entropy.

Poincaré plot A Poincaré plot (see figure 9) is a plot, where every NN interval is plotted against the previous NN interval, thus creating a scatter plot. This plot then allows reseachers to visually search for patterns buried within time series. Poincaré plot can be analysed by fitting an ellipse around the plotted points. After fitting this ellipse, three different non-linear measurements can be derived: S, SD1 and SD2.

S S is the area of the fitted ellipse and it represents the total HRV. It correlates with baroreflex sensitivity (shortened BRS, baroreflex is a mechanism to regulate

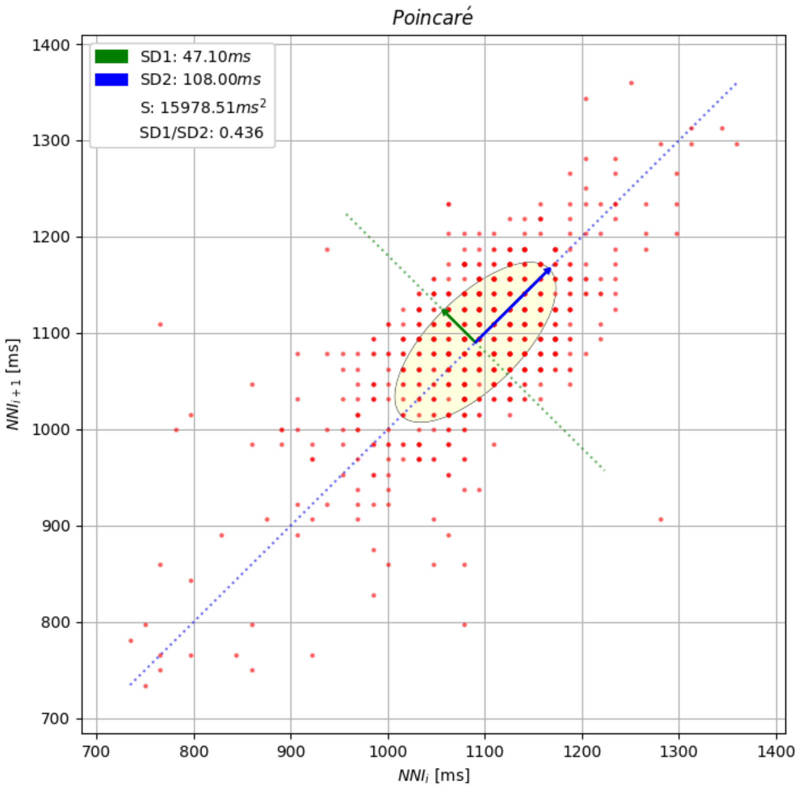

Figure 9: An example Poincaré plot from one of our experiments.

acute blood pressure changes), LF and HF power, and RMSSD [158].

SD1 and SD2 Both SD-metrics are standard deviations (hence the name). SD1 is the standard deviation of the distance of each point from $y = x$ axis, which specifies the width of the ellipse. SD1 measures short-term HRV and it too correlates with BRS and HF power. SD1 is also an identical metric to RMSSD [158]. SD2 is the standard deviation of each point from the $y = x + \overline{NN}$ axis, which specifies the ellipse's length. SD2 measures short- and long-term HRV and correlates with LF power and BRS [18][19][174].

SD1/SD2-ratio The SD1/SD2 ratio, which measures the unpredictability of the NN time series, is used to measure autonomic balance when the monitoring period is sufficiently long and there is sympathetic activation [158]. SD1/SD2 is (negatively) correlated with the LF/HF ratio [10][66].

Sample entropy Sample entropy measures the complexity and irregularity of a time-series [139]. The higher the value is, the more complex and less regular and thus less predictable the signal is. An advantage of sample entropy is that it can be calculated from less than 200 values [97]. As per mathematical definition, sample entropy is the negative logarithm of the probability that if two sequences of length m remain similar (i.e. have a distance less than $< r$) also at the next point $m+1$.

Thus if we have a time series data set of length $N = x_1, x_2, x_3, ..., x_N$, we would define a template vector of length m so that $X_m(i) = x_i, x_{i+1}, x_{i+2}, ..., x_{i+m-1}$ and a distance function $d[X_m(i), X_m(j)], i \neq j$, so then we could define sample entropy as:

$$SampEn = -\log \frac{A}{B} \tag{6}$$

Where A describes the number of template vector pairs which have $d[X_{m+1}(i), X_{m+1}(j)] < r$ and B describes the number of template vector pairs which have $d[X_m(i), X_m(j)] < r$. Usually, we define $m = 2$ and $r = 0.2 * std$, which is what we use in this thesis as well. We could also say that sample entropy measures how much the value of a given data point in the signal $(m+1)$ depends on the previous values (m) of that signal, averaged over the whole signal.

2.5 Psychophysiological metrics in UX

As we discussed in the earlier chapters, the traditional methods to access the experience of the user (such as subjective reports or surveys) are limited. The participants have to be inquired about their experiences and emotions, which in turn interrupts the process of the experience and destroys their flow. Psychophysiological methods offer a possibility to gather data throughout the whole of the experience, which is why their possibilites have not went unseen by HCI researchers. Lately, psychophysiological recordings have been shown to be valuable approaches for measuring valence (a quality for positive and negative emotions) and arousal throughout the process of an experience, thus unfolding whole new possibilities for UX evaluation [51]. Though these methods are numerous, in this thesis, we will only focus on EDA and SCR, because they are our chosen metrics as well.

As *A Guide for Analysing Electrodermal Activity (EDA) Skin Conductance Responses (SCRs) for Psychological Experiments* well states, the EDA is the most useful index of changes in sympathetic arousal [16]. No wonder, that Google Scholar search for "eda" AND "human computer interaction" yields 3690 results. The same is true for HRV: a Google Scholar search for "heart rate variability" AND "human computer interaction" yields 4410 results. We will not go through of all of them, but rather want to point out that the research is wide and on-going. However, as a point of interest we will introduce a few which are particularly relevant for this thesis.

- Research done with EDA:
    1. Cognitive load in HCI correlates with EDA [157]
    2. EDA as a measure of arousal [178][114]
- Research done with HRV:
    1. HRV as a measure of valence [2]

2. MF component as an indicator of mental effort/weak points in software [72]
3. HRV differentiating between rest, engagement and stress [111]

- Research commbining EDA and HRV:

  1. 2D-estimation of emotions with neural networks [184]

Needless to say, psychophysiological methods in UX can help and enhance research and understanding. The why's and how's have been well written down in a review article by Ganglbauer *et al.* [51] in their article *Applying Psychophysiological Methods for Measuring User Experience*. Rather complementary than antagonistic to qualitative subjective methods, psychophysiology can support the analysis of certain crucial situations of an experience that are essential to a development of a good product. Additionally, psychophysiological measurements can provide help uncovering social masking, such as saying positive things because they hesitate to be negative. Furthermore, it's possible to analyse data to look for special situations during the evaluation. Psychophysiological methods can thus alleviate the problems with subjective reports, which are prone to the fact that emotions are not always easy to put into concrete words and details are sometimes forgotten. When it comes to applying psychophysiologial measurements, the same researchers argued that these methods should be applied multimodally, not unimodally. A multi-modal approach is thought to be more accurate and gives a broader variety of results. For example, correlating two simultaneous metrics for valence can validate the results better. The disadvantage is though then the complexity of combining the multiple channels, analysing them and then finally interpreting them. In addition, even though they acknowledged the problems of subjective reports, they argue that the psychophysiological methods should be supplemented by subjective reports to assess the truthfulness and meaning of those measurements.

3 Research material and methods

The aim for this thesis is to help building a proprietary user research system for the case company. The project of building this system consists of two different thesis projects, from which this is the first one. This thesis is about basic research of emotion arising in HCI. The background part introduced a wider perspective of emotion and why it is important for human-computer interaction. In this research part, we utilize explorative data analysis methods in order to see and understand what we are measuring. We want to study the construct validity of psychophysiological HCI methods: i.e. what are the most viable metrics for our case and do the methods measure what they claim to measure? This research experiment consists of two parts: the interview part with two psychological questionnaires and the experimental part with five different tasks.

The motivation for the interview part is to build a quantitative model of the affectivity of the subjects. This has two aims: first, to study their mutual correlations. Second, to study the correlations between these psychological constructs and actual measured physiological reactions. Respectively, the motivation for the experimental part is to design experiments that would simulate well the different categories of situations our users might experience during their computer usage on a normal day in order to gather valid ground truth data. From this data we could then analyse similarities and differences between different tasks and people.

Building the emotional user model for each of the test participants was done in two parts: first with an open interview, where we asked the people about their computer usage at work and at home and what kind of activities give them pleasure and what kind of things annoy them. The second part consisted of two psychological questionnaires: to measure their emotional reactivity we utilized the Affect Intensity Measure -questionnaire [103] and to measure whether they prefer to use the rational or the experiential system the participants filled the Rational-Experiential Inventory-questionnaire, which measures their preferred information processing style [130].

After this the 10 participants were divided randomly into two groups: a test group and a control group. Both did the same experiments, but only the test group was asked after each test about how they felt. There were five experiments in total: four plus one. One for each respective emotional category in the two-dimensional valence-arousal space and one that would only measure cognitive load. The rationale for the control group is that we wanted to make sure that the interviews between tasks wouldn't affect too much the experience or the psychophysiological readings.

3.1 Psychological questionnaires

For psychological questionnaires we wanted a construct which would focus mostly on emotions. Thus many larger and more complex constructs, such as the Big Five or Myers-Briggs Type Indicator, were ruled out immediately. It would be much more simpler - and perhaps valid - to focus only on affect, which even alone is an important individual difference variable [101]. After studying the literature we settled out on survey called Affect Intensity Measure (AIM) developed by Larsen [102] as a measure of a theoretical construct referred to as affect intensity, or the characteristic strength with which people experience their emotions.

AIM consist of 40 different affirmations, such as "when I accomplish something difficult I feel delighted or elated.". Each of these are then rated on a scale of 1 to 6. 1 means "never" and 6 means "always". The subject must answer so that it reflects their own experience. AIM has gathered some attention in research and later research has confirmed that it has good reliability and validity [144]. However, the division of AIM is still debated and at least two different subscales are suggested in addition to the original unimodal scale of Larsen's. Based on exploratory principal component analysis, Weinfurt *et al.* [182] proposed a four factor model with a mediocre fit:

- Positive Affectivity, which measures the intensity of positive emotions
- Negative Intensity, which measures the intensity of experienced negative emotions
- Negative Reactivity, which measures the intensity of reactions to negative emotions
- Serenity, which measures a preference for low-arousal or the rejection of high-intensity positive affections

Another suggestion was developed by Bryant *et al.* [22] with an a priori theory that affect would be experienced with a valence (Positive and Negative) and that this valence would have an experiential component (Intensity) and a behavioural component (Reactivity). In the same paper though they combined the Positive Reactivity and Positive Intensity scales as they were highly correlated. Theoretical explanation was that positive emotions are usually both experienced and displayed with similar intensity due to their social acceptance where as negative emotions are not (your mileage may wary due to your culture). Similarly, only a subset of 27 questions were used, because they explained most of the variance and because some quesions were confusing (reversed). Their model is called Affect Intensity and Reactivity (AIR), and it consists of similarly named three subscales: Positive Affectivity, Negative Reactivity and Negative Intensity. In this model, Serenity is dropped. Although the model's subscales have similar names and they measure the same underlying factor, they are slightly different because AIR doesn't include some of the questions that Weinfurt's 4-factor model does. Interestingly, neither of the papers claim superiority over the other but rather suggested further investigations to the external validity of the construct. Both agree though that if one's objective is to test hypotheses about the different facets of affective experience then a multidimensional model is preferred to the unidimensional one. Thus, in this thesis we shall follow the suggestion of Rubin *et al* [144] to present subscales data from all constructs and not narrow down our observations too early.

In addition, as we have so many times already in this thesis pointed out the intertwining of emotion and experiential cognition, we wanted to tie in this part as well. Thus, we chose also the Rational-Experiential Inventory (REI) survey. It distinguishes whether the subject relies more on rationality or experientiality when making decisions [49]. Seymour Epstein, who developed CEST-theory, was one of the developed of REI4. REI, too, consist of 40 affirmations, such as "I like to rely on my intuitive impressions" and "I have a logical mind", which are rated on a scale of 1 (completely false) to 5 (completely true). It has a good internal consistency and some proof of construct validity outside of Big Five [151].

The rational style is measured with an adapted Need for Cognition (NFC) scale [24] and it emphasizes a conscious and analytical approach. This scale is further divided into to two sub-scales: first, rational ability (RA), which measures the ability to think logically and analytically. Second, rational engagement (RE), which measures the reliance on and enjoyment of thinking in analytical, logical manner.

The experiential style is measured by the Faith in Intuition (FI) scale and it emphasizes a pre-conscious, affective and holistic approach. This scale is also divided into two subscales: experiential ability (EA), which is the ability with respect to one's intuitive impressions and feelings, and experiential engagement (EE), which is the reliance on and enjoyment of feelings and intuitions in making decisions [130].

The Rational-Experiential Inventory is often referred to as REI-40 or REI-10 in literary, depending on whether the short of long form is used. The original measure contained 59 questions and is referred to as the REI-59. However, REI-40 and REI-10 are preferred now as they are considered more valid [130]. In addition, high scores in *both* rational and emotional scales at the same time are considered a marker of emotional intelligence and mental wellbeing [151].

3.2 The measuring device: Empatica E4 wristband

For this research we use Empatica E4, which is a wristband that can measure both EDA and HRV, while worn quite comfortably on the wrist. It has received medical appliance certificate and the manufacturer claims it can obtain clinical quality observation data from subjects [137]. Validation research by McCarthy *et al.* [108] confirms this. Empatica E4 compares very well, achieving similar data quality 85% of the time compared to more bulkier and more expensive laboratory research devices.

The Empatica E4 measures EDA by capturing electrical conductance (inverse of resistance) across the skin. This is achieved by passing a minuscule amount of current between two electrodes in contact with the skin [183](this is called the exosomatic method [16]). The data is measured in microSiemens (µS), the measurement for conductance, and the data is captured at a sample rate of 4 Hz [42].

HRV is calculated from data measured by an photoplethysmography (PPG) sensor, whose functioning principle is also presented in figure 10. A common name in the scientific literature is also blood volume pulse (BVP). The device has two LEDs, one green and one red. These are placed against the skin. Depending on how the blood is oxygenated, the green light is absorbed or reflected back. The more oxygen there is the blood, the more green light is being absorbed. Thus, during a heartbeat there is noticeable dip in the light reflected back. This PPG data is the input signal to the algorithm that detects the heart beats and that provides the IBI signal as output, which can be then used to calculate HRV. The red LED is used as

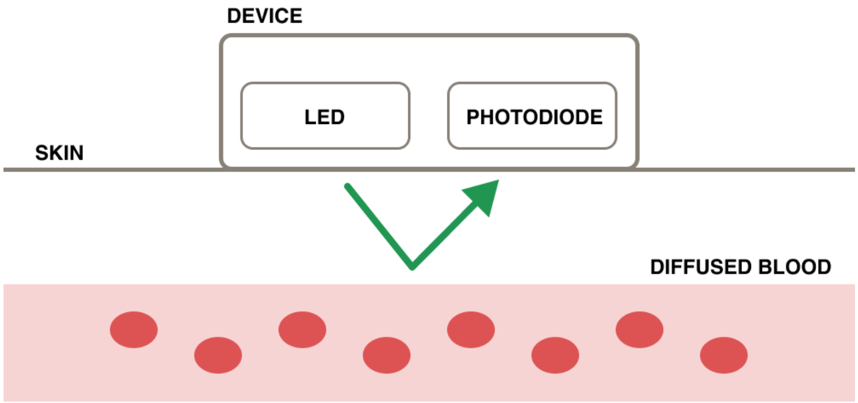

Figure 10: A figure of photoplethysmography. Some of the green light emitted by LED is absorbed by oxygenated red blood cells. The reflected light is then measured at the photodiode.

a reference light level to calculate out motion artifacts. [176]

Wrist measurements are not as good for measuring EDA as palm or fingers would be, nor is it as good for HRV as a ECG would be but it is more practical [45]. To ensure the quality of data Empatica E4's sensors are specifically designed to be more sensitive to measurements from the wrist [183]. The device also includes other sensors to achieve a better signal-to-noise ratio and to remove artifacts from the data. For example, there are a 3-axis acceleration meter and skin temperature meter, whose data can be used to remove artefacts caused by moving the wrist or increasing skin temperature.

3.3 The experiments

In our experiments we try to elicit a similar affective phenomena across multiple subjects. We do this in naturalistic settings (office of the case company) and we research the possibility of using personality as an individual difference variable that could be used later in future models of emotion. From literary research and from our initial workings with Empatica wristband, we've noted that we could possibly detect two different emotional phenomena with our device. We hypothesize that activity for arousal would come mainly from EDA and that valence could be deducted from HRV data. More specific, we hypothesise that arousal levels should correlate with the amount of phasic components of EDA. Tonic component is also of interest as it can show the general level of arousal over extended periods of time [178]. Some

	Positive valence	**Negative valence**
Low arousal	Relaxation	Boredom
High arousal	Excitement	Anxiety

Table 3: The possible combinations of our hypothesized 2-dimensional emotion model.

component of arousal level could possibly be read from HRV readings too.

The hypothesis for HRV being a measure of valence comes from literature that there are indicators in HRV frequency domain metrics that relate to mental load and anxiety. Those emotional states could be described as negative. The same goes for time-domain measurements, as some time-domain HRV data also seems to correlate with mental stress [141]. More specifically, an increase in mental stress should be read as a decreased PNS activity and an increased SNS activity in HRV metrics. Thus, in negative or stressful situations we should measure a decrease in measurements which are correlated with PNS, such as SDSD, RMSSD, pNN50 or SD1. We are proposing that with this data we could probably build a rudimentary 2-dimensional quantitative space of emotional experience. By analyzing Empatica's data we could possibly read experiences with either high or low arousal and positive and negative valence and then these two axes could be combined into four different measurable emotional reactions (see table 3).

We thus devised four different tasks, where each one tries to elicit one distinguishable reaction in this 2-dimensional emotion space. In addition we did one experiment which should elicit high cognitive load component as this is something that does not straight fit perfectly into any region in the 2D-model but is nonetheless interesting as literature says it should be measurable [157]. It would also be of great interest for the continued experiments as the software of case company is used for highly mission critical business cases which require attention and problem solving. Thus finding these moments of cognitive load and noticing any problems in the software would be of great help.

In total, five different tasks were conducted as within-subject variables. The subject group (N=10) was also divided into two subgroups of 5 each, where group 1 was asked after each experiment what kind of emotion did they experience to confirm if our hypothesis about the induced emotions were right. Group 2 acted as a control group where we did not ask the subjects about the tasks, as human interaction can cause major changes in physiological data and could thus introduce noise into our measurements. The subjects were chosen through a voluntary enrollment from the case company. The subjects were not controlled for age nor gender and almost all of them had a similarities in their educational background (IT and university degree or related) as all of them are employees at the case company.

All tasks were conducted at the office of the case company during a normal work day. We wanted to minimize the effect that daily work would have on our experiments and vice versa, so we scheduled the experiments so that subjects had enough free time before and after the experiments. Experiments were conducted either preferably

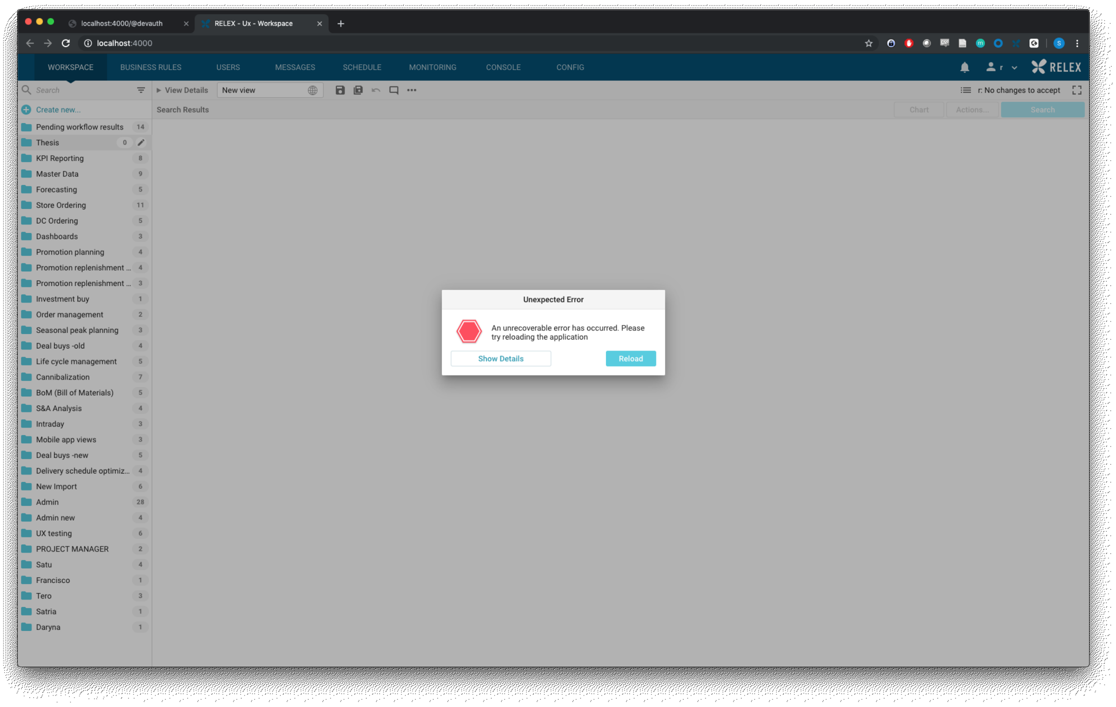

Figure 11: The error that pops up in task 1 when the user tries to save the view.

in the morning or afternoon but never just after lunch, so that general activity levels would be similar. However, we did not control for caffeine intake.

We utilized three different rooms, but they were mostly similar. In all of them, we used curtains to prevent other people or any other office life from distracting our subjects. Room temperature and humidity were kept at constant levels (within the limitations of the automated office air conditioning system). We as researchers also left the experiment room for the duration of the experiments. All the experiments were videotaped and the ones including software usage were also screen recorded so we could observe later what had happened in the room.

3.3.1 Task 1: High Arousal, Negative Valence

In task 1, we elicit an emotion with high arousal and negative valence, shortened HANV. This emotion could also be described as anxiety.

In this task, the subject is asked to use a software to build a certain database view. This is a task that is highly similar to our subjects normal day tasks. However, we've modified the software so that saving the work is impossible in this version (see figures 11 and 12). In addition, we have a 5 minute time limit on the task, which causes extra pressure on the subject.

We hypothesize that the pressure caused by the time limit should cause negative valence and the sudden pop-up error message should show arousal peaks in our data.

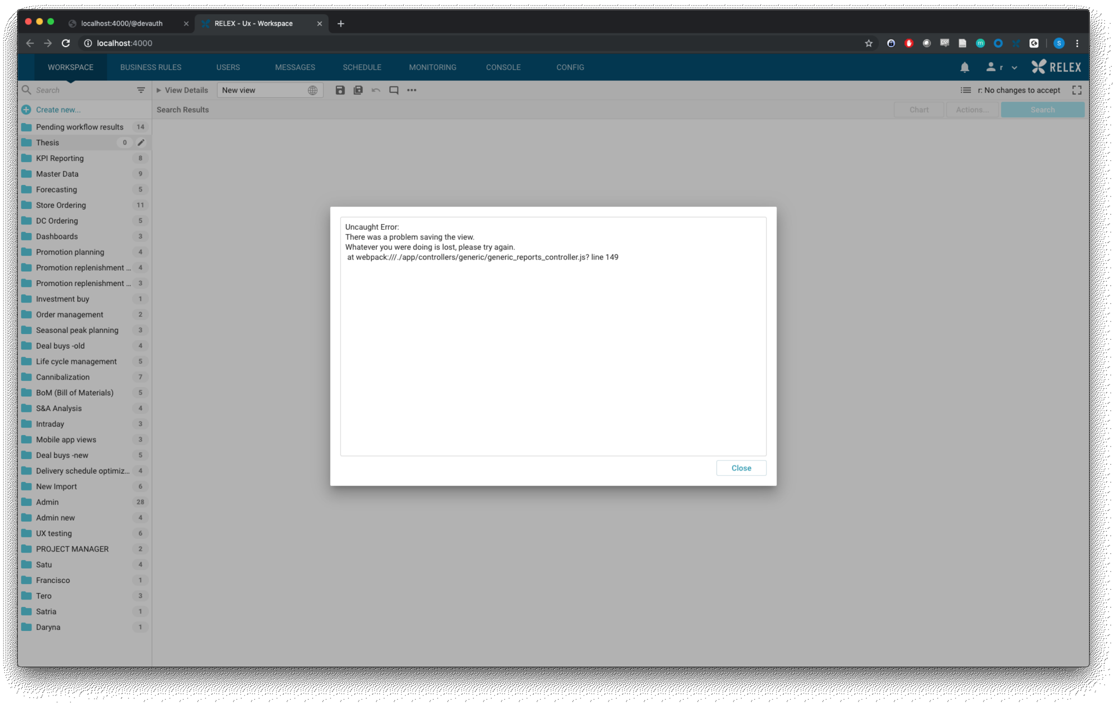

Figure 12: A further image of the error that pops up in task 1 when the user tries to save the view.

3.3.2 Task 2: Low Arousal, Negative Valence

In task 2, we elicit an emotion with low arousal and negative valence, shortened LANV. This emotion could also be described as boredom. In this task, the subject is asked to manually copy a column of 100 random numbers one by one to the nearby column in the spreadsheet (see figure 13). The usage of copy-paste or macros is forbidden. There is no time limit and the subjects are told that they can have breaks during the task to alleviate any time pressure.

This task should thus produce no mentionable arousal, but it should also not be perceived as a pleasant task either.

3.3.3 Task 3: High Arousal, Positive Valence

In task 3, we elicit an emotion with high arousal and positive valence, shortened HAPV. This emotion could also be described as excitement. In this task, the subject is asked to play a computer game named *agar.io*. The idea of the game is very simple: you control a cell with your mouse and you try to eat other cells. Eat more cells and you will get bigger and get more points, but don't get eaten by other players' cells. See figure 14 for a screenshot of the game play.

We only ask the participant to play the game, but we don't give any further advice. We hypothize that this task would elicit the action mode in the subjects, where they don't have any set goals that they're trying to achieve, but they find out by playing how the game works. Thus his game should elicit quite high arousal in

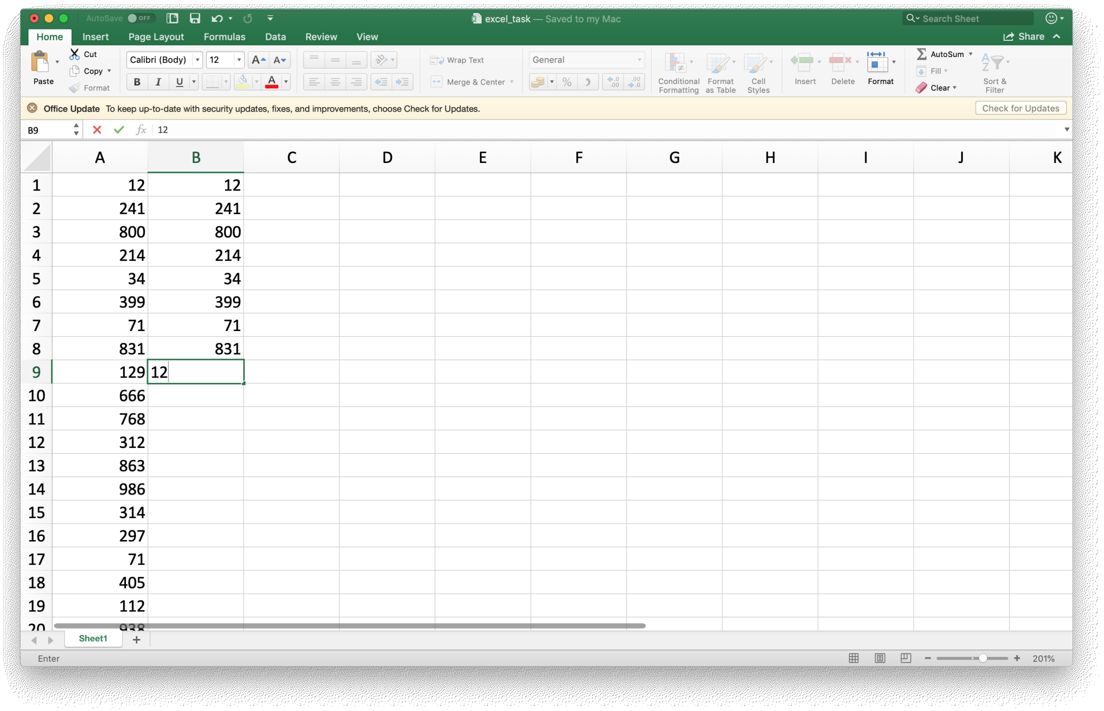

Figure 13: Task 2, the boredom task, where the subject is asked to manually copy the numbers to the next column.

subjects, due to learning a new game and because it requires constant attention and the game state changes quite often. As it is a game and played for fun, it should also be perceived as fun to play, so we should record an emotion with a positive valence.

The duration for this task was 10 minutes with no timer visible for the subject.

3.3.4 Task 4: Low Arousal, Positive Valence

In task 4, we elicit an emotion with low arousal and positive valence, shortened LAPV. This emotion could also be described as relaxation. In this task, we ask the participant to look at a video with aerial drone footage and with soothing music and nature sounds (see figure 15). The duration for this task was 10 minutes with no timer visible for the subject.

3.3.5 Task 5: Cognitive Load

In task 5, we don't elicit any particular emotion, but rather elicit high cognitive load on the participant (shortened CL).In this task the subject is asked to solve three visual puzzles with matchsticks. In this task, there are two different animal figures made with matchsticks and the subject has to move their direction with a certain amount of moves. Please see figure 16 for the tasks. The task requires visuospatial problem solving skills and is by no means trivial. This task should thus elicit high amount of cognitive load in the subjects.

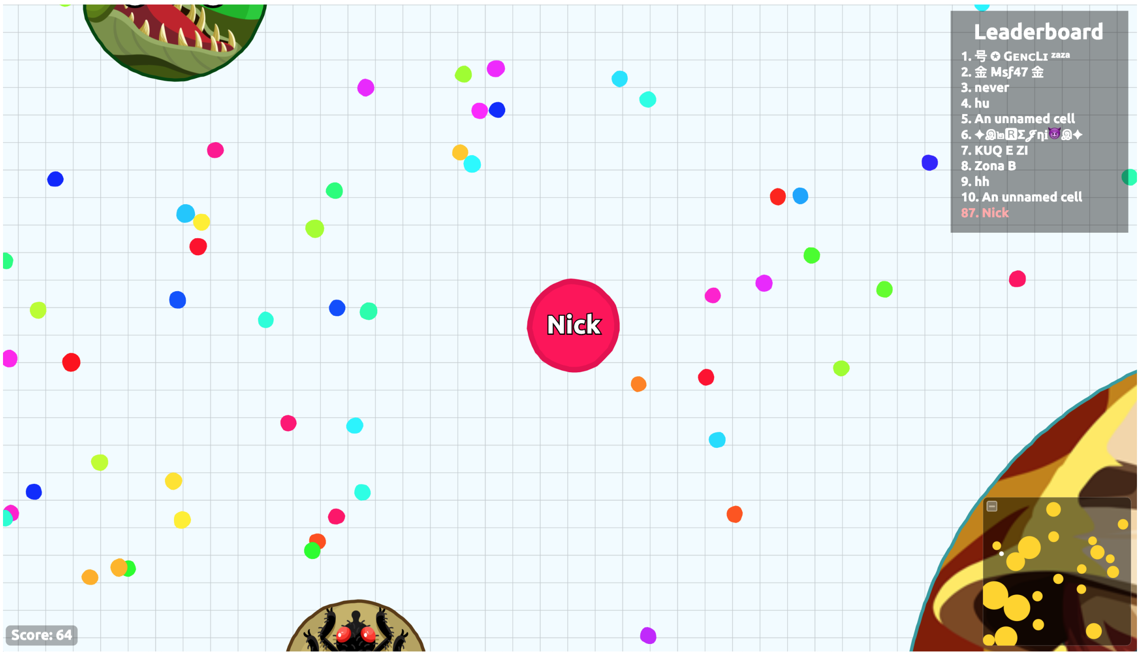

Figure 14: A screen caprure of *agar.io*, the web game that subjects will play in task 3.

For this task we had no time limit, but we were present for the duration of the experiment to answer any further questions by the subject and to change the matchsticks into the correct position for each subtask. The subjects were also given one hint, as we emphasized that the direction of matchsticks don't matter as both ends are considered similar.

3.4 Data analysis

Our data will be analysed in two parts. For EDA we will utilize automatic event detection made by MIT researchers [170] and for HRV we will use our own custom analyzer code.

3.4.1 Analyzing the EDA data

Even though the typical SCR has been well formalized, the detection of them from the data is not trivial. Real world data is noisy, even more so when we're dealing with naturalistic settings. Movement, surrounding electrical noise, or sudden displacement of the electrodes from the skin make the signal vulnerable to noise artefacts. Some of these artefacts can be mistaken for real SCRs. Another problem, is that in usual experiments like these, we need to look at hours of data. Going through these manually is a tedious task and could potentially subject us to even more errors. Luckily, for analyzing EDA, a group of researchers have developed web-based tool (hosted at `http://eda-explorer.media.mit.edu`) that automatically removes noise artefacts and detects events from your Empatica data.[170]

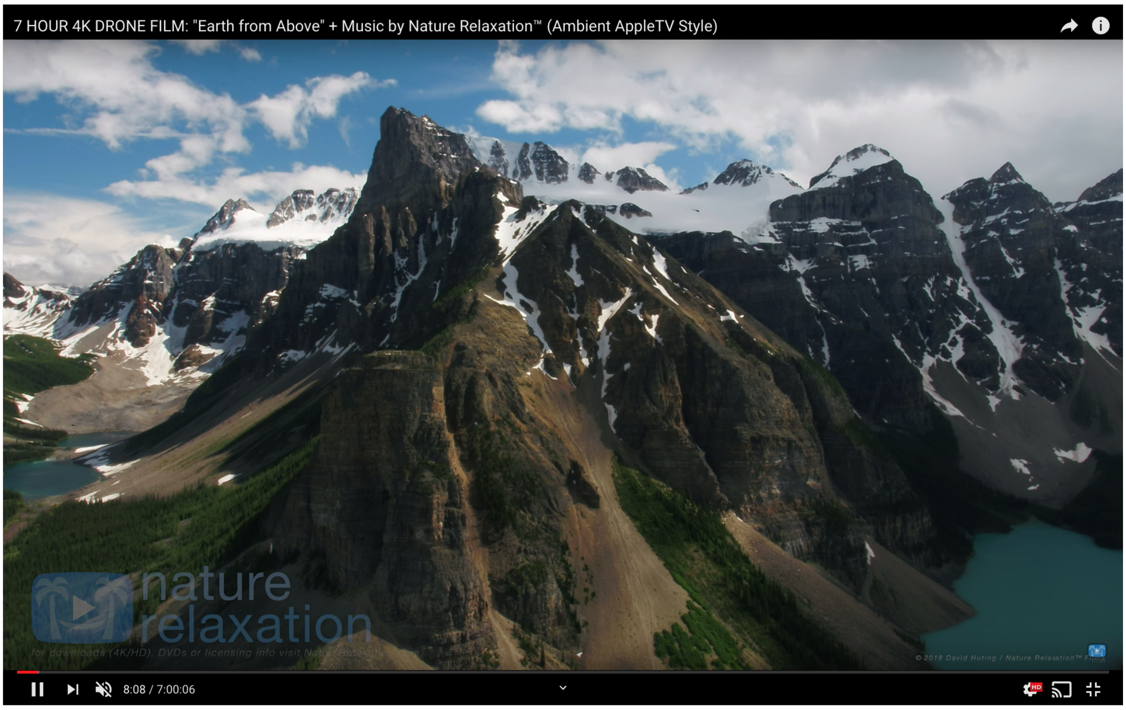

Figure 15: Task 4, a task with a relaxing nature video with soothing music in the background.

The automatic noise and event detection relies on machine learning. The algorithm uses support vector machines and the ground truth for the algorithm was acquired from two respected EDA researchers, who manually labeled true events from noise artefacts. Furthermore, the algorithm has been trained on a dataset gathered on similar device as ours, Empatica's predecessor wrist sensor, the Affectiva Q. For futher details and break down on the algorithm, interested readers are encouraged to read the full paper by Taylor, Jacques *et al.* [170].

To use the EDA Explorer classifier, one has to choose six different values: minimum amplitude, offset, filter frequency, filter order, maximum rise time and maximum decay time. Minimum amplitude is amplitude threshold above which a potential SCR must reach in order to be counted as an SCR. For this, we chose the value 0.01 µS, although historically 0.05 µS has been used as the minimum. This was due to the fact that it was the smallest possible shift that could be detected in paper recorders [16, p. 7]. With modern devices, such as the Empatica E4, much smaller values can be used and 0.01 µS is suggested even in the official recommendations at Empatica's webpage, if the experimental conditions so require [42]. We reasoned, that our experiments - even the highly arousing ones - would not probably elicit major arousal but only small shifts from the baseline so a small threshold value should be used. The same minimum amplitude threshold was also used in a research, which analysed the quality of data recorded from Empatica E4 [45]. Based on that research we also chose the offset value of 0.8 seconds. Filter frequency, filter order, max rise time and max decay time were kept at their default values of 1.0, 6, 4 and 4.

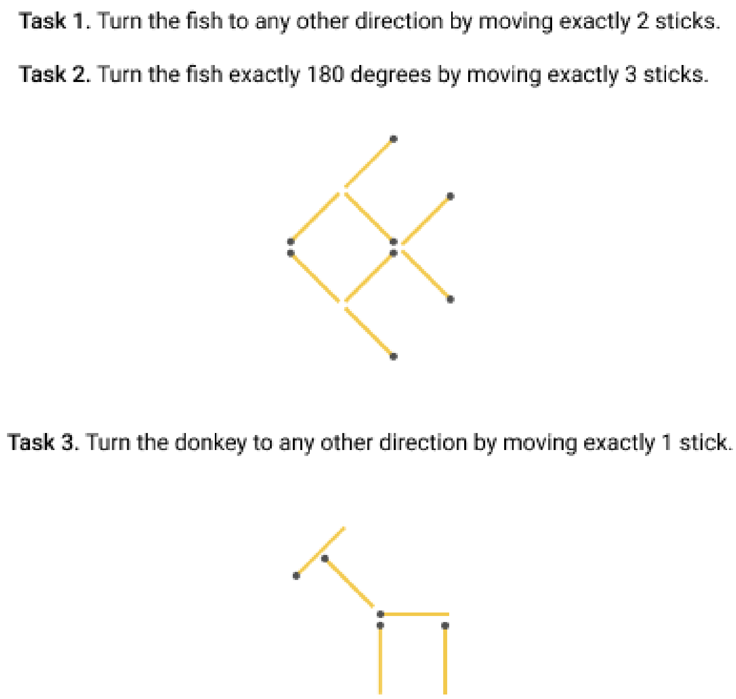

Figure 16: The matchstick tasks of task 5. The reader is encouraged to try them.

After the actual SCRs have been distinguished from noise and we have the results, we still have one more step to consider. A general friction in EDA recordings that needs to be balanced out is the removal of non-specific SCRs [16, p. 6]. That is, the SCRs that are not tied to our experimental manipulation. We thus remove any SCRs from our measurements, that were recorded during the first 30 seconds after the beginning of the experiment and the last 30 seconds before the end of experiment. We arrived to this decision for the following reasons:

1. We need to mark on the data when the experiment starts and when it ends, and this requires a click on the wristband. Just this action or the anticipation could cause SCRs.

2. We must leave the room before the start of the experiment, and this takes some time. That disruption could cause discomfort for the subject and thus cause SCRs.

Min amplitude	**Offset**	**Filter frequency**	**Filter Order**	**Max rise time**	**Max decay time**
0.01	0.8	1.0	6	4	4

Table 4: Our chosen values for the EDA Explorer [170]. The values are based on literature references from [16, p. 7], [45, p. 18] and [42].

3. Marking the end of the experiment also requires a click on the wristband and this action or the anticipation of it could cause an SCR

Removing 30 seconds of data from both ends could alleviate this problem and only leave in our data the SCRs caused by our tasks. This process is also shown in figure 17. After this step, we can finally calculate the amount of SCRs during the experiments.

We report our EDA readings as SCRs per minute, as we are in this case not so much interested what actually caused any single SCR, but their total amount in the task overall. Also, a higher amplitude of SCRs does not necessarily mean a higher arousal level and amplitude levels don't generalise between subjects [37]. In addition, SCRs/min was also the chosen method in Damasio's somatic marker theory paper to measure somatic responses to emotionally charged stimuli [33].

3.4.2 Analyzing the HRV data

To analyze the HRV data, we developed a custom analyzer script with Python (the script can be found from the appendix). Python was the perfect tool for the job

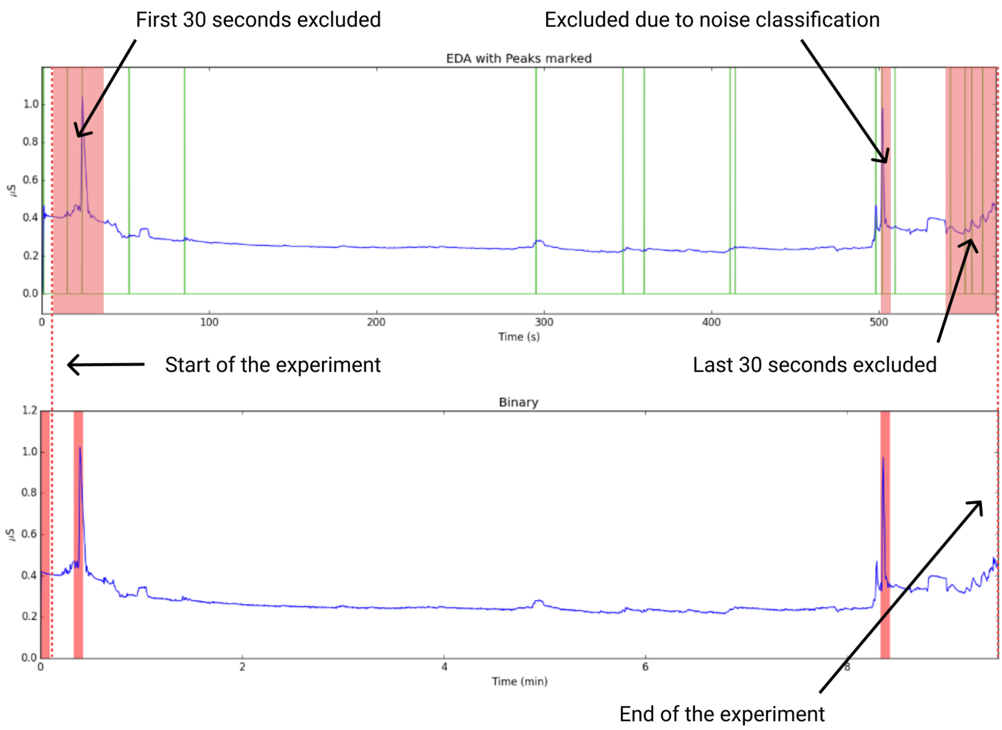

Figure 17: An example of what was included in the EDA data. 30 seconds of data were removed from both ends as well as any events that were classified as noise by EDAexplorer webtool.

due to its multiple toolkits that are made for this kind of work: Pandas for data manipulation and SciPy for signal processing tools and MatPlotLib for plotting. The code to calculate has been developed by the author, but inspiration and help have been found from the following internet sources: *Analyzing a Discrete Heart Rate Signal Using Python* in Paul van Gent's blog [52] and Pedro Gomes' open source *pyHRV* toolkit from GitHub [55]. In the analyser, we use both time and frequency domain methods, and both linear and non-linear metrics.

To study the frequency domain, we create a plot with the original signal, interpolated signal and filtered signal with only the MF-component for easy visual inspection. For the filter, we chose a peak filter. It is the perfect tool for this job, as it can create filters with very narrow bandwidth. We chose 0.15 Hz as the peak frequency with 0 dB attenuation (see figure 18 for a visualisation). Although on the higher limit (remember that MF-component was reported to be between 0.07 - 0.15 Hz in table 2), this was chosen because we noticed that the lower frequencies usually have much higher absolute power than the higher frequencies spillovers from lower frequencies and thus they could introduce noise to our measurements. Also, as HF component is also reported to correlate with emotional arousal, the spillover from that frequency bin was not considered so detrimental. There could definitely be better choices for the peak frequency depending on the task and the subject. Further research and experiments with the filter values are encouraged for future research. But peak filter would probably still be the better choice as normal band-pass filters are harder to implement with such strict specifications. For filter's we also lower our sampling frequency to 1 Hz (from Empatica's standard 4 Hz) as this removes noise and eases the calculations yet is high enough so that our Nyquist frequency remains at 0.5 Hz, well above our interest region. A simple description of the algorithm would go as:

1. Read in IBI.csv -data
2. Create an interpolated signal from this data
3. Filter the interpolated signal with 0.15 Hz IIR-peak filter
4. Plot the original, interpolated and filtered signal on top each other

This should afford for an easy visual comparison of our data. According to [72], MF component should drop when the subject is experiencing mental stress. Now, looking at this created plot we could see when the MF component drops and compare, whether there was noticeable mental stress during that task. Please see figure 20 for an example plot.

In addition, we also create a visual spectrogram with the approximate aforementioned frequency bins (VLF, LF, MF, HF) for visual inspection. To create the spectogram, we create a windowing function (a Hamming window with the size of 16 samples) that slides through the signal with slight increments (incremented by 2 samples, i.e. each consecutive window has an overlap of 14 samples). A fast fourier transform (FFT) is then calculated for these windows. An absolute value is then taken from these to remove the negative frequencies created by the FFT. The powers are then normalised by taking their logarithm and dividing that by the maximum power found in the absolute valued spectrum. These values are then plotted to bins (frequency) on the y-axis and window number (time) on the x-axis. The power is visualised by color mapping. Due to limitations in the code, the frequency bins are not exactly as described by literature but approximately as following: 0.0 - 0.08 Hz, 0.08 - 0.17 Hz, 0.17 - 0.25 Hz, 0.24 - 0.33 Hz, 0.32 - 0.42 and 0.42 - 0.5 Hz. The final plot could thus be interpreted so, that the lowest bin would be the combined VLF and LF bins, the second one the MF bin and the rest would belong to the HF bin. This algorithm can be described with following simplified steps:

1. Read in IBI.csv -data
2. Create an interpolated signal from this data
3. Slice the signal into windows (size = 16, step = 2)
4. Calculate fast fourier transform for each window
5. Take absolute value from the spectrum
6. Normalise spectrum values by taking logarithm and dividing by maximum value
7. Create a plot with frequency bins
8. Categorize each value into these bins

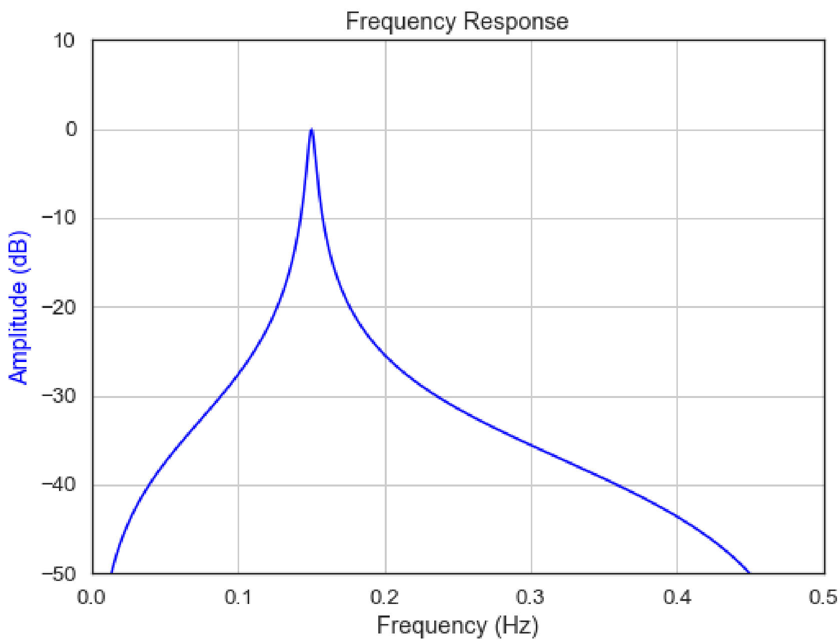

Figure 18: The visualization of our peak filter. The MF-component is reported to exist between 0.07 Hz and 0.15 Hz.

9. Color code the bins (yellow = high power, green = low power)

For the time domain, we calculate the mean IBI, SDNN, SDSD, RMSSD, pNN20 and pNN50. We also calculate sample entropy and draw a Poincaré plot.

For HRV data, we utilize all of the recorded data. There a few reasons for this. First of all, with HRV longer term measurements are better and more valid [141][158]. Second of all, noisy recordings are already removed from our HRV data by Empatica's script [76]. This is not true for EDA. Third of all, we are not so interested in the small moments elicited by small stimuli, but rather on the average recordings and their relative differences. I.e. how do task 1 and task 2 differ in their readings for SDNN?

4 Results

In this chapter, we will go through the results of our experiments. First, we shall recap the results from our subject interviews and the psychological surveys. Those results will be compared to the general population (or to be precise, compare the results to previous research from [22], [102] and [182]) and we will check how those readings correlate with each other.

Second, we will go through our HRV measurements and how they correlated with each other. This was done because HRV metrics where not so clearly or explicitly presented in literature as EDA was. Many different metrics in HRV are correlated with each other and are indicators of the same physiological phenomenon, but some might be better than the other depending on the experimental set-up [159]. We want to see, what was the case in our measurements.

Third, we shall calculate the task specific metrics and compare their metrics and see how they differ between tasks and if the different tasks are thus distinguishable by looking at the different metrics.

Fourth, we will see whether the AIM or REI measurements correlated with the person specific physiological readings.

4.1 User interviews

There were in total 10 subjects. All subjects were employees at the case company and they were chosen through voluntary listing. Three worked as technical project managers, two as designers, one as a product manager, one as data-analyst, one as business consultant, one as information security specialist and one as a marketeer. Six were female and four male. Their age was between 23 and 35. All of them either already had an university degree or were students at a university. All where accustomed to working with computers both at work and at home. Nationalities and first languages varied within subjects. With people who had Finnish as their mother tongue the experiments and interviews were conducted in Finnish. With people who had a mother tongue other than Finnish, the experiments were conducted in English.

4.1.1 Tasks

To assess, whether we had succeeded in inducing the right kind of emotion, we reviewed the recorded tasks and interviews of the test group with two expert reviewers, who both work as user experience specialists at the case company.

From this review we can say that the most succesfully designed task was CL. Every person said in the interview that they really had to think and focus, and that especially the last part was considered by some "almost impossible". The concentration also was clearly visible for every subject in the video recordings.

The next best designed task was HANV. All interviewed subjects seemed engaged in the task and they were focusing their gaze on the task and on the clock. They also displayed frustrated or surprised facial expression after the error popped up on their screen. All these cues would suggest a state of high arousal. In the interviews, all subjects described experiencing stress especially because of the clock. Interestingly, what could be seen on often on subject's faces was the so called "smile of frustration" - a common human behaviour when presented with a frustrating interaction [75]. When asked about what had happened after that pop-up error, majority of subjects reported that they had then realised that the software was designed to break at saving. The task proved problematic for the classification of valence because many subjects said it was neither negative or positive but both: first it was a negative experience because of the time limit and error, but that it then turned positive because they quickly realised that the task was anyway impossible. We would still conclude that the majority of this task would be experienced as negative because every subject would start with the assumption that the task is possible, only later realising it is not.

The LANV task was the next best, somewhat successfully creating an experience of boredom in the subjects. None of subjects seemed really engaged their work and were showing behavioural patterns related to boredom such as drooping, leaning into their hands or sighing. But this task was also not without it's problems. A common pattern we observed and what was reported by the interviewees as well, was that this task was actually sometimes quite arousing. This was because the task was very routine work - manually copying numbers - but still required concentration. Thus subjects needed to "concentrate to be concentrated". However, the lack of any time pressure alleviated this problem, creating definitely less arousal then the HANV task. One subject also described the experience as "pleasant" and "relaxing" because you only needed to do one specific thing without any worry of outside pressure. Furthermore, subjects said that this kind of work would be extremely boring and tedious if it were real work, but because it was part of an experiment, it was much more pleasant. Also, some subjects tried to make this experiment "more engaging" by trying to use the numeric keypad without looking at it or creating a checker function in Excel. These kinds of actions are problematic, because they can cause changes in EDA and HRV. However, it is interesting to note because if somebody needed to make the task more engaging, it would suggest that the task was actually boring - indicating thus a succesfully designed experiment.

The design of positive experiences was more problematic. LAPV, while described by all subjects as relaxation, was prone to some outside problems. First of all, most people did deep breaths, which almost always elicits a skin conductance reaction [16]. One subject also experienced lag in the video playback and one had a sudden pop-up advertisement. One also said that they were thinking a lot about a recent holiday trip, because the video reminded them of that. However, all seemed and sounded very relaxed after the video. Subjects also displayed behaviour such as leaning back, breathing slowly, drooping eyelids - all indicators of descending arousal levels. The most problematic thing in this task was probably that it actually took considerable time for people to relax, especially if they had been exposed to highly arousing stimuli before that. On the same not, a relaxing stimuli is a little bit of oxymoron - watching a video and hearing music will always be somewhat stimulating. This was affirmed by one subject who complained that they did not like the music in the video.

Of all tasks, the most problematic one was HAPV. Even though the high arousal

Subject	Rational (3.39)	- Ability (3.34)	- Engagement (3.44)	Experiential (3.52)	- Ability (3.49)	- Engagement (3.55)
#1	4.35 (+0.96)	3.90 (+0.56)	4.80 (+1.36)	3.40 (-0.12)	3.60 (-0.11)	3.20 (-0.35)
#2	4.45 (+1.06)	4.50 (+1.16)	4.40 (+0.96)	3.55 (+0.03)	3.50 (+0.01)	3.60 (+0.05)
#3	3.55 (+0.16)	3.70 (+0.36)	3.40 (-0.04)	4.05 (+0.53)	4.20 (+0.71)	3.90 (+0.35)
#4	3.50 (+0.11)	3.70 (+0.36)	3.30 (-0.14)	2.85 (-0.67)	3.20 (-0.29)	2.50 (-1.05)
#5	3.95 (+0.56)	3.70 (+0.36)	4.20 (+0.76)	2.90 (-0.62)	2.70 (-0.79)	3.10 (-0.45)
#6	2.95 (-0.44)	2.60 (-0.74)	3.30 (-0.14)	3.40 (-0.12)	3.30 (-0.19)	3.50 (-0.05)
#7	4.30 (+0.91)	3.90 (+0.56)	4.70 (+1.26)	3.15 (-0.37)	2.90 (-0.59)	3.40 (-0.15)
#8	3.85 (+0.46)	3.80 (+0.46)	3.90 (+0.46)	2.70 (-0.82)	3.00 (-0.49)	2.40 (-1.15)
#9	4.25 (+0.86)	4.40 (+1.06)	4.10 (+0.66)	3.45 (-0.07)	3.50 (+0.01)	3.40 (-0.15)
#10	3.50 (+0.11)	3.30 (-0.04)	3.70 (+0.26)	2.30 (-1.22)	2.20 (-1.29)	2.40 (-1.15)

Table 5: The mean scores for the Rational Experiential Inventory -questionnaire. In parentheses is marked the deviation from the mean score of the original paper [130]. The header values show the means of the original paper.

was most likely there, as subjects needed to have a constant vigilance in order to play, the valence was definitely spread out. 2 out of 5 interviewees said they did not like the game and one seemed mostly bored when playing it.

4.1.2 REI-40

The scores of REI-40 questionnaire for each subject can be seen in table 5. In the table we've included the aggregate scale for both rational system and experiential system. In addition the two subscales for both are included: engagement and ability. We've also included the differences to the means reported in the original paper of Pacini and Epstein [130, p. 979]. In table 6 we have our group means and their relation to the reported original means scores in [130, p. 979].

For the whole group, REI-40 scores were slightly above average in the rational processing style and slightly lower than average in the experiential processing style. The biggest difference in scores were $+0.48$ in Rational Engagement and -0.41 for Experiential Engagement. However, all of the differences were still within the standard deviation shown in Pacini's original paper.

4.1.3 AIM

As there are multiple ways to interpret AIM results, we followed the recommendations of authors in [22], [101] and [182] and included all subscales as we are interested in all aspects of affect and we didn't want to narrow down our observations too early. In addition, we did a correlational analysis on our data to see whether the different subscales truly are independent factors. In table 7 we can see the correlation factors between AIR and Weinfurt scales within our data. There are only few major correlations: first, AIR (Combined) correlates strongly (0.8) with Uniform scale and Positive Affectivity structure of Weinfurt (0.8) (this might be explained alone by the fact that most questions are linked to Positive Affectivity structure [182][101]). Second, Weinfurt's Positive Affectivity and AIR's Positive Affectivity are perfectly correlated (1.00) (also not a great surprise, as they are mostly the same questions). Third, both Negative Intensities (0.8) and Negative Reactivities (0.9) are strongly correlated between both systems (almost the same questions). Negative Intensity and Negative Reactivity also correlate with each other's just at the lower end of strong correlation (0.7-0.8) (again, almost the same questions). Serenity seems to be an independent structure, with only weak to moderate correlations with any other dimension in AIR or Weinfurt. Following this analysis and for simplicity's sake, we will remove all Weinfurt's scales, except Serenity, from further analyses.

4.1.4 Correlations between AIM and REI

In table 8 we can see the correlations between REI and AIM surveys. Uniform AIM was moderately correlated with Experiential Engagement and Experiential Ability. Weinfurt's Positive Affectivity and AIR's Positive Affectivity were moderately correlated with Rationality and Rational Ability. Weinfurt's Negative Reactivity and AIR's Negative Reactivity were moderately and negatively correlated with Experiential Ability and AIR's Negative Reactvity also had a moderate negative correlation with Experiential Engagement, too. Weinfurt's Serenity was moderately correlated with Rationality, Rational Ability, Experiential Ability an Experiential Engagement. AIR's Negative Intensity, was moderately and negatively correlated with Rationality and it also had a strong negative correlation with Rational Ability.

4.2 Experiments

To analyse our experimental data, we utilized the methods discussed in the methods chapter for our experimental data. We will go through first the viability of different metrics, their mutual correlations and finally the significant findings.

	Our group's mean	**Mean in [130]**	**Difference**
Rational	3.87 (±0.49)	3.39 (±0.61)	+0.48
- Ability	3.75 (±0.53)	3.34 (±0.66)	+0.41
- Engagement	3.98 (±0.56)	3.44 (±0.67)	+0.54
Experiential	3.18 (±0.50)	3.52 (±0.47)	-0.34
- Ability	3.21 (±0.55)	3.49 (±0.54)	-0.28
- Engagement	3.14 (±0.53)	3.55 (±0.51)	-0.41

Table 6: Our group's and Pacini's and Epstein's [130, p. 979] mean scores and their standard deviations. On the left-most column we have their difference.

	Uniform	Weinfurt (PA)	Weinfurt (NR)	Weinfurt (NI)	Weinfurt (S)
AIR (Combined)	0.8	0.8	0.3	0.0	0.2
AIR (Positive Affectivity)	0.6	1.0	-0.3	-0.5	0.4
AIR (Negative Affectivity)	0.2	-0.5	0.9	0.8	-0.2
AIR (Negative Intensity)	0.2	-0.4	0.7	0.8	-0.4

Table 7: Correlations between our AIM questionnaire results.

4.2.1 Frequency metrics

Although frequency metrics in HRV seemed promising according to literature [158][3][72], we failed to get any good measurements. The major problems lies not in the method, but in our measuring device. Wrist-worn sensors with PPG measuring devices are much more prone to noise as chest worn ECGs. Almost every time when our subject moved, there was a misreading, and reliable metrics could only be acquired from tasks with little to no movement (only LANV). Even though Empatica E4 automatically discarded noisy recordings, it still posed one major problem: interpolation.

To calculate frequency components, the signal would need to be continuous. However, in our data there might be huge gaps, tens of seconds long. This causes major problems for our interpolation, as we can see from figures 20 and 20. This could

	Rational	- Ability	- Engagement	Experiential	- Ability	- Engagement
Uniform	0.3	0.2	0.3	0.3	**0.5**	**0.6**
Weinfurt (Positive Affectivity)	**0.4**	**0.5**	0.2	0.1	0.3	0.2
Weinfurt (Negative Reactivity)	-0.1	-0.4	0.1	-0.1	**0.5**	0.2
Weinfurt (Negative Intensity)	-0.3	**-0.6**	0.0	0.1	0.0	0.3
Weinfurt (Serenity)	**0.5**	**0.5**	0.3	**0.4**	0.2	**0.4**
AIR (Combined)	0.2	0.1	0.1	0.1	**0.5**	**0.4**
AIR (Positive Affectivity)	**0.4**	**0.5**	0.2	0.1	0.3	0.2
AIR (Negative Reactivity)	-0.2	**-0.4**	-0.1	0.1	**0.5**	**0.4**
AIR (Negative Intensity)	**-0.5**	**-0.7**	-0.2	0.0	-0.1	0.3

Table 8: Correlations between AIM and REI.

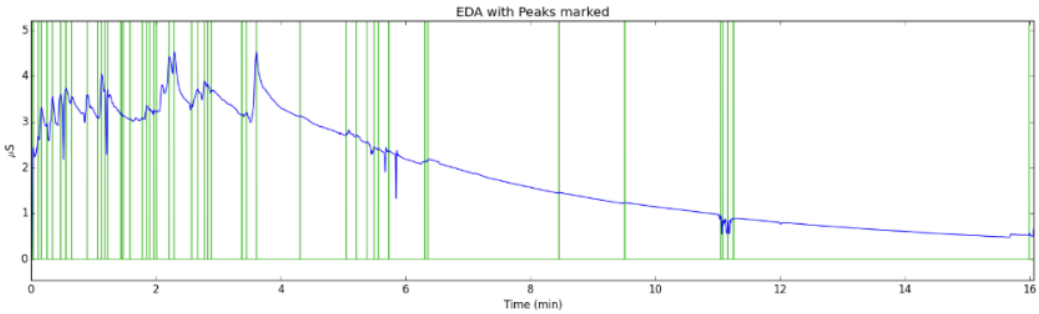

Figure 19: A noticeable drop in tonic signal in the LANV task (relaxation). This is taken from subject nine's data, but similar drops were noticed from other people as well.

of course be counteracted with only looking at those segments, which would have continous measurements. However, this would then pose problematic for comparison of data as HRV measurements are very much influenced by the measuring period. Either way, this kind of work is outside the scope of this thesis.

Dividing the signal into frequency bins and drawing a spectrogram out of them had exactly the same problem. The data had too much noise in it and thus also the spectrogram was rendered very noisy, making any visual comparison pointless. Furthermore, the powers should've been normalised between subjects as there were huge individual variations between subjects.

However, for time domain metrics the loss of data from noisy moments is not a huge loss, as we are anyway more interested in the general readings - not specific distributions at one time. In fact, it can be even an asset, as we can now be sure that all measured interbeat intervals were true ones.

4.2.2 Correlations between our measurements

One aim for this thesis was to find out how the different metrics we calculate from the data correlate. Then we would know what metrics we should focus on in the later, more practical applications of this technology. From figure 22 and from table 9 we can see the correlation matrix between all our measured metrics.

First of all, SCR seems to be an independent metric, which has only weak or no correlation between any other measurements. This is as expected. First of all, it is calculated from a different data source. Second of all this would be as hypothesised by the 2D-emotion model: SCR is correlated to arousal and arousal is a different axis in our proposed emotion model. Third of all, physiologically speaking this also makes sense as SCR (or EDA) is mediated purely by sympathetic nervous system [16, p. 3] where as HRV is always a combination of parasympathetic and sympathetic nervous system activity.

Multiple HRV measurements seem to correlate with each other. SDNN has a strong correlation with both SDSD (0.77), RMSSD (0.77), pNN50 (0.76) and SD1 (0.70). It also has a near perfect correlation with SD2 (0.98). Also, SDSD and RMSSD have a perfect correlation with each other (1.00). Both SDSD and RMSSD have a strong correlation with pNN50 (0.90). SD1 is also perfectly correlated (1.00) with SDSD and RMSSD, but with SD2 only moderately so (0.65). PNN20 has a strong correlation with pNN50 (0.80) and a moderate correlation with SDSD, RMSSD and SD1 (0.64), but only weak correlations with other measurements. Sample entropy displays negative correlations with every metric except SCR and SD1/SD2-ratio. All are weak except with SDNN (-0.52) and SD2 (-0.51). Both S and SD1/SD2 seem like independent metrics as they have only weak correlations with any other metrics.

4.2.3 Skin conductance as a measure of arousal

EDA is recognized as a good index for SNS arousal [32] and it is widely used as a practical measurement for emotional arousal [16, p. 2][89][155]. Our results also indicate that EDA is a good marker for arousal.

When looking at the average values across our subjects and within our experiment tasks, the most arousing task was HANV (high arousal, negative valence), where people had to work with a buggy software under a time pressure, and the least arousing task was LAPV (low arousal, positive valence), where people could relax by

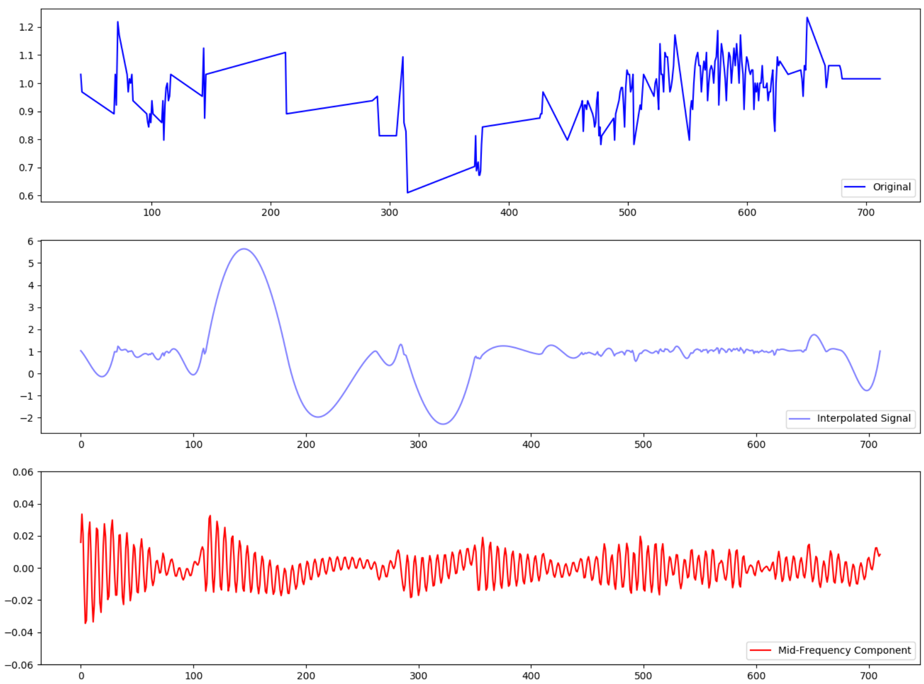

Figure 20: An example HRV frequency measurement. See how the interpolation causes major fluctuations in the interpolation, and thus noise into the amplitudes of those frequencies.

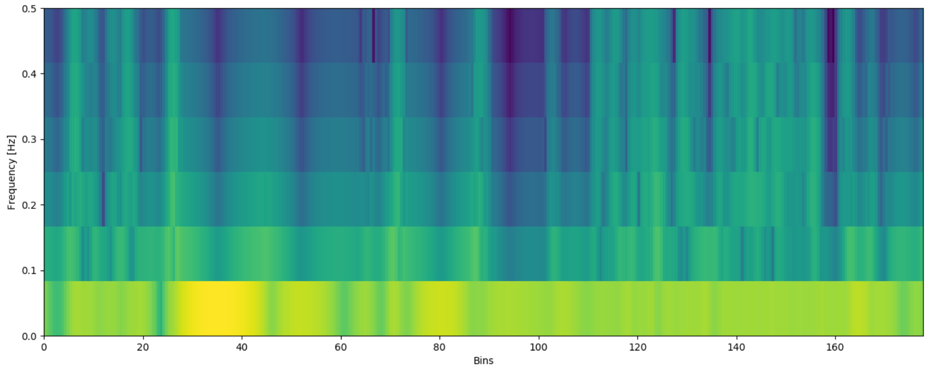

Figure 21: The same tasks as in figure 20, but now with a spectogram. A more yellow color indicates a higher power in that certain power band. The gaps in data lead to erroenous interpolation, which can be seen as a high power in the lower bin and lack of any power in higher bins.

watching an aerial nature video with soothing music.

Interestingly, we also found out that the second most arousing task was not HAPV (high arousal, positive valence), where people would play a game, but LANV (low arousal, negative valence), where people had to do boring and repetitive work with a spreadsheet software. In our recordings valence could be distinguished from SCRs alone so that negative valence had higher amount of SCRs per minute (p-value < 0.01).

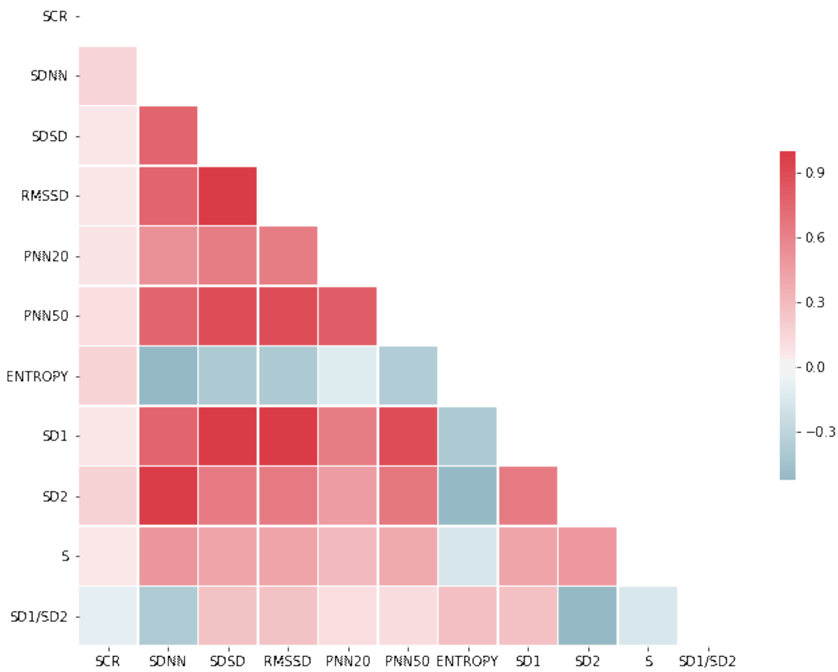

Figure 22: Correlation matrix of our measurements.

It could be that arousal is modulated by valence and negative feelings are experienced as more arousing than positive ones. This kind of effect has been observed in other research as well: in one case EDA was measured between different emotional musical pieces [89]. SCRs were found to be greater with the two more stimulating emotions, fear and happiness, as compared to the two more relaxing emotions, sadness and peacefulness. Other research has also found out that unpleasant sounds and pictures are more arousing than pleasant ones [15][99].

Possibly in our case the two tasks with positive valence were also considered to be more pleasant and thus leading into less arousal. Also, between the two tasks within the same valence the high arousal was distinguishable from low arousal. However, only between the negative valence tasks was this difference statistically significant (p-value < 0.05).

4.2.4 Heart rate variability as a measure of valence

HRV has been researched as a possibility to measure either the arousal dimension [13] or valence dimension [2] or both [168]. However, it is most likely not as good a measurement as EDA is for arousal as HRV can be modulated by both SNS and PNS activity. Our hypothesis was that HRV could thus better be suited for measuring valence as different components of HRV have been linked to a variety of emotional phenomena (see chapters 2.4.2 and 2.5 as well as table 2).

From our measurements the metrics that seem to group valence dimension accordingly were SD1 (and it's equivalent RMSSD), SDSD and pNN50. All these measurements were strongly or perfectly correlated in our data (see table 9 and figure 22), but seem to be correlated in other research results as shown in a review by Shaffer [158]. According to that review article all these measurements reflect short-term variability in the heart rate [158]. All these measurements are also correlated with HF power [92] and HF power is correlated with anxiety [84]. All these measurements point out that they would be more influenced by the PNS, "the-feed-and-breed" nervous system. These measurements would thus indicate that there is more PNS activation in our tasks with negative valence.

	SCR	**SDNN**	**SDSD**	**RMSSD**	**pNN20**	**pNN50**	**Entropy**	**SD1**	**SD2**	**S**	**SD1/SD2**
SCR		0.16	0.07	0.07	0.08	0.11	0.17	0.07	0.18	0.06	-0.09
SDNN	0.16		0.77	0.77	0.53	0.76	-0.52	0.77	0.98	0.50	-0.38
SDSD	0.07	0.77		1.00	0.64	0.90	-0.40	1.00	0.65	0.43	0.26
RMSSD	0.07	0.77	1,00		0.64	0.90	-0.40	1.00	0.65	0.43	0.26
pNN20	0.08	0.53	0.64	0.64		0.80	-0.12	0.64	0.46	0.30	0.11
pNN50	0.11	0.76	0.90	0.90	0.80		-0.37	0.90	0.67	0.39	0.12
Entropy	0.17	-0.52	-0.40	-0.40	-0.12	-0.37		-0.40	-0.52	-0.17	0.28
SD1	0.07	0.77	1.00	1.00	0.64	0.90	-0.40		0.65	0.43	0.27
SD2	0.18	0.98	0.65	0.65	0.46	0.67	-0.52	0.65		0.49	-0.53
S	0.06	0.50	0.43	0.43	0.30	0.39	-0.17	0.43	0.49		-0.16
SD1/SD2	-0.09	-0.38	0.26	0.26	0.11	0.12	0.28	0.27	-0.53	-0.16	

Table 9: The correlations between different metrics.

	Recorded SCRs per minute										
TASK	**#1**	**#2**	**#3**	**#4**	**#5**	**#6**	**#7**	**#8**	**#9**	**Mean**	**StDev**
CL	3.78	0.58	0.22	0.26	6.55	0.45	0.70	4.88	7.00	2.71	2.85
HANV	0.40	0.00	1.10	0.13	0.86	1.45	2.00	7.86	4.00	1.98	2.52
LANV	0.35	0.18	0.00	0.00	0.62	1.20	0.94	6.75	1.27	1.26	2.12
HAPV	0.17	0.06	0.00	0.09	2.11	0.00	0.17	4.63	0.00	0.80	1.59
LAPV	0.00	0.15	0.19	0.10	0.19	0.18	0.00	0.26	0.24	0.15	0.09

Table 10: Our recorded measurements of skin conductance reactions between the experiments. Subject 10 is not included, as they displayed no SCRs during any of the experiments. The table is sorted descending for mean level.

However, this is completely opposite according to literature and what we hypothesized, HRV metrics such as SD1 should be negatively correlated with stress [158][3], not positively. If SD1, RMSSD and SDSD are a markers of PNS activity, and thus correlated with HF power activity (which, in too is an indicator of PNS activity [158]) and our experiments truly elicited stress, then the findings should be opposite. However, the same articles also included, that the indicators of HRV metrics are not so simple to analyse. Especially as HRV measurements in general are not very replicable if the metrics are not normalised and very well controlled [174]. Furthermore, a single parameter indicator might not be enough - after all, HRV is always a by-product of both SNS and PNS activity [158, p. 3]. For example in the paper *Resolving Ambiguities in the LF/HF Ratio: LF-HF Scatter Plots for the Categorization of Mental and Physical Stress from HRV* [141] researchers Von Rosenberg *et al.* point out that:

> "Of all one-dimensional parameters in Mental Stress –, only HFp (*writer's note: HFp is correlated with SD1, RMSSD and SDSD. All are indicators of PNS activity*) indicated an increase in the stress level (i.e., a reduced activity in HFp) until Arithmetic, and a decrease afterwards." [141, p. 10]

	Recorded HRV metrics									
TASK	**SDNN (ms)**	**SDSD (ms)**	**RMSSD (ms)**	**pnn20 (%)**	**pnn50 (%)**	**Sample entropy**	**SD1 (ms)**	**SD2 (ms)**	**S (ms^2)**	**SD1/SD2 (%)**
LANV	80.86	85.14	85.15	35 %	20 %	1.76	60.21	96.23	18703.19	64.92%
HANV	93.83	83.00	82.99	34 %	21 %	1.80	58.81	118.24	22912.17	53.25 %
LAPV	81.92	75.21	75.21	36 %	20 %	1.76	53.12	102.12	27565.37	52.80 %
CL	80.68	73.04	73.04	33 %	17 %	1.99	51.65	101.39	18021.20	52.97 %
HAPV	77.88	69.62	69.62	33 %	18 %	1.76	49.23	98.23	15573.41	51.43 %

Table 11: Mean values for our HRV metrics during each task. The table is sorted descending for SD1/SDSD/RMSSD.

What they are indicating is that even though we can measure mental stress from a singe parameter, in many cases a single parameter might be too simplistic to capture the underlying complexity of HRV and they thus suggest against using only a single parameter. In their different tasks, only HF power was an indicator of stress and even that didn't succeed in all tasks. The same problem applies for ratios such as the LF/HF ratio - the measurement commonly held as a possible metrics for stress - because although it applies two variables, it's dimensionality is still reduced to 1D. A ratio of 2-to-2 is the same as 10-to-10, even though the magnitude is different. But they proposed a solution for this problem. Their suggestion is a novel approach of looking at the HRV through 2-dimensional scatterplots.

We tried a similar approach with our data, plotting our data on a scatterplot where SD1 (as an indicator of HF power or possible PNS activity) was plotted on the X-axis and SD2 (as an indicator of LF power or possible SNS activity). We removed CL task from the the data, as it was considered as neutral in valence. By visual inspection the data seems to separable; however a fit with support vector machines (SVMs) produces only fit whose accuracy is at best a little bit over 60% - not much better than random choice. The plots can be seen in figure 23.

4.2.5 HRV and EDA as measurements of mental load

In the fifth task, cognitive load, we induced cognitive load into the subjects by making them solve a visuospatial problem. According to literature both EDA [128][157][166] and HRV [65][122][147][141] have been reported as possible markers of cognitive load

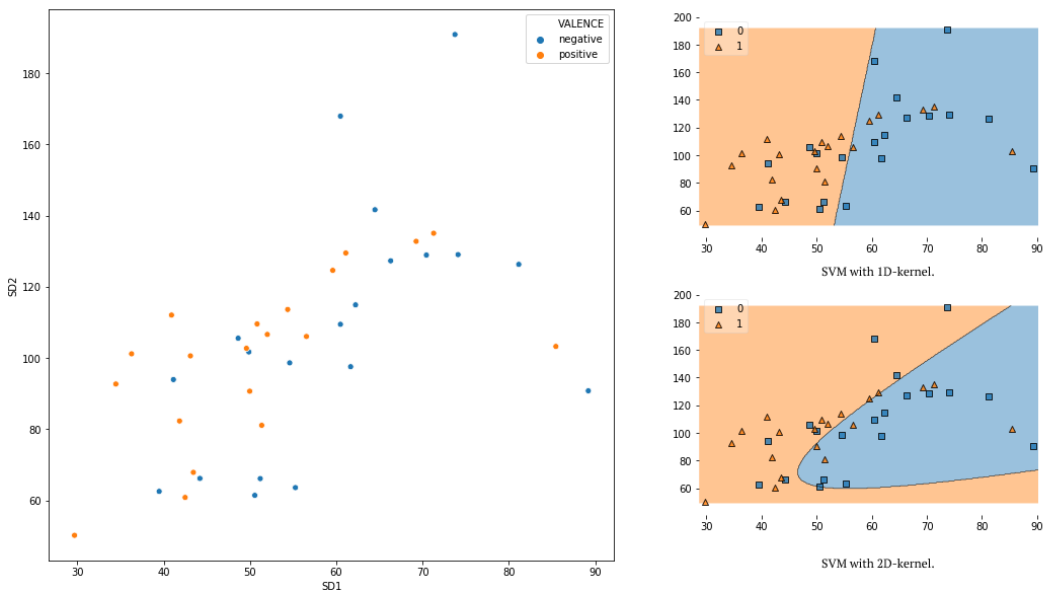

Figure 23: Our HRV data plotted on 2-dimensions. X-axis, or SD1 is an indicator of HF power and Y-axis, or SD2, is an indicator of LF-power. On the right there are suggestions for classification with 1D- and 2D-SVMs.

or mental stress.

If we look at our table 10 with recorded average SCRs/mins, we can see that CL task has the highest arousal. Much higher than any other task. However, it also has the biggest standard deviation. From the table 11 with our HRV measurements we notice that SDSD, RMSSD, SD1 and pnn50 go down. This could mean reduced PNS activity in the heart rate complex and thus increased stress. However, sample entropy is much higher than in any other tasks indicating a more complex and chaotic signal and thus indicating a highly complex heart rate variability which could mean lower stress.

4.3 Correlations with personality

One aim of the thesis was to explore and measure correlations between different psychological constructs and psychophysiological readings. In this subsection, we will go through basic correlations and few of the more interesting findings in detail.

Negative correlation between EDA and AIM From our experiments, we can see a strong negative correlation between mean SCRs/min and Affect Intensity Measure with $r = -0.8$, $p < 0.01$ (see figure 24 and table 12). The biggest contributor of the individual subaxes is Positive Affectivity (both Weinfurt and AIR) ($r = -0.7$ and $p < 0.05$), but they are not as good as predictors as the combined AIR score is.

This could suggest evidence for our hypothesis, where people who report having intense affections do not show them with EDA. The possibility why this would happen is that people who consciously process their emotions do not need to do so unconsciously and could thus have lower SCRs/min and vice versa.

Note should be taken that we removed subject 10 from these calculations as an

	SCRs/min	**SD1**	**SD2**	**Sample entropy**
Uniform	-0.6	-0.1	-0.2	0.0
Positive Affectivity (Weinfurt)	-0.7	0.1	0.0	-0.3
Negative Reactivity (Weinfurt)	0.1	0.1	-0.1	0.5
Negative Intensity (Weinfurt)	0.1	-0.3	-0.3	0.3
Serenity (Weinfurt)	-0.3	-0.2	-0.2	-0.1
AIR	-0.8	0.1	-0.0	-0.1
Positive Affectivity (AIR)	-0.7	0.1	0.0	-0.3
Negative Reactivity (AIR)	0.2	0.1	0.0	0.4
Negative Intensity (AIR)	-0.1	-0.1	-0.1	0.1

Table 12: Correlations between personality traits and psychophysiological metrics.

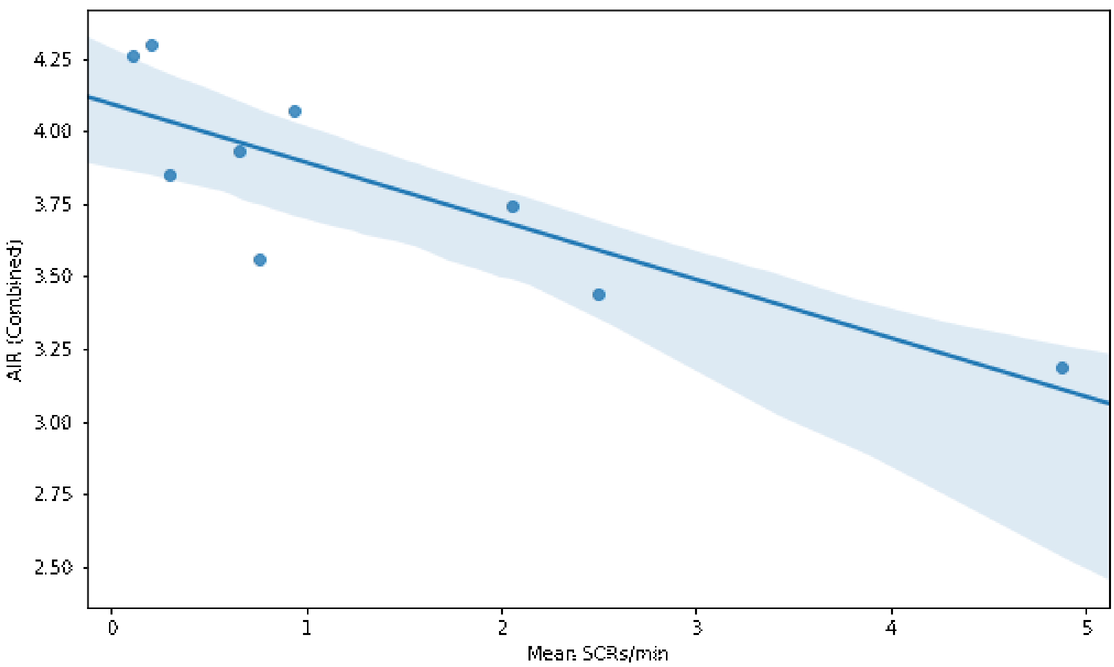

Figure 24: A strong negative correlation ($r = -0.8$ and $p < 0.01$) between AIM (AIR Uniform -scale) score and electrodermal activity (calculated as mean skin conductance reactions per minute between all experiments). The regression line shows also the 95% confidence interval.

outliar as they did not display any skin conductance reactions during any of the experiments.

Correlations between HRV metrics and AIM In our data, we can only see weak to non-existent correlations between AIM subsclales and HRV metrics such as SD1, SD2 and sample entropy. However, if we only look at certain tasks, this does change. These correlations can be seen in tables 13, 14, 15 and 16. Notice, that we now only report AIR subscales and Weinfurt's Serenity because the other constructs were so strongly correlated with each other.

In the HANV task we can see a strong negative correlation between AIR and SCRs/min. In addition, there are moderate negative correlations between SCRs/min and Positive Affectivity and Serenity. With Negative Reactivity and Negative Intensity we have a moderate positive correlations with SD1. Serenity also has a negative and moderate correlation with SD1.

In the LANV task too we can see a strong negative correlation between AIR and SCRs/min. Serenity and Positive Affectivity have a moderate negative correlation as well. But SD1 in this one has a moderate negative correlation with AIR and Negative Intensity.

In the HAPV task, the similar pattern repeats: moderate negative correlations with AIR and Positive Affectivity. However, now we can also observe a moderate positive correlation with SD1 and Negative Reactivity and Negative Intensity. This

Correlations: HANV				
Trait/Metric	**SCRs/min**	**SD1**	**SD2**	**Sample entropy**
AIR	**-0.9**	0.3	-0.3	0.1
Positive Affectivity (AIR)	**-0.7**	-0.1	-0.3	0.2
Negative Reactivity (AIR)	0.2	**0.6**	0.2	-0.3
Negative Intensity (AIR)	-0.2	**0.5**	0.0	0.0
Serenity (Weinfurt)	**-0.5**	**-0.4**	-0.2	0.0

Table 13: Correlations between personality construct subscales and our measurements in the HANV task.

same applies for SD2. Sample Entropy is also moderately and positively correlated with Negative Reactivity.

The LAPV task has exactly the similar correlations with AIR & Positive Affectivity and SCRs/min: negative and moderate. In this task though we observe a strong positive correlation with SD1 and Negative Reactivity and moderate positive correlations with Negative Reactivity and SCRs/min and SD2. In addition Negative Intensity has moderate positive correlations with SD1 and SD2, and Positive Affectivity is negatively and moderately correlated with SD1 and positively and moderately correlated with Sample entropy.

Correlations: LANV				
Trait/Metric	**SCRs/min**	**SD1**	**SD2**	**Sample entropy**
AIR	**-0.7**	**-0.4**	0.1	-0.1
Positive Affectivity (AIR	**-0.6**	-0.2	0.0	-0.1
Negative Reactivity (AIR)	0.2	-0.3	0.1	0.0
Negative Intensity (AIR)	-0.1	**-0.4**	0.0	-0.2
Serenity (Weinfurt)	**-0.5**	0.1	0.1	-0.2

Table 14: Correlations between personality construct subscales and our measurements in the LANV task.

Correlations: HAPV				
Trait/Metric	**SCRs/min**	**SD1**	**SD2**	**Sample entropy**
AIR	**-0.6**	0.3	0.0	0.4
Positive Affectivity (AIR)	**-0.5**	-0.2	-0.4	-0.3
Negative Reactivity (AIR)	0.1	**0.6**	**0.6**	**0.5**
Negative Intensity (AIR)	-0.1	**0.7**	**0.5**	0.4
Serenity (Weinfurt)	-0.2	0.2	0.2	-0.3

Table 15: Correlations between personality construct subscales and our measurements in the HAPV task.

Correlations: LAPV				
Trait/Metric	**SCRs/min**	**SD1**	**SD2**	**Sample entropy**
AIR	**-0.4**	0.0	0.0	0.3
Positive Affectivity (AIR)	**-0.5**	**-0.4**	-0.3	**0.4**
Negative Reactivity (AIR)	**0.5**	**0.8**	**0.6**	-0.1
Negative Intensity (AIR)	0.1	**0.4**	**0.5**	-0.3
Serenity (Weinfurt)	0.0	0.0	0.1	-0.2

Table 16: Correlations between personality construct subscales and our measurements in the LAPV task.

5 Discussion

The conclusions from our results are not straightforward. The first thing we can confirm is that eliciting emotion is possible. However, eliciting *specific and wanted* emotions is hard. Especially when it comes to enjoyable experiences and positive emotions. We don't think that we can say with any confidence that our experiments produced similar experiences across all people. We did dl a pilot experiment with one subject whose interviews confirmed the emotions in our tasks to be as designed. However, from the actual subjects only one interviewee described the experiences to be what we designed them to be.

5.1 Limitations

The only task that seemed to produce similar and consistent response in all of the subjects was the cognitive load task. The high arousal negative valence task was also somewhat consistent. This is a major limitation for the construct validity of our experiments. Thus any higher extrapolation about our experiment's results should not be made and the results should be only interpreted in the context of this study or as a idea toward further research.

Our test subjects were also problematic, especially if we were interested in studying personal differences. Our group was extremely homogeneous, which can be seen already in the REI-scores. Everybody scored high on the Rational axes, as this was an IT company and everybody had university degrees. For the research, this bias on the rational scores prevented any comparisons between "rational" and "experiential" groups. We considered the possibility of running further psychological surveys with multiple people, so that we could then choose respective groups but soon decided against this. We simply didn't have the time and resources to do that kind of large scale surveys in the scope of this thesis. Also, it might've been hard to find suitable people with a high experiential score, given that all the people were from the same IT company and that a random sample already gave us such a huge bias towards rational processing style.

Furthermore, our subjects were not controlled for any of the widely known variables that influence HRV — sex, BMI, smoking habits, physical activity levels, depression, anxiety, and stress [135]. Also, we were all co-workers who knew each other beforehand. To conclude, even though we did control human interaction and other disturbances during the tasks, there are still major concerns regarding the internal validity of our experiments.

However, external validity was controlled better: this is a real-word experiment with real people doing real-life tasks. Although population was not controlled and sample was biased, it was still clearly defined to include only people who are relevant for the case company and who are familiar with the technologies used in the tasks. Thus all the people should have more or less similar backgrounds and for example computer usage proficiency shouldn't affect the task measurements. Pre- and post-test effects were also controlled with random order, similar interview structures and by minimizing human-interaction. Situational factors were kept so low as possible:

similar test times, enough time to settle down before and between tasks and similar rooms. We thus argue, that our experiments also have a good ecological validity because they approximate real-life structure almost perfectly.

However, a real concern is that do our generalize to the individual level? High arousal was distinguishable from low arousal with SCRs per minutes -value, if we looked at the means of all recordings. But if we look at the individual level (see table 17 for the subjective order) things are not so certain. Out of 9 subjects only 3 displayed the same order of EDA as the mean values did (HANV, LANV, HAPV LAPV). But then again, 7 out of 9 had HANV as their highest reading. The other two had highest readings from either LANV or HAPV tasks. Interestingly what we would've expected to be the lowest reading for most people, the LAPV task, was actually the lowest reading for only 4 people. The other subjects had either in LANV or HAPV as their lowest. Subject 2 was a complete turn-around: for them the lowest readings came from HANV and highest from LANV.

Multiple ideas come to mind as a reason for this. Especially after hearing the interviewees by the subjects who were asked to describe their experiences, we realised that we probably did not succeed in eliciting the intended emotional category in all of the subjects. For example, the LANV task was not that simulating for many, causing next to no EDA readings. But what it also required, was the "concentration to concentrate". Most likely this kind of conscious attention arousal would lead into physiological arousal as well. Also, as some people tried to make it more stimulating for them and some just did it as explained, the readings vary due to this fact as well.

In addition, HAPV could've been considered relaxing - after all, it was about playing a computer game which is a favourite past-time for many. And in one case (subject 5) the experience was considered negative, in effect turning this task into one with negative valence as well. This would explain their high readings of EDA as well. Furthermore, although as an designed experience the LAPV task worked as intended, it was problematic in its readings across subjects. Many subjects did deep breaths, which elicit SCRs [16, p. 11, 42]. Also, at least one described a "racing mind" effect, where they were thinking a lot about other work and daily tasks instead of relaxing. This would suggest that actually relaxing takes quite a lot of time for some people. Thus high arousal could still stay as a prevalent mood for this experiment as well,

	#1	**#2**	**#3**	**#4**	**#5**	**#6**	**#7**	**#8**	**#9**
Most SCRs (high arousal)	CL	CL	HANV	CL	CL	HANV	HANV	HANV	CL
	HANV	LANV	CL	HANV	HAPV	LANV	LANV	LANV	HANV
	LANV	LAPV	LAPV	LAPV	HANV	CL	CL	CL	LANV
	HAPV	HAPV	LANV HANV	HAPV	LANV	LAPV	HAPV	HAPV	LAPV
Low SCRs (low arousal)	LAPV	HANV		LANV	LAPV	HAPV	LAPV	LAPV	HAPV

Table 17: The tasks sorted by their personal arousal.

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
High SD1/SD2 (low stress)	CL	HANV	LAPV	HANV	LANV	HANV	LAPV	LANV	LANV	LANV
	HAPV	HAPV	LANV CL	LANV	HAPV	LAPV	LANV	LAPV	LAPV	HAPV
	LANV	LANV	HAPV	HAPV	HANV	LANV	CL	CL	HAPV	HANV
	HANV	CL	HANV	CL	CL	CL	HAPV	HANV	CL	CL
Low SD1/SD2 (high stress)	LAPV	LAPV		LAPV	LAPV	HAPV	HANV	HAPV	HANV	LAPV

Table 18: The tasks sorted by SD1/SD2 readings (which is a supposed indicator of PNS/SNS-ratio and thus an indicator of stress [158].

and thus maybe giving some people higher readings than expected.

However, when looking at the data from subject 7 we have exactly the kind of readings we would expect. The order of EDA readings was HANV, LANV, HAPV and LAPV (see table 17 and the order for SD1/SD2 was LAPV, LANV, HANV and HAPV (see table 18. And in fact, of all the interviewed subjects only subject 7 described the tasks as we had designed them. Although an interesting result, it was still only a single "hit" in a group of misses, and we wouldn't draw any conclusions from this data point.

5.2 Significant findings

We have a few significant findings in our results, and we have listed them with further details here below.

Emotional user personas Although far from complete, quantified emotion user personas are an interesting phenomenon. Even from such a small and homogeneous sample, we can measure distinguishable differences between subjects. These surveys could be used in the future for a basis of "emotional" user personas. User personas are needed and important for the communication and clarification of early designs. This could be a way to clarify the methods and messages [86][117].

Interpretation of AIM and REI Our group preferred the rational information processing style to the experiential style. This difference can of course be explained in many ways. It could be a statistical error or it could be a cultural preference. However, as all the subjects work in a information technology firm, it would not be an unjustified claim to say that the subjects like work that engages their rational side more. Rational problem solving is after all a highly sought-after and needed ability in IT.

EDA seems to be a good marker for arousal EDA produced constant and reliable measurements in all tasks and arousal was distinguishable from EDA within subjects and between tasks. The data was also easy to segment and control only for the duration of the experiment. We also succeeded in controlling the measurements. Data-analysis was straight-forward with the already existing tools. All this would suggest proof for EDA being a good marker for emotional arousal.

On the negative correlation between AIM and EDA As pointed out, we found a statistically significant, very strong negative correlation between AIM score and EDA, and thus support for one of our hypotheses. This could be due to different emotion regulation strategies that people use.

There are multiple kinds of emotion regulation strategies described in literature, but two different kinds of emotion regulation strategies and their differences described by James J. Gross could prove of interest here: they are the the cognitive reappraisal and expressive suppression [62]. In *cognitive reappraisal* strategy, the emotions are reinterpreted cognitively so that their emotional impact changes [105]. I.e. when using of all of your free time to finish a master's thesis, one might view it as a great opportunity to focus on an intellectual problem instead of falling into a deadline-driven panic. In *expressive suppression* strategy, the person inhibits their current exhibition of emotions [59]. A great example would be the case of a "poker face" during a card game. Interestingly, suppression strategy has been found to suppress the display of emotional behaviour, but not the experience. This is especially true for negative emotions.

When used to down-regulate a negative emotion, reappraisal should succesfully reduce both the experiential and expressive components of the emotion. Where as suppression only inhibits the expressive component, but not the experience. Thus maybe AIM is an indicator of the preference between these two strategies. People with a low AIM score, might prefer to use the suppression strategy and people with high AIM score would prefer the reappraisal strategy. A one further indicator of this would be the fact that people in our test group who had a low AIM score still scored very high on the Negative Experience axis. This interesting as this asymmetry has been reported in literature: whereas suppressing negative emotions left intact the experience of negative emotion, suppressing positive emotions decreased the experience of these emotions.

Thus maybe what is happening here is that, people with low AIM scores are using this suppression strategy. Even though they might not express the emotions and thus have a low AIM score, they are experiencing them. While not shown outwardly, this experience still exists and can be seen in the EDA recordings. A possible mechanism could be related to the cognitive load. As the suppression strategy is reported of being more cognitively taxing [62, p. 350], it could lead to increased SCRs as they can be elicited by the increased activity in prefrontal cortex [37, p. 204], an area responsible for inhibition control [136] and executive attention [88].

However, there is little to no reverse-effect in our results: i.e. a high value in negative personality constructs don't predict higher SCRs. Except in one: LAPV and Negative Reactivity.

Intepretation of HRV Now this one is interesting. We can spot the difference from our readings between the valence - but it is not what we would have expected, it is completely the other way around! During high mental stress the HF band and its respective markers, such as SD1/SDSD/RMSSD should decrease. But they didn't during our experiments. This same can be seen with the controversial metric of LF/HF-ratio (which is the inverse of our SD1/SD2). A low LF/HF-ratio should indicate a parasympathetic dominance, and tend-and-befriend behaviour. But what we observe is just the opposite: according to these interpretations our subjects would've been the most stress-free during the boring spreadsheet task (LANV) and the time-pressure error task (HANV). And the most stressed during playing video games for fun!

Now why is this? Multiple ideas come to mind. First of all, our data is incredibly noisy and non-consistent between measurements. The duration of our data and the amount of data points is not at all constant, a major fault as HRV measurements are highly affected by the used time window [158]. Second of all, and as already pointed out: maybe our experiments were badly designed, and they don't actually represent the valence very well? Maybe we are not measuring valence, but some other dimension such as cognitive load or mental stress. Third is the problem of subjectivity. Maybe boring tasks actually are relaxing? After all, we seem to struggle the most when left idle. The human mind always needs something to tackle. Maybe the mundane, repeating task is actually "pleasant" for our nervous system and we can just relax. This could be the case, as the two most bottom tasks were the cognitive load task (CL) and video-game playing task (HAPV). Even though pleasant, video games can definitely increase mental load. Or maybe it is exactly the other way around and for some routine tasks are incredibly taxing? For example, if we look at subject 2 in the table 17, we can see that for them LAPV (the excel task) was the most arousin oneg. Weird at first, but when asked about that later subject 2 described themselves as being highly competent with Excel and using that software a lot in their daily works. Maybe that task turned out extremely frustrating for them and thus leading to high arousal.

5.3 Conclusions

To conclude, we would say that emotion recognition in naturalistic settings is a very hard problem, and this work provides further evidence of that. The uncontrolled setting makes data analysis very convoluted. However, some of that could be filtered out by better experiment design and more rigorous and standardised measurement methods.

The biggest faults in our experimental design was that we didn't properly control the measurement time of HRV and that we knew the people too well. Also our induced emotions by our experiments were probably not what we meant them to be, even though we did a small pilot study. As we noted from the huge variance between readings and personal arrangement of the tasks, we want to say that designing experiences is very hard. Even though you can maybe draw some big lines, you will most likely fail to induce exactly the same kind - or even mostly similar - experience across subjects. We believe we only clearly succeeded in one task, the CL. Only two out of eleven subjects (one in the pilot and one in the research) described all of the tasks as we designed them. Simplifying all possible emotions to a 2D-model already in the beginning of the experiment might also be limiting. All our subjects described very phenomenologically different feelings during their tasks. Thus we would not draw any conclusion about our proposed 2D-model of emotion with EDA as being the marker of arousal and HRV as the marker for valence.

However, we think we can say with some assurance that we succeeded in measuring arousal or sympathetic nervous system activity by measuring EDA. This gives further evidence for EDA being a viable measurement of emotional arousal. EDA as psychophysiocal method for UX seems ripe for application, and it could be used in real-life UX testing or monitoring. Even in naturalistic, poorly controlled settings with a wrist-worn sensor, we could get highly constant data across all but one subject. And from this data we could differentiate between different emotional states, both on general and individual level.

HRV, on the other hand was not consistent. Thus conclusions from our HRV results are much more complicated and vague, especially if we would like to use HRV as a marker of valence. We think this is a two-faceted problem: the complexity of measurement and the complexity of the semantics. First of all, HRV is a complex metric - even by definition. It is lways convoluted by two different neural activations, the SNS and PNS. Ideally, one would only have one system activation as with EDA. Similarly, it has a very poor reproductability in studies [17]. In addition, heart-beat detection from wrist-worn device is a very noisy method. Second of all, the question of valence is hard: both philosophically and practically. We could not distinguish valence from HRV - at least not in the way we expected. We succeeded in grouping them, but results were inconsistent with literature. Thus we were probably measuring something else (low construct validity). Based on newer research the whole idea of grouping HRV into single-axis might be wrong [141]. Arousal is most likely "a general activation level of your (sympathetic) nervous system", and thus quite readily measurable. But valence is something much less tangible and not necessarily something that could be read only from neural activation. It is more context dependent. Maybe it should be thought as a derivative: a change towards a more stable state of being (homeostasis) is considered a positive one, no matter what the starting point is. Where as a step toward a less stable state of being is considered bad (allostasis). And even that is a dubious claim as people arguably can describe biologically frightful or painful experiences as pleasant.

5.4 Interpretations

Due to our problematic results and mistakes in our experimental design, we started to think that we might be looking at emotions from a completely wrong angle. Now most of research - including us - draws too early conclusions about emotions and their respective phychophysiological markers. We are paying a lot of interest into the interaction, and saying that a certain interaction always causes a certain emotional reaction (see figure 25). But this might be too simplistic. We are forgetting that what happens in the head of the user, how they process the information, is probably much more important and relevant than the actual stimuli if we want to measure

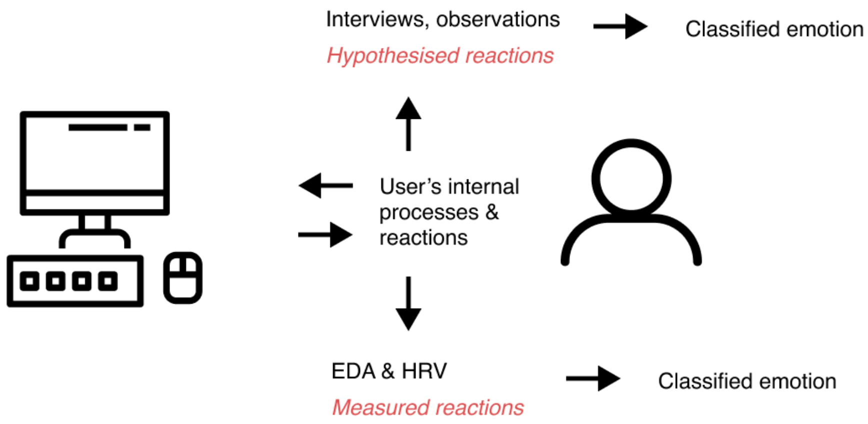

Figure 25: The black-box model of human-computer interaction.

affect. Saying that "this reaction causes that emotion and we can measure it's signals" treats the emotion as a black box system, making studying it hard. This is an inherent problem with subjective qualitative reports such as interviews and observations as well - they treat emotions as responses to certain stimuli (in this case the human-computer interaction).

If we look at the research by Spinoza, Damasio, Herbert, Kahneman and Norman, we should realise that emotion is not a simple reactionary process but a key component of human intellect. That "black box" can be an extremely complex optimization process that involves all the personal memories of the person combined with every possible sensory input coming in. It should not be any wonder, that then inducing any certain specific emotion is hard. We can only control the inputs, and the black box, which the person processes into an output - emotional reaction. Thus, we can not start with saying that "this input will cause that output". A simple stimuli-reaction model might be even wrong if we remember what Damasio said about primary and secondary emotions. Only primary emotions would be constant and similar across all subjects due to our common biological ancestry. The secondary emotions, the more complex ones, would vary from person to person due to their different life stories. Thus a complex stimuli such as human-computer interaction would cause very different emotional reactions in different subjects as everybody would associate different memories with that stimuli. Furthermore, the goals or needs of a user are not necessarily set upon the beginning, but rather constantly re-evaluated and reformed in time by the new sensory input coming in.

Thus we think we might have started with a possible false pretense, that we can induce a certain emotion in the person and that this emotion could be read from the

SCRs and HRV. Where as we should be paying attention to the cognitive processes underlying the emotional processing and understand, at which point the SCR was caused and then only start pondering why this happened. Say, if an SCR might originate from multiple different brain areas, then it could be caused by different cognitive processes as well. Thus depending on that temporal context, also the meaning of SCR changes and it is not sufficient to draw conclusions only from their amplitude or frequency during some task. Just as important would be to understand what was the context - what happened in that user's head that lead to that SCR.

To further explain this, let's look at this problem through the lens of information-processing view of HCI laid out by Don Norman in [126]. A visualisation can be seen in figure 26. Here the user interacts with the computer. The user has a goal, and creates an action sequence and those actions are then fed-forward to the computer. The computer then processes that action and gives a response through the user interface, which is then interpreted by the user. That interpretation is then evaluated by the user, how it is related to their goals and and a new intention to act is then created, thus beginning the cycle again. This framework has been applied to many common best practices in the industry and courses in the university lectures. Norman argued, that much of the problems in HCI arise from two exact points in that cycle: the gulfs of evaluation and execution. Gulf of execution means that even though the user knows what they want, the software doesn't enable it. Either it is not possible or the user doesn't know how to do it in the system. Thus the possibilities of the system didn't meet the intended actions. The gulf of evaluation means whether the system could provide representations that could be directly perceived and interpreted in terms of the expectations and intentions of the user. If the system represents its state poorly or it is otherwise misleading, the user will do false evaluations of what happened.

Now, if we look at much of the research surrounding affect and psychophysiology in HCI research, we could argue that this is the view that dominates the results and the experiments. We think in terms of "cognitive load" or "mental stress" or that "HRV can distinguish usability problems". The underlying suggestion is that it would be those gulfs of executions and evaluation (or some other part of that information processing) that cause these reactions. The emotion rises when something doesn't go according to the user's plan or there is annoying thing such as a pop-up. This is already a much more helpful distinction, because it can then help us classify the SCR reactions or HRV readings into certain categories: was this caused by an interaction problem in the UI during execution? Or was it only mental load induced by a challenging task? Both can of course be negative or positive depending on the context. Mental load during a challenging game play would most likely be welcome, but not something that you want to experience when ordering a delivery pizza.

However, albeit more useful, the problem with this information processing view is that in doesn't really take any stand about emotions and what they are. It still treats them as reactions to problems or internal changes in the user. Furthermore, it treats all humans as more or less similar in this information-processing aspect, leaving little room for personal differences, and it has the claim that all information would be processed - but much of it could be just behavioural automation. After all,

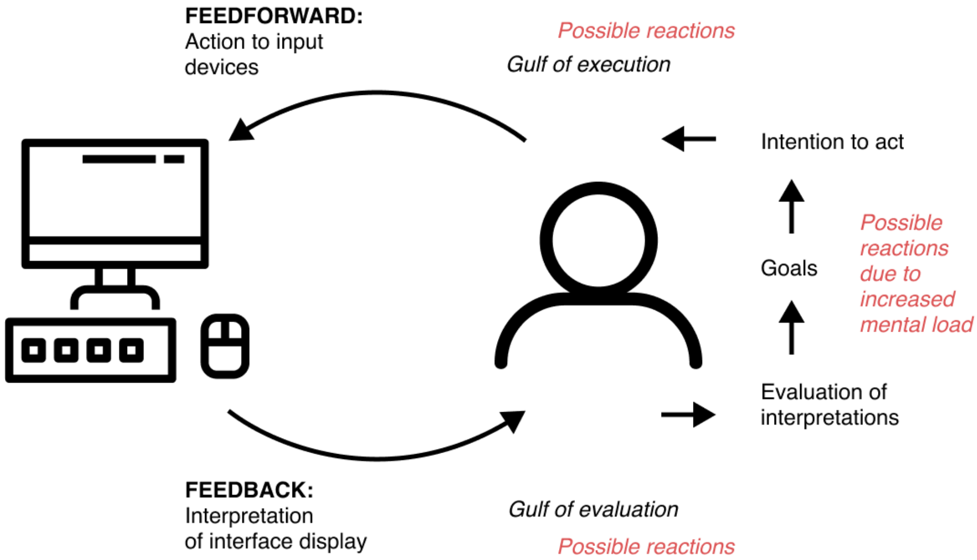

Figure 26: The information processing model of human-computer interaction.

people don't necessarily process information when acting with computers: they start yelling at it, because it failed to do something. From the information processing point of view, this kind of emotional behaviour doesn't make any sense, because if the information would be processed, it should be clear that these kinds of behaviours have absolutely no value - they can even be detrimental. So, we argue that the actual cognitive processing should be something different. For this we will use Don Norman's Three Levels of Affcetive processing framework (remember figure 5. This visualisation can be seen in figure 27.

In this model, which we call the cognitive-affective model of human-computer interaction, the information processing starts already the moment stimuli is sensed by the sensory system and this information is fed to the quickest level - the visceral. This system does snap judgements about the goodness and badness of our surroundings and it also alerts behavioural and emotional reactions in the user. This already can lead to action by the user: say in a game, it might be a sudden change that requires quick reflex actions. Or if watching an emotional movie, it might quickly elicit a feeling of sadness in you and make you cry. This level can already cause SCRs and changes in HRV even without the conscious appraisal or recognition by the user.

The information would then be fed forward to the behavioural level, which is the seat of most human conscious perception and where all the more complex behaviour takes place. Here the SCRs or HRV changes would be caused by for example postural changes or highly complex motor movements. All those could cause notable readings in biosignals, but they wouldn't necessarily be emotionally significant. The emotional judgement has already happened at the visceral level, which is now only affecting our behaviour.

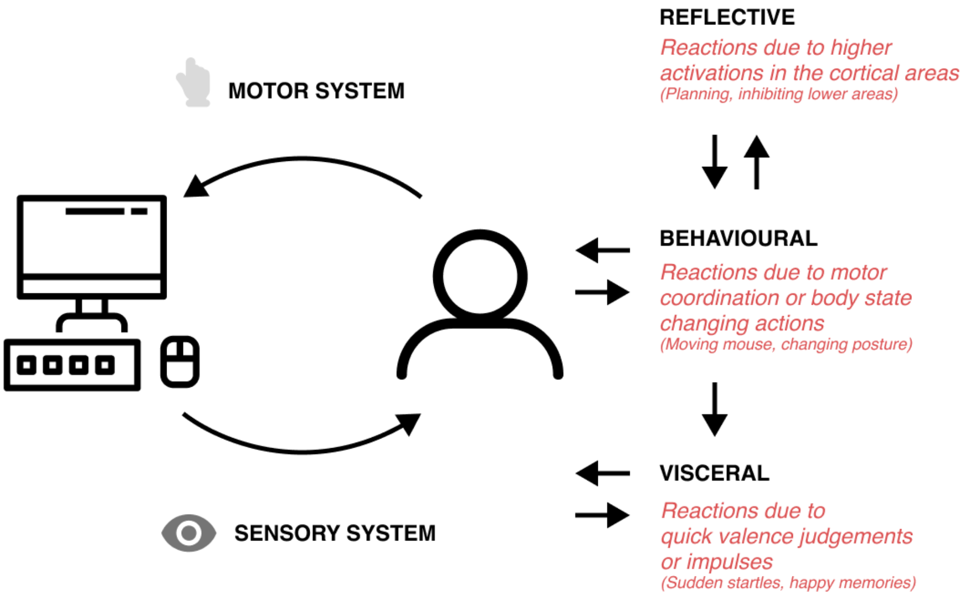

Figure 27: The cognitive-affective model of human-computer interaction.

The information could be then fed-forward even further (but it isn't most of the time) to the reflective level, which is the cognitive and intellectual one and which watches over the behaviour level. By denifition, this reflective level wouldn't cause any SCRs or change is HRV through any behaviour. But in here the changes would be caused for example by increased mental load or increased cognitive load. For example, activation in the frontal cortex would lead to inhibition of other areas and would then increase the general arousal level. In addition, with the combination of visceral level this is the level that could lead the person into extreme emotional thought loops, if an enough salient emotional stimuli was presented. Say, by triggering a traumatic memory or if a sudden computer crash just destroyed your only back-up of your thesis.

Although this study was not specifically designed to create any new theoretical frameworks, it was something that arouse during this exploratory study. We think that this last model, the cognitive-affective model of HCI, would be the most accurate one to describe what happens in the user's head during an emotional reaction. This is due to following reasons. First of all, it is not a completely new theory. It takes existing models from cognitive psychology, design science and neuroscience and fuses them together by their connected points. Second of all, there is neuroscientific evidence showing that EDA reactions are elicited by multiple different brain areas [37] and these areas are roughly equivalent to the Norman's theory of three levels of design [125]:

1. Amygdala (affectional processes) - visceral level

2. Premotor cortex (fine-motor control) - behavioural level

3. Prefrontal cortex (attention & planning) - reflective level

Third of all, it would explain very well the peculiarities in our results. Why did we record so many SCR reactions during the LAPV task for some? Because they hated the music, thus leading to arousal by the visceral level, or because the music evoked happy memories in them, leading to arousal by the reflective level. And why was the Excel task so arousing one? Because even though it doesn't require any cognitive processing from the person, it still required a lot of fine motor control thus causing a lot of SCRs. Furthermore, in one person it started a reflective thought loop.

This model wouldn't also be blind to personal differences in information processing. If a person would score high on the rational processing, they would probably use more reflective processing. And if they were experiential they would probably use more visceral processing. These kind of straight interpretation would of course need to be validated by experiments, and we are not saying it is like that. We are bringing this up, because it shows that Rational-Experiential model is attachable to this theory. Our model would also explain the correlations between Affect Intensity and recorded SCRs. If it truly is so, that people with high AIM score are using cognitive reappraisal strategy, this model would predict low SCRs. If the person automatically translates everything as "good" already in the visceral level, then that visceral level doesn't need to arouse anything in the behavioural level. We are already closer to homeostasis and no drastic changes need to be made. Most SCRs would only be caused by reflection, complex motor movements or if something is truly non-likeable by the user. However, if the person is using suppression strategy (which could be predicted by a low AIM score), then the visceral level is judging much more stimuli as "bad" and alerting the behavioural level to increase it's alertness. But because this is not necessarily beneficial as judged by the reflective level, then it also needs to keep the visceral level in control - which it cannot, but it can only suppress the behavioural level. This is also what suppression strategy suggests: only behavioural level is affected, not the experience - especially in the case of negative emotions [62].

This kind of theoretical framework could be a fruitful future approach to measuring affect in HCI. It dodges the semantical and phenomenal problem of "what is an emotion" by not taking into account what the actual emotion is. We are only interested, that where during emotional processing the emotional reaction and it's respective biosignal happened: visceral, behavioural or reflective? In some cases, this theory could also be more useful than purely utilistic and functional information processing views, because it could take into account contextual information (such as reaction time) and help us thus to categorize reactions (e.g. a quick reaction time would most likely indicate visceral reaction). We could then ask the users what do they think happened there? In addition this model could take personality as a priori into account: personal preferences in individual processing styles could help to normalise base levels of biosignals and to understand further in which stage the emotional reaction happened.

5.5 Recommendations for further research

This work has already led into further research at the case company. We validated what metrics should be used for naturalistic usability studies from biosignals. The best one would be EDA, as it is easy to study in temporal dimension and it is much more resistant to noise and easier to analyse than HRV. We advised against using HRV. In addition, our theoretical background and framework provided insights to what different SCRs could mean in different contexts.

We should still look into HRV readings in other research. We did measure task differences and personal differences with HRV. Our suggestion is to think of it in a different way than just "stress" or valence. It is a complex metric, not a simple indicator. The viability of multidimensional metrics and Poincaré plots should be researched further. A paper about using Poincaré plots to analyse HRV readings was released during summer 2019, when this thesis was in the making, further suggesting a possible methodology that could work for the analysis of HRV [96].

In addition, our statistically significant finding about AIM correlating with EDA seems interesting. Further research studying specifically that and with more controlled groups and bigger sample size should be done.

For automated wrist-based emotion recognition in general, the possibilities of machine learning systems should be researched. The power of machine learning in automated SCR recognition was already shown by the functionality of EDAexplorer [170], and it was extremly helpful for this research as well. The same approach could be used for HRV and for the combination of HRV and EDA. The findings from this thesis could be used for feature engineering. Also, we should not necessarily try straight 2D-fits of data with HRV being one axis and EDA being one axis, but rather using them as weighted features. For example, we could have the frequency of SCRs per minute on one axis and tonic EDA level on one axis. Also, in that case the HRV might benefit from much more complex representations, such as the 2D-space that we described with the support vector machines. The hardest part would be then to understand what measurements actually carry relevant information for any specific emotion. Also, our suggested new framework to understand emotions could be used as a mean to collect ground truths. This would require new experimental designs.

We thus propose a new idea for experimental design (see figure 28 for a visualisation). We shouldn't start with ground truths that are labeled and decided by the researchers. Because emotions are so personal and subjective, the labels should come from the subjects themselves. However, as emotions are at least to some degree unconscious it might be really hard for the subject to elicit a real emotion at will. Thus instead of trying to elicit reactions with already laid out labels, we should either present stimuli or gather data during real-life and wait for a reaction and then try to find out in which information processing phase did it happen, what part of the stimuli caused it, and how did the subjects experience it. By interviewing we could ask the subject, what had happened: what did they feel, what do they think caused that reaction and ask them to describe their internal process. This should be done with as little distraction as possible, because the anticipation [16, p. 3] alone could be enough to cause a reaction. We could use for example an app that would notify the user and ask them to label the event and describe with their own words what happened.

Now then if we would have multiple subjects, we would at some point have simultaneous reactions and similar reactions. For the simultaneous reactions we could use the subjective interviews to understand how the experiences were different and how that shows in the data. Or if we would have found out two very similar reactions at completely different times, we could use the subjective interviews to find out what made that experience so similar between the subjects.

Psychological surveys, such as AIM and REI, could also be used to quantify how people are different (or similar) and to group subjects into different categories. This would also help with the labeling of events, because if there would be a negative correlation with EDA and affect intensity, we would need to normalise the base-levels. E.g. if there would be two subjects, one whose affect intensity is high and one whose affect intensity is low, then their respective emotional reactions during a similar experience would give very different readings. I.e. a low AIM -person (who thus has high EDA readings) could react to stimuli A with 4.7 SCRs/min, and a high AIM -person (who thus has low EDA readings) could react to stimuli B with 0.7 SCRs/min. But because we know from quantified psychological constructs that the baselines for these two groups (high and low AIM) are different and negatively correlated, we could then treat these reactions as equal. Thus with this approach we could gather ground truths for automated emotion recognition system. Instead of static emotion labels for various reactions, we would have context-reaction-pairs for real word data from emotional reactions. In addition, we could use personality as a priori for feature engineering and data normalisation.

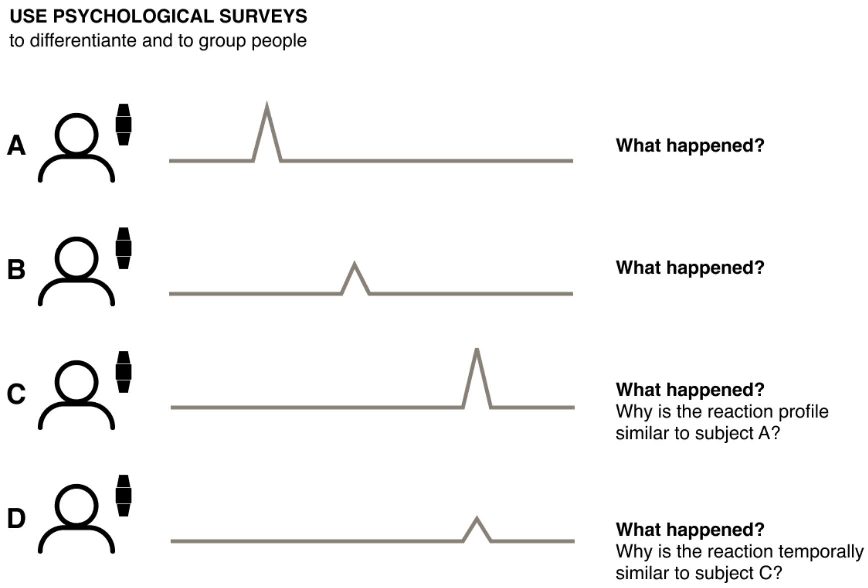

Figure 28: A suggested experimental design.

6 Summary

In this thesis, we went through the theory of emotion from many different fields of science: philosophy, psychology, neuroscience, economics and finally, human-computer interaction. We showed, how they all have very similar constructs and ideas about emotions, yet each one of them has their blind spots. We argued, that the weak point of the study of emotion in human-computer interaction is the way how simplistically it looks at emotions. Much of the research in HCI deals with emotions as constructs that are something very discrete and which can be easily produced by certain stimuli and which are similar across all subjects.

We can't give any exhaustive answer to our first research question "what is an emotion in the context of human-computer interaction", but we can say that emotion is not something simple or easily excited. Emotions are very complex phenomena that arise either from our bodily sensations and then affect our conscious thoughts or from our conscious thoughts and then affect our bodily sensations. They are something unconscious, but something whose presence can be consciously noted. An emotion and especially emotional reaction is always something very personal and unique, but the variety of emotions is quite similar across all subjects; two people might not have the exact same emotional reaction to a certain stimuli, but those people will still have a similar range of emotions.

We must also agree with the literature that emotions are important. Thus to answer our second research question, "why are emotions important for human-computer interaction", we can say the following: emotions are a crucial part of any experience including human-computer interaction. They seem to form a very core component of our decision making. When computers are used in situations that require high impact decision making, then emotional reactions caused by the technology should be taken into account. Furthermore, the development of affective computing or emotion aware technologies is important for any natural human-computer interaction. Human-computer interaction must not be as emotional as human-human interaction is, but it should be taken into account. And this is not a sci-fi vision into the future or robotics, but already an action point for the development of user interfaces. Total user experience is becoming more and more relevant than just pure usability. The computer as it was, was an extremely rational tool, used to solve problems and enhance calculations. The inputs that were taken in include typed in letters and hand movements. With the age of personal computing a change has come: computers are more and more about lifestyle and identity. Inputs include now touch screens and biometrics. Communication is turning from written text into emojis and other affective signals. But at this point we must exactly not necessarily do more, but do better. If we want to include emotions into our interaction with computers, shouldn't it be done so that it increases our human potential, rather than abuses our human limitations? And this is not only an ethical judgement, but a rather honest direction to point out for further research. When our technological capabilities increase and we live more and more in an artificial environment, we must also think about "what should be done" instead of "what is possible". This means, that there should be an increase in the research about ethical implications of emotion aware technologies. In the coming age of ambient intelligence this kind of thinking is increasingly important.

To answer our third research question, we can say that Empatica E4 turned out to be a viable tool for measuring EDA, but not necessarily HRV. The advantage of Empatica was an automated script that removed noisy heartbeats from the data; the disadvantage was that this noisy data covered most of our data. Thus for real-live measurements we doubt the usefulness of Empatica or other PPG heart-beat measurements, especially if we are interested in short-term metrics. Although HRV looks promising, the technology is still not quite there. Major problem is the noisiness of data and the inability to have any reliable short-term metrics with wrist sensors. For scientific and real-time data gathering we would consider an ECG still to be the better choice.

As for the measurement of arousal (RQ4), we agree that EDA is a good choice. EDA has all the things needed, reliability, objectivity and it can be measured unobtrusively. We did not find this to be the case for HRV. But for measuring valence (RQ5), we have a slightly different conclusion. There were differences for valence that could be seen in EDA, but we are not sure whether it was due to different valence judgement or just as side product of arousal or failed experiment design. However, with evolutionary psychology in mind it could be valid to say that negative experiences are more arousing than positive ones. Or it could just be a construct bias, because negative experiences are much more easier to elicit by experimental design then positive ones. There were differences for valence that could be measured with HRV too, but these were opposite to what we were expecting. Thus the answers are inconclusive for HRV. Valence is also philosophically and phenomenologically a fuzzy question (e.g. who is to say, what is good and what is bad?) and thus our research question might not even be meaningful at all.

When it comes to our research question number 6, "how to take personality into account for psychophysiological measurements", we had significant results. Personality as a priori for psychophysiological metrics was a novel idea presented in this thesis. Even though the sample size was small, we did find statistically significant correlations between personality constructs and our measurements, which were also as hypothesised.

This leads to our initial hypotheses. From our experiments and based on literature we would say that our hypothesis H1 is a right one: EDA is an indicator of emotional arousal. Similarly, we can not accept hypothesis H2 as true: HRV is not necessarily an indicator of emotional valence. Even though the hypothesis was suggested and supported by literature, we did not find evidence for ut. However, our measurements with HRV had experimental problems in both experiment designs and measurements. For personality and AIM, we proposed in H3 that "the self-assessed affect intensity (AIM) is negatively correlated with the intensity of psychophysiological metrics". We did find a statistically significant, negative correlation between AIM and EDA, leading us to reject the null-hypothesis and to suggest proof for our hypothesis. People who report having intense affections do not show them with EDA. The possible mechanism for this would be the preferred emotion regulation strategy by the subject. However, we did not find any reverse effect (people with high negative score didn't necessarily predict high SCRs) and there is an internal bias in AIM survey to measure positive affections. These things should be kept in mind and controlled if a further study with a bigger sample size would be conducted.

In addition, inspired by the oddness of our data and the problems within our experiment designs, we suggested a new framework to understand emotions better in the context of human-computer interaction. Instead of classifying emotions early up and dealing with them as black box input-output systems, we should understand their complexity and how they play a crucial part in our decision making and information processing. Then instead of measuring the abstract emotion labels, we should pay more attention to that processing itself and where it happens. The studies would then of course need to be designed differently: we shouldn't start with straight-up emotion categories or theories, and try to group people's experiences into them. Instead we should first gather emotionally significant experiences irrelevant of the stimuli. The psychophysiological measurements from those significant experiences would then be combined with subjective reports. From that data, we could then start to investigate why some reactions are similar or why they happen at the same time. In addition, we could start building up much more specific "emotional profiles" of subjects. If a person is known to experience frustration every time they see an pop-up, we could use that as a ground truth for frustration in further experiments. In addition we could compare their frustration and see how it is different or similar to other people's frustration. This means that the ground truth problem needs to be thought differently. The same problem was already pointed out by Scheirer et al. [149]:

> Assigning these labels or categories to the data is a non-trivial problem, which deserves careful consideration since the class categorisations we shall use to label the data have only been induced not firmly established. In other words, there is uncertainty associated with the class to which the data belongs. There is, for instance, a possibility that a stimulus failed to induce a frustration response, and conversely, that a subject showed a frustration response in the absence of the controlled stimulus due to another uncontrolled stimulus, such as a cognitive event.

To understand emotions further in HCI and to provide the needed rigor into the theory of user experience, we need more theoretical frameworks deal with low-level abstractions of the human cognition to understand how experience forms. That theory needs to be backed up by experimental set-ups that draw inspiration from other sciences. We can't just say that emotion is a classification problem because it necessarily isn't - or at least not one we can easily measure. But we could measure the classification of the different emotional processing phases, a model which does have evidence from cognitive psychology and neuroscience, and get insights to user's internal process through that. Human-computer interaction has done a lot to achieve the needed rigour to achieve the status of scientific discipline, and for that we have received inspiration and knowledge from other fields. We shouldn't choose the easy way out with emotion recognition, but study the topic thoroughly. Furthermore, this could be an opportunity to give back to other scientific fields, as human-computer interaction provides an unique and interesting lens through which we can understand and analyse human experience in our increasingly technological society.

References

[1] *affect | Definition of affect in English by Oxford Dictionaries.* URL: https://en.oxforddictionaries.com/definition/affect (visited on 03/29/2019).

[2] Jenni Anttonen and Veikko Surakka. "Emotions and heart rate while sitting on a chair". In: *In Proceedings of the CHI 2005 SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2005, pp. 491–499.

[3] Bradley M. Appelhans and Linda J. Luecken. "Heart Rate Variability as an Index of Regulated Emotional Responding". en. In: *Review of General Psychology* 10.3 (Sept. 2006), pp. 229–240. ISSN: 1089-2680, 1939-1552. DOI: 10.1037/1089-2680.10.3.229. URL: http://journals.sagepub.com/doi/10.1037/1089-2680.10.3.229 (visited on 08/16/2019).

[4] Leena Arhippainen and Marika Tähti. "Empirical Evaluation of User Experience in Two Adaptive Mobile Application Prototypes". en. In: (), p. 8.

[5] Shuotian Bai, Tingshao Zhu, and Li Cheng. "Big-Five Personality Prediction Based on User Behaviors at Social Network Sites". In: *arXiv:1204.4809 [cs]* (Apr. 2012). arXiv: 1204.4809. URL: http://arxiv.org/abs/1204.4809 (visited on 09/20/2019).

[6] Jay S. Balachandran et al. "PNN10 And PNN20, Markers Of Vagal Tone, Are Decreased In Obese Patients With Obstructive Sleep Apnea Compared To Controls". In: *A73. SLEEP DISORDERED BREATHING AND CARDIOVASCULAR IMPAIRMENT*. American Thoracic Society International Conference Abstracts. American Thoracic Society, May 2011, A2212–A2212. DOI: 10.1164/ajrccm-conference.2011.183.1_MeetingAbstracts.A2212. URL: https://www.atsjournals.org/doi/abs/10.1164/ajrccm-conference.2011.183.1_MeetingAbstracts.A2212 (visited on 09/11/2019).

[7] Gene Ball and Jack Breese. "Relating Personality and Behavior: Posture and Gestures". en. In: *Affective Interactions: Towards a New Generation of Computer Interfaces*. Ed. by Ana Paiva. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 196–203. ISBN: 978-3-540-44559-3. DOI: 10.1007/10720296_14. URL: https://doi.org/10.1007/10720296_14 (visited on 09/20/2019).

[8] Antoine Bechara and Antonio R. Damasio. "The somatic marker hypothesis: A neural theory of economic decision". In: *Games and Economic Behavior*. Special Issue on Neuroeconomics 52.2 (Aug. 2005), pp. 336–372. ISSN: 0899-8256. DOI: 10.1016/j.geb.2004.06.010. URL: http://www.sciencedirect.com/science/article/pii/S0899825604001034 (visited on 08/01/2019).

[9] Frank Beckers, Bart Verheyden, and André E. Aubert. "Aging and nonlinear heart rate control in a healthy population". In: *American Journal of Physiology-Heart and Circulatory Physiology* 290.6 (June 2006), H2560–H2570. ISSN: 0363-6135. DOI: 10.1152/ajpheart.00903.2005. URL:

https://www.physiology.org/doi/full/10.1152/ajpheart.00903.2005 (visited on 09/11/2019).

[10] Soroor Behbahani, Nader Jafarnia Dabanloo, and Ali Motie Nasrabadi. "Ictal Heart Rate Variability Assessment with Focus on Secondary Generalized and Complex Partial Epileptic Seizures". In: 2013.

[11] Hans-Rudolf Berthoud and Winfried L Neuhuber. "Functional and chemical anatomy of the afferent vagal system". In: *Autonomic Neuroscience*. Fever:the role of the vagus nerve 85.1 (Dec. 2000), pp. 1–17. ISSN: 1566-0702. DOI: 10.1016/S1566-0702(00)00215-0. URL: http://www.sciencedirect.com/science/article/pii/S1566070200002150 (visited on 09/10/2019).

[12] George E. Billman. "The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance". In: *Frontiers in Physiology* 4 (Feb. 2013). ISSN: 1664-042X. DOI: 10.3389/fphys.2013.00026. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3576706/ (visited on 02/01/2019).

[13] M.H. Bonnet and D.L. Arand. "Heart rate variability: sleep stage, time of night, and arousal influences". en. In: *Electroencephalography and Clinical Neurophysiology* 102.5 (May 1997), pp. 390–396. ISSN: 00134694. DOI: 10.1016/S0921-884X(96)96070-1. URL: https://linkinghub.elsevier.com/retrieve/pii/S0921884X96960701 (visited on 08/16/2019).

[14] Wolfram Boucsein. *Electrodermal Activity*. en. Google-Books-ID: 6N6rnOEZEEoC. Springer Science & Business Media, Feb. 2012. ISBN: 978-1-4614-1126-0.

[15] Margaret M. Bradley and Peter J. Lang. "Affective reactions to acoustic stimuli." In: *Psychophysiology* 37.2 (2000), pp. 204–215. DOI: 10.1111/1469-8986.3720204.

[16] Jason Braithwaite et al. *A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments*. 2015. URL: https://www.birmingham.ac.uk/Documents/college-les/psych/saal/guide-electrodermal-activity.pdf (visited on 01/30/2019).

[17] Sigrid Breit et al. "Vagus Nerve as Modulator of the Brain–Gut Axis in Psychiatric and Inflammatory Disorders". In: *Frontiers in Psychiatry* 9 (Mar. 2018). ISSN: 1664-0640. DOI: 10.3389/fpsyt.2018.00044. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5859128/ (visited on 09/10/2019).

[18] M. Brennan, M. Palaniswami, and P. Kamen. "Do existing measures of Poincare plot geometry reflect nonlinear features of heart rate variability?" In: *IEEE Transactions on Biomedical Engineering* 48.11 (Nov. 2001), pp. 1342–1347. DOI: 10.1109/10.959330.

[19] Michael Brennan, Marimuthu Palaniswami, and Peter Kamen. "Poincaré plot interpretation using a physiological model of HRV based on a network of oscillators". In: *American Journal of Physiology-Heart and Circulatory Physiology* 283.5 (2002), H1873–H1886. DOI: https://doi.org/10.1152/ajpheart.00405.2000. (Visited on 09/11/2019).

[20] T. E. Brown et al. “Important influence of respiration on human R-R interval power spectra is largely ignored”. In: *Journal of Applied Physiology* 75.5 (Nov. 1993), pp. 2310–2317. ISSN: 8750-7587. DOI: 10.1152/jappl.1993.75.5.2310. URL: https://www.physiology.org/doi/abs/10.1152/jappl.1993.75.5.2310 (visited on 09/11/2019).

[21] Tim Brown and Barry Katz. “Change by Design”. en. In: *Journal of Product Innovation Management* 28.3 (2011), pp. 381–383. ISSN: 1540-5885. DOI: 10.1111/j.1540-5885.2011.00806.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-5885.2011.00806.x (visited on 08/22/2019).

[22] Fred B. Bryant, Paul R. Yarnold, and Laurence G. Grimm. “Toward a Measurement Model of the Affect Intensity Measure: A Three-Factor Structure”. en. In: *Journal of Research in Personality* 30.2 (June 1996), pp. 223–247. ISSN: 00926566. DOI: 10.1006/jrpe.1996.0015. URL: https://linkinghub.elsevier.com/retrieve/pii/S009265669690015X (visited on 09/30/2019).

[23] Ross Buck. “Mood and Emotion: A Comparison of Five Contemporary Views (Book)”. en. In: *Psychological Inquiry* 1.4 (Oct. 1990), pp. 330–336. ISSN: 1047-840X, 1532-7965. DOI: 10.1207/s15327965pli0104_15. URL: http://www.tandfonline.com/doi/abs/10.1207/s15327965pli0104_15 (visited on 03/29/2019).

[24] John T. Cacioppo and Richard E. Petty. “The need for cognition.” en. In: *Journal of Personality and Social Psychology* 42.1 (1982), pp. 116–131. ISSN: 0022-3514. DOI: 10.1037/0022-3514.42.1.116. URL: http://content.apa.org/journals/psp/42/1/116 (visited on 12/19/2018).

[25] Erik Cambria. “Affective Computing and Sentiment Analysis”. In: *IEEE Intelligent Systems* 31.2 (Mar. 2016), pp. 102–107. ISSN: 1941-1294. DOI: 10.1109/MIS.2016.31.

[26] A. J. Camm et al. “Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology”. eng. In: 93.5 (1996), pp. 1043–1065. DOI: 10.1161/âĂŇ01.CIR.93.5.1043. URL: https://air.unimi.it/handle/2434/188038#.XXdydpMzZhE (visited on 09/10/2019).

[27] D. J. Chalmers. *Facing up to the problem of consciousness.* en. Text. Mar. 1995. URL: https://www.ingentaconnect.com/content/imp/jcs/1995/00000002/00000003/653 (visited on 08/27/2019).

[28] Joseph Ciarrochi, Frank P Deane, and Stephen Anderson. “Emotional intelligence moderates the relationship between stress and mental health”. In: *Personality and Individual Differences* 32.2 (Jan. 2002), pp. 197–209. ISSN: 0191-8869. DOI: 10.1016/S0191-8869(01)00012-5. URL: http://www.sciencedirect.com/science/article/pii/S0191886901000125 (visited on 08/01/2019).

[29] Christian Collet et al. “Autonomic nervous system response patterns specificity to basic emotions”. In: *Journal of the Autonomic Nervous System* 62.1 (Jan. 1997), pp. 45–57. ISSN: 0165-1838. DOI: 10.1016/S0165-1838(96)00108-7. URL: http://www.sciencedirect.com/science/article/pii/S0165183896001087 (visited on 09/12/2019).

[30] *Competitive Landscape: Emotion AI Technologies, Worldwide*. en. URL: https://www.gartner.com/en/documents/3970212/competitive-landscape-emotion-ai-technologies-worldwide (visited on 11/22/2019).

[31] “Counting every moment”. In: *The Economist* (Mar. 2012). ISSN: 0013-0613. URL: https://www.economist.com/technology-quarterly/2012/03/03/counting-every-moment (visited on 10/11/2019).

[32] Hugo D. Critchley. “Review: Electrodermal Responses: What Happens in the Brain”. en. In: *The Neuroscientist* 8.2 (Apr. 2002), pp. 132–142. ISSN: 1073-8584. DOI: 10.1177/107385840200800209. URL: https://doi.org/10.1177/107385840200800209 (visited on 09/16/2019).

[33] Damasio Antonio R. et al. “The somatic marker hypothesis and the possible functions of the prefrontal cortex”. In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 351.1346 (Oct. 1996), pp. 1413–1420. DOI: 10.1098/rstb.1996.0125. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rstb.1996.0125 (visited on 05/13/2019).

[34] Antonio Damasio. *Descartes’ error*. Random House, 1994. ISBN: 978-84-9892-166-3.

[35] Antonio Damasio. *Looking for Spinoza: Joy, sorrow, and the feeling brain*. Houghton Mifflin Harcourt, 2003. ISBN: 978-0-434-88345-5.

[36] Antonio Damasio. *The feeling of what happens: Body and emotion in the making of consciousness*. Houghton Mifflin Harcourt, 1999. ISBN: 978-0-15-601075-7.

[37] Michael E. Dawson, Anne M. Schell, and Diane Filion. “Chapter 8: The electrodermal system”. In: *Handbook of psychophysiology*. 3rd edition. Vol. 2. 2007, pp. 200–223.

[38] Liqiong Deng et al. “User experience, satisfaction, and continual usage intention of IT”. In: *European Journal of Information Systems* 19.1 (Feb. 2010), pp. 60–75. ISSN: 0960-085X. DOI: 10.1057/ejis.2009.50. URL: https://orsociety.tandfonline.com/doi/abs/10.1057/ejis.2009.50 (visited on 09/20/2019).

[39] Rene Descartes. *Discourse on the Method: Of Rightly Conducting One’s Reason and of Seeking Truth in the Sciences*. en. Google-Books-ID: 7Ck-LvOBHnr4C. The Floating Press, Dec. 2009. ISBN: 978-1-77541-687-6.

[40] René Descartes. *Passions of the Soul*. en. Google-Books-ID: Ca9aBAAAQBAJ. Hackett Publishing, Dec. 1989. ISBN: 978-1-62466-198-3.

[41] *Donald Norman's Three Levels of Design*. en. URL: https://www.interaction-design.org/literature/article/donald-norman-s-three-levels-of-design (visited on 10/08/2019).

[42] *E4 data - EDA Expected signal*. en-US. URL: http://support.empatica.com/hc/en-us/articles/360030048131-E4-data-EDA-Expected-signal (visited on 07/29/2019).

[43] R Edelberg. "Mechanisms of electrodermal adaptations for locomotion, manipulation, or defense." In: *Progress in physiological psychology* 5 (1973), pp. 155–209.

[44] P. Ekman, R. W. Levenson, and W. V. Friesen. "Autonomic nervous system activity distinguishes among emotions". en. In: *Science* 221.4616 (Sept. 1983), pp. 1208–1210. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.6612338. URL: https://science.sciencemag.org/content/221/4616/1208 (visited on 09/12/2019).

[45] N M Enewoldsen. "Analysis of the quality of electrodermal activity and heart rate data recorded in daily life over a period of one week with an E4 wristband". en. In: (), p. 97.

[46] Seymour Epstein. "Cognitive-Experiential Self-Theory: An Integrative Theory of Personality". en. In: *The Relational Self: Theoretical Convergences in Psychoanalysis and Social Psychology*. Google-Books-ID: ELc63AwqAcAC. Guilford Press, Mar. 1991. ISBN: 978-0-89862-558-5.

[47] Seymour Epstein. "Integration of the Cognitive and the Psychodynamic Unconscious". en. In: *American Psychologist* (1994), p. 16.

[48] Seymour Epstein. "The self-concept revisited: Or a theory of a theory." en. In: *American Psychologist* 28.5 (1973), pp. 404–416. ISSN: 1935-990X, 0003-066X. DOI: 10.1037/h0034679. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/h0034679 (visited on 12/19/2018).

[49] Seymour Epstein et al. "Individual differences in intuitive–experiential and analytical–rational thinking styles." en. In: *Journal of Personality and Social Psychology* 71.2 (1996), pp. 390–405. ISSN: 1939-1315, 0022-3514. DOI: 10.1037/0022-3514.71.2.390. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.71.2.390 (visited on 12/19/2018).

[50] S. Freud. "Repression". In: *The Standard Edition of the Complete Psychological Works of Sigmund Freud Volume XIV (1914-1916): On the History of the Psycho-Analytic Movement, Papers on Metapsychology and Other Works,* (1915). URL: https://www.pep-web.org/document.php?id=SE.014.0141A (visited on 09/20/2019).

[51] Eva Ganglbauer et al. "Applying Psychophysiological Methods for Measuring User Experience". In: *Possibilities, Challenges, and Feasibility. Proc. User Experience Evaluation Methods in Product Development workshop*. 2009.

[52] Paul van Gent. *Analyzing a Discrete Heart Rate Signal Using Python. A tech blog about fun things with Python and embedded electronics.* en-US. URL: http://www.paulvangent.com/2016/03/21/analyzing-a-discrete-heart-rate-signal-using-python-part-2/ (visited on 07/31/2018).

[53] Michael D. Gershon. "The Enteric Nervous System: A Second Brain". In: *Hospital Practice* 34.7 (July 1999), pp. 31–52. ISSN: 2154-8331. DOI: 10.3810/hp.1999.07.153. URL: https://doi.org/10.3810/hp.1999.07.153 (visited on 09/12/2019).

[54] Daniel Goleman and Cary Cherniss. *The emotionally intelligence workplace.* 2001.

[55] Pedro Gomes, Hugo Silva, and Petra Margaritoff. *pyHRV - Open-Source Python Toolbox for Heart Rate Variability.* original-date: 2018-10-20T00:14:50Z. July 2019. URL: https://github.com/PGomes92/pyhrv (visited on 07/08/2019).

[56] Catharina C. Grant et al. "Importance of Tachogram Length and Period of Recording during Noninvasive Investigation of the Autonomic Nervous System". en. In: *Annals of Noninvasive Electrocardiology* 16.2 (2011), pp. 131–139. ISSN: 1542-474X. DOI: 10.1111/j.1542-474X.2011.00422.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1542-474X.2011.00422.x (visited on 09/10/2019).

[57] Daniel Graziotin, Xiaofeng Wang, and Pekka Abrahamsson. "Happy software developers solve problems better: psychological measurements in empirical software engineering". en. In: *PeerJ* 2 (Mar. 2014), e289. ISSN: 2167-8359. DOI: 10.7717/peerj.289. URL: https://peerj.com/articles/289 (visited on 11/22/2019).

[58] John Howard Griffin. *Black Like Me: The Definitive Griffin Estate Edition, Corrected from Original Manuscripts.* en. Google-Books-ID: dH4eAwAAQBAJ. Wings Press, 2004. ISBN: 978-0-930324-72-8.

[59] James J. Gross. "Antecedent- and response-focused emotion regulation: divergent consequences for experience, expression, and physiology." In: *Journal of personality and social psychology* 74.1 (1998), pp. 224–237. DOI: 10.1037/0022-3514.74.1.224.

[60] James J. Gross. "The Emerging Field of Emotion Regulation: An Integrative Review". en. In: *Review of General Psychology* 2.3 (Sept. 1998), pp. 271–299. ISSN: 1089-2680. DOI: 10.1037/1089-2680.2.3.271. URL: https://journals.sagepub.com/doi/abs/10.1037/1089-2680.2.3.271 (visited on 03/29/2019).

[61] James J. Gross. "The Future's So Bright, I Gotta Wear Shades". en. In: *Emotion Review* 2.3 (July 2010), pp. 212–216. ISSN: 1754-0739, 1754-0747. DOI: 10.1177/1754073910361982. URL: http://journals.sagepub.com/doi/10.1177/1754073910361982 (visited on 03/29/2019).

[62] James J. Gross and Oliver P. John. “Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being”. In: *Journal of Personality and Social Psychology* 85.2 (2003), pp. 348–362. ISSN: 1939-1315(Electronic),0022-3514(Print). DOI: 10.1037/0022-3514.85.2.348.

[63] Paul Grossman and Edwin W. Taylor. “Toward understanding respiratory sinus arrhythmia: Relations to cardiac vagal tone, evolution and biobehavioral functions”. In: *Biological Psychology*. Special Issue of Biological Psychology on Cardiac Vagal Control, Emotion, Psychopathology, and Health. 74.2 (Feb. 2007), pp. 263–285. ISSN: 0301-0511. DOI: 10.1016/j.biopsycho.2005.11.014. URL: http://www.sciencedirect.com/science/article/pii/S0301051106001876 (visited on 09/11/2019).

[64] Dainora Grundey. *Experiental Marketing vs. Traditional Marketing: Creating Rational and Emotional Liaisons with Consumers*.

[65] R. Gutierrez-Osuna and J. Choi. “Using Heart Rate Monitors to Detect Mental Stress”. In: *2009 Sixth International Workshop on Wearable & Implantable Body Sensor Networks Conference (BSN 2009)(BSN)*. 2009, pp. 219–223. ISBN: 978-0-7695-3644-6. DOI: 10.1109/BSN.2009.13. URL: doi.ieeecomputersociety.org/10.1109/BSN.2009.13 (visited on 02/01/2019).

[66] Przemysław Guzik et al. “Correlations between Poincaré Plot and Conventional Heart Rate Variability Parameters Assessed during Paced Breathing”. In: *The Journal of Physiological Sciences* advpub (2007), pp. 0702020009–0702020009. DOI: 10.2170/physiolsci.RP005506.

[67] Marc Hassenzahl. “The Thing and I: Understanding the Relationship Between User and Product”. en. In: *Funology 2: From Usability to Enjoyment*. Ed. by Mark Blythe and Andrew Monk. Human–Computer Interaction Series. Cham: Springer International Publishing, 2018, pp. 301–313. ISBN: 978-3-319-68213-6. DOI: 10.1007/978-3-319-68213-6_19. URL: https://doi.org/10.1007/978-3-319-68213-6_19 (visited on 01/29/2019).

[68] Marc Hassenzahl. “User experience (UX): towards an experiential perspective on product quality”. en. In: *Proceedings of the 20th International Conference of the Association Francophone d’Interaction Homme-Machine on - IHM ’08*. Metz, France: ACM Press, 2008, p. 11. ISBN: 978-1-60558-285-6. DOI: 10.1145/1512714.1512717. URL: http://portal.acm.org/citation.cfm?doid=1512714.1512717 (visited on 10/17/2019).

[69] Marc Hassenzahl, Sarah Diefenbach, and Anja Göritz. “Needs, affect, and interactive products – Facets of user experience”. en. In: *Interacting with Computers* 22.5 (Sept. 2010), pp. 353–362. ISSN: 09535438. DOI: 10.1016/j.intcom.2010.04.002. URL: https://academic.oup.com/iwc/article-lookup/doi/10.1016/j.intcom.2010.04.002 (visited on 09/03/2019).

[70] Marc Hassenzahl, Robert Kekez, and Michael Burmester. “The importance of a software’s pragmatic quality depends on usage modes”. en. In: (2002), p. 3.

[71] Simon Herbert. "Theories of bounded rationality". In: *Decision and Organization* 1.1 (1972), pp. 161–176.

[72] Károly Hercegfi. "Heart Rate Variability Monitoring during Human-Computer Interaction". en. In: *Acta Polytechnica Hungarica* 8.5 (2011), p. 20.

[73] Morten Hertzum and Torkil Clemmensen. "How do usability professionals construe usability?" en. In: *International Journal of Human-Computer Studies* 70.1 (Jan. 2012), pp. 26–42. ISSN: 1071-5819. DOI: 10.1016/j.ijhcs.2011.08.001. URL: http://www.sciencedirect.com/science/article/pii/S1071581911001030 (visited on 11/22/2019).

[74] Wilhelm Hofmann, Malte Friese, and Fritz Strack. "Impulse and Self-Control From a Dual-Systems Perspective". en. In: *Perspectives on Psychological Science* 4.2 (Mar. 2009), pp. 162–176. ISSN: 1745-6916, 1745-6924. DOI: 10.1111/j.1745-6924.2009.01116.x. URL: http://journals.sagepub.com/doi/10.1111/j.1745-6924.2009.01116.x (visited on 11/13/2019).

[75] M. E. Hoque, D. J. McDuff, and R. W. Picard. "Exploring Temporal Patterns in Classifying Frustrated and Delighted Smiles". In: *IEEE Transactions on Affective Computing* 3.3 (July 2012), pp. 323–334. ISSN: 1949-3045. DOI: 10.1109/T-AFFC.2012.11.

[76] *How is IBI.csv obtained?* en-US. URL: http://support.empatica.com/hc/en-us/articles/201912319-How-is-IBI-csv-obtained- (visited on 09/30/2019).

[77] Kenneth Hugdahl and Arne Öhman. "Effects of instruction on acquisition and extinction of electrodermal responses to fear-relevant stimuli". In: *Journal of Experimental Psychology: Human Learning and Memory* 3.5 (1977), pp. 608–618. ISSN: 0096-1515(Print). DOI: 10.1037/0278-7393.3.5.608.

[78] T. P. Hutchinson. "Statistics and graphs for heart rate variability: pNN50 or pNN20?" en. In: *Physiological Measurement* 24.3 (June 2003), N9–N14. ISSN: 0967-3334. DOI: 10.1088/0967-3334/24/3/401. URL: https://doi.org/10.1088%2F0967-3334%2F24%2F3%2F401 (visited on 09/11/2019).

[79] *Innostuimme mittaamaan askeliamme, untamme ja sykettämme, ja nyt laitteet kertovat meille, miten pitäisi elää – "Emmekö osaa enää tehdä mitään itse?" kysyy professori.* fi. Feb. 2019. URL: https://www.hs.fi/teknologia/art-2000005986221.html?share=82c2ccb28e80c0b570b792b3d83fa90f (visited on 10/11/2019).

[80] Carroll Izard. *Emotions in Personality and Psychopathology.* en. Google-Books-ID: N3bjBwAAQBAJ. Springer Science & Business Media, Nov. 2013. ISBN: 978-1-4613-2892-6.

[81] William James. "What is an Emotion?" In: *Mind* 9.34 (1884), pp. 188–205. ISSN: 0026-4423. URL: https://www.jstor.org/stable/2246769 (visited on 08/26/2019).

[82] William James. "What is an emotion?" In: *Mind* os-IX (1984), pp. 188–205. DOI: 10.1093/mind/os-IX.34.188.

[83] W. Jänig. “Autonomic Nervous System”. en. In: *Human Physiology*. Ed. by Robert F. Schmidt and Gerhrad Thews. Berlin, Heidelberg: Springer Berlin Heidelberg, 1989, pp. 333–370. ISBN: 978-3-642-73831-9. DOI: 10.1007/978-3-642-73831-9_16. URL: https://doi.org/10.1007/978-3-642-73831-9_16 (visited on 01/30/2019).

[84] Peter Jönsson. “Respiratory sinus arrhythmia as a function of state anxiety in healthy individuals”. In: *International Journal of Psychophysiology* 63.1 (Jan. 2007), pp. 48–54. ISSN: 0167-8760. DOI: 10.1016/j.ijpsycho.2006.08.002. URL: http://www.sciencedirect.com/science/article/pii/S0167876006002236 (visited on 01/31/2019).

[85] C. G. Jung. *The Archetypes and the Collective Unconscious*. en. Google-Books-ID: hmXfBQAAQBAJ. Routledge, 1959. ISBN: 978-1-317-53461-7.

[86] Plinio Thomaz Aquino Junior and Lucia Vilela Leite Filgueiras. “User modeling with personas”. en. In: *Proceedings of the 2005 Latin American conference on Human-computer interaction - CLIHC '05*. Cuernavaca, Mexico: ACM Press, 2005, pp. 277–282. ISBN: 978-1-59593-224-2. DOI: 10.1145/1111360.1111388. URL: http://portal.acm.org/citation.cfm?doid=1111360.1111388 (visited on 09/30/2019).

[87] Daniel Kahneman. *Thinking, fast and slow*. Farrar, Straus and Giroux, 2011. ISBN: 978-0-374-27563-1. URL: http://www.worldcat.org/isbn/9780374275631 (visited on 01/25/2019).

[88] Michael J. Kane and Randall W. Engle. “The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective”. en. In: *Psychonomic Bulletin & Review* 9.4 (Dec. 2002), pp. 637–671. ISSN: 1531-5320. DOI: 10.3758/BF03196323. URL: https://doi.org/10.3758/BF03196323 (visited on 08/28/2019).

[89] Stéphanie Khalfa et al. “Event-related skin conductance responses to musical emotions in humans”. en. In: *Neuroscience Letters* 328.2 (Aug. 2002), pp. 145–149. ISSN: 03043940. DOI: 10.1016/S0304-3940(02)00462-7. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304394002004627 (visited on 08/16/2019).

[90] Jieun Kim, Ahreum Lee, and Hokyoung Ryu. “Personality and its effects on learning performance: Design guidelines for an adaptive e-learning system based on a user model”. In: *International Journal of Industrial Ergonomics* 43.5 (Sept. 2013), pp. 450–461. ISSN: 0169-8141. DOI: 10.1016/j.ergon.2013.03.001. URL: http://www.sciencedirect.com/science/article/pii/S0169814113000334 (visited on 09/20/2019).

[91] Julian Kiverstein and Mark Miller. “The embodied brain: towards a radical embodied cognitive neuroscience”. English. In: *Frontiers in Human Neuroscience* 9 (2015). ISSN: 1662-5161. DOI: 10.3389/fnhum.2015.00237. URL: https://www.frontiersin.org/articles/10.3389/fnhum.2015.00237/full (visited on 10/08/2019).

[92] Robert E. Kleiger, Phyllis K. Stein, and J. Thomas Bigger. "Heart Rate Variability: Measurement and Clinical Utility". en. In: *Annals of Noninvasive Electrocardiology* 10.1 (2005), pp. 88–101. ISSN: 1542-474X. DOI: 10.1111/j.1542-474X.2005.10101.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1542-474X.2005.10101.x (visited on 09/11/2019).

[93] Emma Komulainen et al. "The Effect of Personality on Daily Life Emotional Processes". en. In: *PLOS ONE* 9.10 (Oct. 2014), e110907. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0110907. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0110907 (visited on 08/22/2019).

[94] Michal Kosinski et al. "Manifestations of user personality in website choice and behaviour on online social networks". en. In: *Machine Learning* 95.3 (June 2014), pp. 357–380. ISSN: 1573-0565. DOI: 10.1007/s10994-013-5415-y. URL: https://doi.org/10.1007/s10994-013-5415-y (visited on 09/20/2019).

[95] Sylvia D. Kreibig. "Autonomic nervous system activity in emotion: A review". In: *Biological Psychology*. The biopsychology of emotion: Current theoretical and empirical perspectives 84.3 (July 2010), pp. 394–421. ISSN: 0301-0511. DOI: 10.1016/j.biopsycho.2010.03.010. URL: http://www.sciencedirect.com/science/article/pii/S0301051110000827 (visited on 09/12/2019).

[96] Chi'e Kurosaka, Hiroyuki Kuraoka, and Shinji Miyake. "Poincaré Plot Indexes of Heart Rate Variability: Pattern II Responses and Mental Workload". en. In: *HCI International 2019 - Posters*. Ed. by Constantine Stephanidis. Communications in Computer and Information Science. Springer International Publishing, 2019, pp. 238–243. ISBN: 978-3-030-23528-4.

[97] Tom Kuusela. - *Methodological Aspects of Heart Rate Variability Analysis*. en. Apr. 2016. DOI: 10.1201/b12756-7. URL: https://www.taylorfrancis.com/ (visited on 09/11/2019).

[98] E. Láng and N. Szilágyi. "Significance and assessment of autonomic indices in cardiovascular reactions." eng. In: *Acta physiologica Hungarica* 78.3 (1991), pp. 241–260. ISSN: 0231-424X. URL: http://europepmc.org/abstract/med/1814167 (visited on 02/01/2019).

[99] Peter J. Lang, Margaret M. Bradley, and Bruce N. Cuthbert. "Emotion and Motivation: Measuring Affective Perception". en-US. In: *Journal of Clinical Neurophysiology* 15.5 (Sept. 1998), p. 397. ISSN: 0736-0258. URL: https://journals.lww.com/clinicalneurophys/Citation/1998/09000/Emotion_and_Motivation__Measuring_Affective.4.aspx (visited on 08/16/2019).

[100] Peter J. Lang et al. "Looking at pictures: Affective, facial, visceral, and behavioral reactions". en. In: *Psychophysiology* 30.3 (May 1993), pp. 261–273. ISSN: 0048-5772, 1469-8986. DOI: 10.1111/j.1469-8986.1993.tb03352.x. URL: http://doi.wiley.com/10.1111/j.1469-8986.1993.tb03352.x (visited on 01/30/2019).

[101] Randy J. Larsen. "Affect Intensity". en. In: *Handbook of Individual Differences in Social Behavior*. Google-Books-ID: VgcGZ5sCEcIC. Guilford Press, 2009, pp. 241–254. ISBN: 978-1-59385-647-2.

[102] Randy J. Larsen. "Theory and measurement of affect intensity as an individual difference characteristic". PhD thesis. US: ProQuest Information & Learning, 1985.

[103] Randy J. Larsen and Ed Diener. "Affect intensity as an individual difference characteristic: A review". In: *Journal of Research in Personality* 21.1 (Mar. 1987), pp. 1–39. ISSN: 0092-6566. DOI: 10.1016/0092-6566(87)90023-7. URL: http://www.sciencedirect.com/science/article/pii/0092656687900237 (visited on 12/19/2018).

[104] R. S. Lazarus. *Psychological stress and the coping process*. Psychological stress and the coping process. New York, NY, US: McGraw-Hill, 1966.

[105] Richard S. Lazarus and Elizabeth Alfert. "Short-circuiting of threat by experimentally altering cognitive appraisal". In: *The Journal of Abnormal and Social Psychology* 69.2 (1964), pp. 195–205. ISSN: 0096-851X(Print). DOI: 10.1037/h0044635.

[106] Robert W. Levenson. "The Autonomic Nervous System and Emotion". en. In: *Emotion Review* 6.2 (Apr. 2014), pp. 100–112. ISSN: 1754-0739. DOI: 10.1177/1754073913512003. URL: https://doi.org/10.1177/1754073913512003 (visited on 09/12/2019).

[107] Robert W Levenson, Karl Heider, and Paul Ekman. "Emotion and Autonomic Nervous System Activity in the Minangkabau of West Sumatra". en. In: *Journal of Personality and Social Psychology* 62.6 (1992), pp. 972–988.

[108] Tao Lin and Atsumi Imamiya. "Evaluating Usability Based on Multimodal Information: An Empirical Study". In: *Proceedings of the 8th International Conference on Multimodal Interfaces*. ICMI '06. New York, NY, USA: ACM, 2006, pp. 364–371. ISBN: 978-1-59593-541-0. DOI: 10.1145/1180995.1181063. URL: http://doi.acm.org/10.1145/1180995.1181063 (visited on 02/01/2019).

[109] *MacArthur SES & Health Network | Research*. July 1999. URL: https://www.macses.ucsf.edu/research/allostatic/parasym.php#autonomic (visited on 09/12/2019).

[110] P. D. MacLean. *The Triune Brain in Evolution: Role in Paleocerebral Functions*. en. Google-Books-ID: 4PmLFmNdHL0C. Springer Science & Business Media, Jan. 1990. ISBN: 978-0-306-43168-5.

[111] V. Magagnin et al. “Heart rate variability and respiratory sinus arrhythmia assessment of affective states by bivariate autoregressive spectral analysis”. In: *2010 Computing in Cardiology*. Sept. 2010, pp. 145–148.

[112] Takahiko Masuda et al. “Placing the face in context: Cultural differences in the perception of facial emotion”. In: *Journal of Personality and Social Psychology* 94.3 (2008), pp. 365–381. ISSN: 1939-1315(Electronic),0022-3514(Print). DOI: 10.1037/0022-3514.94.3.365.

[113] David Matsumoto. “Cultural Influences on the Perception of Emotion”. en. In: *Journal of Cross-Cultural Psychology* 20.1 (Mar. 1989), pp. 92–105. ISSN: 0022-0221. DOI: 10.1177/0022022189201006. URL: https://doi.org/10.1177/0022022189201006 (visited on 08/27/2019).

[114] John McCarthy and Peter Wright. “Technology as experience”. en. In: *interactions* 11.5 (Sept. 2004), p. 42. ISSN: 10725520. DOI: 10.1145/1015530.1015549. URL: http://portal.acm.org/citation.cfm?doid=1015530.1015549 (visited on 09/03/2019).

[115] Rollin Mccraty and Fred Shaffer. “Heart Rate Variability: New Perspectives on Physiological Mechanisms, Assessment of Self-regulatory Capacity, and Health Risk”. en. In: *Global Advances in Health and Medicine* 4.1 (Jan. 2015), pp. 46–61. ISSN: 2164-9561. DOI: 10.7453/gahmj.2014.073. URL: https://doi.org/10.7453/gahmj.2014.073 (visited on 09/11/2019).

[116] Albert Mehrabian, Andrew L. Young, and Sharon Sato. “Emotional empathy and associated individual differences”. en. In: *Current Psychology* 7.3 (Sept. 1988), pp. 221–240. ISSN: 1936-4733. DOI: 10.1007/BF02686670. URL: https://doi.org/10.1007/BF02686670 (visited on 08/27/2019).

[117] Tomasz Miaskiewicz and Kenneth A. Kozar. “Personas and user-centered design: How can personas benefit product design processes?” en. In: *Design Studies* 32.5 (Sept. 2011), pp. 417–430. ISSN: 0142694X. DOI: 10.1016/j.destud.2011.03.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S0142694X11000275 (visited on 09/30/2019).

[118] J. E. Mietus et al. “The pNNx files: re-examining a widely used heart rate variability measure”. en. In: *Heart* 88.4 (Oct. 2002), pp. 378–380. ISSN: 1355-6037, 1468-201X. DOI: 10.1136/heart.88.4.378. URL: https://heart.bmj.com/content/88/4/378 (visited on 09/11/2019).

[119] Pat Moore and Charles Paul Conne. *Disguised: A True Story*. English. W Publishing Group, 1985. ISBN: 0-8499-0516-8.

[120] Nicholas Negroponte. *Being Digital*. English. United States: Alfred A. Knopf, 1995. ISBN: 0-679-43919-6.

[121] Weiting Ng and Ed Diener. “Personality differences in emotions: Does emotion regulation play a role?” In: *Journal of Individual Differences* 30.2 (2009), pp. 100–106. ISSN: 2151-2299(Electronic),1614-0001(Print). DOI: 10.1027/1614-0001.30.2.100.

[122] Peter Nickel and Friedhelm Nachreiner. "Sensitivity and Diagnosticity of the 0.1-Hz Component of Heart Rate Variability as an Indicator of Mental Workload". en. In: *Human Factors* 45.4 (Dec. 2003), pp. 575–590. ISSN: 0018-7208. DOI: 10.1518/hfes.45.4.575.27094. URL: https://doi.org/10.1518/hfes.45.4.575.27094 (visited on 01/31/2019).

[123] Jakob Nielsen. *Usability Engineering*. en. Google-Books-ID: DBOowF7LqIQC. Elsevier, Nov. 1994. ISBN: 978-0-08-052029-2.

[124] Don Norman. *3 ways good design makes you happy*. en. URL: https://www.ted.com/talks/don_norman_on_design_and_emotion (visited on 09/19/2019).

[125] Don Norman. *Emotional Design: Why We Love (Or Hate) Everyday Things*. 2004. ISBN: 0-465-05135-9.

[126] Don Norman. *The design of everyday things: Revised and expanded edition*. Basic books, 2013. ISBN: 978-0-465-05065-9.

[127] Don Norman, Jim Miller, and Austin Henderson. "What you see, some of what's in the future, and how we go about doing it: HI at Apple Computer". en. In: *Conference companion on Human factors in computing systems - CHI '95*. Denver, Colorado, United States: ACM Press, 1995, p. 155. ISBN: 978-0-89791-755-1. DOI: 10.1145/223355.223477. URL: http://portal.acm.org/citation.cfm?doid=223355.223477 (visited on 08/27/2019).

[128] Nargess Nourbakhsh et al. "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. event-place: Melbourne, Australia. New York, NY, USA: ACM, 2012, pp. 420–423. ISBN: 978-1-4503-1438-1. DOI: 10.1145/2414536.2414602. URL: http://doi.acm.org/10.1145/2414536.2414602 (visited on 10/03/2019).

[129] Antti Oulasvirta, Esko Kurvinen, and Tomi Kankainen. "Understanding contexts by being there: case studies in bodystorming". en. In: *Personal and Ubiquitous Computing* 7.2 (July 2003), pp. 125–134. ISSN: 1617-4909. DOI: 10.1007/s00779-003-0238-7. URL: https://doi.org/10.1007/s00779-003-0238-7 (visited on 08/22/2019).

[130] Rosemary Pacini and Seymour Epstein. "The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon." en. In: *Journal of Personality and Social Psychology* 76.6 (1999), pp. 972–987. ISSN: 1939-1315, 0022-3514. DOI: 10.1037/0022-3514.76.6.972. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.76.6.972 (visited on 12/19/2018).

[131] Rosalind W. Picard. *Affective Computing*. MIT press, 2000.

[132] Rosalind W. Picard. "Affective computing: challenges". en. In: *International Journal of Human-Computer Studies* 59.1-2 (July 2003), pp. 55–64. ISSN: 10715819. DOI: 10.1016/S1071-5819(03)00052-1. URL: https://linkinghub.elsevier.com/retrieve/pii/S1071581903000521 (visited on 10/16/2019).

[133] Robert Plutchik and Hope R. Conte, eds. *Circumplex models of personality and emotions.* Circumplex models of personality and emotions. Washington, DC, US: American Psychological Association, 1997. ISBN: 978-1-55798-380-0. DOI: 10.1037/10261-000.

[134] Stephen W. Porges. "Emotion: An Evolutionary By-Product of the Neural Regulation of the Autonomic Nervous System". en. In: *The Integrative Neurobiology of Affiliation.* Google-Books-ID: Im8FJJzjJUsC. MIT Press, 1999. ISBN: 978-0-262-53158-0.

[135] Daniel S. Quintana et al. "Heart rate variability is associated with emotion recognition: Direct evidence for a relationship between the autonomic nervous system and social cognition". In: *International Journal of Psychophysiology* 86.2 (Nov. 2012), pp. 168–172. ISSN: 0167-8760. DOI: 10.1016/j.ijpsycho.2012.08.012. URL: http://www.sciencedirect.com/science/article/pii/S0167876012005843 (visited on 08/16/2019).

[136] Charlotte L. Rae et al. "The Prefrontal Cortex Achieves Inhibitory Control by Facilitating Subcortical Motor Pathway Connectivity". In: *The Journal of Neuroscience* 35.2 (Jan. 2015), pp. 786–794. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.3093-13.2015. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4293423/ (visited on 08/28/2019).

[137] *Real-time physiological signals | E4 EDA/GSR sensor.* en-us. URL: https://www.empatica.com/research/e4 (visited on 03/01/2019).

[138] William Revelle and Klaus R. Scherer. "Personality and Emotion". In: *Oxford companion to emotion and the affective sciences* (2009), pp. 304–306.

[139] Joshua S. Richman and J. Randall Moorman. "Physiological time-series analysis using approximate entropy and sample entropy". In: *American Journal of Physiology-Heart and Circulatory Physiology* 278.6 (June 2000), H2039–H2049. ISSN: 0363-6135. DOI: 10.1152/ajpheart.2000.278.6.H2039. URL: https://www.physiology.org/doi/full/10.1152/ajpheart.2000.278.6.h2039 (visited on 09/11/2019).

[140] Christian Rohrer. "When to Use Which User-Experience Research Methods". en. In: (), p. 7.

[141] Wilhelm von Rosenberg et al. "Resolving Ambiguities in the LF/HF Ratio: LF-HF Scatter Plots for the Categorization of Mental and Physical Stress from HRV". In: *Frontiers in Physiology* 8 (June 2017). ISSN: 1664-042X. DOI: 10.3389/fphys.2017.00360. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5469891/ (visited on 06/13/2019).

[142] Philipp V. Rouast, Marc T. P. Adam, and Raymond Chiong. "Deep Learning for Human Affect Recognition: Insights and New Developments". In: *IEEE Transactions on Affective Computing* (2019). arXiv: 1901.02884, pp. 1–1. ISSN: 1949-3045, 2371-9850. DOI: 10.1109/TAFFC.2018.2890471. URL: http://arxiv.org/abs/1901.02884 (visited on 11/22/2019).

[143] Jean-Claude Roy, Henrique Sequeira, and Bernard Delerm. "Neural Control of Electrodermal Activity: Spinal and Reticular Mechanisms". en. In: *Progress in Electrodermal Research*. Ed. by Jean-Claude Roy et al. NATO ASI Series. Boston, MA: Springer US, 1993, pp. 73–92. ISBN: 978-1-4615-2864-7. DOI: 10.1007/978-1-4615-2864-7_7. URL: https://doi.org/10.1007/978-1-4615-2864-7_7 (visited on 09/12/2019).

[144] David C. Rubin, Rick H. Hoyle, and Mark R. Leary. "Differential Predictability of Four Dimensions of Affect Intensity". In: *Cognition & emotion* 26.1 (2012), pp. 25–41. ISSN: 0269-9931. DOI: 10.1080/02699931.2011.561564. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3564960/ (visited on 09/30/2019).

[145] Tomas Rytel. "Emotional Marketing Concept: The New Marketing Shift in the Postmodern Era". English. In: *Verslas: teorija ir praktika* 1 (2010), pp. 30–38. ISSN: 1648-0627. URL: https://www.ceeol.com/search/article-detail?id=235689 (visited on 11/22/2019).

[146] Pertti Saariluoma and Isto Maarttola. "Stumbling blocks in novicee building design". In: *Journal of Architectural and Planning Research* 20.4 (2003), pp. 344–354. ISSN: 0738-0895. URL: https://www.jstor.org/stable/43030671 (visited on 08/22/2019).

[147] L. Salahuddin et al. "Ultra Short Term Analysis of Heart Rate Variability for Monitoring Mental Stress in Mobile Settings". In: *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Aug. 2007, pp. 4656–4659. DOI: 10.1109/IEMBS.2007.4353378.

[148] Stanley Schachter and Jerome Singer. "Cognitive, social, and physiological determinants of emotional state". In: *Psychological Review* 69.5 (1962), pp. 379–399. ISSN: 1939-1471(Electronic),0033-295X(Print). DOI: 10.1037/h0046234.

[149] Jocelyn Scheirer et al. "Frustrating the user on purpose: a step toward building an affective computer". en. In: *Interacting with Computers* 14.2 (Feb. 2002), pp. 93–118. ISSN: 0953-5438. DOI: 10.1016/S0953-5438(01)00059-5. URL: https://academic-oup-com.libproxy.aalto.fi/iwc/article/14/2/93/758896 (visited on 10/16/2019).

[150] Amy M. Schmitter. *17th and 18th Century Theories of Emotions*. English. 2016. URL: https://plato.stanford.edu/entries/emotions-17th18th/LD2Descartes.html#WonGen (visited on 08/26/2019).

[151] Nicola Schutte et al. "Experiential and rational processing styles, emotional intelligence and wellbeing". en. In: *Australian Journal of Psychology* 62.1 (Mar. 2010), pp. 14–19. ISSN: 0004-9530. DOI: 10.1080/00049530903312865. URL: http://doi.wiley.com/10.1080/00049530903312865 (visited on 08/28/2019).

[152] Mark S. Schwartz and Frank Andrasik. "Cardiorespiratory biofeedback". en. In: *Biofeedback, Fourth Edition: A Practitioner's Guide*. Google-Books-ID: mqgHDgAAQBAJ. Guilford Publications, Mar. 2017, pp. 196–213. ISBN: 978-1-4625-3194-3.

[153] Marshall H. Segall, Donald T. Campbell, and Melville J. Herskovits. "Cultural Differences in the Perception of Geometric Illusions". en. In: *Science* 139.3556 (Feb. 1963), pp. 769–771. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.139.3556.769. URL: https://science.sciencemag.org/content/139/3556/769 (visited on 08/01/2019).

[154] Henrique Sequeira and Jean-Claude Roy. "Cortical and Hypothalamo-Limbic Control of Electrodermal Responses". en. In: *Progress in Electrodermal Research*. Ed. by Jean-Claude Roy et al. NATO ASI Series. Boston, MA: Springer US, 1993, pp. 93–114. ISBN: 978-1-4615-2864-7. DOI: 10.1007/978-1-4615-2864-7_8. URL: https://doi.org/10.1007/978-1-4615-2864-7_8 (visited on 09/12/2019).

[155] Henrique Sequeira et al. "Electrical autonomic correlates of emotion". en. In: *International Journal of Psychophysiology* 71.1 (Jan. 2009), pp. 50–56. ISSN: 01678760. DOI: 10.1016/j.ijpsycho.2008.07.009. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167876008007630 (visited on 08/16/2019).

[156] Alexander Serebrenik. "Emotional Labor of Software Engineers". en. In: (), p. 7.

[157] C. Setz et al. "Discriminating Stress From Cognitive Load Using a Wearable EDA Device". In: *IEEE Transactions on Information Technology in Biomedicine* 14.2 (Mar. 2010), pp. 410–417. DOI: 10.1109/TITB.2009.2036164.

[158] Fred Shaffer and J. P. Ginsberg. "An Overview of Heart Rate Variability Metrics and Norms". In: *Frontiers in Public Health* 5 (Sept. 2017). ISSN: 2296-2565. DOI: 10.3389/fpubh.2017.00258. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5624990/ (visited on 06/05/2019).

[159] Fred Shaffer, Rollin McCraty, and Christopher L. Zerr. "A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability". English. In: *Frontiers in Psychology* 5 (2014). ISSN: 1664-1078. DOI: 10.3389/fpsyg.2014.01040. URL: https://www.frontiersin.org/articles/10.3389/fpsyg.2014.01040/full (visited on 09/10/2019).

[160] N. Sadat Shami et al. "Measuring affect in hci: going beyond the individual". en. In: *Proceeding of the twenty-sixth annual CHI conference extended abstracts on Human factors in computing systems - CHI '08*. Florence, Italy: ACM Press, 2008, p. 3901. DOI: 10.1145/1358628.1358952. URL: http://portal.acm.org/citation.cfm?doid=1358628.1358952 (visited on 08/28/2019).

[161] Michelle N. Shiota et al. "Feeling good: Autonomic nervous system responding in five positive emotions." en. In: *Emotion* 11.6 (2011), pp. 1368–1378. ISSN: 1931-1516, 1528-3542. DOI: 10.1037/a0024278. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0024278 (visited on 09/12/2019).

[162] Herbert A. Simon. *Models of man; social and rational.* Models of man; social and rational. Oxford, England: Wiley, 1957.

[163] Herbert A. Simon. "Motivational and emotional controls of cognition." en. In: *Psychological Review* 74.1 (1967), pp. 29–39. ISSN: 1939-1471, 0033-295X. DOI: 10.1037/h0024127. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/h0024127 (visited on 07/10/2019).

[164] Herbert A. Simon. *The Sciences of the Artificial.* en. Google-Books-ID: k5Sr0nFw7psC. MIT Press, Sept. 1996. ISBN: 978-0-262-26449-5.

[165] Ruth M. Sladek, Malcolm J. Bond, and Paddy A. Phillips. "Why don't doctors wash their hands? A correlational study of thinking styles and hand hygiene". In: *American Journal of Infection Control* 36.6 (Aug. 2008), pp. 399–406. ISSN: 0196-6553. DOI: 10.1016/j.ajic.2007.11.002. URL: http://www.sciencedirect.com/science/article/pii/S0196655308000655 (visited on 09/26/2019).

[166] Joonwoo Son and Myoungouk Park. "Estimating Cognitive Load Complexity Using Performance and Physiological Data in a Driving Simulator". en. In: (), p. 4.

[167] Benedictus de Spinoza. *Ethics.* en. Google-Books-ID: FJrOf7k44NMC. Wordsworth Editions, 2001. ISBN: 978-1-84022-119-0.

[168] Christian Stickel et al. "Emotion Detection: Application of the Valence Arousal Space for Rapid Biological Usability Testing to Enhance Universal Access". en. In: *Universal Access in Human-Computer Interaction. Addressing Diversity.* Ed. by Constantine Stephanidis. Vol. 5614. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 615–624. ISBN: 978-3-642-02706-2 978-3-642-02707-9. DOI: 10.1007/978-3-642-02707-9_70. URL: http://link.springer.com/10.1007/978-3-642-02707-9_70 (visited on 08/16/2019).

[169] Akihiro Tanaka et al. "I Feel Your Voice: Cultural Differences in the Multisensory Perception of Emotion". en. In: *Psychological Science* 21.9 (Sept. 2010), pp. 1259–1262. ISSN: 0956-7976. DOI: 10.1177/0956797610380698. URL: https://doi.org/10.1177/0956797610380698 (visited on 08/27/2019).

[170] S. Taylor et al. "Automatic identification of artifacts in electrodermal activity data". In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Aug. 2015, pp. 1934–1937. DOI: 10.1109/EMBC.2015.7318762.

[171] *The Definition of User Experience (UX)*. en. URL: https://www.nngroup.com/articles/definition-user-experience/ (visited on 12/21/2018).

[172] Manfred Thüring and Sascha Mahlke. "Usability, aesthetics and emotions in human–technology interaction". fr. In: *International Journal of Psychology* 42.4 (2007), pp. 253–264. ISSN: 1464-066X. DOI: 10.1080/00207590701396674. URL: https://onlinelibrary.wiley.com/doi/abs/10.1080/00207590701396674 (visited on 08/20/2019).

[173] Renée M. Tobin et al. "Personality, emotional experience, and efforts to control emotions". In: *Journal of Personality and Social Psychology* 79.4 (2000), pp. 656–669. ISSN: 1939-1315(Electronic),0022-3514(Print). DOI: 10.1037/0022-3514.79.4.656.

[174] Mikko P. Tulppo et al. "Vagal modulation of heart rate during exercise: effects of age and physical fitness". In: *American Journal of Physiology-Heart and Circulatory Physiology* 274.2 (Feb. 1998), H424–H429. ISSN: 0363-6135. DOI: 10.1152/ajpheart.1998.274.2.H424. URL: https://www.physiology.org/doi/full/10.1152/ajpheart.1998.274.2.H424 (visited on 09/11/2019).

[175] Ken Umetani et al. "Twenty-Four Hour Time Domain Heart Rate Variability and Heart Rate: Relations to Age and Gender Over Nine Decades". en. In: *Journal of the American College of Cardiology* 31.3 (Mar. 1998), pp. 593–601. ISSN: 0735-1097, 1558-3597. DOI: 10.1016/S0735-1097(97)00554-8. URL: http://www.onlinejacc.org/content/31/3/593 (visited on 09/11/2019).

[176] *Utilizing the PPG/BVP signal*. en-US. URL: http://support.empatica.com/hc/en-us/articles/204954639-Utilizing-the-PPG-BVP-signal (visited on 03/01/2019).

[177] David E Vaillancourt and Karl M Newell. "Changing complexity in human behavior and physiology through aging and disease". In: *Neurobiology of Aging* 23.1 (Jan. 2002), pp. 1–11. ISSN: 0197-4580. DOI: 10.1016/S0197-4580(01)00247-0. URL: http://www.sciencedirect.com/science/article/pii/S0197458001002470 (visited on 09/11/2019).

[178] R. D Ward and P. H Marsden. "Physiological responses to different WEB page designs". In: *International Journal of Human-Computer Studies*. Applications of Affective Computing in Human-Computer Interaction 59.1 (July 2003), pp. 199–212. ISSN: 1071-5819. DOI: 10.1016/S1071-5819(03)00019-3. URL: http://www.sciencedirect.com/science/article/pii/S1071581903000193 (visited on 09/16/2019).

[179] R. D. Warda and P. H. Marsdenb. *Affective computing: problems, reactions and intentions*. 2004.

[180] David Watson and Kristin Naragon-Gainey. “Personality, Emotions, and the Emotional Disorders”. en. In: *Clinical Psychological Science* 2.4 (July 2014), pp. 422–442. ISSN: 2167-7026. DOI: 10.1177/2167702614536162. URL: https://doi.org/10.1177/2167702614536162 (visited on 08/22/2019).

[181] David Weibel, Bartholomäus Wissmath, and Fred W. Mast. “Immersion in Mediated Environments: The Role of Personality Traits”. In: *Cyberpsychology, Behavior, and Social Networking* 13.3 (June 2010), pp. 251–256. ISSN: 2152-2715. DOI: 10.1089/cyber.2009.0171. URL: https://www.liebertpub.com/doi/abs/10.1089/cyber.2009.0171 (visited on 09/20/2019).

[182] Kevin P. Weinfurt, Fred B. Bryant, and Paul R. Yarnold. “The Factor Structure of the Affect Intensity Measure: In Search of a Measurement Model”. In: *Journal of Research in Personality* 28.3 (Sept. 1994), pp. 314–331. ISSN: 0092-6566. DOI: 10.1006/jrpe.1994.1023. URL: http://www.sciencedirect.com/science/article/pii/S0092656684710233 (visited on 09/30/2019).

[183] *What should I know to use EDA data in my experiment?* en-US. URL: http://support.empatica.com/hc/en-us/articles/203621955-What-should-I-know-to-use-EDA-data-in-my-experiment- (visited on 03/01/2019).

[184] Sun K. Yoo et al. “Neural Network Based Emotion Estimation Using Heart Rate Variability and Skin Resistance”. en. In: *Advances in Natural Computation*. Ed. by Lipo Wang, Ke Chen, and Yew Soon Ong. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2005, pp. 818–824. ISBN: 978-3-540-31853-8.

```
#
# HRV ANALYSER SCRIPT
# Created by Niklas Strengell for this thesis
#
# Move this script to the folder with the IBI-file and run it
    ↪ trough terminal with "python hrv-analyzer.py"
# Remember to have all the needed packages installed
# Requires Python 3
#
# The algorithms are also described step-by-step in the thesis,
# but you can also find the descriptions and corresponding numbers
    ↪ in here.
#


from scipy.interpolate import interp1d
from scipy import signal
from scipy.signal import butter, lfilter
import pandas as pd
import numpy as np
import matplotlib as mpl
import matplotlib.pyplot as plt
import nolds
from scipy.signal import iirpeak

#
# FREQUENCY-DOMAIN METHODS
#

def interpolate(dataset):
        # Turn dataset into a list
        x_list = dataset.TIMESTAMP.values.tolist()
        y_list = dataset.IBI.values.tolist()

        # Interpolate
        x_new = np.linspace(x_list[0],x_list[-1],int(x_list[-1]))
        f = interp1d(x_list, y_list, kind='cubic')
        y_interpolated = f(x_new)

        return y_interpolated

def peak_filter(f0, Q, fs):
        # f0 = frequency to be retained
        # Q = quality factor
```

```
        # fs = sampling frequency
        b, a = iirpeak(f0, Q, fs=fs)
        return b, a

# 1) Read in the IBI.csv data from Empatica
df = pd.read_csv('IBI.csv')
df.columns = ['TIMESTAMP','IBI']

# Create the peak filter, values can be changed
# but keep fs preferably the same.
f = 0.5
fs = 1
b, a = peak_filter(0.15, 30, fs)

# 2) Interpolate the signal
interpolated = interpolate(df)
# 3) filter the interpolated signal
filtered = lfilter(b, a, interpolated)

# 4) Draw three different plots:
# original signal, interpolated signal and filtered signal.
fig = plt.figure(1, figsize=(16,12))

plt.subplot(311)
plt.plot(df.TIMESTAMP, df.IBI, label="Original", color='blue')
plt.legend(loc=4)

plt.subplot(312)
plt.plot(interpolated, color='Blue', alpha=0.5, label='Interpolated␣
    ↪ Signal')
plt.legend(loc=4)

plt.subplot(313)
plt.plot(filtered, color='Red', label='Mid-Frequency␣Component')
plt.ylim(-0.06,0.06)
plt.legend(loc=4)

# Finally save the plots into one image.
fig.savefig("1_HRV-plot.png")

#
# TIME-DOMAIN METHODS
#

# This calculates the RR-values - i.e. peak-to-peak differences -
```

```
# (and squared RR-values) and appends them into a list.
def calc_RR(dataset):
       RR_diff = []
       RR_sqdiff = []

       for item in range(0, len(dataset) - 1):
              RR_diff.append(dataset[item+1] - dataset[item])
              RR_sqdiff.append(math.pow(dataset[item+1] - dataset[
                 ↪ item], 2))

       return RR_diff, RR_sqdiff

# Using RR and RR^2 we can calculate other metrics
# which are then returned as an array
def calc_measurements(ibi, RR_diff, RR_sqdiff):
       measurements = {}

       measurements["mean_ibi"] = np.mean(ibi)
       measurements["sdnn"] = np.std(ibi)
       measurements["sdsd"] = np.std(RR_diff)
       measurements["rmssd"] = np.sqrt(np.mean(RR_sqdiff))
       NN20 = [x for x in RR_diff if (x>20)]
       NN50 = [x for x in RR_diff if (x>50)]
       measurements["nn20"] = NN20
       measurements["nn50"] = NN50
       measurements["p_nn20"] = float(len(NN20)) / float(len(RR_diff
          ↪ ))
       measurements["p_nn50"] = float(len(NN50)) / float(len(RR_diff
          ↪ ))

       return measurements

# The results are printed into the console
def print_measurements(measurements):
       print('------')
       print ("Mean␣IBI:", round(measurements["mean_ibi"], 2))
       print("SDNN:", round(measurements["sdnn"], 2))
       print("SDSD:", round(measurements["sdsd"], 2))
       print("RMSSD:", round(measurements["rmssd"], 2))
       print("pNN20,␣pNN50:", round(measurements["p_nn20"],2), round
          ↪ (measurements["p_nn50"],2))

# Pandas selection for the timestamp column
X = df['TIMESTAMP']
```

```
# Scale the data by 1000 to get milliseconds
Y = df['IBI'] * 1000

import math

RR_diff, RR_sqdiff = calc_RR(Y)
measurements = calc_measurements(Y, RR_diff, RR_sqdiff)
#print('------')
print_measurements(measurements)

#
# FRQUENCY DOMAIN METHODS: Spectrogram
#
# The steps 1 and 2 are already done in previous parts.

from scipy import fftpack
from skimage import util

# 3) Slice the signal into windows (size = 16, step = 2)
M = 16
slices = util.view_as_windows(interpolated, window_shape=(M,), step
    ↪ =2)

win = np.hanning(M + 1)[:-1]
slices = slices * win
slices = slices.T

# You can use these, if you want to check the size of you sliced
    ↪ signal
#print('------')
#print('Shape of slices: ', slices.shape)

# 4) Calculate fast fourier transform for each window
spectrum = np.fft.fft(slices, axis=0)[:M // 2 + 1:-1]
spectrum = np.abs(spectrum)

# You can change these to fit the bins of your image to the way you
    ↪  want
# Thes values create 6 bins between 0.0Hz and 0.5Hz.
rate = 4
N = interpolated.shape[0]
L = N / rate

f, ax = plt.subplots(figsize=(16, 6))
```

```
# 5) Take absolute values from the spectrum to negate the negative
    ↪ values created by fft.
S = np.abs(spectrum)
# 6) Normalise spectrum values by taking logarithm and dividing by
    ↪ maximum value
S = 20 * np.log10(S / np.max(S))

# 7) Create a plot with frequency bins and
# 8) Categorize each value into these bins
# 9) Color code the bins. We used 'viridis', but this can be
    ↪ changed.
ax.imshow(S, origin='lower', cmap='viridis', extent=(0, L, 0, rate /
    ↪ 8))
ax.axis('tight')
ax.set_ylabel('Frequency␣[Hz]')
ax.set_xlabel('Bins');

plt.savefig('2_Bins.png');

#
# NON-LINEAR METHODS: Poincar
#

# Scale the data to milliseconds
df.loc[:,'IBI'] *= 1000

# Poincare plot is calculated by plotting the difference of an
    ↪ interval to a previous interval,
# so spread the array into two arrays with "one-off" indexing
nn = df.iloc[:,1:].values
x1 = df.iloc[:-1,1:].values
x2 = df.iloc[1:,1:].values

# Calculate the standard deviation for both x1 + x1 and x1 - x2
# These account to the two axes in the Poincare plot
sd1 = np.std(np.subtract(x1,x2) / np.sqrt(2))
sd2 = np.std(np.add(x1,x2) / np.sqrt(2))

# Area of the ellipse
area = np.pi * sd1 * sd2

# Prepare the plot
fig = plt.figure(figsize=(8,8))
fig.tight_layout()
ax = fig.add_subplot(111)
```

```
ax.set_title(r'$Poincar\acute{e}$')
ax.set_ylabel('$NNI_{i+1}$␣[ms]')
ax.set_xlabel('$NNI_i$␣[ms]')
ax.set_xlim([np.min(nn) - 50, np.max(nn) + 50])
ax.set_ylim([np.min(nn) - 50, np.max(nn) + 50])
ax.grid()
ax.plot(x1, x2, 'r%s' % 'o', markersize=2, alpha=0.5, zorder=3)

# Compute the mean, aka the center of the plot
nn_mean = np.mean(nn)

# Draw the ellipse
ellipse = mpl.patches.Ellipse((nn_mean, nn_mean), sd1 * 2, sd2 * 2,
    ↪ angle=-45, fc='k', zorder=1)
ax.add_artist(ellipse)
ellipse = mpl.patches.Ellipse((nn_mean, nn_mean), sd1 * 2 - 1, sd2 *
    ↪ 2 -1, angle=-45, fc='lightyellow', zorder=1)
ax.add_artist(ellipse)

# Draw the vectors
arrow_head_size = 5
na = 4
a1 = ax.arrow(nn_mean, nn_mean, (-sd1 + na) * np.cos(np.deg2rad(45)),
    ↪  (sd1 - na) * np.sin(np.deg2rad(45)),
                      head_width=arrow_head_size, head_length=
                          ↪ arrow_head_size, fc='g', ec='g', zorder
                          ↪ =4, linewidth=1.5)
a2 = ax.arrow(nn_mean, nn_mean, (sd2 - na) * np.cos(np.deg2rad(45)),
    ↪ (sd2 - na) * np.sin(np.deg2rad(45)),
                      head_width=arrow_head_size, head_length=
                          ↪ arrow_head_size, fc='b', ec='b', zorder
                          ↪ =4, linewidth=1.5)
a3 = mpl.patches.Patch(facecolor='white', alpha=0.0)
a4 = mpl.patches.Patch(facecolor='white', alpha=0.0)
ax.add_line(mpl.lines.Line2D(
                      (min(nn), max(nn)),
                      (min(nn), max(nn)),
                      c='b', ls=':', alpha=0.6))
ax.add_line(mpl.lines.Line2D(
                      (nn_mean - sd1 * np.cos(np.deg2rad(45)) * na,
                          ↪ nn_mean + sd1 * np.cos(np.deg2rad(45)) *
                          ↪ na),
                      (nn_mean + sd1 * np.sin(np.deg2rad(45)) * na,
                          ↪ nn_mean - sd1 * np.sin(np.deg2rad(45)) *
```

```
                     ↪ na),
                   c='g', ls=':', alpha=0.6))
ax.legend([a1, a2, a3, a4], ['SD1:␣%.2f$ms$' % sd1, 'SD2:␣%.2f$ms$' %
   ↪  sd2, 'S:␣%.2f$␣ms^2$' % area, 'SD1/SD2:␣%.3f' % (sd1/sd2)],
   ↪ framealpha=1)

plt.savefig('3_Poincare.png');

#
# NON-LINEAR METHODSS: Sample entropy
#

dim = 2
tolerance = np.std(nn, ddof=-1) * 0.2
sampen = float(nolds.sampen(nn, dim, tolerance))
print('------')
print("Sample␣entropy:␣%.3f" % sampen)
print('------')
```