



MSc thesis
Computer Science

EDAMUX

A method for measuring User Experience

Svante Häggblom

December 9, 2019

FACULTY OF SCIENCE
UNIVERSITY OF HELSINKI

Supervisor(s)

Prof. T. Männistö

Examiner(s)

Prof. A.-P. Tuovinen

Contact information

P. O. Box 68 (Pietari Kalmin katu 5)
00014 University of Helsinki, Finland

Email address: info@cs.helsinki.fi

URL: <http://www.cs.helsinki.fi/>

Tiedekunta — Fakultet — Faculty		Koulutusohjelma — Utbildningsprogram — Study programme	
Faculty of Science		Computer Science	
Tekijä — Författare — Author			
Svante Häggblom			
Työn nimi — Arbetets titel — Title			
EDAMUX A method for measuring User Experience			
Ohjaajat — Handledare — Supervisors			
Prof. T. Männistö			
Työn laji — Arbetets art — Level		Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
MSc thesis		December 9, 2019	57 pages
Tiivistelmä — Referat — Abstract			
<p>Background: User experience (UX) is seen as an important quality of a successful product and software companies are becoming increasingly interested in the field of UX. As UX has the goal to improve the experience of users, there is a need for better methods in measuring the actual experience. One aspect of UX is to understand the emotional aspect of experience. Psychophysiology studies the relations between emotions and physiology and electrodermal activity (EDA) has been found to be a physiological measurement of emotional arousal.</p> <p>Aims: The aim of this thesis is researching the utility of measuring EDA to identify moments of emotional arousal during human-computer interaction. By studying peaks in EDA during software interaction we expect to find issues in the software that work as triggers or stimuli for the peaks.</p> <p>Method: We used the design science methodology to develop EDAMUX. EDAMUX is a method to unobtrusively observe users, while gathering significant interaction moments through self reporting and EDA. A qualitative single-case study was conducted to evaluate the utility of EDAMUX.</p> <p>Results: We found that we can discover causes of bad user experience with EDAMUX. Moments of emotional arousal, derived from EDA, was found in conjunction with performance issues, usability issues and bugs. Emotional arousal was also observed during software interaction where the user was blaming themselves.</p> <p>Conclusions: EDAMUX shows potential in discovering issues in software that are difficult to find with methods that rely on subjective self-reporting. Having the potential to objectively study emotional reactions is seen as valuable in complementing existing methods of measuring user experience.</p> <p>ACM Computing Classification System (CCS) Human-centered computing → Human computer interaction (HCI) → HCI design and evaluation methods → User studies Human-centered computing → Interaction design → Interaction design process and methods → User centered design</p>			
Avainsanat — Nyckelord — Keywords			
User Experience, Electrodermal activity, Design science, Case study			
Säilytyspaikka — Förvaringsställe — Where deposited			
Helsinki University Library			
Muita tietoja — övriga uppgifter — Additional information			
Software Systems specialisation line			

Contents

1	Introduction	1
2	Background	3
2.1	User experience (UX)	3
2.2	Usability	5
2.3	User involvement in software development	5
2.4	Measuring user experience	7
2.5	Emotions	8
2.5.1	Definition of emotion	8
2.5.2	The dimensional and discrete model of emotion	10
2.6	Psychophysiology	11
2.6.1	Measuring physiological changes	12
2.6.2	Electrodermal activity (EDA)	12
2.7	Measuring physiological changes in UX	13
3	Statement of the problem	17
3.1	Background and need	17
3.2	Purpose of the study	18
3.3	Research questions	18
4	Research material and methods	19
4.1	Design science	19
4.2	Background	21
4.3	EDAMUX	23
4.4	Case study for evaluating EDAMUX	24
4.4.1	Case company	24
4.4.2	Equipment and tools	26
4.4.3	Participant selection	29
4.4.4	Data collection and setting	30

4.4.5	Data analysis	31
4.4.6	Validation	34
5	Results	35
5.1	Software related emotional reactions	37
5.2	Validation sessions	38
6	Discussion	41
6.1	Evaluation of research method	45
6.2	Limitations	47
6.3	Conclusions	48
6.4	Recommendations for future research	49
	Bibliography	51

1 Introduction

It was a dark and stormy night. This might seem like a stupid way to start a thesis. Yet, if we look at it critically it is not a stupid sentence at all, as it truly creates a setting. A setting where an experience can take place. Without any specific words relating to feelings or thoughts, we are transported into our own experience of that dark and stormy night. The sentence itself relies on the imagination of the reader and it is accepted that every reader will have a different experience while reading the sentence. Exactly the same words, but always a different experience. A similar setting, yet a multitude of emotional reactions. The same applies to user experience, since it is unique for everyone. The product stays the same, but the experience does not. You can not know the emotional reactions you will elicit, but you might be able to measure them.

The software industry is becoming increasingly interested in the field of user experience (UX) and good UX is seen as an important quality of a successful product (Lallemand et al., 2015). But when we are designing for an experience, we get lost in the age old problem of translation: The intended experience is transformed and the perceived experience might not be the intended one. The designer should design for the end user, but the designer rarely is the end user of the product. Because UX has the goal of improving the experience of the end users, there is a growing need of methods for objectively measuring and understanding the actual experience (Hassenzahl, 2005). As one aspect of UX is understanding the emotional aspect of an experience (Hassenzahl and Tractinsky, 2006), this thesis set out to research emotions during software interaction.

Psychophysiology studies the relations between emotions and physiology, and electrodermal activity (EDA) has been found to be a reliable measurement of emotional arousal. Peaks in EDA data has been shown to occur in conjunction with stimuli that elicit emotions (Dawson et al., 2000) and EDA can be measured continuously and unobtrusively. Emotionally induced physiological reactions might therefore give us an objective glimpse into the subjective experiences of the user, potentially transforming the field of UX.

This thesis develops this kind of method utilising the methodology of design science. In order to study the feasibility and utility of using EDA to measure the user experience

we developed EDAMUX. It is a method for doing observation of users while gathering significant moments of interaction through EDA and self reporting. The method was developed to be unobtrusive, easily applied and useful in the software development. EDAMUX was evaluated with a qualitative single-case study in order to get an in-depth understanding of the utility of the method in a naturalistic environment.

The thesis follows the following structure: Section 2 presents the reader with background related to the research, defining the key concepts of UX, usability, emotions and psychophysiology. Section 3 motivates the need and purpose of EDAMUX and presents the research questions. Section 4 describes design science as a research methodology and explains how EDAMUX was developed. This section also motivates the selection of case study as the evaluation method and presents how data was collected and analysed. Section 5 presents the raw results from the research and section 6 discusses the results in context of the research questions. The last section also discusses the limitations of the research and presents the conclusions.

2 Background

This chapter has the aim of giving the reader a comprehensive description of the topic domain, the key concepts and the previous research in order to understand this thesis. This chapter will begin by presenting user experience, how it relates to human-computer interaction and its' role in software development. Thereafter emotions and physiology will be discussed and electrodermal activity (EDA) will be presented as a way of measuring physiological changes. Lastly some existing research on applying physiological measurements to understand user experience will be discussed.

2.1 User experience (UX)

Donald A. Norman is generally seen as the father of UX, as his life work has been dedicated to steering the focus away from the computer and towards the human using it. In his book *The Design of Everyday Things* he discusses design, the complexity of it and how to actually solve the right challenges with design (Norman, 2013). This book has worked as an inspiration for many practitioners of design and has revolutionized how we approach the field of human-computer interaction. This book is generally seen as the birth of UX. We highly recommend this book to the reader if there is an interest towards understanding the complexity of design and UX.

User experience (UX) is a widely used term in the software industry, but reluctantly used by the academic society because of its' ambiguity. The academic community still prefers to reference activities in the user experience domain by more established terms like human-computer interaction, user-centered design and usability. As user experience as a term and profession has been widely adopted in the industry, there is an evident need to use this term also in academia. Therefore we will have to discuss the term itself and the terminology surrounding it in order to study the concepts in the user experience domain.

When asking people what UX means to them there are always different answers coming out of their minds (we highly encourage the reader to answer this question them self). And why would the answer even be the same? The user experience is a personal trait and specifically refers to the thing that can not be easily shared or explained: the

experience itself. Nevertheless this ambiguous term, UX, has become a profession, a field of study and a property we value in the things we use. Good UX is something we want!

After having brought up the controversies of UX to the reader we can get a little deeper into discussing the actual definition of the term. The International Organization for Standardization (ISO) defines the user experience as follows:

User's perceptions and responses that result from the use and/or anticipated use of a system, product or service.

Note 1 to entry: Users' perceptions and responses include the users' emotions, beliefs, preferences, perceptions, comfort, behaviours, and accomplishments that occur before, during and after use.

Note 2 to entry: User experience is a consequence of brand image, presentation, functionality, system performance, interactive behaviour, and assistive capabilities of a system, product or service. It also results from the user's internal and physical state resulting from prior experiences, attitudes, skills, abilities and personality; and from the context of use.

[ISO 9241-11:2018, 3.2.3]

As we can see, also the ISO definition of UX is very broad and it encompasses concepts like emotions, behaviours as well as internal and physical states. These will be discussed in more detail later in the thesis. The definition in itself mentions the user's perceptions and responses during both use and anticipated use of a system, product or service. This shows that even without having used a system, product or service the user experience of it has already begun. There are so many aspects that affect the wholesome of the user's perceptions and responses that actually creating good UX is a tall order.

As UX is such a broad concept it is no surprise that there has been a lot of critique towards it and it has even been described as "old wine in new bottles" (Hassenzahl, 2008). This critique is definitely deserved as there is no real consensus on what the actual "doing" part of UX is, or how good UX is done. Still, UX is seen as something new, both in the industry and the academic world (Lallemand et al., 2015).

As UX can and should be discussed and researched in multiple domains, in this thesis we will narrow it down and research UX in the software domain.

2.2 Usability

It has been noted that UX and usability are partly overlapping in their definitions and rely on each other. Therefore good user experience is difficult to achieve if the usability of a system is poor (Hertzum et al., 2012).

In simplest terms, software usability is the ease of use, remembrance and learnability of a system. This means that the aim of usability is to ease the human computer interaction so that the user can communicate better with the software system. On top of this usability can have goals like effectiveness, efficiency and satisfaction (Gupta et al., 2017).

Hertzum and Clemmensen (Hertzum et al., 2012) divided usability into two categories: Utilitarian and experiential usability. The utilitarian category contains the more traditional usability metrics like effectiveness and efficiency, whereas the experiential category relates primarily to the user's self and consisting of, for example, stimulation and identification. They found that usability professionals are focusing more on the utilitarian aspect of usability, indicating that goal-related considerations about task performance are central to their thinking about usability. This finding also suggest that usability professionals have less tools and methods for evaluating the experiential aspects of systems.

The experiential usability category and the user experience have a lot in common and it shows that the division between usability and UX is not clear. Therefore we can conclude that by studying UX we can get insights about usability and vice versa. The involvement of users of the software is nevertheless a key concept for both practices.

2.3 User involvement in software development

There are multiple methodologies and approaches of involving users in the process of software development. Kujala (Kujala, 2003) suggest that instead of classifying particular development approaches we could instead classify the main approaches to user involvement. These main approaches are user-centered design, participatory design, ethnography and contextual design. The approaches are closely linked together and are partly overlapping, but nevertheless all of them include a rationale explaining why to involve users and a methodology on how to do it.

User-centered design aims for the design and development of useful and usable products by following a generally accepted set of principles:

1. Early focus on users and tasks.
2. Empirical measurement.
3. Iterative design.

The user involvement is included in these principles, especially in the second principle which implies that users should carry out real work with simulations and prototypes and reactions should be observed, recorded and analyzed.

The participatory design builds on democratically planning and designing new business practices and interfaces. Users together with designers analyze the organizational requirements to support both individual and organizational needs.

Ethnography is a sociological approach that describes human activities and culture with a focus on the social aspects of human co-operation. Many methods in design have been developed applying the principles of ethnography in the processes. The principles are:

1. It takes place in natural settings.
2. It is based on the principle of holism, that is, particular behaviours must be understood in the respective context.
3. It develops descriptive understanding in contrast to prescriptive.
4. It is grounded in a member's point-of-view.

Contextual design is an approach where the idea is to study and design things holistically. The work processes are studied and the design is not only aimed at the software, but also role structures, supporting tasks, possible automation and the elimination of unnecessary steps in the process. The approach includes a general philosophy of visiting end users.

As user experience isn't limited to the interaction between the human and machine, there are a lot of factors to take into account when specifying, designing and developing software. Many of these factors are affected by the environment and the context that

the software is going to be used in and therefore a part of the evaluation needs to happen in the real context of use.

Observing users in the natural context of use is seen as valuable practice of gathering implicit and non-verbal needs of a system (Kujala, 2003; Christensen et al., 2010). One challenge with field studies is the vast amount of collected raw data and the analysis of the data is extremely time consuming and difficult. Therefore there is a need for more time efficient methods of doing field studies and contextual inquiries (Kujala, 2003).

2.4 Measuring user experience

User experience is a term, which has been touted as important for the development of any product (Hassenzahl, 2005), but it is unfortunately also a term, which lacks theoretical framework and empirical investigation. It is something that designers and researchers claim they can do, but not something that they can define. Measuring UX or part of it requires definitions, in order to know what to measure (Fenton, 1997).

Part of the problem might be that the experience is always subjective. Thus, there will almost always be mismatches between the designer's intention and the user's experience. Hence, evaluation of the designs will be needed, both in the requirement phase and validation phase. Subjective interviews, requirement lists and usability tests have traditionally been utilised for this and they have dominated the industry (Thüring et al., 2007). However, UX covers much more than just that. Instead of pure utility, UX takes an experiential perspective on product quality (Hassenzahl, 2008).

If there is an intention to improve the UX of a product, there is a need to measure some aspect of it. Measurement can help us to understand the effect of our design and development on the user experience and give us a better understanding of relationships among activities and entities. The measurements allow us to control what is happening in our software development projects and therefore gives us the possibility to predict how changes in processes and software will affect the user experience. Measurements also encourages us to improve our processes and products (Fenton, 1997, p. 16).

Already early on, academia has encouraged practitioners to engage in empirical UX research to advance the understanding of UX (Hassenzahl and Tractinsky, 2006). UX research can be approached from three perspectives: addressing the human needs beyond the instrumental, understanding the affective and emotional aspects of interaction, and

lastly, to comprehend to nature of the experience (Hassenzahl and Tractinsky, 2006). In this thesis, UX is approached and measured through that second aspect: the understanding of emotion.

2.5 Emotions

In UX, emotion can be understood in two different ways. First, emotions can be understood as a consequence of product use (Kim et al., 1998; Desmet, 2002; Hassenzahl, 2005) or as antecedents of product use and evaluative judgements (Singh et al., 1999). However, emotion is hard to put into words and important little details can be missed and important moments lost (Ganglbauer et al., 2009). Psychophysiology has been suggested as a possibility to solve this gap in knowledge (Ganglbauer et al., 2009; Mandryk et al., 2006), as it can measure emotional processes that cause changes in the human physiology.

As this thesis researches the understanding of emotion in measuring the user experience, there is a need for some background to support the research. This section gives a short overview on the research into emotions, how they relate to physiology and why electrodermal activity (EDA) is a feasible measurement of emotional arousal. As emotion is a diffuse concept in science, this section starts by discussing emotions and presents one way of organizing the related terms in affective science. The two main emotional models are then presented as well as how emotions affect the physiology. EDA is then described in more detail and lastly some research about applying psychophysiological measurements in the field of UX is presented.

2.5.1 Definition of emotion

Emotions play an important role in the life of people and in everyday life the word *emotion* is used as an umbrella term to describe a diversity of phenomena. Everything from playing an exciting game to funerals elicit emotions. Although we know emotions exist, the scientific community are yet to give an answer to what exactly emotions are.

In order to study emotions we need to have some coherence in the terms we use and try to find some way of navigating this "conceptual and definitional chaos" (Buck, 1990). In this thesis a widely adopted definition and organization of the terms related to emotions will be used, which is illustrated in 2.1 (Gross, 2010).

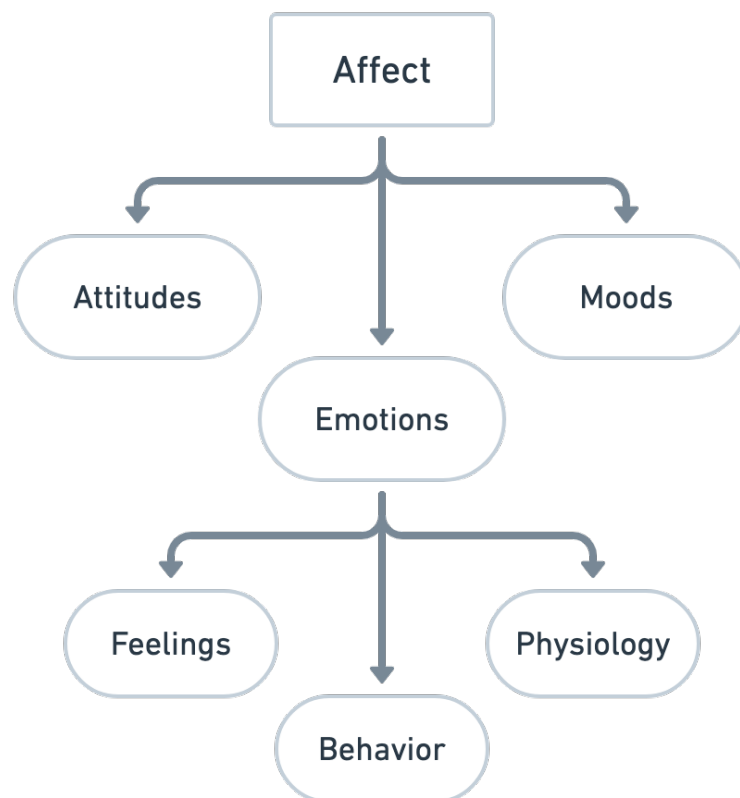


Figure 2.1: Semantic organisation of the key terms in affective science (Gross, 2010)

The definition puts *affect* on the highest level in the definitional hierarchy. Affect is used as an umbrella term to cover attitudes, emotions and moods. *Attitudes* are relatively stable beliefs about the goodness or badness of someone or something (Frijda, 1994). *Moods* are less stable than attitudes and are not directed towards any object or person (Siemer, 2005). *Emotions* are the shortest lived phenomena of the three and represent efficient modes of adaptation to the environment (Levenson, 1999).

When expressing emotions people often refers to *feelings*, e.g. feeling sad. A feeling is the conscious perception of an emotion, that is the personal experience of the wholesome of psychological and physiological changes that emotions cause. Emotions have also been shown to affect our decision making and behavior, for example people tend to take lower risks when they are in positive states in comparison to negative states (George et al., 2016).

The human body has evolved to stay in an optimal state to ensure our long-term survival, this state is called homeostasis. Levenson proposes (Levenson, 1999) that

”negative” emotions work as a ”temporary antidote” for this state of homeostasis, to ensure our short-term survival. Negative emotions, such as anger, fear and disgust, influences the autonomic nervous system, that in turn regulates internal body temperature, blood pressure and volume and different aspects of the body’s chemistry. This prepares the human body to be in an optimal state to properly react to the events in the environment that has caused these emotions. These emotions that put the body out of the state of homeostasis are also thought contribute to a number of physical illnesses, if the states of these emotions are prolonged.

Positive emotions also cause changes in our physiology, but no correlation has been found that they contribute to any physical illnesses. It has been proposed that positive emotions are a way of returning to homeostasis and ”undoing” the effects of negative emotions (Levenson, 1999). This proposal also finds support from a study, that measured recovery time from fear-related arousal. Positive emotions clearly hastened the speed of returning to baseline levels of cardiovascular arousal (L. Fredrickson et al., 1998).

2.5.2 The dimensional and discrete model of emotion

The dimensional and discrete theories are the two most common approaches to modeling emotions (Gendron et al., 2009). Even though they have different theories in how emotions form, they are just two ways of approaching the same concept.

The most common discrete emotion theory proposes six basic emotions: happiness, sadness, fear, anger, disgust and surprise. The theory proposes that these emotions are biologically inherited and can not be broken down into constituent psychological elements (Hamann, 2012).

Dimensional emotion theories presents emotions as combinations of different dimensions. Most commonly used is the two dimensional approach, which presents emotions based on valance (pleasantness) and arousal (emotion strength). The dimensional model has proven to be an attractive theory to use, as it is applicable across multiple domains (Hamann, 2012).

If the goal of an experiment is to derive specific emotions, the discrete model gives a good base in finding the best way to measure this. If the goal is to derive the strength of the positive/negative reactions to something, the dimensional model might give the granularity needed. The figure 2.2 illustrates how the basic emotions can be represented

within a dimensional framework.

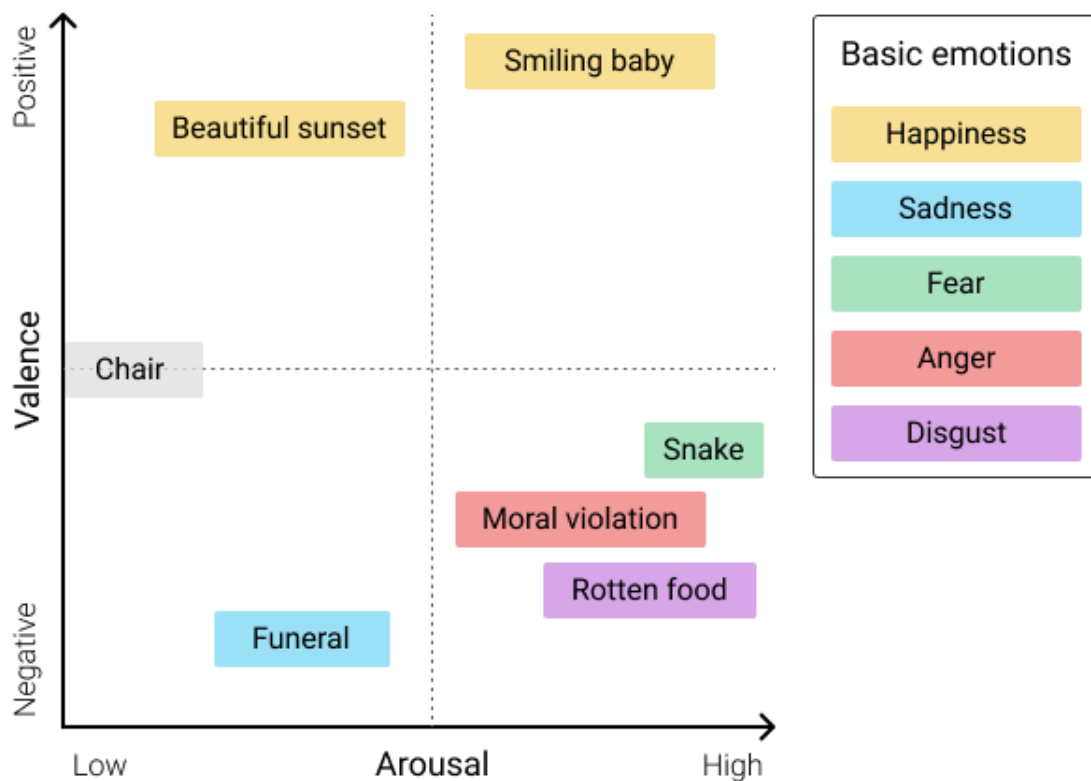


Figure 2.2: Representing basic emotions within a dimensional framework. (Hamann, 2012)

2.6 Psychophysiology

Psychophysiology is the study of psychological processes in the intact organism as a whole by means of unobtrusively measured physiological processes (Furedy, 1983). This is the field of study for understanding the relations between emotions and physiology.

Recognizing emotions in ourselves and others is a natural artifact in humans and we use this information to adapt to our environment and to the people around us. As the emotions play a key part in our interaction with our environment, there is a continuous interest in having ways to measure these phenomena. The difficulty lies in the complexity of emotions and the fact that we still can't answer the question on what emotions really are. As we know that emotions cause changes in the physiology, the measuring of physiological changes is an attractive approach to get insights on the emotional reactions of people.

The physiological states of arousal are regulated by the sympathetic and the parasymp-

pathetic subdivisions of the autonomic nervous systems. Parasympathetic activation occurs when the body needs to relax and conserve bodily energy, whereas sympathetic activation stimulates increased metabolic output to deal with external challenges (Poh et al., 2010).

2.6.1 Measuring physiological changes

Sympathetic arousal is closely linked to emotion, but also increases during cognitive load and in preparation and execution of energetic movements (Critchley, 2002). Measuring physiological changes caused by the sympathetic nervous system can therefore give us an insight in changes of emotional arousal (Poh et al., 2010). Another advantage in measuring physiological reactions is that it is robust against possible artifacts of human social masking, where people for some reason want to hide their real emotions and feelings (Wagner et al., 2005).

The two main approaches for observing physiological changes are measurements of the heart rate or the electrodermal activity (EDA). Measurement of both of these can be done continuously and unobtrusively and measurement equipment is commercially available. Therefore these measurements have been of interest in the field of user experience and multiple studies have been done in utilizing physiological measurements in human-computer interaction (Lin et al., 2006; Hercegfı and Pászti, 2009; Hercegfı, Pászti, et al., 2009; Mount et al., 2012; Xu et al., 2012).

The heart rate reacts to both sympathetic and parasympathetic activation of the nervous system (Poh et al., 2010), which makes analysing the data in the context of emotional arousal very complex. Because EDA reacts physiologically only to sympathetic nervous system activation, this physiological measurement was selected for this thesis.

2.6.2 Electrodermal activity (EDA)

EDA (also referred to as skin conductance or galvanic skin response) refers to the electrical properties of the skin. The changes in EDA is caused by eccrine, a sweat gland innervated by sympathetic nerves which react to emotional arousal (Dooren et al., 2012). There are eccrine sweat glands across the body, but the highest densities are found on the palms and soles.

As the skin is the only organ that is purely innervated by the sympathetic nervous system and the sympathetic nervous system reacts to emotional arousal, the measurement of EDA is a feasible way to measure the changes in arousal. A considerable advantage with measuring EDA is that this can be done continuously with equipment that are unobtrusive, e.g. with a wrist band (Garbarino et al., 2015). Another advantage with measuring physiological reactions is that it is robust against possible artifacts of human social masking, where people for some reason want to hide their real emotions (Wagner et al., 2005).

Measuring EDA gives as an insight of the emotional arousal of a person, but no specific emotions can be derived from these measurements. The measurement only gives us an one dimensional look into the emotional state of the person and without any other context it isn't possible to derive the valance of the state. This means that a strong feeling of happiness can not necessarily be distinguished from e.g. disgust or anger. Figure 2.3 visualizes the emotional arousal in the context of valance and discrete emotions.

EDA is usually measured on the palm, soles or the wrist by putting two electrodes against the skin. A constant small electric current is fed between the electrodes and the conductance between them is measured in μS (microsiemens) (Poh et al., 2010). By analyzing EDA data there is a possibility to identify peaks, or so called skin conductance responses (SCRs). An SCR is characterized by a sudden rise of EDA by at least $0.01 \mu S$ followed by an exponential decay and lasts between 1-5 seconds (see figure 2.4) (Taylor et al., 2015). An SCR generally occurs after some stimuli is presented to a person and is a way of recognizing moments of potential emotional arousal.

2.7 Measuring physiological changes in UX

Measuring emotions during usability tests has been an attractive concept in order to gather requirements of software and understand the user experience better. As mentioned earlier emotional arousal can be measured through the physiology of a person. Having the physiological data might help identify aspects that the users fail to mention, either because they think it is insignificant or they blame themselves for the problems they have with the system (Ward et al., 2004). Several studies have been done measuring physiological changes, mainly heart rate, in the UX domain and the studies have generally been done in controlled usability testing settings.

Studies have found that measuring the physiology, especially the heart rate, can give in-

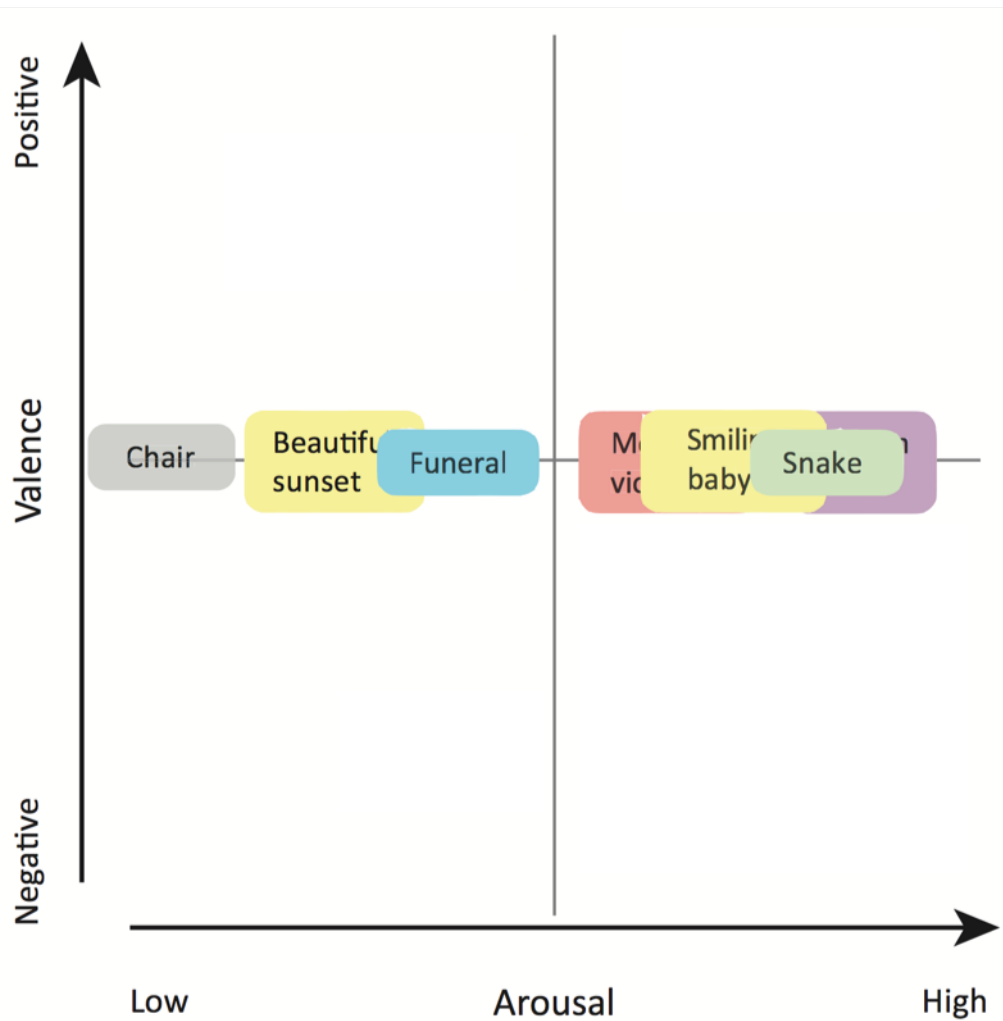


Figure 2.3: Visualization of emotional arousal

sights of aspects like mental effort (Lin et al., 2006; Hercegi and Pászti, 2009; Hercegi, 2018) and stress (Ceniza et al., 2016). The measurement of emotional arousal during software interaction has also been researched by some papers (Stickel et al., 2009; Foglia et al., 2014; Liu et al., 2016) and EDA has been used in all of these studies as one of the physiological measurements.

A common approach in measuring physiological changes during human-computer interaction is to have a clinical environment where all reactions can be tied to the actual interaction with the software (Ward et al., 2003; Ward et al., 2004; Tuch et al., 2009). Even though this approach will derive reactions to the software, it doesn't reflect the interaction in a natural environment.

Measuring physiological changes in a natural environment brings up the problem of

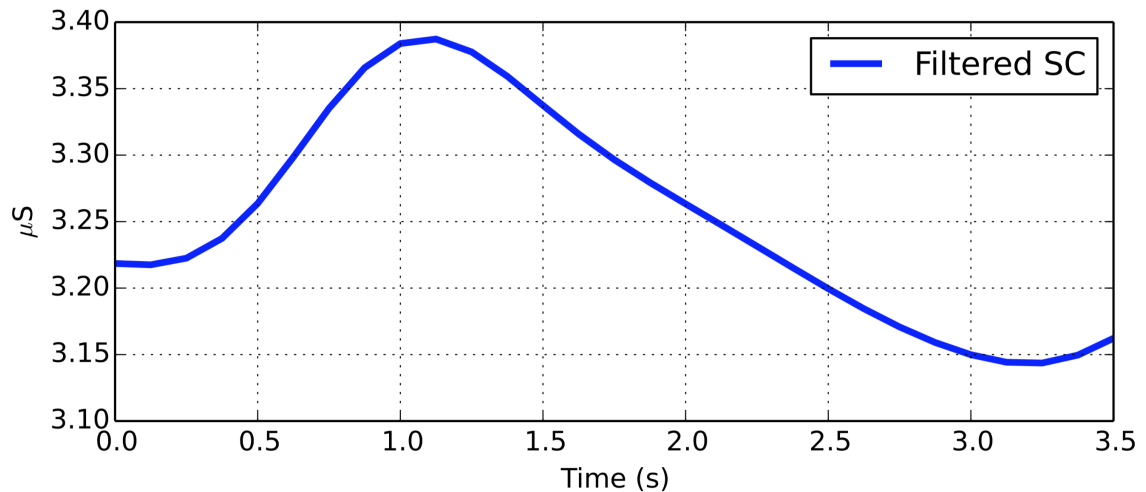


Figure 2.4: EDA signal during an SCR (Taylor et al., 2015)

not being able to tie the emotional arousal to specific features and events. The user might take a sip of coffee with bad milk in it and have a strong emotional reaction to that, thus not reacting to the software that is being evaluated (Ward et al., 2003). Having contextual data, like video recordings, might help us determine the stimuli of a reaction.

As physiological measurement can give us an insight of emotional arousal we might be able to distinguish situations that cause emotional responses. As the specific emotions the user is experiencing is not possible to determine from the physiological data, the supporting contextual data is important in analysing interaction data. Not being able to derive the specific emotion might be of less importance when measuring UX, the fact that the user experienced emotional arousal at a specific time could work as a trigger for further investigation of the specific interaction that occurred during that moment (Tuch et al., 2009).

3 Statement of the problem

As technological devices are more and more an intrinsic part of our lives and as the number of different applications is growing, people and companies have more options to choose from, when it comes to software. This has made user experience (UX) a growing factor in having a successful product and software companies are becoming increasingly interested in this field (Hassenzahl, 2008). User experience has become a profession and covers everything from user research, usability testing and prioritization of development to operative design and development. As UX has the goal of improving the experience of users, there is a growing need for better methods in measuring and understanding the actual experience (Hassenzahl, 2005).

3.1 Background and need

The current way of developing software has a lot of methods in measuring and testing usability and doing high quality user research, which has produced a lot of great products to this world. The usability measurements have indeed improved software a lot and basic heuristics have been developed to more easily build usable products. Methods, like subjective interviews, requirement lists and usability tests, are all aiming to improve the software and yield insights into the UX of software (Thüring et al., 2007). The issue with all these methods are that they rely heavily on the users ability to formalise their experience. Therefore these methods are always going to include some degree of subjectivity and bias.

To complement the existing methods of improving and measuring UX, there is a need for better methods of objectively studying the experience, or more specifically the emotional aspect of the experience (Hassenzahl, 2005). As one of the perspectives of UX is the understanding of emotion (Hassenzahl and Tractinsky, 2006) and because electrodermal activity (EDA) has proven to be way of measuring emotional arousal (Dooren et al., 2012), there is potentially a way of utilizing EDA for measuring UX more objectively. The understanding of when and why there is emotional arousal during software interaction, could help us to identify and improve the weak parts of software, when it comes to the user's experience.

3.2 Purpose of the study

In response to this need, this thesis will develop an artifact in the design science paradigm with the aim of measuring the user experience in the natural context of use. The artifact being developed is EDAMUX, a method for doing observation of users while gathering significant moments of interaction through EDA and self reporting. From the EDA data, moments of emotional arousal will be identified from skin conductance responses (SCRs), which have been shown to be occurring in conjunction with stimuli that elicit emotions (Dawson et al., 2000).

3.3 Research questions

This thesis has the aim of answering the following research questions:

RQ1 How can we get a useful number of SCRs during software interaction?

RQ2 What insights about user experience can be gathered from SCRs during software interaction?

RQ3 What is the utility of EDAMUX in software development?

RQ1 refers to how the actual identification of SCRs from EDA should be done in order to get a useful number of interaction moments to study. **RQ2** studies the interaction moments during SCRs more closely to analyse what kind of software interaction works as stimuli for the SCRs. **RQ3** evaluates EDAMUX as a whole and discusses the utility of the method in software development.

4 Research material and methods

This thesis applies the design-science paradigm in order to create an artifact for solving the challenges of measuring user experience. This chapter starts by discussing design science as a framework and how it is applied in this research. Secondly some background is presented related to the creation of the artifact and the problems it tries to solve. The artifact itself, EDAMUX, is then presented and described. The last parts of the chapter describes how and in what context EDAMUX was evaluated and how the data was collected and analysed.

4.1 Design science

Design science is a framework for doing research in the information systems discipline. It is a paradigm that seeks to extend the boundaries of human and organizational capabilities by creating new and innovative artifacts. The research in this thesis adopts the framework presented in a research essay (Hevner et al., 2004) by developing an artifact and evaluating it. To give the reader a good overview of the complexity of the information systems research, in which design science is included, the figure 4.1 was added to bring clarity.

The two main processes of design science are building an artifact and evaluating it. The aim of this process is to improve the understanding of the problem in order to improve both the quality of the product and the design process. This build-and-evaluate loop is optimally iterated in order to fine-tune and improve the artifact, but in this research no iteration was done because of the impeded resources of a masters' thesis. Nevertheless the utility of the created artifact can be evaluated and the shortcomings of it can be improved in further research.

Table 4.1 presents a set of guidelines that were used in conducting the research of EDAMUX and the discussion chapter of the thesis contains a section that explains how the guidelines were realized.

Guideline	Description
Guideline 1: Design as an Artifact	Design-science research must produce a viable artifact in the form of a construct, a model, a method, or an instantiation.
Guideline 2: Problem Relevance	The objective of design-science research is to develop technology-based solutions to important and relevant business problems.
Guideline 3: Design Evaluation	The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods.
Guideline 4: Research Contributions	Effective design-science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies.
Guideline 5: Research Rigor	Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact.
Guideline 6: Design as a Search Process	The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.
Guideline 7: Communication of Research	Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences.

Table 4.1: Guidelines and their descriptions for design science (Hevner et al., 2004)

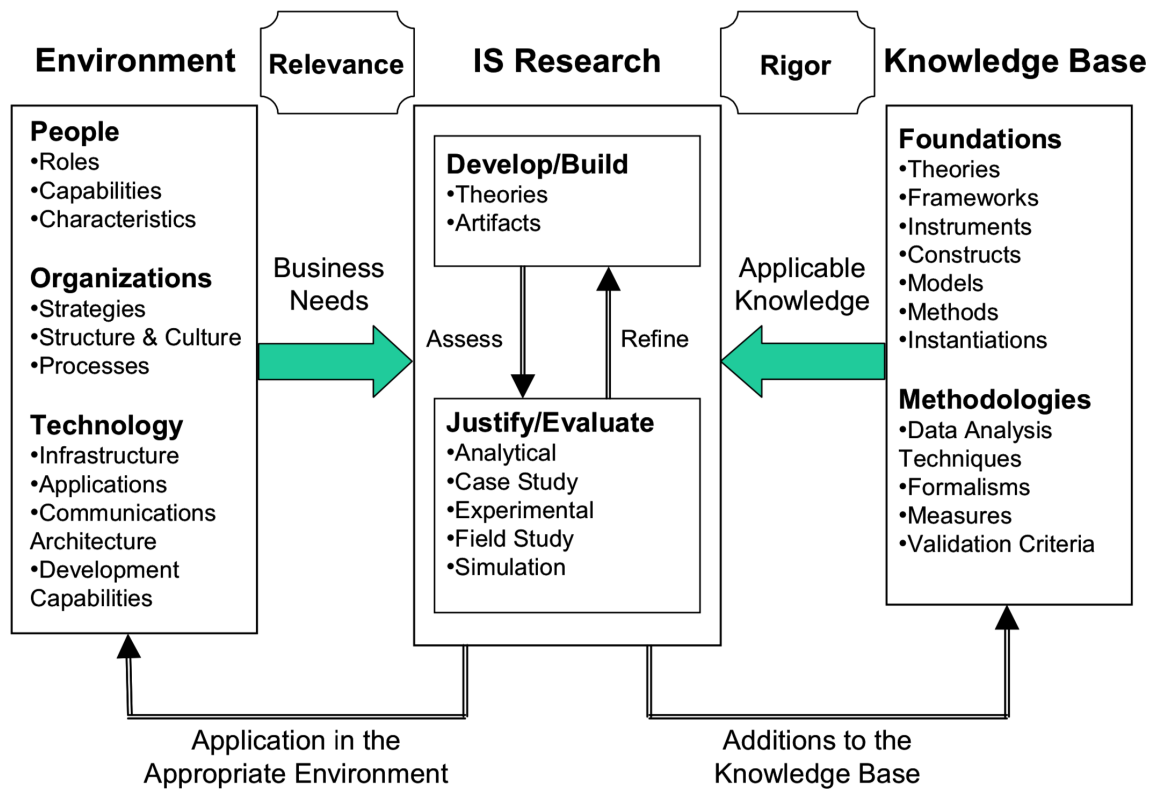


Figure 4.1: Information systems research framework (Hevner et al., 2004)

4.2 Background

EDAMUX stems from a method of remote usability evaluation and having the user report critical incidents that occur during software interaction (Castillo, 1997). The, so called, user-reported critical incident method involves users located in their own working environment, doing their everyday tasks and reporting critical incidents without direct interaction with evaluators. It was found that users can identify, report and rate the severity level of their own critical incidents, but that users report the incidents a good while after they occur.

Pitkänen et al. (Pitkänen, Pitkäranta, and Nieminen, 2012) continued the research of the user-reported critical incident method and developed a method called user-triggered usability testing. A dedicated device, called UXblackbox, was developed to handle the capturing of interaction data.

Early experiences of the user-triggered usability testing has shown promising results in identifying both positive and negative moments of software interaction during usability tests. From looking at the recordings of the negative moments, experts in usability have

been able to identify issues in the software. Pitkänen et al. have nevertheless identified that some clearly problematic or good events aren't reported at all (Pitkänen and Pitkäranta, 2014), which indicates that not every meaningful moment is reported by the user.

The author also has past personal experience utilizing the UXblackbox in a project where the usability of a new feature was evaluated. The evaluation was done at two different customers and a total of 6 users were observed for a week doing their everyday work with said feature. The observation was done in the real work environment, without an evaluator present and the users could freely report positive and negative moments of software interaction. The observation resulted in a lot of improvements to the design of the feature based on the user reported moments. Because of the huge amount of interaction data to analyse, the analysis was focused only on the user reported moments. Because of the limitations of user reported moments, e.g. self-blaming or the thought that something is too insignificant to report, a lot of interaction data was not analysed. This left the author with a feeling that there could be a way of recognising more significant moments of software interaction and this way get more data points of interaction to study. A deeper insight into the actual user experience was still lacking.

As one perspective of UX is understanding the affective and emotional aspects of interaction (Hassenzahl and Tractinsky, 2006) the building of EDAMUX was focused towards measuring the emotional aspect of software interaction. The idea was to utilize psychophysiology in measuring emotional arousal and through these measurements identify significant moments of software interaction.

Multiple studies have been done measuring the heart rate and EDA during software interaction and emotional arousal has been addressed in some of them (Stickel et al., 2009; Foglia et al., 2014; Liu et al., 2016). The issue with these studies is that the physiological reactions from multiple people are averaged and therefore the uniqueness and temporality of a specific experience is lost.

As EDAMUX has the aim of observing the user in the natural environment we aimed to find research related to physiological measurements during software interaction in a naturalistic setting. We found that the Budapest University of Technology and Economics has developed a method called INTERFACE (Hercegfi, Pászti, et al., 2009), which does exactly that. INTERFACE identifies meaningful events from moments of mental effort, derived from the heart rate measurements of the user, and have even been successful in identifying relative weak points of human-computer interaction. EDAMUX aims to

identify moments of emotional arousal from EDA in a naturalistic setting and study the human-computer interaction during these moments.

Usability issues has also been identified by measuring physiological changes during human-computer interaction. The following list lists some examples of usability that have been found through psychophysiological measurements:

- Non-optimal response times of the system (Dabrowski et al., 2011)
- Non-optimal input devices and methods (Mount et al., 2012)
- Images that the user assumes to be hot-links (Hercegi and Pászti, 2009)
- Scroll bar that is difficult to find (Hercegi and Pászti, 2009)
- Bad design solution for choosing a time period in a list view (Hercegi, Pászti, et al., 2009)

These findings indicate that EDAMUX also might be able to identify usability issues as the trigger for emotional arousal.

4.3 EDAMUX

EDAMUX (Electrodermal activity measurement in user experience), a method that combines parts of the different methods described earlier, will be developed in this thesis. EDAMUX aims to be unobtrusive, easily applied and useful in the software development. The method aims to identify meaningful moments during software interaction through self reported events as well as events derived from EDA and analyze the software interaction during these moments. Some of these meaningful moments are expected to be translated into issues in the software, elicit needs for improvement and help with prioritizing the development. It is also expected that more complex insights about the users workflows, frustrations and moments of cognitive load can be communicated to the design team of the software.

Some important characteristics of EDAMUX is that it is applied in the users' own working environment and doesn't require the users to do any artificial tasks, only their normal work. The setting therefore resembles Castillo's work on remote usability evaluation (Castillo, 1997). EDAMUX also gives the possibility to the users to report

meaningful events with a two-button console, to give the user experience experts data points to study. On top of this, EDA is recorded and analyzed to find SCRs that can be tied to moments that cause emotional arousal or cognitive load in the users. This way the unreported moments that has been mentioned (Pitkänen, Pitkäranta, and Nieminen, 2012) could potentially be found anyway.

A complete list of the data gathered with EDAMUX is:

- Screen recording
- Mouse clicks
- Keyboard presses
- Self reported moments
- EDA of the user

4.4 Case study for evaluating EDAMUX

The evaluation of the developed artifact, EDAMUX, was done based on a qualitative single-case study. We chose the case study as the evaluation method since we wanted to get an in-depth understanding of the utility of EDAMUX in a naturalistic environment. The qualitative single-case in this research refers to studying the utility of EDAMUX with one user. The reason for studying only one user lies in the nature of experience, as we assume that the experience is unique for everyone. The physiology of people is also unique and therefore physiological reactions differ from person to person (Dawson et al., 2000). Having only one user to study enabled us to study the user experience for that specific user more deeply and relieved us from the challenges of interpersonal differences. The studied user was selected purposefully in order to make sure they elicit the SCRs required for discovering the moments of emotional arousal.

4.4.1 Case company

The company, at which EDAMUX is researched, is a technology and software development company in the retail and supply chain domain. The company has over 250 customers and around 700 employees and sells business-to-business SaaS (Software as

a service) solutions. The company sells multiple solutions, but this thesis will focus on a service and software solution that is the core of the company.

The software in question is called Planning cloud and does demand forecasting and supply chain optimization and is integrated to the ERP (Enterprise resource planning) software of the customer. The software is business critical for the customer, as it determines what products should be ordered, when and where they should be delivered and gives forecasts on their demand. The Planning cloud calculates forecasts for all product store combinations every night and based on these gives proposals on what should be ordered to meet the forecasted demand, while minimizing waste and inventory. The software automates a big part of the ordering for retailers, the forecasts are extremely accurate, especially for products with steady demand, and therefore automation of the ordering process is possible.

As the Planning cloud is a business critical software for the customer, the release process of the software is quite rigorous. The software is tested thoroughly in the product delivery pipeline, but also at the customer's end and therefore the software is updated only around twice a year at the customer. This creates a pretty big feedback loop for the features and improvements developed in the software. The development process of a UI feature in Planning cloud follows roughly the following process:

1. Validation of the need
2. Design & Development
3. Code review & Usability testing
4. Release management
5. Customer delivery
6. Gathering feedback

EDAMUX could possibly be useful in steps 1, 3 and 6 in the development process, as these are all steps in the development where involvement of the end users of the software brings value. Observation with EDAMUX could also be added as a step before 1. Validation of the need, to proactively discover parts of the existing software that needs improvement. By observing users working in the software and focusing the research on the self reported moments and SCRs, improving and prioritizing improvement of existing features could be done more easily.

For the last step in the development process, EDAMUX could be applied to end users using the newly developed feature. This would make gathering feedback more proactive and would not only rely on proactive feedback by the users.

4.4.2 Equipment and tools

This section will present the equipment used for the EDA measurement and the interaction recording, as well as the software tool used for analysing SCRs.

Empatica E4

An Empatica E4 was used for the EDA measurement in this research. The Empatica E4 is a wearable wireless multisensor device for real-time computerized biofeedback and data acquisition. The device is a wristband that has been designed to be worn during daily activities and is not likely to interfere with everyday activities. Four sensors are embedded in the Empatica E4 and one of these sensors is the EDA sensor. The terminal part of the EDA sensor is composed of two silver-coated electrodes and a small alternating current is applied to the skin through the electrodes. The EDA sensor can measure conductance in the $[0.01,100] \mu S$ range with a sampling rate of 4 Hz. This section is based on an article about the Empatica E3 (Garbarino et al., 2015), but the EDA sensor is the same in the E4.

UXblackbox

The UXblackbox (Pitkänen, Pitkäranta, and Nieminen, 2012) was used for capturing the contextual interaction data in the research. The UXblackbox is a capturing device connected between wired connections of user interface devices and the computer (illustrated in figure 4.2). A complementary two-button console for tagging positive and negative moments of interaction is also included in the UXblackbox setup.

The UXblackbox records the following information on the removable memory connected to the device:

- Screen recording of the display connected to the device
- Mouse events (clicks of the primary or secondary mouse buttons)

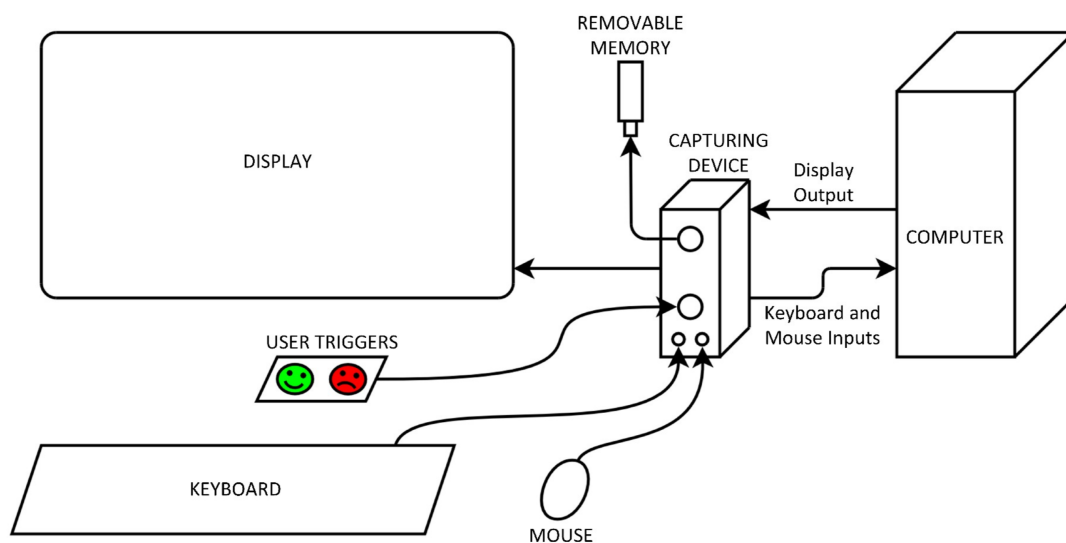


Figure 4.2: UXblackbox setup for desktop computer environment (Pitkänen, Pitkäranta, and Nieminen, 2012)

- Keyboard input
- User reported moments through the two-button console

The user can control when the UXblackbox starts and stops recording and every recording consists of video files and subtitle files containing the interaction data (mouse events, keyboard input and user reported moments). The UXblackbox automatically splits up the recording of the users screen onto video clips of 30 seconds, so there will be less corruption in case issues occur in the recording equipment. Every session the user records also creates a log file that contains all interactions that are recorded as well as exact timestamps on when the recording has started and ended.

EDA Explorer

EDA explorer is a project to help researchers extract meaningful information from EDA data (Taylor et al., 2015) and is hosted at <https://eda-explorer.media.mit.edu>. The research in this thesis utilises the EDA peak detection script (available at <https://github.com/MITMediaLabAffectiveComputing/eda-explorer>) for extracting SCRs from the EDA data. The script takes an EDA file as input and requires four parameters to be configured for the SCR detection. The parameters are:

- Minimum amplitude
- Offset
- Max rise time
- Max decay time

The minimum amplitude refers to the amplitude a potential SCR must reach in order to be counted as an SCR. The offset determines the minimum on how long the derivative of the EDA signal must be positive before a peak and how long it must be negative after the peak. The max rise time is the maximum number of seconds between the start of the SCR and the peak of the SCR and the max decay time is the maximum number of seconds for the EDA signal to decrease to 50% of the amplitude after the peak. Figure 4.3 illustrates the different features of an SCR used in the peak detection script.

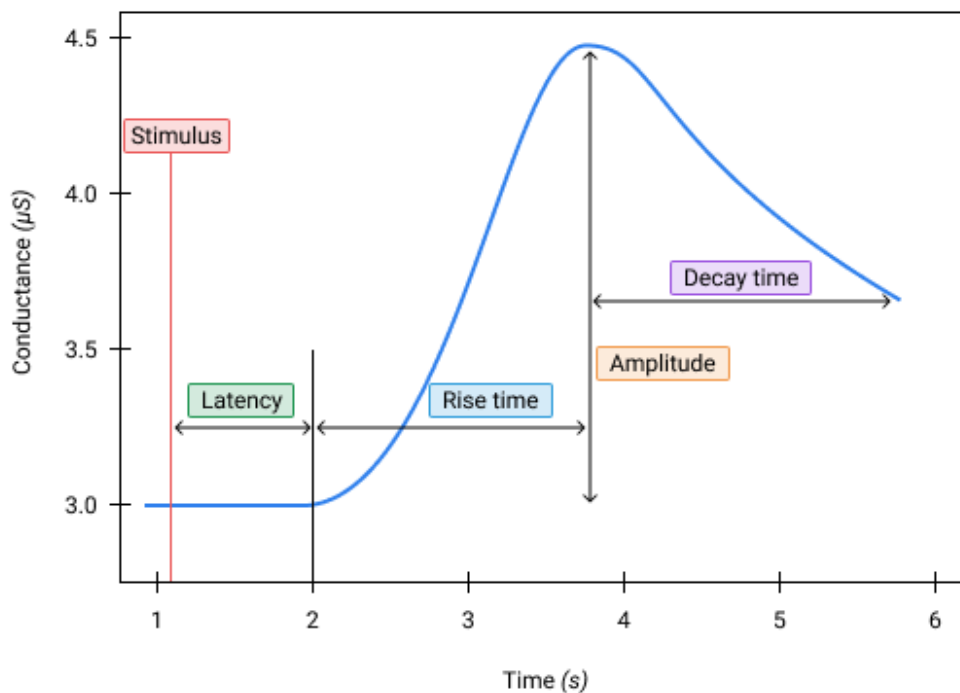


Figure 4.3: The features of an SCR (Dawson et al., 2000)

4.4.3 Participant selection

As planning of the supply chain is growing more centralized, the users of the Planning cloud are generally sitting in the same building. The work in forecasting and ordering has, through automation, become more and more exception based, focusing the time and effort of the end users to parts of the supply chain that requires human decision making. The Planning cloud is mainly used with different views that display the exception data and gives the user a possibility to make decisions in software.

To create the views and to configure the Planning cloud to work according to the user needs, there are a lot of internal user of the software inside the case company. For the sake of researching the utility of EDAMUX a decision was made to study an internal user, as they also are end users of Planning cloud and more easily accessible for the research.

The team selected for the research was the team that creates content and configures the software to be possible to use "out-of-box". The team mainly works on creating views and workflows that will suffice the basic needs of the customers, but as they all are very experienced users of the software a lot of their work also revolves around solving difficult challenges for the customer projects. As the work is done in the Planning cloud itself, it creates a good possibility to study the software interaction in a naturalistic environment. The work is also done in the most recent stable version of the software, which make the research even more relevant for the development.

Three potential candidates where selected for the research. Out of these three one was selected for the case study. The selection of the candidate was done by gathering EDA data from the wrist during a normal workday. The day of this research was selected to, as closely as possible, resemble a normal workday, containing normal software work and meetings.

The EDA data was then analysed using the EDA-Explorer (Taylor et al., 2015). The data was cleaned from reactions during meetings as the research is aiming to study software interaction. The candidate with the most skin conductance responses per minute was selected for the case study. This was done to make sure the candidate wasn't part of the "nonresponder" group that don't give any measurable SCRs (Dawson et al., 2000) and to give best possible ground to find reactions related to the software interaction.

The results from this study can be seen in Table 4.2 and as we can see all of the subjects

Subject	SCRs/min
Subject 1	0.4
Subject 2	1.8
Subject 3	3.6

Table 4.2: Skin conductance responses per minute during workday (the subject selected for the study highlighted)

show activation in EDA derived with EDA-Explorer. Subject 3 had clearly the most SCRs during a workday and was therefore selected for the case study.

4.4.4 Data collection and setting

The case study was conducted with the selected user and a UXblackbox was installed to the users work station. The users workstation involved two screens situated side by side and one of these screens was recorded, as well as the users mouse and keyboard interactions. On top of this the users EDA was recorded with the Empatica E4 wristband and the user’s self reported moments were recorded with a two-buttoned remote. Figure 4.4 visualises the setup for the research.

The user was instructed to use the software under research (Planning cloud) on the screen that was being recorded and to report any significant moments of interaction by pressing either the “:)” or “:(”-button on the remote. The user was also instructed to give written comments on the screen if they felt it was necessary for understanding the situation.

The control of what and when to record was fully in the hands of the user, as the idea of EDAMUX is to be as unobtrusive and non-disturbing as possible. The user was nevertheless instructed to record during the usage of the Planning cloud software. The researcher was not present during the data collection, not to affect the user’s interaction with the software.

The case study gathered data during four work days with the aim of validating the feasibility of EDAMUX in gathering insights of user experience in the real context of work. As there was data collected from two sources, the UXBlackbox and the Empatica E4 wristband, the data had to be synchronized in order for analysis to take place. The internal clocks of both devices were synchronized in the beginning of the research in order to make sure that the data was going to be aligned. As the UXblackbox

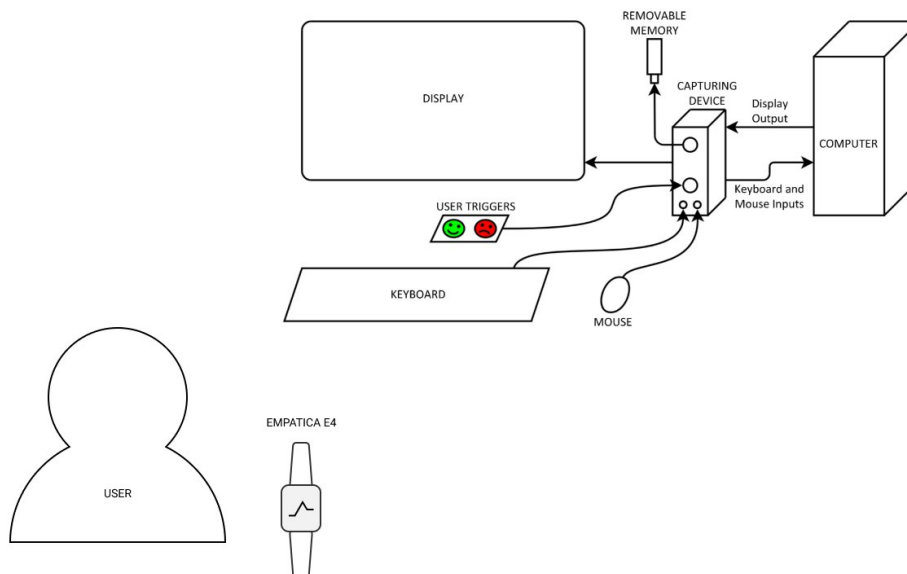


Figure 4.4: EDAMUX setup for desktop computer environment

automatically splits up the recording of the users screen to video clips of 30 seconds, the clips were merged together using a video editing tool, so that every recording session had its own video file.

4.4.5 Data analysis

The raw EDA data gathered with the Empatica E4 wristband was analyzed with the EDA peak detection script developed by Taylor et al. (Taylor et al., 2015). Three different peak files per session were created from the data, varying the minimum amplitude of the peak detection algorithm. As the minimum amplitude of SCRs vary from person to person (Dawson et al., 2000), we chose this parameter as the variable in affecting the number of detected SCRs. The selected minimum amplitudes for the study were $0.01 \mu S$, $0.2 \mu S$ and $0.5 \mu S$, in order to get a broad perspective on how many SCRs the different minimum amplitudes produce. The rest of the parameters for the algorithm were as follows:

- Offset = 1s
- Max rise time = 4s
- Max decay time = 4s

These values were the default settings of the peak detection algorithm (Taylor et al., 2015) and they have been found to be good values in detecting SCRs (Dawson et al., 2000).

At this point a decision was made to analyse the data with the 0.2 and 0.5 μS minimum amplitudes, as the 0.01 μS minimum amplitude produced way too many data points to be useful (see figure 4.5).

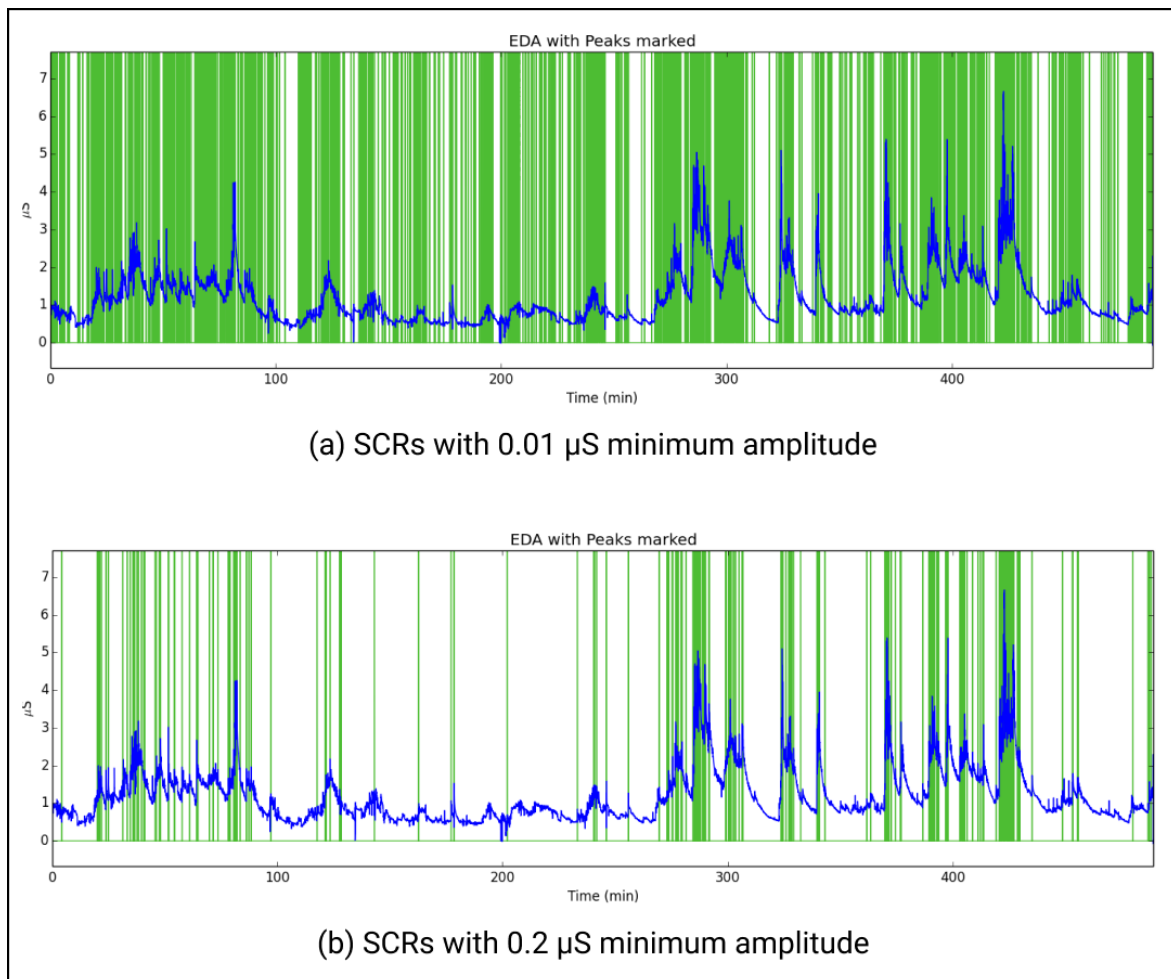


Figure 4.5: SCRs derived with different minimum amplitudes during a workday (SCRs marked with green markers).

A script was created to synchronize the SCR data with the recording of the UXblackbox. The script takes a log file from the UXblackbox and an SCR file as input and produces a file that lists all the self reported moments and the SCRs during a recording. All events have a relative timestamp to the start of the video, so that it is easy to find the corresponding interaction related to the moment in question. Figure 4.6 lists all the intermediate and final files and describes the data flow used in the analysis.

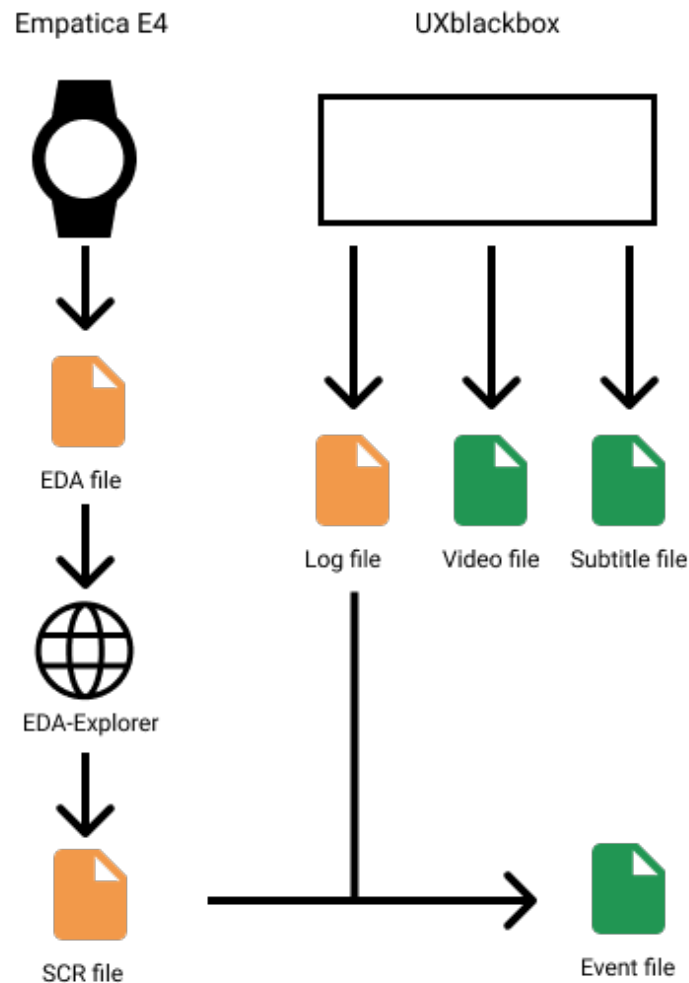


Figure 4.6: The data flow of the analysis

The video of a session was analysed based on the moments in the $0.2 \mu S$ event-file and the moments were automatically categorized to self reported moments and SCRs. The SCRs were analyzed in the following way:

1. Determine if interaction is happening in the software under research

2. Analyse if there is an observable cause to the SCR
3. Categorize the moment to cognitive load or emotional reaction
4. Validate moments with the user

The first step of the SCR analysis was done by looking at the video and seeing if there was any observable interaction with the software during a 5 second window before the event (Setz et al., 2010). The second step was a subjective analysis of the situation as a whole and observing any potential causes to the SCR in the interaction with the software. In the third step a division of the moment was done, based on the screen recording the SCR was categorized into cognitive load or emotional reaction. The constructs of emotional reaction and cognitive load where used in the following way:

Emotional reaction A potential stimuli in the software interaction is observed in conjunction with the SCR (e.g. loading indicator)

Cognitive load The SCR occurs during software interaction, but no potential stimuli for it is observed.

4.4.6 Validation

To increase the internal validity of the research, the emotional reactions related to the interaction of the software under research were validated. The results from the SCR analysis were validated by showing the moments of interaction during an SCR to the user and discussing the causes of the reaction.

5 Results

The case study gathered a total of nearly 19 hours of interaction and EDA data divided into four recordings during three days of work. From this data we could observe that different minimum amplitudes produced very different numbers of SCRs and that SCRs were divided unevenly during workdays. SCRs relating both to emotional reactions and cognitive load were observed, from the emotional reactions we could observe reactions related to performance, usability and bugs.

The user spent the time pretty evenly between the usage of Planning cloud (the software under research) and other work activities (see table 5.1). Recordings 3 and 4 were recorded during the same day. One recording had to be left out of the analysis, as the user forgot to start the EDA recording on the wristband.

	Planning cloud interaction [h : min]	Other work activities [h : min]	Total [h : min]
Recording 1	2:17	1:37	3:54
Recording 2	3:15	4:23	7:38
Recording 3	2:25	1:42	4:07
Recording 4	2:46	0:23	3:09
Total	10:43	8:05	18:48

Table 5.1: Total durations of recordings.

There were only three self reported events across the whole case study and two of these were referring to the same bug in the software. No SCRs were observed in conjunction with the self reported moments. The reasons for this will be discussed further in the discussion section of the thesis and the further results will address the EDA and SCRs.

The research questions for the thesis focused on the EDA data and the SCRs derived from it. The EDA data was analyzed with three different minimum amplitudes of the peak detection algorithm, but the $0.01 \mu S$ minimum amplitude was scrapped after the first recording as it produced over 1000 SCRs for a two and a half hour interaction video. This value was too high for any further analysis, as analysing nearly 5 SCRs/min is the same as looking at the video without any event data at all. The rest of the data was analyzed with the $0.2 \mu S$ and $0.5 \mu S$ minimum amplitudes.

	Planning cloud interaction	Other work activities	Total
Recording 1	22	153	175
Recording 2	61	231	292
Recording 3	2	28	30
Recording 4	80	112	192
Total	165	524	689

Table 5.2: Amount of 0.2 μS SCRs during recordings.

Table 5.2 shows the total amount of 0.2 μS SCRs during the recordings and how they were divided between the interaction with Planning cloud and other activities during work. As we can see in the table only a quarter of the total amount of SCRs happened during the interaction of the software under research. We could nevertheless observe SCRs during software interaction in a work environment.

Table 5.3 shows the numbers and frequencies of SCRs during Planning cloud interaction and as we can see the number of SCRs was about a third for the 0.5 μS minimum amplitude versus the 0.2 μS . So by changing the minimum amplitude we can affect the amount of data points to study.

To address the differences between the different minimum amplitudes we looked at what moments would be lost when using the higher 0.5 μS minimum amplitude. As we can see in table 5.4 we would find only around 18% of the emotional reactions caused by the software by using the higher 0.5 μS minimum amplitude and around 35% of the cognitive load reactions. By increasing the minimum amplitude we lost more of the emotional reactions triggered by the software in comparison with the moments of cognitive load. This was seen as a too big a loss of data points for the analysis of the causes of emotional reactions to the software.

	SCRs		SCRs/min	
	0.2 μS	0.5 μS	0.2 μS	0.5 μS
Recording 1	22	8	0.16	0.06
Recording 2	61	17	0.31	0.09
Recording 3	2	0	0.01	0.00
Recording 4	80	26	0.48	0.16
Total & Average	165	51	0.26	0.08

Table 5.3: Number and frequencies of SCRs during Planning cloud interaction.

	Emotional reactions		Cognitive load reactions	
	0.2 μS	0.5 μS	0.2 μS	0.5 μS
Recording 1	8	2	14	6
Recording 2	12	0	49	17
Recording 3	0	0	2	0
Recording 4	20	5	60	21
Total	40	7	125	44

Table 5.4: Division between emotional and cognitive load reactions for different minimum amplitudes.

5.1 Software related emotional reactions

The different causes of the emotional reactions in the software can be seen in table 5.5 and the performance category stands for 70% of all SCRs related to emotional reactions to the software. The performance category contains all SCRs that happened in conjunction with any kind of loading or where the user had to pause and wait in order to continue the work. Figure 5.1 shows one example of a situation where an SCR was measured and was categorized as a performance related emotional reaction.

Two of the usability issues that caused emotional reactions occurred during interaction with a chart modal in the software. One of the issues related to showing the exact value of a data point in the chart by hovering the mouse over it, the hover area was too small for this to be easily achievable. The other issue was related to inadequate feedback to switching between graphs for different products, the user was presented with the wrong information in the header of the modal before the chart had loaded completely.

One usability issue was related to selecting an action from a drop-down with a 4-level

	Performance	Usability issue	Bug	Other
Recording 1	6	1	1	0
Recording 2	9	2	0	1
Recording 3	0	0	0	0
Recording 4	13	1	3	3
Total	28	4	4	4

Table 5.5: Causes of emotional reaction in the software (0.2 μS minimum amplitude)

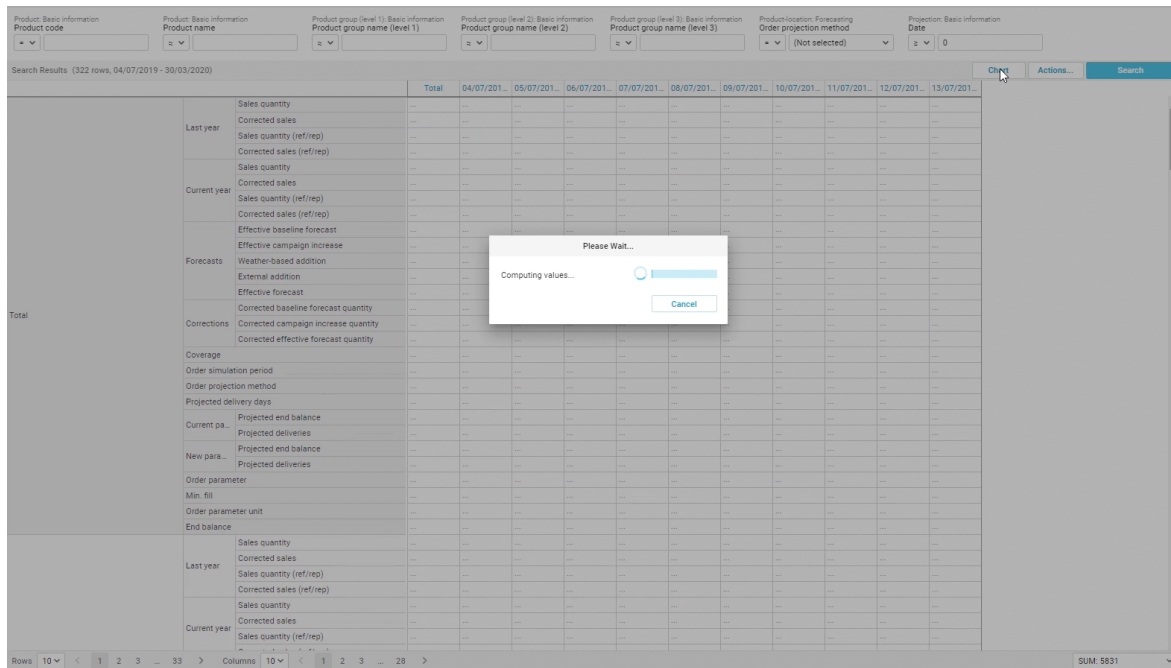


Figure 5.1: Example of a situation that caused a performance related SCR.

deep hierarchy. The last usability issue related to resizing elements in the UI.

Three of the bugs that caused emotional reactions were in the same chart modal described earlier, two of them relating to an issue where the modal doesn't fit the screen and has to be manually resized. One bug was related to unexpected behaviour of the UI when some values were dragged and dropped.

The other category of the causes of emotional reactions contains two situations where the user is frustrated with the displayed data and is frantically moving the mouse around while explaining the situation to a colleague. One reaction is caused by the user making a typo and the last one is a positive reaction to an action finishing faster than expected.

5.2 Validation sessions

During the validation sessions the recordings were shown to the user and the moments categorized as emotional reactions were discussed. In total there were three one hour sessions for analysing the data.

It was effortless for the user to remember and reflect on what they were doing at all the moments that were shown. One observation from these sessions was that the

categorization of SCRs was accepted by the user and the user could give more context to the emotional reactions observed from the SCRs. The user also mentioned to be unsure of how much this could help the development as they said: "I don't know how helpful this is, as I have observed some users of our software and they don't use it even closely as I do. I use charts a lot!"

We aimed to study the most recent version of the software and specifically the configuration and building of the use processes to make the software usable out of the box. The interaction observed during the case study was nevertheless only related to helping out and solving issues in a customer specific project. This resulted in observation of a slightly older version of the software and a significant part of the interactions were happening in parts that are not being developed anymore. The user also mentioned that it was a very stressful time, as there was a lot of temporal pressure to get the issues solved for the customer. Some SCRs were categorized as cognitive load during the sessions as the user said that they are most probably reactions to erroneous data and unexpected results and therefore not directly caused by the software, but by the work itself. The work the user was doing was solving these anomalies in the data.

Another comment by the user was that all emotional reactions except one were related to negative experiences. They mentioned that maybe these SCRs are more prevalent when doing stressful work, especially related to the emotional reactions caused by performance. During the loading and waiting for the software the user mentioned to be questioning themselves and thinking about what they did wrong. "Is it my fault that the loading takes so long? Did I do something wrong?" were a couple of thoughts the user reported to be thinking during these moments.

6 Discussion

In this chapter we will discuss and interpret the results from the case study. We will start by looking at the results more deeply in the context of the research questions and move on to discuss the research in the context of the guidelines of design science. Thereafter we will discuss the limitations of the research and present the conclusions.

We could already observe in the participant selection that all three potential participants elicited SCRs in the work environment, but we had no way of validating if any of these were caused by the software interaction. In the case study we studied the interaction with a particular software and by analyzing the data we could conclude that the interaction indeed does elicit SCRs in the work environment. This finding is in line with existing research into the topic of EDA measurement (Dawson et al., 2000).

RQ1: How can we get a useful number of SCRs during software interaction?

During the participant selection phase we could observe that the different participants had significantly different frequencies of SCRs (see table 4.2). This was expected as there are differences between the physiology and reactions of people. There have been multiple approaches for generalizing the results of different people. This has been done by compensating for the interpersonal differences (Johannes et al., 2014), trying to control the setting and tasks (Hercegfi, 2018) or by having dynamic triggers for recognizing emotionally arousing events (Westerink et al., 2009). In this thesis we did not try to compensate for interpersonal differences, instead we chose to study only one subject. By accepting that the user experience always is unique we could tweak the recognition of significant moments for this specific user and get an insight about their experience.

By changing the minimum amplitude of what is recognized as an SCR we could affect the number of SCRs and the number of interaction moments to study. The minimum amplitude of $0.01 \mu S$ produced way too many SCRs to be useful in the analysis of the interaction data, as too much data makes us choose nothing over something. For the minimum amplitude of $0.01 \mu S$ there probably are also a lot of SCRs that are not caused by any stimulus and typically the minimum amplitude of an event related SCR ranges from 0.2 to $1.0 \mu S$ (Dawson et al., 2000).

We compared the minimum amplitudes of 0.2 and 0.5 μS to assess the differences of the software interaction moments during the SCRs. We found that increasing the minimum amplitude caused a bigger loss of SCRs categorized as emotional reactions in comparison to cognitive load. As we were interested in the potential issues in the software we argue that the minimum amplitude of 0.2 μS is more useful than 0.5 μS for analysing the interaction data for this specific user.

RQ2: What insights about user experience can be gathered from SCRs during software interaction?

After getting to the conclusion that the 0.2 μS minimum amplitude for recognising SCRs was the most useful one for this specific user, we studied the moments of software interaction more deeply. In this thesis we divided the SCR reactions during software interaction into either emotional reactions and cognitive load reactions. This division is arbitrary as it has been found that SCRs occur during physical activity, emotional arousal, cognitive load, mental effort and stress (Dawson et al., 2000; Xu et al., 2012; Ceniza et al., 2016). Still the division of reactions is of minor interest here, as the emotional reaction category contains the reactions that can be tied to something occurring in the software. These are the moments we studied more deeply and we leave the deeper analysis of the other moments for future research.

From the emotional reactions we could observe 4 different categories of stimulus or triggers for the SCRs (number of occurrences in the parenthesis):

- Performance (28)
- Usability issue (4)
- Bug (4)
- Other (4)

The performance category stood for 70% of all SCRs that were caused by some trigger or stimulus in the software. This category contains all the moments where the software is loading something and the user is forced to wait until the loading is finished. This result is not surprising, as it has been found that increases in inter-task delay result in more anxiety and that the increase in anxiety is not due to the stressful nature of the task to be performed (Dabrowski et al., 2011). The user also reported to be questioning themselves and thinking about what they did wrong during the waiting for the loading to finish.

It is an interesting finding that we encountered moments where the user is blaming themselves by looking at the SCRs, as these are moments where the user is not blaming the software. This is significant because these are moments of potential shortcomings of the software that are not reported by the users, or as Don Norman puts it: "Because everyone perceives the fault to be his or her own, nobody wants to admit to having trouble. This creates a conspiracy of silence, where the feelings of guilt and helplessness among people are kept hidden." (Norman, 2013, p. 61). For further understanding of this subject we highly encourage the reader to read chapters *Blaming the Wrong Things* and *Falsely Blaming Yourself* from *The Design of Everyday Things* (Norman, 2013, pp. 59-68).

Usability issues and bugs were also observed in conjunction with SCRs and these would not have been noticed without having the moments of potential interest to study. Arguments can of course be made that these SCRs are false positives and not necessarily reactions to the software and we do not try to contradict these arguments. Instead, we look at the interaction data during the moments of SCRs and observe if we can find any issues in the usability of the software; If we find usability issues or bugs in the software we can fix, does it matter how we recognize them? Ward et al. seems to agree with this aswell: "Assuming that a reaction of some kind has definitely been identified in the physiological signals, its actual meaning can still be far from clear. In usability testing, we might not be too concerned with this, it may be sufficient to know that something occurred to alter arousal levels or some other aspect of a user's physiology, and steps can then be taken to find out what that something is." (Ward et al., 2004).

The part that was not researched more deeply was the reactions without any clear stimulus in the software, but still happened during software interaction. These reactions were broadly categorized as cognitive load reactions, even though they might contain emotional reactions, stress reactions and physical movement. Without any deeper analysis of these reactions it was still observed that these reactions happened more or less in clusters, where the user seemed to be focusing in on something or was working in a graph. These clusters of SCRs could potentially give an understanding of which parts and workflows of a software requires more resources from the user and could work as a way of helping select parts of a software to study more deeply.

RQ3: What is the utility of EDAMUX in software development? Having discussed the previous research questions, which looked more deeply into how SCRs

should be recognised and what specific moments could be derived from them, it is time to look at the bigger picture. As evaluating the utility of a method is extremely complex, we want to note that many of our interpretations might be interpreted differently, but nevertheless a discussion about the subject should be had.

First of all EDAMUX is a method of observing users in their natural environment and recognizing moments that might be of interest from the perspective of usability, performance or the overall structure of the software. As the method studies the interaction between a user and a software, there is a need to have some, at least partly, working version of a software to study. It can either be a prototype or a full version of the software, which denotes to what the intention of the observation study is. Are we evaluating the user experience and utility of a prototype or are we trying to identify weak parts of a bigger entity? We argue that EDAMUX can bring value to both of these, as long as there is interaction to study.

Through analysing interactions during the moments recognized by EDAMUX, issues that affect the user experience can be derived. These issues can then work as input to the design team of a software company, highlighting parts of software interaction that still needs refining and development. Especially recognizing situations where the user blames themselves for not being able to use the software, could highlight problems that are otherwise difficult to uncover.

To discuss when in the software development process EDAMUX could be utilized, we will abstract the processes into the human-centered design process as described by Norman (Norman, 2013, pp. 221-236) and discuss EDAMUX in context of this. This process of design iterates four steps:

1. Observation
2. Idea generation (ideation)
3. Prototyping
4. Testing

By looking at these steps it can easily be argued that EDAMUX could be utilized in the observation and testing steps in the process, so let us look at these steps more closely.

Observation is the initial research in understanding the nature of the problem itself, by observing would-be (or current) users in their natural environment. This will optimally create a deep understanding of the goals, motives and true needs of the people and will help shape the problem definition for the idea generation (Norman, 2013, pp. 222-224).

EDAMUX could be utilized in this process by observing the current software in use, if the software development aims to develop a replacing software. By using the method in observing the current software, bad moments of interaction could be identified to help understanding the issues that needs to solved in the software-to-be. Also good moments of interaction of interaction could be identified to help understanding what the current software does well and therefore what the users will at least expect from the replacing software.

The same applies if an older version of a software is in use and the development aims to develop this software further. EDAMUX could also here be utilized to gather understanding and needs for the idea generation for the further development.

Testing refers to validating the prototype built based on the observation and idea generation phase in the design process. It is done to ensure that the problem is well understood and to ensure that the new design meets the needs and abilities of those who will use it (Norman, 2013, pp. 228-229).

As the created prototype generally is a limited set of features, usability testing is often applied to validate if the prototype solves the intended challenges. There could be potential in using EDAMUX in more controlled settings, as the method gathers moments and insights into the user experience that are not reported by the users. This could result in discussions with the user about the moments that were identified through SCRs to get a better understanding of how well the prototype performed. Further research should be done applying EDAMUX in usability testing.

6.1 Evaluation of research method

As the research of this thesis followed the guidelines of design science as presented by Hevner et al. (Hevner et al., 2004), this section will present and discuss the research in context of the guidelines (see table 4.1).

Guideline 1: Design as an Artifact The research produced an artifact in form

of a method called EDAMUX. EDAMUX is a method of conducting observation of software interaction in the natural environment and analysing the interaction data based on moments gathered from self-reporting and EDA.

Guideline 2: Problem Relevance The construction and development of EDAMUX aimed to solve the challenge of measuring user experience and to utilize the measurements to recognize weak parts of the software. This is a relevant problem in the software development business, as good user experience is a selling point of software and a measurement of quality.

Guideline 3: Design Evaluation A case study was chosen as the evaluation method of the artifact, as the case study made it possible to study EDAMUX in depth in the business environment. The case study showed that EDAMUX is applicable as a method of observation and that it is possible to recognize weak parts of the software through SCRs derived from EDA.

Guideline 4: Research Contributions The case study of EDAMUX yielded in contributions to the field of psychophysiological measurements in user experience. Recognizing moments of self-blame during software interaction utilizing EDA can be seen as a significant finding in the field, even though the finding needs to be verified by future research.

Guideline 5: Research Rigor Thorough research of the literature in the field of user experience, usability, user research, emotions and physiology was done in order to build the artifact. The authors previous experience of doing observation with the UXblackbox also played a crucial role in developing EDAMUX. A case study was done to validate EDAMUX and steps were taken to ensure the validity of the research and the limitations chapter lists the identified aspects that might affect the validity of the case study.

Guideline 6: Design as a Search Process As measuring user experience can be seen as wicked problem (Hevner et al., 2004), it is infeasible to prove that EDAMUX is the optimal solution to this problem. Nevertheless studying the existing literature steered the selection EDA as physiological measurement, because of its' reliability in measuring activation of the sympathetic nervous system. Even though all aspects of why EDAMUX works are not explicitly clear, we could anyway conclude that the first iteration of EDAMUX yielded insights in the domain of measuring user experience.

Guideline 7: Communication of Research The research of EDAMUX has been

presented to the company that has developed the UXblackbox and to the case company. Both of the presentations had technology-oriented as well as management-oriented people present. Interest was shown by practitioners and investors during the presentations towards applying EDAMUX in their own work as well as continuing the research in different fields, e.g. entertainment technology. There was especially an interest towards objectively recognizing emotionally arousing moments from EDA. This thesis should work as a sufficient technological description of EDAMUX in order for practitioners to take advantage of the benefits of the method in their own work.

6.2 Limitations

This section will discuss the observed limitations of the research and factors that might affect the validity of the results.

The aim of the research was to study the most recent version of the software, to make sure the findings are relevant for the software development. Nevertheless we ended up mainly observing software interaction in parts of the software that are not being developed anymore, which caused us not having much relevant feedback to the software development. This does affect the validity of the results about the utility of EDAMUX in software development.

As EDAMUX gathered both self-reported moments as well as moments derived from the EDA of the user, both of these should have been present in the evaluation. There was a total of three self-reported moments, which led us to focus only on the moments derived from the SCRs. The reason for having so few self-reported moments are two-fold: The user already gives feedback continuously to the software development and the user felt that there was no big issues to report during the interaction. The lack of self-reported moments affects the evaluation of the utility of EDAMUX in software development.

The users work environment should not be affected by using EDAMUX, but because of the limitations of the UXblackbox the screen resolution of the users screen was affected by the research. Forcing the user work on a lower screen resolution causes a new variable to be introduced in the evaluation. This was not taken into account when evaluating the data as the effects of the lower screen resolution are unpredictable.

In the analysis of the EDA data and SCR extraction, noise was not taken into account.

Therefore there is no certainty that an SCR in this research is not caused by noise and there might be false positives in the study. This is a factor that affects the validity of the data analysis.

The categorization of SCRs used in this thesis is as well limited. The construct validity of emotional reaction and cognitive load does not fully cover all potential constructs that could be used to describe an SCR. Therefore the validity of the exact reactions that the user experienced might not be fully valid.

6.3 Conclusions

Because of the growing interest in user experience in the software industry, a need has been identified for improved ways of developing products with better user experience. An intrinsic part of creating better user experience is the deeper understanding of the emotional component of people, as emotions play an important role in any experience. If we could, at least, partly measure emotions we could better understand the user experience, and through these measurements improve the user experience of a product.

Psychophysiological measurements has been shown to react to emotions and especially EDA has been used to measure the activation of the sympathetic nervous system, which has been closely linked to emotional arousal. By measuring EDA we should therefore be able to partly measure the user experience.

To answer this need we developed EDAMUX, a method for unobtrusively observing a user, while gathering significant interaction moments through self reporting and EDA. To validate the utility of EDAMUX we conducted a case study where we applied this method of observation to the real world. We gathered three days of interaction data during the usage of a software under development, while recording the electrodermal activity of the user. We then configured the recognition of SCRs from the EDA data to match the physiology of the user and used these SCRs to derive meaningful moments of software interaction.

We found that we can recognise moments of bad user experience based on the SCRs and we could derive the reasons for these reactions from the software interaction recordings. The moments where the user was forced to wait for the software to load something were the most common stimuli for the SCRs during software interaction. Some usability issues and bugs were also observed as stimuli to the SCRs.

Through the validation sessions with the user it was found that some SCRs occurred during moments where the user was blaming themselves for the software not working as they expected. This can be seen as a significant finding as the moments of self-blame are very difficult to find through methods that rely on self-reporting. Further research should be done to validate this finding.

6.4 Recommendations for future research

EDAMUX should be further researched with users that do not have a direct way to give feedback on the software they are using, this way we should be able to get more self-reported moments on top of the moments derived from EDA. Research into correlations of self-reported moments and SCRs should also be done, as this was not possible in the setting this research was done in. EDAMUX should also be researched in more controlled settings, like usability tests, to evaluate the utility of interaction moments derived from SCRs in this context.

Further research in the EDA measurement and the SCRs should also be done, especially in the context of recognising false positive SCRs and filtering noise. This would make the data analysis of EDAMUX more efficient and focusing the time of the researcher to more relevant interaction moments.

To round up this thesis we want to leave the reader with some thoughts of a potential future that might be around the corner. People are more and more using personal devices that measure different physiological changes in themselves and EDA might soon be part of these measurements. What are the possibilities and risks of a world where subjective experiences can be measured objectively on a large scale?

Bibliography

- Buck, R. (1990). “Mood and Emotion: A Comparison of Five Contemporary Views (Book)”. In: *Psychological Inquiry* 1.4, pp. 330–336. ISSN: 1047-840X. DOI: [10.1207/s15327965pli0104_15](https://doi.org/10.1207/s15327965pli0104_15). URL: http://dx.doi.org/10.1207/s15327965pli0104_15.
- Castillo, J. C. (1997). “THE USER-REPORTED CRITICAL INCIDENT METHOD FOR REMOTE USABILITY EVALUATION”. en. PhD thesis. Virginia Tech.
- Ceniza, A. M. and Cantara, A. D. (2016). “Stress Sensor Prototype: Determining the Stress Level in using a Computer through Validated Self-Made Heart Rate (HR) and Galvanic Skin Response (GSR) Sensors and Fuzzy Logic Algorithm”. en. In: *International Journal of Engineering Research* 5.03, p. 11.
- Christensen, L. and Frøkjær, E. (2010). “Distributed Usability Evaluation: Enabling Large-scale Usability Evaluation with User-controlled Instrumentation”. In: *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*. NordiCHI '10. event-place: Reykjavik, Iceland. New York, NY, USA: ACM, pp. 118–127. ISBN: 978-1-60558-934-3. DOI: [10.1145/1868914.1868932](https://doi.org/10.1145/1868914.1868932). URL: <http://doi.acm.org/10.1145/1868914.1868932> (visited on 03/09/2019).
- Critchley, H. D. (2002). “Review: Electrodermal Responses: What Happens in the Brain”. en. In: *The Neuroscientist* 8.2, pp. 132–142. ISSN: 1073-8584. DOI: [10.1177/107385840200800209](https://doi.org/10.1177/107385840200800209). URL: <https://doi.org/10.1177/107385840200800209> (visited on 10/30/2017).
- Dabrowski, J. and Munson, E. V. (2011). “40 years of searching for the best computer system response time”. en. In: *Interacting with Computers* 23.5, pp. 555–564. ISSN: 0953-5438. DOI: [10.1016/j.intcom.2011.05.008](https://doi.org/10.1016/j.intcom.2011.05.008). URL: <https://academic.oup.com/iwc/article/23/5/555/662569> (visited on 04/02/2018).
- Dawson, M., Schell, A., and Filion, D. (2000). “The electrodermal system”. In: pp. 200–223. DOI: [10.1017/CB09780511546396.007](https://doi.org/10.1017/CB09780511546396.007).
- Desmet, P. (2002). “The basis of product emotions”. In: *Product Experience*.
- Dooren, M. van, Vries, J. J. G. (-J. de, and Janssen, J. H. (2012). “Emotional sweating across the body: Comparing 16 different skin conductance measurement locations”. In: *Physiology & Behavior* 106.2, pp. 298–304. ISSN: 0031-9384. DOI: [10.1016/j.physbeh.2012.05.011](https://doi.org/10.1016/j.physbeh.2012.05.011).

- [physbeh.2012.01.020](http://www.sciencedirect.com/science/article/pii/S0031938412000613). URL: <http://www.sciencedirect.com/science/article/pii/S0031938412000613> (visited on 10/23/2017).
- Fenton, N. E. (1997). *Software metrics : a rigorous and practical approach*. 2. ed., revised printing. PWS Publishing Company. ISBN: 978-0-534-95425-3.
- Foglia, P. and zanda, M. (2014). “Towards Relating Physiological Signals to Usability Metrics: A Case Study with a Web Avatar”. In: *WSEAS Transactions on Computers* 13, p. 624.
- Frijda, N. H. (1994). “Varieties of affect: Emotions and episodes, moods, and sentiments.” In: URL: <http://dare.uva.nl/search?smode=showBag;docsPerPage=1;startDoc=1> (visited on 10/09/2017).
- Furedy, J. J. (1983). “Operational, analogical and genuine definitions of psychophysiology”. en. In: *International Journal of Psychophysiology* 1.1, pp. 13–19. ISSN: 0167-8760. DOI: [10.1016/0167-8760\(83\)90021-1](https://doi.org/10.1016/0167-8760(83)90021-1). URL: <http://www.sciencedirect.com/science/article/pii/0167876083900211> (visited on 11/18/2019).
- Ganglbauer, E., Schrammel, J., Deutsch, S., and Tscheligi, M. (2009). “Applying Psychophysiological Methods for Measuring User Experience”. In: *Possibilities, Challenges, and Feasibility. Proc. User Experience Evaluation Methods in Product Development workshop*.
- Garbarino, M., Lai, M., Bender, D., Picard, R., and Tognetti, S. (2015). “Empatica E3 - A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition”. In: pp. 39–42. DOI: [10.1109/MOBIHEALTH.2014.7015904](https://doi.org/10.1109/MOBIHEALTH.2014.7015904).
- Gendron, M. and Barrett, L. F. (2009). “Reconstructing the Past: A Century of Ideas About Emotion in Psychology”. In: *Emotion review* 1.4, pp. 316–339. ISSN: 1754-0739. DOI: [10.1177/1754073909338877](https://doi.org/10.1177/1754073909338877). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2835158/> (visited on 10/23/2017).
- George, J. M. and Dane, E. (2016). “Affect, emotion, and decision making”. In: *Organizational Behavior and Human Decision Processes*. Celebrating Fifty Years of Organizational Behavior and Decision Making Research (1966-2016) 136.Supplement C, pp. 47–55. ISSN: 0749-5978. DOI: [10.1016/j.obhdp.2016.06.004](https://doi.org/10.1016/j.obhdp.2016.06.004). URL: <http://www.sciencedirect.com/science/article/pii/S074959781630365X> (visited on 10/16/2017).
- Gross, J. J. (2010). “The Future’s So Bright, I Gotta Wear Shades”. In: *Emotion Review* 2.3, pp. 212–216.
- Gupta, D., Ahlawat, A. K., and Sagar, K. (2017). “Usability Prediction & Ranking of SDLC Models Using Fuzzy Hierarchical Usability Model”. In: *Open Engineering* 7.1,

- pp. 161–168. DOI: [10.1515/eng-2017-0021](https://doi.org/10.1515/eng-2017-0021). URL: <https://www.degruyter.com/view/j/eng.2017.7.issue-1/eng-2017-0021/eng-2017-0021.xml?format=INT> (visited on 11/15/2019).
- Hamann, S. (2012). “Mapping discrete and dimensional emotions onto the brain: controversies and consensus”. In: *Trends in Cognitive Sciences* 16.9, pp. 458–466. ISSN: 1364-6613. DOI: [10.1016/j.tics.2012.07.006](https://doi.org/10.1016/j.tics.2012.07.006). URL: <http://www.sciencedirect.com/science/article/pii/S1364661312001726> (visited on 10/23/2017).
- Hassenzahl, M. (2005). “The Thing and I: Understanding the Relationship Between User and Product”. In: vol. 3, pp. 31–42. DOI: [10.1007/1-4020-2967-5_4](https://doi.org/10.1007/1-4020-2967-5_4).
- (2008). “User experience (UX): Towards an experiential perspective on product quality”. In: vol. 339, pp. 11–15. ISBN: 978-1-60558-285-6. DOI: [10.1145/1512714.1512717](https://doi.org/10.1145/1512714.1512717).
- Hassenzahl, M. and Tractinsky, N. (2006). “User experience - a research agenda”. en. In: *Behaviour & Information Technology* 25.2, pp. 91–97. ISSN: 0144-929X, 1362-3001. DOI: [10.1080/01449290500330331](https://doi.org/10.1080/01449290500330331). URL: <http://www.tandfonline.com/doi/abs/10.1080/01449290500330331> (visited on 05/17/2019).
- Hercegfı, K. (2018). “Heart Rate Variability Monitoring during Human-Computer Interaction”. In:
- Hercegfı, K. and Pászti, M. (2009). “A complex physiology-based empirical usability evaluation method in practice”. en. In: *Periodica Polytechnica Social and Management Sciences* 17.2, p. 57. ISSN: 1416-3837, 1587-3803. DOI: [10.3311/pp.so.2009-2.01](https://doi.org/10.3311/pp.so.2009-2.01). URL: <https://pp.bme.hu/so/article/view/1603> (visited on 04/02/2018).
- Hercegfı, K., Pászti, M., Tóvölgyi, S., and Izsó, L. (2009). “Usability Evaluation by Monitoring Physiological and Other Data Simultaneously with a Time-Resolution of Only a Few Seconds”. en. In: *Human-Computer Interaction. New Trends*. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 59–68. ISBN: 978-3-642-02573-0 978-3-642-02574-7. DOI: [10.1007/978-3-642-02574-7_7](https://doi.org/10.1007/978-3-642-02574-7_7). URL: http://link.springer.com/chapter/10.1007/978-3-642-02574-7_7 (visited on 04/02/2018).
- Hertzum, M. and Clemmensen, T. (2012). “How do usability professionals construe usability?” en. In: *International Journal of Human-Computer Studies* 70.1, pp. 26–42. ISSN: 1071-5819. DOI: [10.1016/j.ijhcs.2011.08.001](https://doi.org/10.1016/j.ijhcs.2011.08.001). URL: <http://www.sciencedirect.com/science/article/pii/S1071581911001030> (visited on 11/15/2019).

- Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004). “Design science in Information Systems research”. English. In: *Mis Quarterly* 28.1. WOS:000220207500005, pp. 75–105. ISSN: 0276-7783.
- Johannes, B. and Gaillard, A. W. K. (2014). “A methodology to compensate for individual differences in psychophysiological assessment”. In: *Biological Psychology* 96, pp. 77–85. ISSN: 0301-0511. DOI: [10.1016/j.biopsycho.2013.11.004](https://doi.org/10.1016/j.biopsycho.2013.11.004). URL: <http://www.sciencedirect.com/science/article/pii/S0301051113002263> (visited on 03/27/2018).
- Kim, J. and Moon, J. Y. (1998). “Designing towards emotional usability in customer interfaces—trustworthiness of cyber-banking system interfaces”. en. In: *Interacting with Computers*. HCI and Information Retrieval 10.1, pp. 1–29. ISSN: 0953-5438. DOI: [10.1016/S0953-5438\(97\)00037-4](https://doi.org/10.1016/S0953-5438(97)00037-4). URL: <http://www.sciencedirect.com/science/article/pii/S0953543897000374> (visited on 11/18/2019).
- Kujala, S. (2003). “User involvement: a review of the benefits and challenges”. In: *Behaviour & Information Technology* 22.1, p. 1. ISSN: 0144929X. DOI: [10.1080/01449290301782](https://doi.org/10.1080/01449290301782). URL: <http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=9234269&site=ehost-live&scope=site> (visited on 03/14/2019).
- L.%20Fredrickson and L. Fredrickson, B. (1998). “Positive Emotions Speed Recovery from the Cardiovascular Sequelae of Negative Emotions”. In: *Cognition & Emotion* 12.2, pp. 191–220. ISSN: 0269-9931. DOI: [10.1080/026999398379718](https://doi.org/10.1080/026999398379718).
- Lallemant, C., Gronier, G., and Koenig, V. (2015). “User experience: A concept without consensus? Exploring practitioners’ perspectives through an international survey”. In: *Computers in Human Behavior* 43, pp. 35–48. ISSN: 0747-5632. DOI: [10.1016/j.chb.2014.10.048](https://doi.org/10.1016/j.chb.2014.10.048). URL: <http://www.sciencedirect.com/science/article/pii/S0747563214005718> (visited on 05/16/2019).
- Levenson, R. W. (1999). “The Intrapersonal Functions of Emotion”. In: *Cognition and Emotion* 13.5, pp. 481–504. ISSN: 0269-9931. DOI: [10.1080/026999399379159](https://doi.org/10.1080/026999399379159). URL: <http://dx.doi.org/10.1080/026999399379159>.
- Lin, T. and Imamiya, A. (2006). “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, pp. 364–371. ISBN: 978-1-59593-541-0. DOI: [10.1145/1180995.1181063](https://doi.org/10.1145/1180995.1181063). URL: <http://doi.acm.org/10.1145/1180995.1181063> (visited on 04/02/2018).
- Liu, S., Zhang, B.-Y., and Liu, C. (2016). “Research on the Application of GSR and ECG in the Usability Testing of an Aggregation Reading App”. en. In: *International*

- Journal Bioautomation* 20.3, pp. 359–372. ISSN: 1314-1902, 1314-2321. URL: <https://doaj.org> (visited on 04/06/2018).
- Mandryk, R. L., Inkpen, K. M., and Calvert, T. W. (2006). “Using psychophysiological techniques to measure user experience with entertainment technologies”. In: *Behaviour & Information Technology* 25.2, pp. 141–158. ISSN: 0144-929X. DOI: [10.1080/01449290500331156](https://doi.org/10.1080/01449290500331156). URL: <https://doi.org/10.1080/01449290500331156> (visited on 03/27/2018).
- Mount, W. M., Tuček, D. C., and Abbass, H. A. (2012). “A Psychophysiological Analysis of Weak Annoyances in Human Computer Interfaces”. en. In: *Neural Information Processing*. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 202–209. ISBN: 978-3-642-34474-9 978-3-642-34475-6. DOI: [10.1007/978-3-642-34475-6_25](https://link.springer.com/chapter/10.1007/978-3-642-34475-6_25). URL: https://link.springer.com/chapter/10.1007/978-3-642-34475-6_25 (visited on 03/27/2018).
- Norman, D. (2013). *The design of everyday things: Revised and expanded edition*. Basic books.
- Pitkänen, J. and Pitkäranta, M. (2014). “Käytettävyyden arviointi ja käytettävyydestäuksen soveltaminen terveydenhuollon tietojärjestelmien valinnassa”. fi. In: *Finnish Journal of eHealth and eWelfare* 6.2-3, pp. 103–110. ISSN: 1798-0798. URL: <https://journal.fi/finjehew/article/view/45256> (visited on 09/18/2019).
- Pitkänen, J., Pitkäranta, M., and Nieminen, M. (2012). “Usability Testing in Real Context of Use: The User-triggered Usability Testing”. In: *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*. NordiCHI '12. New York, NY, USA: ACM, pp. 797–798. ISBN: 978-1-4503-1482-4. DOI: [10.1145/2399016.2399153](http://doi.acm.org/10.1145/2399016.2399153). URL: <http://doi.acm.org/10.1145/2399016.2399153> (visited on 01/30/2018).
- Poh, M. Z., Swenson, N. C., and Picard, R. W. (2010). “A Wearable Sensor for Unobtrusive, Long-Term Assessment of Electrodermal Activity”. In: *IEEE Transactions on Biomedical Engineering* 57.5, pp. 1243–1252. ISSN: 0018-9294. DOI: [10.1109/TBME.2009.2038487](https://doi.org/10.1109/TBME.2009.2038487).
- Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., and Ehlert, U. (2010). “Discriminating Stress From Cognitive Load Using a Wearable EDA Device”. In: *IEEE Transactions on Information Technology in Biomedicine* 14.2, pp. 410–417. ISSN: 1089-7771, 1558-0032. DOI: [10.1109/TITB.2009.2036164](https://doi.org/10.1109/TITB.2009.2036164).

- Siemer, M. (2005). “Mood-congruent cognitions constitute mood experience.” In: *Emotion* 5.3, pp. 296–308. ISSN: 1931-1516(Electronic),1528-3542(Print). DOI: [10.1037/1528-3542.5.3.296](https://doi.org/10.1037/1528-3542.5.3.296).
- Singh, S. N. and Dalal, N. P. (1999). “Web home pages as advertisements”. en. In: *Communications of the ACM* 42.8, pp. 91–98. ISSN: 00010782. DOI: [10.1145/310930.310978](https://doi.org/10.1145/310930.310978). URL: <http://portal.acm.org/citation.cfm?doid=310930.310978> (visited on 11/18/2019).
- Stickel, C., Ebner, M., Steinbach-Nordmann, S., Searle, G., and Holzinger, A. (2009). “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*. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 615–624. ISBN: 978-3-642-02706-2 978-3-642-02707-9. DOI: [10.1007/978-3-642-02707-9_70](https://doi.org/10.1007/978-3-642-02707-9_70). URL: https://link.springer.com/chapter/10.1007/978-3-642-02707-9_70 (visited on 03/27/2018).
- Taylor, S., Jaques, N., Weixuan Chen, Fedor, S., Sano, A., and Picard, R. (2015). “Automatic identification of artifacts in electrodermal activity data”. en. In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Milan: IEEE, pp. 1934–1937. ISBN: 978-1-4244-9271-8. DOI: [10.1109/EMBC.2015.7318762](https://doi.org/10.1109/EMBC.2015.7318762). URL: <http://ieeexplore.ieee.org/document/7318762/> (visited on 09/13/2019).
- Thüring, M. and Mahlke, S. (2007). “Usability, aesthetics and emotions in human–technology interaction”. fr. In: *International Journal of Psychology* 42.4, pp. 253–264. ISSN: 1464-066X. DOI: [10.1080/00207590701396674](https://doi.org/10.1080/00207590701396674). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1080/00207590701396674> (visited on 11/15/2019).
- Tuch, A. N., Bargas-Avila, J. A., Opwis, K., and Wilhelm, F. H. (2009). “Visual complexity of websites: Effects on users’ experience, physiology, performance, and memory”. In: *International Journal of Human-Computer Studies* 67.9, pp. 703–715. ISSN: 1071-5819. DOI: [10.1016/j.ijhcs.2009.04.002](https://doi.org/10.1016/j.ijhcs.2009.04.002). URL: <http://www.sciencedirect.com/science/article/pii/S107158190900055X> (visited on 11/02/2017).
- Wagner, J., Kim, J., and Andre, E. (2005). “From Physiological Signals to Emotions: Implementing and Comparing Selected Methods for Feature Extraction and Classification”. In: *2005 IEEE International Conference on Multimedia and Expo*, pp. 940–943. DOI: [10.1109/ICME.2005.1521579](https://doi.org/10.1109/ICME.2005.1521579).

- Ward, R. D. and Marsden, P. H. (2003). “Physiological responses to different WEB page designs”. In: *International Journal of Human-Computer Studies*. Applications of Affective Computing in Human-Computer Interaction 59.1, pp. 199–212. ISSN: 1071-5819. DOI: [10.1016/S1071-5819\(03\)00019-3](https://doi.org/10.1016/S1071-5819(03)00019-3). URL: <http://www.sciencedirect.com/science/article/pii/S1071581903000193> (visited on 11/02/2017).
- Ward, R. D. and Marsden, P. H. (2004). “Affective computing: problems, reactions and intentions”. In: *Interacting with Computers* 16.4, pp. 707–713. ISSN: 0953-5438. DOI: [10.1016/j.intcom.2004.06.002](https://doi.org/10.1016/j.intcom.2004.06.002). URL: <https://academic.oup.com/iwc/article/16/4/707/690397> (visited on 11/02/2017).
- Westerink, J., Ouwerkerk, M., Vries, G. J. d., Waele, S. d., Eerenbeemd, J. v. d., and Boven, M. v. (2009). “Emotion measurement platform for daily life situations”. In: *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pp. 1–6. DOI: [10.1109/ACII.2009.5349574](https://doi.org/10.1109/ACII.2009.5349574).
- Xu, J. and Montague, E. (2012). “Psychophysiology of the passive user: Exploring the effect of technological conditions and personality traits”. In: *International Journal of Industrial Ergonomics* 42.5, pp. 505–512. ISSN: 0169-8141. DOI: [10.1016/j.ergon.2012.07.007](https://doi.org/10.1016/j.ergon.2012.07.007). URL: <http://www.sciencedirect.com/science/article/pii/S0169814112000686> (visited on 04/02/2018).

