OULUN YLIOPISTO
UNIVERSITY of OULU

A Review of Using EEG and EMG Psychophysiological Measurements in User Experience Research

University of Oulu
Information Processing Science
Master's Thesis
Tewodros Belete Taffese
Date 24.05.2017

Abstract

Application of psychophysiological measurement tools in user experience studies is continuously growing. The development, on one hand, originates from the advantages the tool's capabilities bring to the user experience world. On the other hand, the continuous development of systems and devices and the corresponding need for more interactive systems motivate the user experience domain to navigate possible options of improvement.

The thesis reviews the use of electroencephalography (EEG) and electromyography (EMG) measurements in user experience research. Both types of measurements come to the UX domain with their unique advantages. Facial EMG is the most suitable tool to monitor the valence of emotion as well as aspects such as stress and focus. The use of EMG is dependent on the context of use and it is important to take those aspects into consideration when taking it into use. In addition EMG can be used as part a user experience design component. Its application could be implemented as interaction or biofeedback mechanisms.

From the emotion evaluation point of view, EEG is mostly used with respect to attention and withdrawal towards a system. Various aspects of emotion such as engagement, excitement, attention and focus can be evaluated using the tool. EEG can also be used as part of a user experience design component. Applications could be as a brain computer interface or feedback mechanisms for systems that can adjust themselves on the basis of such data.

The study is conducted as a narrative literature review. Background study was extensively done on user experience, psychophysiology, emotion studies, the nervous system and corresponding measurement tools. On the basis of continuous knowledge development, previous studies were assessed, classified, summarized and synthesized. Advantages, limitations, use contexts, application areas and future potentials of EEG and EMG were studied extensively. Results of the study have implications on user experience research domain, as well as industries that intend to produce products with rich user experience.

Foreword

I would like to express my utmost gratitude to my supervisor University teacher, Ph.D. Dorina Rajanen, for the excellent guidance and assistance throughout the research. I would also like to thank University teacher, Ph.D. Mikko Rajanen, for his comments and reviews.

Above all, I would like to thank my mom Ejigayehu Assefa (Ejigeye), for making me the man I am today. Without restrictions, I owe her everything good about me. I would also like to thank my lady, Zelalem Temesgen (Babyshu), for being by my side through thick and thin.

Finally, yet importantly, I would like to thank my family and friends, for always making tough times bearable and good times enjoyable.

Tewodros B. Taffese

Oulu, May 24, 2017

Abbreviations

BVP	Blood Volume Variation
ECG	Electrocardiography
EDA	Electro dermal Activity
EEG	Electroencephalography
EMG	Electromyography
ERP	Event Related Potential
GSR	Galvanic Skin Response
HCI	Human Computer Interaction
HR	Heart Rate
HRV	Heart Rate Variation
ICA	Independent Component Analysis
IOT	Internet of Things
IS	Information Systems
NS	Nervous System
PMs	Psychophysiological Measurements
SCL	Skin Conductance Level
SLR	Systematic Literature Review
UX	User Experience

Contents

1.   Introduction

For decades, there have been various research efforts on improving the usability and user experience (UX) of different products and systems. Accordingly, different forms of user experience evaluations techniques were developed. However, evaluations mostly focused on self-report evaluations, at post usage level, through questionnaires and surveys. This approach is prone to social biases and misses out on the intuitive aspects of emotions (Pentus, Mehine, & Kuusik, 2014). Emotion is a complex concept and may be too hard for users to put into words. In addition, in some cases, users' experience could be interrupted when they get questioned about their experiences during usage (Ganglbauer, Schramme, Deutsch, & Tscheligi, 2009).

Psychophysiological measurement (PM) tools have the capability to monitor a person's emotions through the process of experience (Ganglbauer et al., 2009). Knowledge and understanding of emotional states elicited by a stimulus are essential in developing products that are rich in user experience. Such tools may not directly affect the user experience, but they improve the way in which we measure it, leading to a better understanding and knowledge. The knowledge gained along the process helps in developing products with rich experience. Some of the psychophysiology measurement tools have the capability of being integrated into a product design as input/output interfaces. Instructions from users can be communicated via different gestures that can be read and interpreted by such tools. However, the latter usage is mostly deployed in novel, experimental applications rather than complete products.

Various researchers have conducted studies on psychophysiological measurement tools in a wide range of application domains. One of such domains is the Human Computer Interaction (HCI), in usability and user experience studies. Hazlet (2003) studied user frustration using electromyography (EMG), Arndt et al. (2016) reviewed the use of EEG to measure the quality of experience when viewing pictures, videos and listening to audio, Lottridge (2008) studied emotional responses as a measure of human performance through a two dimensional tool of valence and arousal, Laufer & Nemeth (2007) studied skin conductance level (SCL) as a measure of emotion and biofeedback mechanism in the context of gaming. The motivation of this study arises from the context of psychophysiological measurement tools in user experience studies. Studies in these regards are gaining more attention due to improved understanding of their potentials and respective improvement in the measurement tools' applicability. Improvements made to EEG tools could be an example of such advances made in simplifying complexity of the tools and improving their affordability.

Usability laboratory at the University of Oulu has purchased EEG and EMG tools for user experience studies. Their accessibility in the premises and the wide range of use potential is a major inspiration for this thesis. Our project aims to give a solid foundation in context of previous studies conducted with such tools and the wide range of possibilities for researchers to make use of. The results of the thesis can be used by researchers who want to make UX studies using various PM methods in general, and EEG and EMG tools in specific. We want to create a reference point where compiled understanding of previous implementations and studies of such tools can be found. We aim to improve the understanding of research possibilities and limitations of using psychophysiology in UX research. We have designed research questions that we believe will help us achieve our goals.

The research aims to give concise answers to the questions; how *can EEG and EMG psychophysiological measurements be used in user experience research*? We intend to elaborate the main research questions through answering;

1) What limitations exist and how they are overcome?
2) What are the advantages in comparison with traditional methods: subjective self-reports through questionnaires, interviews and focus groups and observational video analysis?
3) How are psychophysiological measurement tools used in user experience studies?
    a. As part of UX design or
    b. As a UX evaluation tool.
4) What applications areas are researched?
5) What are the possible future research directions?

A narrative literature review is believed to be the best approach in order to answer the research questions. The study of psychophysiological measurements in the context of user experience studies is a relatively new domain and literature review can help lay a foundation for advancing knowledge. Accordingly, the study is divided in chapters that intend to answer those specific questions. Each chapter reviews previous studies conducted with respect to the topic critically. The study followed three fundamental steps through the research process; first topic relevance and context establishment, second continuous inclusion of relevant material for review and third, synthesizing the reviewed studies in a sensible manner. More details of the methods used, the process, research categories, and validation of methods is discussed in the research methods chapter (Chapter 3).

The thesis is structured in nine chapters, four of which are directly related to the research questions we want to address. Accordingly, chapter 2 focuses on laying a theoretical foundation regarding the concept of user experience studies, emotion & motivation, the human nervous system, psychophysiological measurements with more focus on EEG and EMG. Chapter 3 explains the research method used in this study and how it is validated. Advantages and limitations of psychophysiological methods with the focus on EEG and EMG (RQ1, RQ2), is reviewed in chapter 4. On the basis of the context of use established at the background study, chapter 5 discusses the two main formats psychophysiological tools are used in user experience studies (RQ3). Application areas and future directions are reviewed in the following chapters 6 & 7 (RQ4, RQ5). A summed up view of the review is discussed in chapter 8, followed by the conclusions in chapter 9.

2. Theoretical Background

Before going into details of the psychophysiological measurements, it is important to understand the research perspective. The focus here is reviewing the use of psychophysiological measurements for user experience studies. As a background study, we first discuss user experience and usability from the perspective of our study, then emotion, motivation and its meaning are discussed with the intention to give an idea of what we measure and how they are related to psychophysiology, the human nervous system is elaborated to give a view of where psychophysiological measurements (PMs) tools lie in the human anatomy, and finally psychophysiological measurement tools are explained with more emphasis on EEG and EMG (Figure 1). User experience studies and where the psychophysiology fits in is explained justifying the need for PM tools in UX. The need, in this case may also originate from the potentials of what PMs could supply us with in UX evaluations, or from the evolution of user experience studies themselves.
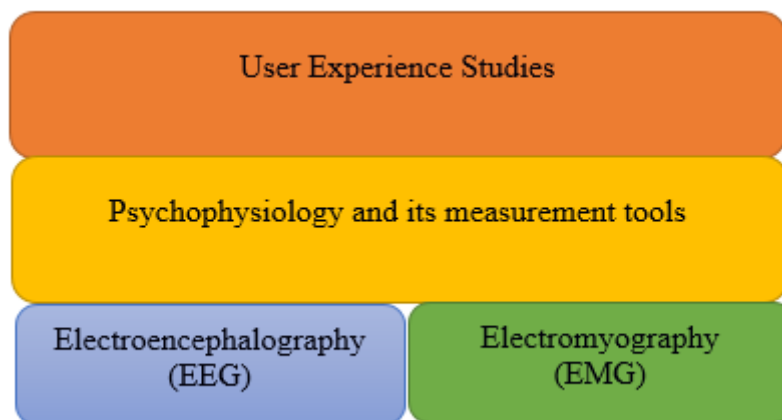
Figure 1. Overall thesis architecture

Overall, as an outcome, the project intends to make scientific contributions in terms of understanding the role of psychophysiology in UX studies. This result aims facilitate the research community with a platform and/or an idea of conducting further studies (Figure 1).

2.1 User experience (UX)

Usability is an important aspect of software and/or product development that has continuously grown in significance. The field of usability has evolved to become a major component of software development. Usability of a product and the emotional experience it supplies users with, is a critical part of its success. Usability has had various definitions in software development. One definition of usability refers to how efficiently software is utilized by users (Hedberg, Iivari, Rajanen, & Harjumaa, 2007; Zhao & Deek, 2006).

HCI community has defined usability as being "*the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use*" (ISO/IS 9241-11, 1998).

On the same standard human-centred design is defined as an "*approach to systems design and development that aims to make interactive systems more usable by focusing on the use of the system and applying human factors/ergonomics and usability knowledge and techniques*" and user experience is

"*person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service*" (ISO/IS 9241-11, 1998).

The overall theme is that when users use a software product or an interactive system, they should be able to interact with minimal difficulty and carry out their tasks effectively without any significant problems.

Usability does not necessarily mean an aesthetically pleasant looking interface (Boivie, I., Gulliksen, J., & Göransson, B., 2006), but a contextual relation with the users. Hedberg et al. (2007) argued that usability should always be from the perspective of the user intended to use it. In other words, a usable software should allow the intended users be able to complete the desired task efficiently. Other scholars relate usability with aspects such as, "learnability", "efficiency", "memorability", "error", and "satisfaction" (Nielsen J., 1993; Iivari, Hedberg, & Kirves, 2008). A usable software should be easy to learn and use, intuitive to remember, and free from errors or providing a way for easy recovery when errors occur.

Hassenzahl & Tractinsky (2006) draw a very clear distinction of the trends from HCI and interaction designs towards user experience. The authors give a picture of what UX is and what it will become. In their view, the UX concept is a combination of three predominant perspectives; it deals with addressing human needs beyond the instrumental (i.e., includes holistic, aesthetic, hedonic dimensions), it stresses the affective and emotional aspect of interaction (i.e., subjective, positive antecedents and consequences), and it is dependent on the nature of experience (i.e., dynamic, unique, complex, situated). They argue that when the three perspectives meet/intersect, we can say a technology fully captures a user experience (Figure 2).
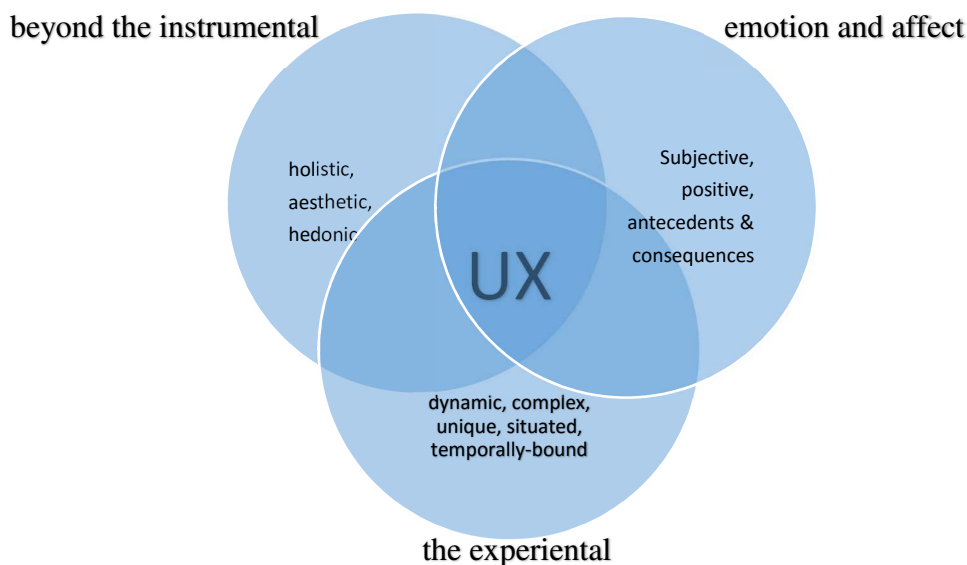
Figure 2. Facets of UX (source: Hassenzahl and Tractinsky, 2006)

Hassenzahl and Tractinsky (2006) demonstrated the range of topics in UX, its approaches, and results, and defined UX as a multidimensional construct spanning three axes: the user' internal state, the system's characteristics, and the context of use and interaction between user and system: "*UX is about technology that fulfils more than just instrumental needs in a way that acknowledges its use as a subjective, situated, complex and dynamic encounter. UX is a consequence of a user's internal state*

(predispositions, expectations, needs, motivation, mood, etc.), the characteristics of the designed system (e.g. complexity, purpose, usability, functionality, etc.) and the context (or the environment) within which the interaction occurs (e.g. organizational/social setting, meaningfulness of the activity, voluntariness of use, etc.)". When studying UX, PM tools can be employed to measure and monitor the internal state of users (such as mood, emotions, and cognitive load) when interacting and using a system. The aim is to evaluate the internal state of users during the interaction and use of a system, to identify potential problems with the interface, and ultimately to improve the usability and user experience of the systems. Typically, the PM tools are employed in a laboratory setting; however, with the advancement of wireless technologies, the application of psychophysiological research extends also to field and real-life situations of product use.

In the field of computer science, there has been a continuous process of trying to understand the user perspective of things. The areas of user centred design, human-computer interaction, and user experience are results of such efforts to fill the gap between user's feelings and how products cater to it, or what type of emotions products generate. Hornbæk (2006) identifies, among key research issues, the agenda of improving the usability of interactive systems. The author proposes alternative methods to the traditional evaluations to obtain more valid indicators of usability. In other words, it is not sufficient to only think about improving usability, but also the way we measure it. Accordingly, the understanding of people's emotional responses is a valuable knowledge base not only in directly improving usability, but also in improving the approach we take in improving it.

The most commonly used techniques to get user feedbacks is, for example, through post experience questionnaires. Such approaches have limitations in a sense that they do not tell the whole story of the user experience and are also biased with social masking (Ganglbauer et al., 2009; Lottridge, 2008). Users may only remember their peak emotions, may not want to give negative reviews, and there could be some emotions they cannot put into simple words. Such subjective measures do have their usages and are important, but do not paint a complete picture. Such limitations lead to the need for improved techniques of user experience studies.

Mandryk, Atkins & Inkpen (2006) define the standard methodologies in three categories: subjective self-reports through questionnaires, interviews & focus groups, and objective report through observational video analysis. The authors explain that subjective report is good for understanding attitude of the user, but also state that subjects are bad at self-reporting in certain situations. The novelty of an application especially in the case of entertainment environment could potentially skew subjective responses of the users. Observational analysis using videos is also time consuming and rigorous experience with few consulting firms that specialize in observational analysis of entertainment technologies.

Mandryk et al. (2006) also suggest that techniques such as heuristic evaluations are useful in uncovering usability issues within a play environment but fail in evaluating playability of an entertainment technology, missing out on analysis of overall experience. In addition, this technique is usually conducted using experts and when it comes to novel applications, there are no experts. Think aloud techniques are also unfitting to play environments due to the disturbance to the player and the impact they have on game play. The study presents the current user experience methodologies in an X/Y graph; a qualitative to quantitative X-axis and a subjective to objective Y-axis graph. They state that there is a knowledge gap on the Objective-Quantitative quadrant, since all traditional methods fail to address it fully. Accordingly, psychophysiological tools are expected to fill that gap in the areas of user experience studies.

Psychophysiological measurements are gradually becoming handy in facilitating an understanding of user's emotions practically during the process of usage (Ganglbauer et al., 2009). Such aspects of psychophysiological measurements make them a critical means of improving product quality and the

experience it generates. Some research even suggests that such tools will facilitate a human computer interaction without the use of a physical interface such as keyboards and mouse (Culpepper, 2003). Such changes however, do not only need the development of devices and tools that can read our emotional responses, but also implies that the way we think about our actions should change to convey our intentions to the computer more clearly. There is ongoing research in this direction, for example, on a brain signal based interface in vehicles (He, Bi, Lian, & Sun, 2016).

It is important to note that PMs have their own limitations and should be used with complimentary tools, or at least with an understanding that limitations exist in them as well. The major drawback has been that PMs, in most cases, are found to be costly and complex (Ganglauber et al., 2009; Pentus et al., 2014). In addition, some papers also argue that PMs are not capable enough to draw on discrete emotions. The studies claim that discrete emotions are complex in nature and vary culturally and individually, leading to difficulty in designing a discrete model for them (Ganglauber et al., 2009).

To overcome the limitations of PMs, researchers argue that it is important to use a mixture of PMs with traditional methods. Hornbæk (2006) argues that it is vital to consider both subjective and objective measurements when dealing with user experiences. Ganglauber et al. (2009) claim that PMs should be viewed as complementary to subjective studies and they should be used in a bi-modal format with traditional methods. However, what all researchers agree with is the fact that PMs are very vital tools, which supply us with methods to uncover social masking, make continuous (uninterrupted) measurements, and address emotions and experiences as they occur (Lottridge, 2008; Ganglauber et al., 2009; Al-Husain, Kanjo, & Chamberlain, 2013; Hornbæk, 2006).

The possibility of being able to study emotions during the process of usage (in real-time), rather than post experience via questionnaires together with the understanding and knowledge of emotions that can drive us into developing systems that target a complete experience, are the main driving forces of studies regarding PMs in the UX context.

2.2 Emotions

Studying user experience and emotion measurement tools may not be complete without the understanding of emotions themselves. This section gives an overall picture of what emotions are and how they relate to using a media and or an interactive system.

Emotions have had different interpretations across different times and eras through the study of behavioural science and psychology. Affective neuroscience is concerned with trying to understand the concepts of emotions and moods in the neurobiological and psychological levels and their interface. This is accomplished via the usage of neuroimaging, behavioral experiments, electrophysiological recordings, animal and human lesion studies, and animal and human behavioral experiments. Studies of affective emotions dates to Darwin (1872), who made two key findings in his study of human emotions in humans and animals. His two key findings were that 1. Animal emotions are homologues for human emotions, and 2. There are limited sets of fundamental emotions that exist across species, those emotions include anger, fear, surprise, and sadness. His perception of the fact that emotions need a cause and are not only feels is a key concept. However, it was William James, who is known to be the father of psychology, who laid out an introduction into a physiological approach of emotions (James, 1890). He claims that the essence of emotions is a physiological response. When we encounter a stimulus, certain physiological response follows that elicits an emotion based on perception.

James' view of physical phenomenon as a cause for emotions was supported by Carl Lange who was doing the same investigation in parallel. Their collaboration led to the James-Lange theory in 1885. Their theory concluded that bodily changes code different emotions. Although their assumptions were controversial, the emphasis they put on the link between psychology and physiology remains to be

significant in affective neuroscience, more specifically, the view that body concomitants of emotions can change the intensity of an experience. Even though emotions will later be proved not to originate from physical movements, there is a relationship between physiological responses and emotions felt. (Dalglish T., 2004.)

The James-Lange theory was later challenged by Cannon-Bard theory (1920), which states that physiological responses alone cannot be the root causes of emotion. They believed that physiological emotions are too slow and their artificial hormonal is not sufficient to generate emotion. Physical response and human emotion occur simultaneously after a brain pattern activation with respect to stimulus (Dalglish T., 2004). They conducted their research through surgical separation of the viscera from the brain in animals. Their study marked two beneficial points: 1) the usage of animal emotions as traces for human emotions as noted by Darwin, and 2) the use of surgical brain lesions to understand emotion. Their thinking is that changes due to the operation must reflect on the processes that are related to the lesion parts of the brain.

The next major advancement was the dual mode theory and conceptualization that cognition is an important element of emotion (Miller 1959 as cited in Cacioppo, Tassinary, & Berntson, 2007). The dual mode conceptualization states that emotions can be broken down into two states, such as action/withdrawal, appetitive/aversive, or approach/withdrawal. This thinking led to and/or is inclusive of a two-state emotion conceptualized based on positive to negative dimensions (Dalglish, 2006). This concept is an essential starting point for the notion of valence and modern day measurements of it.

Further studies have led to the concept of multi-system models, which states that certain groups of discrete emotions are underpinned by various neural systems in the brain. These studies originate from human lesion and functional neuroimaging studies that associate emotions such as fear and disgust to the activation of different nervous systems. However, further studies suggest that more brain regions than anticipated are involved in the processing of emotion. Even though there is a lack of consistency about a definite conclusion about the roles of individual regions of the brain on emotion processing, there is an impressive body of knowledge developing. (Cacioppo, Tassinary, & Berntson, 2007.)

There have been disagreements and arguments on the definitions of emotion. Kivikangas (2015) states that the main error originates from the attempt in giving emotion a generalized, single definition. In a review of previous studies, Kivikangas (2015) compiles emotions theories into four major categories. The first conceptualization views emotion as discrete emotional states, i.e., anger, disgust, fear, happiness, sadness, and surprise. These emotional states are biologically fixed and are universal to all humans, and few other animals as well. A combination of discrete emotions creates complex emotions. Complex emotions are also influenced by culture and environment in a person's make up (Ekman, P. & Cordaro D., 2011). The second theory states that emotions are unique to each unique situation. They are constructed in the moment of experience and are attributed, conceptualized, and contextualized resulting in a unique emotion towards a stimulus (Cunningham, W. A., Dunfield, K. A. & Stillman, P. E. 2013).

The third conceptualization views emotion as a dimensional construct. In this view, emotion is a combination of psychological dimensions such as valence (pleasant/positive to unpleasant/negative axis) and arousal (anxiety to boredom axis). This two-dimensional approach states, for example, that two different emotions such as happiness and sadness, could have the commonality in their level of intensity, which is the arousal level. Happiness has high arousal and positive valence, whereas sadness has high arousal but negative valence. Accordingly, emotions do not differ from each other categorically, but rather through the central dimensions of valence and arousal states (Kivikangas, 2015).

The fourth concept or theory states that the mind evaluates/appraises a stimulus according to a set of criteria in order to produce a certain response. It evaluates a stimulus and parts of its content against biologically predefined appraisal values. Interacting together, these appraisals define the emotion (Kivikangas, 2015).

For this study, the most important aspect we need to understand is that there are no emotions without physical expression (James, 1890), and that the physiological measurements help us understand and identify emotions as they occur continuously during a task or exposure to complex stimuli, for example, when individuals are interacting with a software application or digital interface. Although differences exist between concepts of emotion, there is a lot less disagreement when it comes to the measurement tools. The explanation in this section was intended to give the basic understanding of emotion and some of its components. Latter sections will be referring to valence, arousal, and frontal asymmetry approach/withdrawal attributes of emotion that are typically measured by employing PM tools. Hence, a brief introduction to the overall picture was felt to be needed.

2.3   The nervous system

The ability of humans to feel and act on a certain event based on their emotions is a critical element of life (Potters & Bolls, 2012). Every minute-by-minute interactions and actions of life are related to our varying emotions, with different levels of intensity. Our capacity to experience emotions through interactions with different systems and events is heavily linked with our mental ability to perceive, process, and act on them (Potters & Bolls, 2012). There is a constant interaction between the human body and mind that is responsible for the emotional experience (Larsen et al., 2008). These interactions and activities between the human body and mind, are the ones used for evaluation of emotions, when talking about PM tools.

The interactions of the mind and body dictate how we feel about things we come across with. Zimmermann (2016) explains that signals between sections of the body are carried through by a collection of nerves and neurons that make up the nervous system. The nervous system (NS) is composed of two elements, the central and the peripheral nervous systems. The central nervous system comprises of the brain and the spinal cord. Psychophysiological measurements that focus on this section of the nervous system intend to directly measure and index specific brain activities (Figure 3). Electroencephalography (EEG) belongs to this category and intends to directly measure emotions and cognitive processes through direct measurement of brain activities during an activity (Potters & Bolls, 2012).
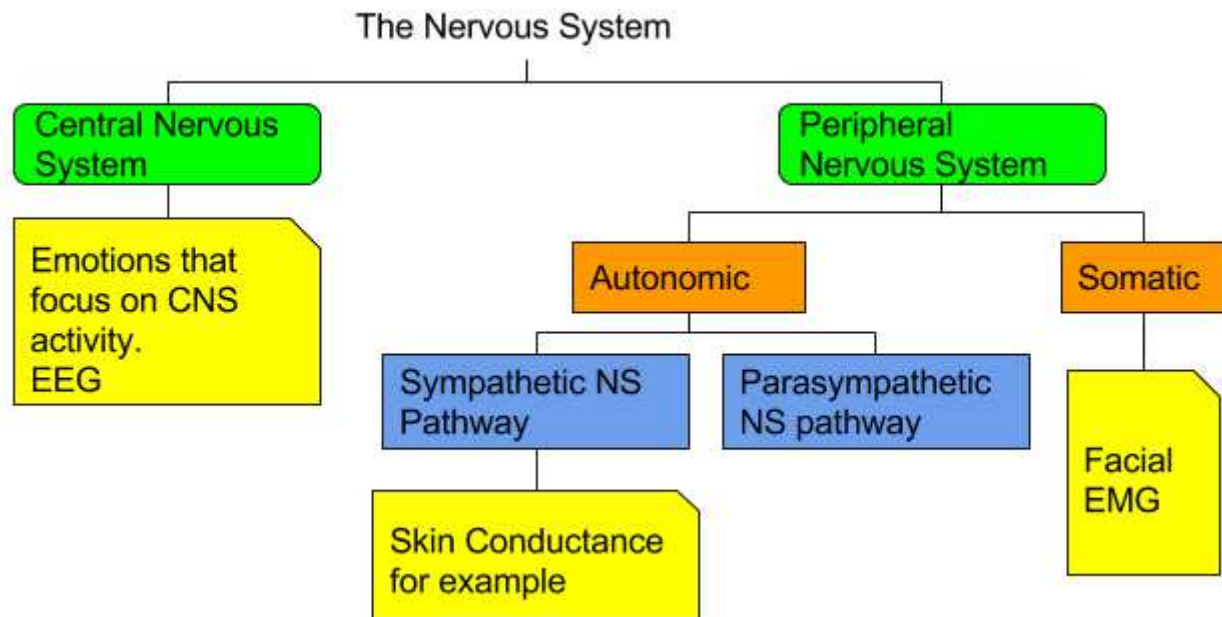
Figure 3. Overview of the nervous system setup and its relation to emotions measurement tools.

The peripheral NS is divided into autonomic and somatic branches. The somatic branch is responsible for innervating skeletal muscles. It is part of the nervous system that is responsible for the external, more voluntary processes (Potters and Bolls, 2012, p. 127). Facial electromyography (EMG) falls in this category; it measures activities underlying facial muscles in emotional experiences (Figure 3). The autonomic branch is responsible for the internal world which are composed of more automatic processes. It innervates organs and glands through two pathways; the sympathetic nervous system and the parasympathetic. Sympathetic NS is related to the facilitating energy expenditure, more like in the fight or flight situations. They are represented by responses such as increase in heart rate (HR), blood pressure (BP), dilated pupil, and skin conductance or electro dermal activity (EDA) (Potter and Bolls p.107, 2012). Parasympathetic NS, on the other hand, facilitates energy storage, mainly in situations that call for rest and digest. This state is represented by a calming, and relaxing feeling, for example, slowing of HR, BP, digestion activation and contraction of pupils. This section gave a brief understanding of emotions with respect to their mapping with the nervous system and the psychophysiological measurements that can be used to capture them. In the next sections, we describe in more detail the psychophysiological measurements.

2.4 Psychophysiological measurements (PMs)

Psychophysiology is a branch of psychology that deals with the interaction between the mind (psyche) and body (physiology). It is a scientific study that deals with such interactions and correlations. It is a multidisciplinary topic with the inclusion of scientists and researchers from various disciplines. In a nutshell, it is a section of psychology that deals with physiological changes that are instigated by psychological stimuli. Cacioppo and Tassinary (1990; as cited in Ravaja, 2004) sum up the definition of psychophysiology and its purpose as:

> "Psychophysiology concerns the study of cognitive, emotional, and behavioral phenomena as related to and revealed through physiological principles and events. As a discipline psychophysiology not only addresses fundamental questions regarding human processes (e.g., mind–body relationships, organismic–environmental transac-

tions) but also provides a conceptual perspective and a methodological armamentarium that cuts across aspects of the biological, behavioral, and social sciences." (Cacioppo and Tassinary (1990; as cited in Ravaja, 2004))

Psychophysiology and its wide range of application potentials are quite intriguing. Thus, there have been tools and techniques developed to monitor physiological responses to mental processes. Such tools include; the electroencephalograms (EEGs), magnetic resonance imaging (MRI), electromyography (EMG), electrodermal activity (EDA), cardiovascular systems and computerized axial tomography scans. (Encyclopaedia, Culpepper & Keller et al., 2003, Potter & Bolls, 2012).

The first instance of using the PMs dates back to 1933, when Dysinger and Ruckmick measured the skin resistance of participants watching films, while dipping their two fingers in a box. The pulse rate data was also monitored simultaneously. However, the study had too much variation in reactions and its conclusion was that impact of film was dependent on individuals' mental lives and should be seen on an individual psychophysiological make up basis (Dysinger & Ruckmick, 1933). The results, bound to its limitations, adhere to several theoretical concepts in modern psychology.

Studies in the early days did not materialize to fulfil the main ethics of science, in the context of understanding why things happen and how things add up to forming a certain effect (Sparks, 2002). It is this realisation that led scientists to seek for a deeper understanding of a broader context that includes wider area of disciplines including psychology, linguistics, computer science, anthropology, and philosophy. Such transition evolved into the direction of information processing approach, in the 1970s (Potter & Bolls, 2012). Overall, it is through such evolutions that the study of psychophysiology and psychophysiological measurements came to their current form. Since the focus of this research is on selected measurement tools (namely EEG and EMG) and their use context, in the following, we focus on the PM tools applied to user experience concepts.

As noted in Section 2.2 (Emotions), although there are differences in the concepts of emotion, there is a basic common understanding regarding the measurement of emotion. Evidences of the neurophysiological basis of emotions, because of psychophysiological and neuroscientific research advancements, paved the way to focus on the neural perspective of emotion instead of the subjective experiences alone (Lang P. J., 2014). The increased availability and development of neuroscientific techniques have also initiated interests from Information Systems (IS) scholars who have begun studying the potentials of neuroscience in IS research (Riedl, Banker, Benbasat, Davis, & Dennis, 2009). The theories of applying neurosciences into the studies of Information Systems is also referred to as NeuroIS. The concept of NeuroIS also falls under the category of applying psychophysiological methods, neuroscience, and neurophysiology in the context of information systems studies in general and user experience in specific.

2.5 Electroencephalography (EEG)

Electroencephalograms (EEG) record the electric activity of a human's brain through electrodes that are attached to the scalp. Traditionally the tool was used for medical purposes in diagnosis and management of various brain related disorders. For example, in the case of epilepsy, abnormal neuronal activity can be measured using EEG analysis (De Smedt, 2016). This feature might also assist in tracking such patients' brain activities while they are carrying out their daily tasks.

EEG's capabilities also include monitoring the changes of emotions as shown by the brain signals. Potter and Bolls (2012) explain that EEG is a direct measure of central nervous system activity that is initiated by information processing in the brain. Skroumpelou, Mavros, and Smith (2015) point out that EEG is a brain imaging tool that supplies us with an insight of a human's affective state of mind. It is this capability that identifies EEG as a unique PM tool, in a sense that other methods such as

EMG and skin conductance level measure physical/peripheral reflexes of emotions. Researchers then draw inferences about those reflexes and their related process in the central nervous system (Potter & Bolls, 2012). Researchers in the areas of computer science have noted the potentials of EEG and have been working to exploit its benefits such as in the development of brain computer interfaces and in the evaluation of usability and UX.

The understanding that EEG is vital in giving direct measures of emotion and cognitive processes was identified by media researchers. Measuring brain activity of individuals, while they are viewing different television advertisements was one of the earliest PM research (Potter & Bolls, 2012). However, it is not without limitations that EEG came to the media research sector. One of the drawbacks of the EEG was that it is complex to implement, since there are a lot of interfering signals through simple gestures such as a blink of an eye (Ganglbauer, 2004). Moreover, discrete emotions are difficult to differentiate and an experience may contain more than one emotion (Ganglbauer, 2004). Such limitations have been identified by other researchers as well who point out that independent component analysis (ICA) is needed to distinguish between signals. For example, Culpepper & Keller (2003) used ICA techniques in their study on enabling computer decision based on EEG input. The authors identified EEG as a valuable tool for human computer interaction and their main aim was to validate the use of ICA as an algorithm that removes unwanted artefacts without losing useful data in the HCI context. The authors suggest that such possibility paves the way for EEGs applicability in a real-time control systems environment.

EEG complexity makes the work a bit more tedious for media researchers who may need further mathematical knowledge. Such difficulties of signal interference were also noted by Gupta et al. (2016) in their pursuit to design an affect recognition system for user centric multimodal indexing. Their work also included the usage of EEG, electrocardiography (ECG), and galvanic skin response (GSR). The problem of signal contamination is not unique to EEG, but occurs to all physiological measurements. However, nowadays some tools come with built-in software for signal differentiation and analysis. Such transitions and/or improvements are paving the way for more research that focus on the emotional signals and how such knowledge could be used in improving user experience.

Another limitation with EEG and other PM tools was that they may affect the users comfort with wires and electrodes attached to their bodies (Ganglbauer et al., 2004; Wang, Gwizdka, Chaovalitwongse, 2015). The authors claim that most PM methodologies, including EEG, restrict participants in terms of moving, feeling, and interacting with a system. Movement can also create interfering signals in the form of heart rate and electro dermal activities, which add to the first issues identified in previous paragraphs. They suggest that the least susceptible for such interferences and more viable for measuring negative or positive valence of emotion is facial electromyography (EMG), which monitors our facial muscles movements.

The other issue with the EEG was portability and price, when it comes to using the tool on research and development (Pentus et al., 2014). Ganglbauer et al. (2004) also supports the notion that one of the limitations of PM tools is their expensive prices. This idea is validated by various studies including Potter & Bolls (2012), who state that financial limitations together with lack of technical expertise has limited the wide usage of EEG tools in most university media labs. It is also suggested that the common usage of EEG tools was bulky and expensive, limiting its usage in the context of real world applications (Wang et al., 2015). However, the latter authors also elaborate that the emergence of wireless EEG acquisition systems, such as Emotiv, paves the way for overcoming the price and portability issues.

The complexity and handiness of EEG tools is being continuously simplified. For example, Emotiv Epoc is an EEG portable headset device which is light and a lot more convenient to use than its predecessors. The device is increasingly becoming portable and affordable, in a suitable manner for

research (Pentus et al., 2014). Such portable devices have made it easier to conduct researches in terms of comfort, but have also increased the deployment areas and use context of the technology. As a result, such portable tools are being used in wider range of research sectors such as urban planning, assisted living for elders and people with disabilities, patient monitoring, gaming industries, and in HCI studies (Skroumpelou et al., 2015, Evdokimos et al., 2015, De Smedt & Menschaert, 2016).

2.6 Electromyography (EMG)

Electromyograms record electronic signals that are related to muscle activities in the respective body parts (Costanza, Inverso, & Allen 2005; Potter & Bolls, 2012). The recorded values have been used as primary psychophysiological indicators for valence and arousal (activation) of emotional experiences across various research contexts.

Facial EMG is the recording of electronic signals that are related to muscles on the face. There are four facial muscle groups that are mainly associated with valence and arousal in PM studies. Those muscles are the zygomaticus major (the muscle around cheekbones that lifts the lip in case of smiling), the corrugator supercilii (the muscle that is responsible for a frown pulling the inner half of eyebrows downwards), the orbicularis oculi (a muscle that close the eyelids), and frontalis (responsible for moving the eyebrows upwards). (Partala et al., 2005; Potters & Bolls, 2012, p. 125; Branco, Firth, Encarnação, & Bonato, 2005.)

As stated in the previous section, EMG is found to be a very reliable PM tool. It is one of the tools that is extensively used by researchers (Potters & Bolls, 2012). EMG is carried out by monitoring the voltage amplitude around those muscle areas. This is related to how our muscles are bundled together. Each muscle has striated fibres bundled together, to make up motor units. Each motor unit is innervated by neuron axons. When an action potential travels through these neuron axons, all the striated fibres are activated causing a muscle contraction (Potters & Bolls, 2012; Costanza et al., 2005). The neural signals are generated prior to the actual smile. However, these signals are not strong enough to be identified without signal amplifications. Hence a signal amplifier of 50 to 100K is needed. It is also important to note that these signals are prone to interferences from different electrical noises, including equipment such as computers that may be running during a recording. However, this issue could be minimized through using band pass filters that exclude signals other than the desired facial muscle movement activities, soothing, and rectification of electronic signals. Some of the amplification devices are equipped with such tasks, making the lives of a media researcher easier (Potter & Bolls, 2012).

The signals collected using EMG may not be hundred percent accurate in telling that every smile represents a pleasant experience and every frown of otherwise. Hence, in most cases, judgments of emotions are evaluated by using a combinations muscle reflexes (zygomatic with frontalis for example) that help identify a more precise indication of those expressions (Branco et al., 2005). But in a general context, the zygomaticus major and the corrugator supercilii are major indicators of valence variation, between a positive and negative or neutral state of emotion. These values in turn correspond to the smile-frown events during an experience, or in real time (Partala, Surakka, & Vanhala, 2005).

The most important attribute of EMG has been identified as its ability to measure variance of valence, capability to detect invisible muscle activity and/or movements, and its zero susceptibility to illumination and different visual approach limitations (Ganglbauer et al., 2009; Costanza et al., 2005; Potters & Bolls, 2012; Partala et al., 2005). In most cases, the facial EMG is used in combination with other PM tools such as skin conductance, blood pressure, respiration, and EEG (Ganglbauer et al., 2009; Costanza et al., 2005).

The potentials of EMG, similarly to EEG and other PM tools, are immense. Researchers look for a future where such tools could be used as input/output tools, reducing and/or possibly removing the need for physical devices such as keyboard, mouse, or even touch screen in some context. Muscle movement driven task executions, facial movement and/or expression such as eye blink, expression driven error reporting, or even moving virtual objects with brain signals is a consideration (Costanza et al., 2005; Partal, Surakka, & Vanhala, 2005; Konstantinidis et al., 2015).

In summary, this chapter explained what is meant by psychophysiological measurements in general and EEG and EMG in specific, emotions and its measures, and user experience briefly as the context of our research. Each chapter itself is a broad area with a capacity of independent study and coverage in their own right. However, it is important to stick to the main goal of the current study which is to understand the usage, use context, benefits and drawbacks of psychophysiological measurements for user experience research, thus the level of detail in this chapter was limited to the purpose of this study. In the next chapter, we describe the research method employed to answer the project goals.

3.   Research Methods

This work was started due to a purchase of EEG and EMG tools by the University of Oulu for user experience evaluation studies. As stated in the introductory section, the project intends to give a solid foundation regarding studies conducted using such tools and the wide range of possibilities for researchers to make use of. The results of the thesis can be used by researchers who want to make UX studies using various PM methods in general, and EEG and EMG tools in specific. We want to create a reference point where compiled understanding of previous implementations and studies of such tools can be found. We aim to improve the understanding of research possibilities and limitations of using psychophysiology in UX research.

To reach the intended goals, few research questions were designed extracted from background studies. Accordingly, the aim is to find an explanation to the following research questions: What limitations exist and how they are overcome? What are the advantages in comparison with traditional methods: self-reports, observations, usability testing? In what contexts are these methods used: as part of UX design or UX evaluation? What applications areas are researched: game, mobile shopping, and e-commerce? And what future research directions are important to follow, in order to advance the field?

Studies of psychophysiological measurements for user experience evaluations are relatively in an early stage of development and topics are distributed across different forms of digital services. Early studies were on the usage of such tools in media research (Potter & Bolls, 2012), mainly about how people consumed media resources and how those stimuli affect observers. In order to address the research questions and reach the goals intended, a narrative literature review was selected as the optimal method of study. Literature reviews are key components on science in laying a foundation of advancing knowledge. However, there is a lack of its usage in Information Systems studies and the need for it apparent (Watson & Webster, 2002). MIS quarterly launched the MISQ Review with the main objective of advancing the knowledge base in IS. The nature of our study and the lack of similar studies, that compile previous studies of a similar topic in an UX context, is one motivator behind this work.

Pursuing a theoretical study rather than practical, we had two general choices of conducting our review. One option is to conduct a systematic literature review (SLR), and the second a narrative/traditional literature review. After a careful view of our goals and the method's suitability to our study, a narrative literature review was selected.

The method selection depends on the purpose of the thesis. Generally, it is claimed that SLR provides a standardized method for literature reviews through replicability, transparency, objectivity, rigour and unbiasedness. In a sense, the method follows a step by step procedure/protocols of rigorously searching studies, documenting all the steps and analysing their data based on findings of the papers. Boell & Cecez-Kecmanovic (2015) carefully study this perception/assumption of SLRs and their applicability in the IS domain. They claim that the SLR's perceived superiority over traditional literature review is misleading and that its applicability is heavily dependent on the purpose of a study. They studied the method through examining its origins, the prescribed protocols, the perceived assumptions and critically evaluating the claims and implications. Overall, they argue that, in the context of IS, SLR leaves little room for a researcher's personal input or analysis and is focused too much on searching rigor and how the process of identifying the studies is reproducible by others (Boell & Cecez-Kecmanovic, 2015). Such an approach may be suitable for literature reviews that focus on a specific purpose of summarizing clearly identifiable evidence from earlier research. The authors also note that such limitation is clearly understood in the field of medicine, where SLR originated from, and is neglected in its adaptation in the field of IS. In the latter field, it is usually perceived as offering a

general approach to conducting literature reviews. Conducting this thesis and searching for thesis works that make use of literature review in the Degree programme of information processing sciences, it was evident that SLR is being seen as the main method to conduct literature reviews. Almost all the Master's thesis studies I found were conducted using SLRs and finding a single paper fully conducted with a narrative literature review was difficult.

The general view of Boell & Cecez-Kecmanovic (2015) is that SLR is being welcomed in the field of IS without sufficient reflection on their assumptions, limitation, and implication in the domain. Maclure (2005) also noted the threat of SLRs in the context of IS, stating that if they are implemented without a thorough analysis of their advantages and limitations. For example, in SLR, database searches are key processes and it assumes that good terms and search strings can be defined in the beginning of the study (Boell & Cecez-Kecmanovic, 2015). However, terms and definitions of a study are not definitely finite and cannot be fully known prior to thorough reading of the topic. For instance, in this study, terms and definitions of psychophysiological measurements and its definitions as NeuroIS is found through the process of review and reading. The concept of categorization of the PM tools use context is also driven through the process of reading. The latter of which led to searches of studies focusing on the relevant contexts defined. Especially when the topic is new, as is the case with this study, terms and definitions are not strictly defined.

A summarized view of the distinctions and commonalities of SLRs and narrative literature reviews from the perspective of Boell & Cecez-Kecmanovic (2015) is shown in table 1. The focus here is not to compare reviews, but to show the choice validation and the basis of it. As stated previously the purpose of this thesis is to create a reference point where compiled understanding of previous implementations and studies of psychophysiological tools in the context of UX can be found. Hence the aim is to improve the understanding of research possibilities and limitations of using psychophysiology in UX research. This resonates with the idea of traditional literature reviews where the purpose is to examine previous research and develop an understanding of the phenomenon under examination (Table 1). In our case the phenomenon is the increased use of PM tools in user experience evaluation studies and the corresponding advancements in the tools use context and application.

	Systematic literature reviews (SLRs)	**Traditional literature reviews**
Description	Is a protocol based approach to literature review which explicitly prescribes the steps and processes for searching, selecting and validating studies and summarizing their results (Atkins and Louw, 2000; Okoli and Schabram, 2010). 'Systematic' assumes that the review process follows explicit procedures and rules as a guarantor that the literature review is 'objective', 'scientific', 'transparent', 'replicable', 'rigorous'.	Traditional literature review is a creative process through which a researcher identifies and examines prior research and develops increasing understanding of a phenomenon under examination and in the process, constructs the relevant body of knowledge. Various guidelines for literature reviews assist researchers in identifying, reading, analyzing, interpreting, mapping, classifying and critically assessing the literature and writing a literature review (Webster and Watson, 2002; Levy and Ellis, 2006; Wolfswinkel et al.,2013; Boell and Cecez-Kecmanovic 2014).
Purpose and aims	Seeks to provide answers to specific questions, such as what works, what works best, how one variable relates to another, questions regarding a certain hypothesis. SLR aims to provide evidence that answers such questions. The presentation and of the literature review and its outcomes fol-	Traditional reviews typically address a topic (a phenomenon or research problem), aiming to develop comprehensive understanding and critical assessment of knowledge relevant for the topic. Reviews involve a dialogical interaction among the researcher and the literature (Wright Mills, 1978 [1959]) and can cross disciplinary boundaries. Depending on their purpose literature reviews

	low a specific schema that is similar across different SLRs (e.g., Williams et al., 2009; Gräning et al., 2011; Amrollahi et al., 2013).	can be very different in their structure and presentation: it may build upon an existing classification framework (Schultze and Leidner, 2002), focus on an analysis of researched variables (Lacity et al., 2011), categorize and summarize streams of earlier research (Leonardi and Barley, 2010) and so on.
Literature selection-relevance criteria	Relevance criteria are defined in advance: studies are included/excluded based on the validity of their findings (external and internal validity), which is ideally replicable by others. It is assumed that the validity of a study and its findings can be assessed abstractly based on research design (and method used) only, using 'the credibility hierarchy of research designs' (e.g., randomized controlled trials being most credible) (Morrell, 2008).	Relevance criteria are gradually developed as the review progresses and researcher's understanding advances. It is assumed that inclusion/exclusion of studies in the final literature review is based on researcher's insight and contextually sensitive judgment of studies' relevance and contribution. The validity of a study is assessed based on both explicit and tacit knowledge, standpoint and worldview, and merits of research design and execution (Hart, 1998).
Role of researchers	SLRs presume and encourage minimal reliance on researcher's interpretation, judgement and discretion in order to eliminate subjectivity and bias. Given the procedural nature of the review process the role and influence of the researcher in the resulting literature review is minimized (see e.g., Staples and Niazi, 2007; Jalali and Wohlin, 2012).	It is assumed that a researcher develops better understanding of a phenomenon by extensive and in-depth reading, interpreting, and comparing, contrasting, classifying and questioning publications. A researcher uses judgement and discretion to classify, critically assess and determine importance of individual studies and their contributions to knowledge (see e.g., Schultze and Leidner, 2002; Leonardi and Barley, 2010; Lacity et al., 2011).
Literature review content	SLRs regulate what counts as evidence and prescribe a particular knowledge extraction and summation process: the results from selected studies are combined and aggregated assuming they address exactly the same phenomenon and the aggregation of results is meaningful. Aggregated evidence aims to provide conclusive assessment regarding the research question.	A literature review is expected to provide an overview, classification, comparison and mapping of previous research results, as a basis for critical assessment of existing knowledge (potentially revealing weaknesses and poor understanding of a phenomenon or problematizing approaches, theories or claims). Critical assessment of literature is essential for developing an argument that certain areas or topics warrant further research.

Table 1. Comparison of SLRs to Traditional LR (source: Boell & Cecez-Kecmanovic, 2015)

The aim of this study matures and develops along the process of understanding and increased knowledge gain. Based on previous studies analysis, the topic of PM's applicability in the user experience studies domain was managed to be classified and summarized into two categories: PM as part of UX design and PM as a UX evaluation tool, sections 5.1 and 5.2. This also resonates more with the aim and purpose of conducting traditional literature reviews than SLRs (Table 1).

Inclusion of studies and their analysis relies heavily on the researcher's insight and judgment rather than predefined inclusion/exclusion criteria as is defined in SLRs. Boel & Cecez-Kecmanovic (2015) claim that the notion of using predefined methods and tools contradict to the original idea of SLRs in

medicine. In their early design, SLR's goal was to avoid dominance of studies from major journals and improve the diversity and richness of studies (Boel & Cecez-Kecmanovic, 2015). The authors assert that the method misses out on studies through 'indeterminacy of language' in their predefined search term usage, and via their inclusion/exclusion criteria based on titles and abstracts during initial screening. Whereas in traditional reviews the authors state that inclusion/exclusion of studies in the final review is based on researcher's insight and contextual judgment of their relevance and contribution.

In SLRs, results from various studies are collected, combined and aggregated to provide a conclusive assessment regarding a research question. This method applies for specific kind of literature reviews in studies regarding a well specified question, with conditions where a summation of evidences can be implemented (Boell & Cecez-Kecmanovic, 2015). For this study, the intention is not to answer specific questions but rather to give an understanding on a phenomenon of psychophysiological methods being implemented in user experience studies. The area is multidisciplinary and in its early stages of implementation. Hence, the idea is that this work will be valuable in supplying the perspective of PMs in UX research, what we mean by them, how they are applied, what are their limitations and advantages, what is being done now, and what the future implications are. Based on our goal, the corresponding approaches needed to give an understanding of our topic, and the comparative analysis described in this section; a narrative literature review was selected as the most suitable approach.

3.1 Literature review process

The study included three general, fundamental process stages; 1. Topic relevance and context establishment, 2. Continuous inclusion of new and relevant studies, and 3. Synthesizing the reviewed studies.

1. The first step includes an early review of papers to establish the context of study and make sure the focus is relevant. This step was conducted through the original background study, while the goal was to conduct an experiment using EEG and EMG tools. The background study for the experiment was used as the basis for the review structure, conceptual design and selection of research focus and research question. It is also important to note that the background study was not only focused on the two tools, EMG and EEG, but briefly reviewed others such as skin conductance level (SCL), electrodermal activity (EDA) and heart rate monitoring (HR). This step included a thorough reading across the disciplines of emotions, psychophysiology, psychophysiological tools, and usability, HCI and user experience studies within the context of our topic.

2. During the process of the study, it was important to continuously check if there are new papers relevant to our study. Even though the time frame for this study is not that long, as with the case of a PhD study, it was still important to check if there are new concepts and perspectives out there. In addition, the study is a relatively new topic and focus on current studies is mandatory and quite natural. This approach has helped in finding newer and sometimes more relevant studies such as the review on using physiology in quality of experience (Arndt et al., 2016) and a short review and primer on EMG in HCI application (Ravaja, Cowley, Torniainen, 2016) to mention few.

3. Relate the findings of our study to others and specify their implications in theory, practice and research. The study analyses previous studies and tries to point out the common links between them in defining limitations, advantages, potentials and future directions of PM tools in user experience studies to answer the research questions.

3.2 Finding research material

In this paper, different methods were used to find relevant research material. The two main methods of search were *electronic searching* and *references of references*. In addition, other research papers that I might have missed were also added through my supervisor. In literature review publication papers, one advantage of having colleagues review your paper is to identify important papers a work might have missed (Watson & Webster, 2002). Electronic searching is quite naturally the most common approach at this day and age, since most recent publications are found in digital formats across various databases. Accordingly, the first approach was to search for relevant studies using various key strings, which evolves through the process of study.

Through the electronic search various studies were found. Selection of papers is made based on topic, with more emphasis given to those more specific to psychophysiological tools with respect to HCI and/or user experience studies. Along the process, we found that most usage and implementation of PM tools seems to be on media research studies and the gaming world (Ravaja, 2004; Kivikangs 2015). One reason could be that the range of emotions in gaming is quite wide, with the role of emotions in games being important for the game success and player engagement. In addition, the usage and application of such tools may be easier in a game context given the abundance of stimuli and psychophysiological responses. Such game studies provide a strong knowledge basis and understanding of PM tools and their applications both as interfaces and/or usability and UX evaluation tools. Accordingly, most references for this study are related to applications of psychophysiological research in games, media, and other HCI application areas. Moreover, studies of applications of NeuroIS in UX evaluations of different information systems are also reviewed. Although there were many studies on psychophysiological tools in gaming research, as the review evolves, interestingly there were also few papers found that talk about measuring constructs such as learnability using EEG tools (Stickel, Fink & Holzinger, 2007).

As mentioned in the previous paragraph, the other search method used in this study is *references of references*. Since it is very important to stick to the focus of study and not always easy to find relevant articles, it was vital to use the method of checking referenced materials. Sometimes the referenced materials were used because they were mentioned as vital, or contain information that needs mentioning in our study as well. In few cases, the articles in the study themselves recommend other papers for details of descriptions or definitions. I also mentioned few of the papers found through references of papers under review, in the next subsection.

All papers were reviewed through reading the abstract, introduction, a few further paragraphs and conclusion sections before being fully reviewed and included. The study is multidisciplinary including psychology, physiology, user experience studies, and HCI. As a result, it was very vital to narrow down focus on the main goal. In fact, the whole nature of the human computer interaction studies is rooted in multiple disciplines, i.e. in the cognitive science of psychology and human factors, in the applied science of engineering and in computer science (Mandryk et al., 2006).

3.3 Research paper categories

Watson & Webster (2002) state that a complete literature review is not limited to one research methodology, one set of journals and or geographic region. Accordingly, this study covers various forms of previous works spread across the globe ranging from Asia and Europe to America.

Review of this thesis is based on three main categories of literature:

1) Previous reviews and overviews of psychophysiological research (e.g., in media research and in IS). The book *Psychophysiological Measurement and Meaning: Cognitive and Emotional Processing of Media* (Potter and Bolls, 2012) was used as a basic conceptual framework for setting the background of this study. This book lays a foundation in the understanding of emotions, the concept of psychophysiology, definitions and descriptions of psychophysiological measurement use in media research. Valuable references mentioned in the book are also searched, reviewed and included in the study. Ravaja (2004) and Caccioppo, Tassinary, & Berntson, (2007), for example, are results of references of references from Potter and Bolls (2012) book.

The *Handbook of Psychophysiology* (Caccioppo, Tassinary, & Berntson, 2007) is also another book that is used in defining basic concepts and view in the study of emotion. The book includes various studies used in this review, which were vital in defining key concepts and definitions.

2) Books, articles and websites clarifying the basic theories and methodologies.

3) Research articles relevant to our review: various studies were covered in this literature review. The primary studies are conducted through various research methods. The main focus is that the topic is relevant, method is validated, and results and findings are justified. The goal is not to cover every study conducted in the topic, but to critically study the most relevant ones (to our topic) and give an insight to readers about what is being done in the area of psychophysiological measurement tools as methods of user experience study by highlighting advantages, limitations, use contexts, application areas and future directions.

The findings of the literature review are describes in the next chapters as follows; chapter 4 discusses advantages and limitations of psychophysiological measurement tools in user experience studies, chapter 5 reviews the context of use, while chapter 6 and 7 discuss application areas and future directions respectively. Discussion is made in chapter 8 and main conclusions are described in chapter 9.

4. Advantages and Limitations of Psychophysiological Measurements in UX

It is important to understand that psychophysiological measurements (PM) have both advantages and limitations.

4.1 Advantages

Psychophysiological methods come into the user experience scenery with a strong basis of added value. Most of the advantages of PMs originate from their capabilities to overcome the limitations of traditional user experience evaluation methods. A deeper understanding of the human emotion and their true experience cannot be achieved with the sole use of self-reports and/or post usage questionnaire evaluations. Researchers claim that there are a lot more emotions that occur without users' conscious knowledge or awareness. They argue that the true implications of media usage on the user's deeper emotion or psychology, can only be achieved through the usage of psychophysiological measures and tools. (Potters & Bolls, 2012.)

Ganglebauer et al. (2009) also support the usage of psychophysiological tools through identifying the weaknesses and limitations of traditional methods and proposing the application of PM tools to improve user experience evaluations. The authors claim that asking participants of their emotions and experiences, interrupts the process of experience and their opinion is prone or susceptible to social biases. They claim that the PM tools will help unmask social masks, help end-products become more than functional and target them to address aspects of experiences such as aesthetics, beauty, playability that will enrich user experience with interactive systems. The availability of different PM tools and their potentials in measuring the human emotion, more specifically on the scale of positive/negative valence and arousal levels is explored by the authors.

Park (2009) claim that psychophysiology study tools are quantitative and observational. With the right analysis, combination of use, and proper use context could give a more honest and real measure of emotions than traditional methods. The perspective in this study is in the notion that traditional methods are prone to questionnaire and social bias. In addition, the responses rely heavily on the user's memory which could be misleading in interpretation. The perception of PM tools as very vital means to measure user emotions is supported by various studies (e.g., Laufer & Nemeth, 2007; Constanza, Inverso & Allen 2005; Potters & Bolls, 2012; Kivikangas, Channel, Ekman & Salminen 2011). Most of the studies note that psychophysiological measurements tools are not contaminated by user's answering styles, social masking, questionnaire wording limitations, participant's memory and/or conscious knowledge of experience and observer bias (Potters and Bolls, 2012; Kivikangas et al., 2011, Constanza, Inversio, & Allen 2005). Most of the studies also note that the benefits of measuring emotions automatically, continuously and through the process of experience are key advantages.

The most common psychophysiological tools in UX research are electroencephalography (EEG), electro dermal activities (EDA), blood volume pressure (BVP), heart rate (HR) and electromyography (EMG). Ganglbauer et al. (2009) sum up the role of each tool as; EEG is very interesting, in principle, for its capability of measuring brain activity directly, EDA is the most valid method to measure arousal in interactive systems, BVP as a valid measure of arousal, heart rate (HR) as a metric for valence measurement, and EMG as the most appropriate and precise tool to measure muscle activities, or representations of emotions that are not noticeable with video accurate or other observation mechanisms. The study claims that facial EMG is a reliable and viable tool for measuring the positive and negative states of emotion. The valence perspective of heart rate variation measurements was studied

by Anttonen & Surakka (2005). In their study, they showed that HR decelerated in response to emotional stimulation and it decelerated the most in response to negative stimuli as compared with responses to positive and neutral stimuli. In addition, in the case of a positive experience decelerating only during the first two seconds and after that the heart rate began to return towards the baseline.

Facial EMG as a viable measure of positive and negative valence compared to visual facial recording mechanisms is also noted by Partala, Surakka and Vanhala (2005). The study explains that EMG is free from problems of illumination, view point variations and visual recognition errors compared to methods such as video recordings. Their study mainly focuses on estimation of emotion from facial expressions with a focus on facial EMG as a measurement tool. The implementation of the tool along with other cases, will be discussed in latter stages of this study. It is also important to note that the study conducted by Ganglbauer et al. (2009) claims that PM tools, in their state at the time of study, should be used together with subjective measures to be able to generate a more accurate and valid result.

In addition to the benefits that originate from the potentials of PM tools and the weaknesses of traditional methods, some studies also note the trends of technology itself and the timely demands it brings as one initiative for the focus on improved and reliable means of user experience measurement (Constanza, Inversio, & Allen, 2005). The authors explain that the advancement with mobile devices is leaving behind the times of old input/output interfaces such as mouse and keyboard that mimic the desktop era. They mention smart glasses as perfect example of a case where an alternate means of interaction than physical manipulation of mobile phones could be needed. Their study is of those that sees psychophysiological measurement tools as potential interfaces for future applications and devices. The experiment studies an EMG based input device for subtle and intimate interaction in a mobile context. The system was also expected to be an interface that can cause minimal social disturbance. The corresponding application was an EMG wearable device for mobile application around the arm, which can be used as an input interface through commands via subtle muscle activities. The study is on an experimental basis with the intention to validate the possibility of EMG-based input, through subtle muscle movement. Issues of its application in the real and complex social context, will demand further improvements and applications. More details of such applications will be discussed in chapter 5 regarding the use of the PM tools in the design and development of new interfaces; here it is important to note that PM tools are also seen advantageous and timely, not only as valuation techniques, but also as parts/components of future applications and products.

It is important to understand that most of the studies do not solely focus on advantages gained by PM tools. They describe various limitations and weaknesses of PM tools, most of which are common across authors, while others are mentioned in few cases. Due to the intention of this section, we mainly covered the advantages of PM tools. However, the next section intends to give a picture of the other side of the coin.

4.2 Limitations

Psychophysiological measurement tools do not only come with their advantages and their application is merely a walk in the park. They come with different types of limitations in the context of price, complexity, device format, and time cost when compared to traditional usability evaluation methods (Kivikangas, Chanel, Cowley & Salminen, 2011). Most of limitations are commonly reported across authors, while others are mentioned in few cases.

Most authors note the price and complexity limitation of PM tools, with few variations in the interpretation and magnitude. Most researchers claim that most PM tools are quite expensive and their prices have limited the advancement in the studies and implementations of the technology (Potter & Bolls, 2012; Ganglbauer et al., 2009; Pentus et al., 2014, Kivikangas et al., 2011; Ravaja, 2004). Most

of the high-end products such as fMRI and EEG are quite expensive to be used in user experience studies. Park (2009) claim that the limitations of psychophysiological measures as tools of HCI are the reasons for its relative scarcity. They claim that shortage of sufficient knowledge regarding the relations of psychology and physiology, expensive equipment and data being dependent on interpretation as key factors of limitations in the HCI context. The later concept is the fact that the measurements are context dependent and the resulting data could be misleading if researchers take the measures as fully objective. The authors argue that most psychology studies are carried out in simple stimulus context which could be misleading if directly implemented in HCI where there could be numerous causes of stimulus in a computer environment. Alternative explanations and context awareness are necessary to make sure we get accurate results. Like other studies, they also conclude that no tools are fully perfect and it is best if used in combination with other research methods.

Price limitation is gradually being addressed to a certain extent. For example, smaller EEG tools such as Emotiv Epoc and Neurosky are the common mentions as affordable tools and commonly being used for user experience studies (Arndt, Brunnström, Cheng, Engelke, Möller & Antons; 2016; Pentus et al., 2014). Epoc Emotiv EEG tool is the one purchased by the University of Oulu and an initiator of this study as well. We need to bear in mind that even though these devices are cheaper, they come at a cost of a less reliable data precision and noisy signals than those devices used in clinical applications (Arndt et al., 2016). Konstantinidis et al. (2015) calls Epoc Emotiv and Neurosky Mindwave headsets as inexpensive, off-the-shelf devices that pave the way for pervasive EEG signalling in smart home environments. They studied the potential of using the headsets as potential devices for addressing the challenges of Internet of Things (IOT) and trends of real time EEG signalling availability. They propose that the integration of such tools of connectivity and EEG signalling could assist enriching the Ambient Intelligence approach. Details of the implementation and the context of application will be reviewed in chapter 5. The mention here is to validate the efforts and improvements made to address the price limitation issues.

Another limitation of PM tools is their complexity and the resulting need for expertise. The complexity level varies across tools and their applications, and the limited number of studies conducted is also a reflection of its complexity (Potters and Bolls, 2012). EEG, for instance is considered to have a huge potential as a direct measure of brain activity with an excellent temporal resolution. However, in addition to the brain activity, there are activities read, that are originated by other causes than the stimulus, such as signal variations due to a blink of an eye (Potters and Bolls, 2012; Arndt et al., 2016). Accordingly, various techniques of analysis such as Event Related Potentials (ERPs) and Independent Component Analysis (ICA) are needed to distinguish events that occur due to a stimulus. ERPs need to be generated via smoothing out the unwanted signals by averaging several trials of a single recording (Arndt et al., 2016). In addition, Arndt et al. (2016) also claim that a highly-controlled experiment set up is needed since EEG generates a lot of noise. This limitation also raises a question in usage of the tool in combination of other tools that are susceptible to noise. Complexity issues in signal interpretation also exist in EMG. One key limitation could be in terms of distinguishing between a real smile/frown and a fake one. There are studies that suggest that people fake their responses to the contrary of their true feelings when facing people of higher order, which in turn leads to the wrong signal being gathered (Ravaja et al., 2015). The authors claim that facial EMG may be affected by display rules (learned rules dictating how people manage their emotional expressions based on social context). Examples of such cases and intriguing results will be discussed further in follow up sections. To this level, it is sufficient to know that such limitations do exist and are noted.

The complexity of PM tools is not only on the signal analysis and interpretation but also on the implementations themselves. Such cases affect the usage in two forms: one is in needing careful implementation of experiments adding some level of procedures and caution, and the second is in the device format where users are told to wear electrodes in somewhat uncomfortable scenario specially when thinking about everyday use (Constanza, Inverso & Allen, 2005; Kivikangas, Chanel, Cowley &

Salminen, 2011; Potters & Bolls, 2012; Ravaja, Cowley, & Torniainen, 2016). For example, steps such as skin abrasion, gelling electrodes, proper placement of electrodes need caution in implementation. Almost all PM tools have their own awkwardness in their placements. Facial EMG places electrodes on the face, ECG puts them on the chest (mostly used in hospital settings, but the best location for accuracy), arms and legs (lesser recognition), skin conductance places them on the palms (makes usage of hands on keyboard inconvenient) and soles (awkward location, and lesser accuracy), and eye tracking limits head movement.

Given the complexity of the PM tools, another limitation is the time consumption due to the preparation and analysis time taken to ensure the experiment is conducted properly, as well as the data analysis process and the corresponding steps it needs are followed correctly and carefully. Compared to other methods where users are asked to fill out questionnaire, and/or they are observed/video recorded while using/interacting with a system, time consumption could be considered a limitation. However, it is important to note that the limitations do not diminish the usage of the tools. On the contrary, provided the knowledge for the limitations, all types of data gathered will have their own value. For example, in a behavioural study to understand how people behave under social circumstances, we may actually seek for the false smiles or frown a person might display.

In a nutshell, the limitations of PM tools could be listed as: complexity in terms of implementation and interpretation, cost, susceptibility to interference, personnel expertise, device maintenance and time consumption. Some issues could be similar across all devices, while others are specific to certain tools and measures.

5. Psychophysiological Measurements Used in UX Research

As mentioned earlier in this study, PM tools are used in the user experience research context in two major ways. One way is that PMs are a part of the product design; PMs being components themselves. The other way is using PMs as an evaluation method of user experience. This chapter comprises of two independent sections exploring the context of PM tools' usage in the user experience domain.

5.1 Psychophysiological measurements in UX design

When we think about PM tools as components of a product design, we should bear in mind that they are, as well, intended to indirectly improve user experience. Be it as an I/O interface for people with physically impairments, as part of a biofeedback/affective feedback systems in game play, prediction tools to know how a user will react to certain experiences, or avatars mediating their emotions to elicit a certain type of emotion, they are improving the user experiences and facilitating ease of use in various contexts. Various studies that make use of PM tools as an interaction component have been conducted. Constanza, Inverso and Allen (2005) studied the potentials of EMG as an input interface. Laufer & Németh (2007) studied skin conductance response (SCR) to measure user stress level and use it in games to predict user actions, Naik, Kumar, Singh, & Palaniswami (2006) studies the potentials of hand gesture recognition as an input mechanism using EMG, Culpepper & Keller (2003) studies the potentials of EEG as an interaction tool.

Some of the papers aim to combine the concepts of psychophysiology measures as indicators of human emotion and the idea of the tools being used as design components together. By using this approach, games, software, and other products become more pleasant or exciting. This is made possible on the basis biofeedback measures provided by the PM tools during usage. Laufer & Németh (2007) attempted to anticipate stress levels in game playing using skin conductance level (SCL), heart rate variations (HRV), and heart rate intervals. In their study, the data were used to teach artificial neural networks to learn when the user will take an action. Their findings indicate that the artificial neural networks could predict a user action 2 seconds before they occur. Skin conductance level has a long latency that could last from 5-10 seconds and the game under research had to be compliant to the limitation so that users do not get frustrated of delays. The study is limited to a simple game Jungle Swing, and users carried out single actions of jumping. A more complex game will come with more difficulties in distinguishing between multiple actions and the corresponding predictions. It is, however safe to say that the indications and possibilities of this study give an indication of the possibilities PM supplies us with in product design.

Prediction of user behavior and understanding of usage patterns and anticipation of variations are studied in other sectors such as mental health. For example, ambulatory monitoring of brain activity for epileptic patients while they carry out their normal day to day tasks with an EEG headset is experimented with in a prototype level (Gyselinckx et al., 2010). The application helps to study abnormal brain activities that may help predict epilepsy patterns for ambulatory services.

Partala, Surakka and Vanhala (2005) also studied the possibilities of emotional experience estimation using facial EMG. They go as far as claiming that methods based on facial expressions can help teach computers become well versed with emotional intelligence and bodily communication that human beings are well versed with in communicating with each other. In the study, they state facial EMG to be efficient in distinguishing positive and negative valence of emotion by 70% for pictures and 80% for videos. The main aim of the research was on the improvement of user experiences in computer

systems. Understanding of users' emotions in real time, using EMG in this case, potentially will enable a computer to make usability decisions and try and fix the problems that seem to elicit negative emotions. The experiment was carried out with 10 test subjects, 5 male and 5 females. The participants observed 52 images that were selected from the International Affective Picture Systems and 6 Videos that were prepared by two actors, acting a negative, neutral and strong positive expressions.

Before conducting the experiment using EMG, the participants rated the stimuli from the scale of 0 to 9; 0 being very negative, 5 neutral and 9 very positive. The subjects then saw the 52 picture stimuli, followed by viewing the videos with an EMG tool monitoring them. EMG readings were analyzed using three estimation models. One based on the zygomatic major activity where, if the score is above zero indicating an estimated positive, and if it is below zero indicating a negative emotion. Second model was based on the corrugator supercilii activity, where above zero indicates a negative feeling and below zero a positive feeling. And a third model based on baseline corrected score of subtracting the corrugator supercilia activity value from the zygomatic major activity values. If the score is above zero, then the experience is positive, and below zero as a negative emotion. In the end, the results of the EMG analysis are compared to the ratings the participants gave for the stimuli. The corresponding result showed a 60% to 80% accuracy between the three modes. The mixed mode, where the corrugator activity is subtracted from the zygomatic activity, had the highest accuracy rate of all the modes. (Partala et al., 2005.)

Overall, the study by Partala, Surakka & Vanhala (2005) was valuable in a sense that it gave a good validation and indication of EMG as a measure for user experience evaluation, through direct comparison with the traditional methods. However, one intriguing question could be that, if we keep comparing accuracy based on traditional methods, how can we claim that the PM tools are more effective or supply more information with regards to user experience? Bearing in mind that the study is aimed at using EMG as a component of a user interface, supplying the computer with information that helps it understand the human emotion, the comparison to the user's ratings maybe vindicated. Meaning that the focus is not in the comparison of the methods but on the possibility of enabling or increasing interactivity through physiology. However, even in such a case, if we are aiming for emotionally intelligent systems, we need to look for implementing experiments that are not only highly accurate when compared to a person's perceived state of emotion, but also the emotions that exist subconsciously. In addition, the study only proves EMG as emotion detecting tools with a potential for being used as a feedback system for computers but does not deliver a sample application or prototype. They note the possibilities and potentials based on what knowledge could be gathered using the tool. The contribution of the study is in its showcasing the possibility of EMG tools detecting emotions and their potential as an interaction mechanism.

Psychophysiological measurement tools were also researched as input/output interfaces with the purpose of creating a true hands free experience. This is facilitated by the rapidly growing devices and their increased connectivity. Constanza, Inverso and Allen (2005) proposed an EMG based wearable interface for mobile devices. The study effectively showcases an application that reliably recognizes subtle gestured without calibration and intensive training across users. The experiment tried to simulate a real-world environment where the subjects could walk around with their arms relaxed. Four main tasks, excluding three familiarizations and one walking with no contraction steps, were conducted. The four main components were intended to represent corresponding distinctions of contraction levels: "generic", "short", "long" and "mixed". The familiarization phases had a 15-minute time frame and users were instructed to conduct contractions at the sound of an audio through their headphones. Workload tests were conducted after each walk to ensure that the tasks do not generate a high demand that could add to a mobile environment where users do not have enough attention to spend on interfaces and interactions, while paying attention to their surroundings. The intention was to allocate the corresponding levels of contraction to initiate different recognition algorithms. There were relatively some difficulties in detecting long and short contractions, the algorithms only detected a

small range of contractions. Some participants were also frustrated at failures to produce contractions. Contrary to the findings some participants claimed they felt long contractions were more difficult than short ones. The project overall showed that algorithms could be used to identify subtle gestures, which in turn could be used as commands to initiate an activity in a mobile device. The study, however, does not include a complete system that integrates and delivers a single application scenario where EMG is used as an input interface.

Like Constanza et al. (2005), other researchers also studied the possibilities of surface EMG as an interaction facilitator between human and a computer. Naik et al. (2006) studied the use of surface EMG to study the human hand gesture as a means of interaction with a computer. Their motivation is initiated through identifying hand gesture recognition as a promising potential as a means of interaction with computers. The authors chose surface EMG because it is easy to record and non-invasive. Study of prior knowledge in anatomy was used, as a guidance, for a proper placement of electrodes. Muscle groups were divided into four subgroups corresponding to wrist, muscle, and finger flexions. Four channels of EMG recorded during hand actions that required not greater than four independent muscles. Independent component analysis (ICA) was used to distinguish between muscle activities. ICA is suitable when the numbers of recordings are same as or greater than the number of sources. The aim of the study was to test the use of ICA for separation of EMG signals and distinguishing between different hand gestures and actions.

The experiments were repeated on two different days (Naik et al., 2006). Three hand actions were performed and repeated 12 times at each instance. Signal sampling rate of 1024 samples/second and each hand movement lasting a duration of 2.5 seconds each. The study claims to have been able to effectively detect the studied hand gestures with an accuracy of 100%. However, it is important to keep in mind that, the sample size was quite small. In addition, this study only verified that the method works in a minimal level of activity. The authors also acknowledged the limitations and suggested further study in larger sample size settings with different action scenarios. Overall, the study was a good starting point in the progress towards more intelligent interactive systems.

Neuroberry, a platform for pervasive EEG signaling in the IOT domain was also proposed by Konstantinidis et al. (2015). The study introduces an IOT concept that makes use of EEG devices (Emotiv Epoc and Neurosky Mindwave), a low-cost Raspberry Pi 2 and the CAC-framework for making data available, in JSON format, on the internet for use on different contexts. The authors claim that the architecture they propose of EEG signals being connected to the Internet domain through their designed architecture, is the first of its kind. Prior studies only attempted to facilitate connection via mobile phones. The research conducted data acquisition from the headsets and streaming of the data to the CAC-framework via the Raspberry Pi. Real time data transmission was ensured through dropping overlapping data packages. The study intends to improve the knowledge base regarding the smart home facilities for healthy and active aging paradigms.

A study by Styliadis, Konstantinidis, Billis, and Bamidis (2014) focuses on the use of EEG in a gaming environment for elders to assess the features that encourage the elderly individuals to interact with the gaming environment daily. Knowledge of motivation factors that initiate elders to engage in a gaming environment daily is then used to generate decision support systems (Styliadis et al., 2014). The gaming environment could be improved per the user's affective state based on such biofeedback. In this architecture, the EEG devices are intended to be used as biosensors together with video cameras. Accordingly, physical and mental data are recorded through the process of interaction. When talking about gaming environment, it is not to mean games that are only fun to play. However, these experiments are conducted on games that are categorised to be serious games that elicit cognitive and emotional activities, which in turn create a pleasant experience for the elders. Identifying and addressing senior users' needs is one of the challenges in smart home systems. Addressing the cognitive and emotional usability issues in elderly users is noted to be a key point in validating smart home

systems as effective (Styliadis et al., 2014). The authors also claim that the most critical factor in game designs for elders is that the gaming environment and scenarios should encourage the elders to engage in and use the system. They note that this is a key element of affective computing where computers systems are intended to explore both the physical and mental states of a user.

Monitoring the brain activities within the game environment gives an understanding of what satisfies users, such as elders (e.g., Styliadis et al., 2014). This knowledge in turn will assist in developing a decision support system and improve it into becoming satisfying and engaging, developing the health states of elders. Even though this study focuses on smart homes and elders in a gaming environment, the proposed architectures and designs are potentially applicable to other user experience improvement cases as well. The ability to read users' brain activities in real time, using/sending such data in an accessible and interpretive format to the internet, and finally generating or developing improvements based on feedback are essential elements of usability improvement.

The experiment architecture includes sensors that monitor the physical activities (cameras, Kinect), mental activities (EEG headsets, Mindwave or Emotiv Epoc), intelligent monitoring systems for data gathering, and data analysis to produce near real-time affective states that can possibly send biofeedback information for the gaming environment to manipulate and adjust the gaming experience accordingly (Styliadis et al., 2014). In a nutshell in the proposed architecture, the users physical and mental activities and emotional states are recorded, the data is then transferred to an intelligent system for analysis and from this analysis a quantified information of the user behaviour can be fed back to the gaming environment where adjustments to the game will be made based of the user behaviour. The experiment showcases a case where an EEG recording can be used to gather emotion information from a user in real time and the data can be used to target and elicit a user emotion in a game experience. The study does not include a complete application, but the information presented is sufficient as a base for developing applications that may fulfil the claim. The possibility of collecting the emotion data, and transferring it in common platform such as the CAC-framework in a readable format (JSON), supplies coders with the possibility to manipulate the data and develop applications that make use of the extracted information. In the context of the research, the main goal is improving the well-being of the elderly and the games are serious games intended to initiate mental and physical activities for the users.

5.2 Psychophysiological measurements in UX evaluations

There are a good number of studies that either research on the potentials of PM tools for user experience evaluation and few more that carry out actual evaluations using the technology. Both approaches intend to make the potentials of PM tools in the context of user experience studies a reality. In this section, the focus is on the usage of PM tools as a measurement method of user experience evaluations.

Park (2009) studies psychophysiology as a tool for HCI research in a detailed manner going through different tools, their respective advantages/disadvantages, success rate and their application areas. I recommend this paper for those who would like to get a brief insight of what we mean by psychophysiology measurement tools and their application in HCI. It is an easy read with valuable information as the basis for selection. The study reviews skin conductance, electromyography, heart rate, and eye tracking as main areas of study. There are mentions of other tools such as EEG and fMRI, but they are not studied in detail. The study identifies the limitations of these tools, as described in section 4.2, but in a tool-specific manner. Eye tracking, for example, limits head movement putting the user in an unnatural tension, facial EMG, SCL, and HR may cause awkward feelings through the placement of electrodes and the controlled experiment set up, and the data gathered is susceptible to misinterpretation since the causes of emotion variation are sometimes hard to distinguish. Hence the authors claim that researchers should keep in mind these limitations and prepare well to minimize

them. They also stress that it is good to use the tools in combination which will make help come up with a more valid result and insight. The limitations are there to give readers about the cautions they should make when using PM tools. However, the research points out that it is important and valid to include PM tools as usability measurement tools which will help assist in producing better products that generate an improved user experience. The main point is in selecting the right tools and/or combination tools, for the right product or study one may be interested in.

Concerns regarding the interferences caused by unrelated (to an experiment) activities, such as movement, during psychophysiological measurements, was also noted by Ganglbauer et al. (2009). They claim that PM tools are not ideal for studies that include having people in motion. However, there was a contrary notion by Constanza et al., (2005) who claim that EMG can be used in creating a subtle user interface in a mobile condition. The latter study was in the context of EMG's usage as an interface. The fact that the focus was in the tool's capability to detect a muscle contraction may not have stressed on the strength of stimuli to emotion unique relations. The muscle movements, in this case, were intentional which might make distinguishing the signal, as what it was, a little bit easier. Of course, the studies also vary in the sense that the latter one was focused on an EMG tool on the upper arm, while the mentions of EMG in the first paper was facial-EMG. However, it is safe to say that the perception of interference and unwanted signal activation exist and it is good to be wary of it.

There are various research works that aim to measure user experience and usability constructs, using PM tools. The degree of success and validation varies across studies, but there is sufficient work to suggest the possibilities and potentials of psychophysiology in the UX evaluation context. Studies of distance perception using EMG (Skroumpelou et al., 2015), product package decision making in supermarkets using emotion evaluations (Pentus et al., 2014), EEG-based learnability assessment (Stickel et al., 2007) research and showcase the possibilities and limitations of psychophysiology measurements for UX studies. Based on the evaluation constructs, the reviewed papers were divided into two categories; 1) those that focus on measuring attributes of emotion and 2) those that focus on measuring usability constructs of a product or system. It is important to note that both perceptions have intersecting points and their distinction is not deep rooted.

5.2.1 Evaluating emotional states

Application areas of PM in the context of user experience studies is also applied in urban environment studies for example (Skroumpelou et al., 2015). The study evaluates how users perceive a space and how their perception affects their walking behavior. The idea is to study the relation be-tween emotional states (affective states) evaluated by a tool and its relation to the user's perception of the distance traveled. The study was conducted using a mobile EEG tool, mainly Emotiv Epoc together with traditional post-task interviews and questionnaires. The participants walk wearing the headsets and carrying a GPS recording device. The authors were interested to know how various events on the road can affect the overall perception of the distance traveled by the participants. An Event Related approach (ERP) was used to study the brain activities with respect to annotated events that happened during the walk. Results of the study are intended to supply urban designers with a better knowledge of the human experience in walking so that they keep these perceptions in mind when designing urban environments. The results showed that roads that generated stressful and tiring experience to be the perceived as the longest. There are also relations found between what is tiring and stressful, and what could be the possible problem on the road, such as lack of light and shops. Although the study is not complete and the authors suggested to continue their work to make more solid findings, it is sufficient to see the possibilities of PM tools, EEG in this case, as evaluates of the human emotion and perception in various contexts. The finding can be used on its own, or as a starting point for other studies.

Branco et al. (2005) have also studied user perception but in a software usage context in their case. This study, for example, could easily have been included in section 5.2.1 for its focus on usability of software. However, the constructs measured and the intriguing results regarding valence of emotion put more weight for it to be included in this section. The study evaluates human facial expressions that are caused due to event occurrences at the user experience level of software. Facial EMG was used to monitor user's facial muscle activities. Results of the measurements were used to draw a pattern between the facial expression and the task difficulty levels of the software under study. The researchers focused specifically on difficulty levels encountered on a task to task basis. Corrugator, frontalis, and zygomatic muscle activities were monitored with the focus on finding out which activities occur during difficulties. Out of the tree activities, the frontalis was dropped due to the some of the trials included interference. The study manages to identify a direct and significant relationship between the increases in activity of facial muscles with the task difficulty level.

One strange result detected on the difficulty-to-muscle activity relation by Branco et al. (2005) was a contradicting result for the zygomatic activity. While the muscle is responsible for initiating a smile, it appeared responsive for a negative feeling in the study. This result calls for caution when distinguishing whether a smile as a real positive smile or the contrary. The authors back their finding relating it to a similar relation found by van Reekum C. (2000, cited in Branco et al., 2005) in computer game settings. Kivikangas et al. (2011) also found a similar result in an experiment in the gaming sector; that included a racing and a first-person shooter game. They found a contrary positive zygomatic response for a person's own death and a negative response for the death of an opponent's character. However, in a later study, Ravaja, Cowely & Torniainen (2016) justified the finding claiming the origins of the contrary results could be results of emotion coping. People tend to smile even though the outcome of the game elicited a contrary feeling. As a result, based on this evidence and others, the latter authors suggest that facial EMG does not always index internal emotions themselves but measure the representations of emotion bearing facial expressions. In the gaming context or any other setting where the social masking is expected, caution should be taken to distinguish actual representations of emotion.

Overall, Ravaja, Cowely & Torniainen (2016) described the use of facial EMG in a review that studies advancements made in the use of the tool in HCI domain. In addition to the false positives in the zygomatic muscle activities during a play environment, the researchers also mention another situation where the corrugator muscle might also indicate a false positive. They mention a case where creative planning tasks elicited a lower corrugator muscle activity when compared to a routine task, contrary to what was expected. The expectation was that in planning tasks that need focus, the frowning muscle would increase indicating attention and focus. However, the creative element, in this case, seems to create a more pleasant experience indicated by a lower corrugator muscle activity. This indicates that understanding the context of an experiment is quite vital when planning to conduct experiments using facial EMG in the study of user experience research. Overall, the study notes that the context of use, social influences, game set up (in the context of play environments, as in with a computer vs. with a person), and other aspects should be taken into consideration when using facial EMG. When implemented with a clear understanding of the issues surrounding the tools, they potentially enable us with a valuable resource to be used as an evaluator and interface in HCI.

Study of EMG as a user experience evaluation tool in a play environment was also conducted by Mandryk et al. (2006). The study conducted an experiment in three conditions; subjects playing with a co-located friend, with a co-located stranger and with a computer. The authors define a two-dimensional diagram for the current method to identify the lack of quantitative objective evaluation techniques, which could be covered by physiological measurement tools. The aim of the study was to develop a user experience evaluation model in play/entertainment environments. The methods used are GSR, ECG (for heart rate HR related to emotional activity could be used for a measure of valence in combination with other indicators, and HRV related with mental effort), and facial-EMG. Twenty-

four male participants (12 pairs) between the ages of 18-27, frequent gamers and computer users on a daily basis were the subjects of the study. Six of the pairs were very experienced or somewhat experienced, three pairs were neutral in their experience, and the other three pair were somewhat inexperienced. The stranger was a 29-year-old gamer, fixed constant and instructed to match the participant's level of play as much as possible. Experienced players played at a higher difficulty level. In each pair, players were treated as though physiological data was being collected from both. Before each experimental condition, participants rested for 5 minutes while listening to a CD containing nature sounds, so that physiological measures return to baseline. After each condition, subjects rated the condition using a 1 to 5 Likert scale. Questions were in the form of "This condition was boring," and same goes for challenging, frustrating, exciting and fun. After completing the experiment, the users were asked to fill questionnaires. They were asked to answer in a retrospective about which condition was fun, most exciting, and most challenging.

Data were inspected and erroneous samples were corrected. All data signals were smoothed and then normalized. The normalized GSR, HR, EMG smiling and EMG frowning signals were used as inputs to a fuzzy logic model. The data was then modeled into two parts to generate user emotions. Arousal and valence (AV) values were computed and then the values were used to generate emotional representations of boredom, challenge, excitement, frustration, and fun. To generate the model, only half of the participants' data, one for each play condition order, was used. The other six was reserved for validating the model. GSR was used to model arousal and was modulated by the HR data. If the HR showed a contradictory level, then arousal was altered, otherwise, it was maintained. Valence was modeled using the EMG signals with EMG smiling as positive valence and EMG frowning for negative valence. Five emotions; boredom, challenge, excitement, frustration, and fun were modeled using the AV values. The AV to emotion model has two inputs of arousal and valence and five outputs. The five outputs are modeled in the AV space in four states, very low, low, medium and high. The subjects rated their emotions based on the five represented emotions as well. (Mandryk et al., 2006.)

The effectiveness of the model was analyzed using the data from six participants not used in generating the model. The model was applied to the data and the time series of each emotion were averaged to compare modeled emotion to the subjective response. Analysis revealed players were having more fun when playing against a friend than a stranger. And playing against a stranger was marginally more fun than playing against a computer. The result revealed similar trends in both the subjective and the modeled analysis for the emotions of boredom, excitement, and fun. However, for the case of 'challenge,' the correlation indicated an inverse relation. The assumption was that a player's arousal would increase when the challenge increases. On the contrary, some participant's comments revealed a strategy to attempt and relax when challenged so that they improve their performance. Hence it is important to study how participants handle challenge in a game. More work needs to be done before challenge can be modeled in such context. (Mandryk et al., 2006.)

Although there were some similar trends of emotion between conditions in both measures, the intensity somehow varied. This is significantly seen in the frustration level, which is a lot more intense in the subjective measures than the modeled measure. However, the authors note that this could be because frustration could be relative. That is the model assumed the maximum and minimum possible emotion levels. In a game environment, since the overall experience is fun and excitement, the relative frustration percentage may appear low. If you compare the frustration in a game to frustration when running out of gas on a highway, the insignificance may be understandable. Same goes to boredom as well. On the subjective side, the players could pick the highest frustration level. More so because the questions asked were general; as in for example, "This condition was frustrating." Had it been described in a more frustrating scenario, the response might have differed. Hence, one conclusion could be that the model shows the real intensity of frustration level, which is relatively low, while the subjective response could be high. (Mandryk et al., 2006.)

Modeled emotion from physiological data is very powerful and it can be observed continuously and can objectively provide a quantitative metric of user experience within a play condition. Modeled emotion pinpoints moments in time when a user's frustration was changing. This is particularly beneficial when there is no baseline or comparative condition. Researchers and developers can uncover individual moments when a user begins to get stressed, starts having fun, or becomes bored. This information could be used as an evaluative tool or could be used to dynamically adapt game settings (e.g., difficulty level) to keep players engaged, preventing them from becoming frustrated or bored. Continuous monitoring is vital in entertainment technologies since the measure of success is determined by the process of play, not the outcome of playing. (Mandryk et al., 2006.)

Mandryk et al., (2006) also suggest that PM evaluation methods can also be used in combination with other evaluation methods as well. For example, with video observation tools, where researchers can identify interesting features such as increase or decrease in an emotional state, and then investigate the corresponding time frames in a video recording instead of examining the entire video. This will drastically reduce the time required to qualitatively examine video of user interaction with entertainment technologies. The authors suggest future work should be done in representing other emotional states, and would also like to see if the model can be generalized in interactions with other play technologies, for example, to study user behavior in ubiquitous play. If a generalized model can be reached, then the modeled emotion can be used to dynamically adjust play environment to keep users engaged. This model can be adapted to analyze user's emotional responses in other interactive systems as well. This is a very valuable study in a sense that it supplies us with a very clear model of emotion representation and analysis using GSR, HR, and EMG measures. However, since the intensity of emotion is dependent on the perspective of what we interact with, it could be difficult to have a standard model across different contexts. Hence numerous modeling and studies based on application properties is mandatory. Adoption of the model into different contexts is possible. The modeled emotions also revealed differences between play conditions, in a condition where subjective reports failed to reach significance, validating the practical advantages of physiological tools.

EMG as a measure of emotion experience was also studied by Hazlett (2003). The focus of his study is to investigate the use of facial EMG to provide a continuous measure of the user's emotional state. Facial EMG was recorded while female users performed five tasks to one of two websites. Frustration Index scores were developed from the corrugator EMG data by calculating a percentage score of a pre-task baseline. The study also notes the selection of emotion as a special element area for CHI 2003 conference indicating its importance and relevance as a major tool for user experience studies. They also note that a successful HCI is reflected via a pleasant human emotional experience.

Hazlett (2003) conducted using key assertions from previous studies, as a basis of the experiment; 1) the user's level of frustration has been considered a pervasive problem in using information and computing technology occurs on a frequent basis, 2) the corrugator muscle activity which provide a sensitive index of the exerted mental effort in the task performance, 3) Facial EMG as being a validated, as it has been proven, measure of positive and negative emotions responses to media usage.

Even though corrugator supercilii and zygomatic major activities and eye tracking data were monitored, the researchers only reported the corrugator activity data with respect to user frustration levels during usage of a web site. Like the experiment conducted by Branco et al. (2005), participants rated the difficulty level of emotion they felt regarding the stimuli, but post-usage in this case. However, the data gathered is not only evaluated with respect to user's ratings, but also the expected value corresponding to each website's difficulty levels. Attributes such as the number of pages, steps it takes to complete a task, correctly completing tasks that are expected to raise frustration levels. The user's proficiency level was also taken into consideration in two categories, as in novice and professional users. The gathered data from corrugator activity were in line with the expectations, for instance, the activities were higher in incorrectly completed tasks than correctly completed ones, or

novice user's difficulty to navigate around were indicated with higher activities in the corrugator muscle areas.

As a result, Hazlett (2003) concluded that the exploratory study shows the validity and usefulness of facial EMG methods for measuring the ongoing emotional state of computer users. A dynamic understanding of human-computer interaction can be achieved through gathering moment to moment emotional experience of the user which can be temporally related to interface events, navigation logs, and other user behavior. Hence a better usable system could be developed by user experience evaluations conducted using PM tools such as EMG. The study also states that the zygomatic muscles could also be used in a similar manner but for detection of positive feelings. However, the experiment did not include reporting regarding this muscle activities. The study is a good start and contains important information that adds to the growing agreement that PM tools could be used to study user experience. Although their limitation to one muscle component may be a drawback since there are studies that state a more reliable data if gathered through combinations of muscle activities than sticking to one component only.

5.2.2 Evaluating usability constructs

An interesting read that draws a clear relation between psychophysiological tools and user experience research is the work by Frey, Muhl, Lotte & Hachet (2013). They do not only study the potential of PM in HCI, but they also draw relations between constructs of usability and the possible measurement tools for them. Even though the focus of the study is on constructs of usability that are measured using EEG, since the study starts from identifying constructs and how they are measured, it gives an idea of what is suitable for which case as well. The study could be an excellent read for people who are interested in EEG as a measure of usability, but also other PM tools are sufficiently mentioned. As a general conclusion, the study claims that workload, attention, vigilance, fatigue, error recognition, emotions, engagement, flow and immersion as recognizable by EEG. Out of those recognizable, workload, attention and emotion assessment are found to be those that benefit from the tool the most. In addition, they believe through the measure of error recognition, a construct that can only be studied using neuroimaging techniques, usability and comfort can be improved. Their study is quite interesting in a sense that it draws a clear relation between aspects of user experience and usability with psychophysiological tools. Such format is mandatory if researchers want to facilitate the basis for other studies that will go directly into evaluations, with minimal need for studying about the tools in details.

A study of linking usability constructs with PM tools was also conducted by Stickel et al. (2007) in examining a product's learnability using EEG. Their study focuses on assessing the ease with which users can make use of a certain software. How fast can a user become familiar with a soft-ware and get to carry out their tasks? In complex software, learnability becomes of a more significance since products that are hard to use become expensive as training consumes time and money. The study bases its analysis on the Yerkes-Dodson law that consists of two components. The first law states that performance increases with arousal until a certain level, a task dependent optimum, is achieved. At the optimum level, when arousal level increases to a point of stress, performance starts to decreases. Hence the second part of the law states that optimum level of arousal decreases with increasing task difficulty. Although the focus was on EEG, the authors briefly evaluate other PM tools as well. They claim HR is a more suitable tool when dealing with interaction tasks than during memory tasks. They state that HR is found to be barely related to task difficulty maybe indicating a parasympathetic withdrawal. On the other hand, SCL was quite high when dealing with difficult interaction tasks to the point of indicating a sympathetic excitation associated with the Fight-Flight-Freeze (FFF) activation. The researchers state that SCL may be a good indicator of an overall usability of a software.

The FFF activation can be seen by an EEG recording in the form of a high beta and gamma brain waves (Stickel et al., 2007). This activation rises the level of cognitive alertness (arousal) but decreases a certain type of higher cognitive performance such as articulation, learning, decision making, and interpretation of paralinguistic signals. Based on Yerkes-Dodson law, the level of arousal is correlated with cognitive performance when using information systems. EEG recording has the capability to show brainwaves related to the level of arousal. The hypothesis of the study was that "changes in the EEG recording of a subject during a usability test can be used for assessing the learnability of the software used". A software was considered learnable in the case of an alpha wave dominance, whereas the high dominance of beta and gamma was considered as barely learnable. The experiment by Stickel et al. (2007) was conducted on 10 participants. The subjects were divided into two groups; fast learners and slow learners. The study attempted to study the trends of learning between the two categories, based on the alpha/beta brainwave activities. There were two test phases called 'control scenario' and 'learning scenario.' In the controlled scenario, the learning trends (alpha and beta signal activity) of both the 'poor' and 'top' players had the same trend of learning with a varying level of a signal. Top players showed higher signal activity than the poor ones, but the pattern of activity was mostly similar.

In the learning scenario, however, there was a varying pattern in the beta wave activities between top players and poor ones. Overall, there was some level of link seen with the level of signal activity and the task difficulty. In the learning session, the activities increase with the process of learning and playing. However, the turning point comes in what could be a resignation phase when repetitious tasks appear, and the learning process becomes insignificant. In their conclusion, the authors seem to focus on defining the differences between the poor and top performers, instead of their original hypothesis regarding the relationship between learning and the alpha/beta brainwave signals. It was also confirmed by themselves that the research does not fully comply with their original hypothesis regarding the dominance of the alpha waves of good learners. Accordingly, they intend to conduct further studies that may help in making a clear distinction regarding the relationship between brainwave signals and a software's learnability, as well as clarifying the type of relationship they possess. In a nutshell, the study validates that the inclusion of PM tools in the existing usability testing scenarios can help assist improving the quality of software. With the introduction of new sensor and input devices, applications of PM tools as biological usability testing components and input/output interfaces will emerge. (Stickel et al., 2007.)

EEG was also used in a more natural setting, where reading experience on paper was compared to reading on a tablet computer (Rajanen, Salminen, Ravaja, 2015). In that study, the approach motivation measured using frontal asymmetry was higher in the paper reading condition. In addition, the study showed that different factors such as behavioral disposition, experience with the tablet and reading style had moderating effects on the approach motivation measured by frontal asymmetry. Thus, frontal asymmetry can be used both as a measure of behavioral dispositional especially when used in a reference task such as baseline recording, and as measure of emotional state during a task experience, in which case frontal asymmetry can be influenced by other factors interacting with the task experience. (Rajanen et al., 2015.)

Cernea, Olech & Ebert (2012) also studied measuring subjectivity using Emotiv Epoc EEG tools. The concept of subjectivity in the study was about feelings and states of minds of users during usability testing. The initiation was from the limitations of traditional usability testing, which conduct studies based on post-experience questionnaires and interviews. The study presents EEG headsets as supplementary tools for subjectivity testing. Two scenarios with the capacity to generate multiple emotional responses were designed and carried out with 12 participants. Values of the neuro-headsets were compared with results of questionnaires, interviews and video log analysis conducted during and post experience. Emotiv Epoc headset software supplies users with a classification of signals as expression, emotion, and commands. The experiment design focused of studying engagement, excitement,

satisfaction and frustration. Satisfaction and frustration are not part of the tool's framework but were rather generated using combinations of levels of excitement and engagement with different facial expressions.

During the experiment, participants wore the headsets always. The games get paused in between mainly when a spike in the emotion parameters is seen. Time is registered, the user then is supplied with questionnaires regarding all four emotions. For the sake of comparison, the game gets stopped in random times where nothing happens and questionnaires are conducted. After game completion, post-experience interviews were conducted as well. For most emotion cases, the comparison be-tween the device detection level and the questionnaire results were quite close with 0.32 units' average difference (on a 1-5 Likert scale). One exception was excitement detection in one of the two games users played, where the margin went a little over 1. The authors, based on post-experience interviews, speculate that this might be from the perception of the construct by the users. There were issues of false positives and false negatives from the EEG signals. The basis of this 'false' was thought to originate from training level and state of mind of the users. False detections only mean that there have been unwanted detections of other activities. The device recordings are influenced by other activities from the skin, muscle, and nerve. To address their assumption of false values from shortage of train-ings, the authors conducted another experiment which includes user training. The attempt that in-cludes training, where participants tune the sensitivity of the system with the supervisor, resulted in a 12% improvement in correctly detecting facial expressions. Improvement in false positives was made through tuning the detection into muscle groups, instead of specific muscles. The authors note establishing a user-specific baseline for each subject and each emotion as well as the usual deviation levels from this baseline could help improve detection levels. (Cernea et al., 2012.)

Arndt et al. (2016) presents a compiled review of studies conducted using EEG to measure the quality of experience in a video, audio and image usage. They state the need for neurophysiological tools, where there is not room for bias since emotion data is directly collected for the user, as vital in ad-dressing emotions that are not describable at the conscious level. The continued demand and contin-uous competition for services such as video streaming, calls for an improved and continuous evalua-tion on the quality of experience. The study reviews experiments conducted using EEG and identify two distinct methods where emotion can be measured with the tool. Firstly, data can be analyzed concerning a short and distinct event that elicits an event-related potential (ERP). Here the amplitude of the ERP's component can vary with the level of quality perceived by the user. The component under study, in this case, is the P300 component. Secondly, data can be studied using an analysis of the frequency band power, this refers to frequency components such as alpha and theta waves. The authors suggest that this is especially interesting when drawing a conclusion about the mental state of participants, or to describe the change in the mental state between conditions.

For the analysis of ERP, a small set of electrodes can be sufficient, usually 8 electrodes, should be distributed along the central line. While for the hemispheric analysis, equally distributed electrodes over the right and left hemisphere are advisable (Arndt et al., 2016). Since EEG signals originate from multiple sources and are potentially overshadowed by unrelated activities to the ones we are interested in, smoothing out the original signal through averaging multiple ERP recordings is necessary. Hence, usually 20-30 trials at minimum are needed for an average ERP per stimulus class. According to the study, such issues could also be one of the drawbacks that limit the number of studies using EEG.

From the reviewed papers by Arndt et al. (2016), in studies of video and images, highly distorted stimuli were indicated by a high amplitude P300, on the ERP extracted around the onset of distortion. In studies using audio stimuli, heavily distorted sounds were also represented by higher P300 ampli-tudes. In addition, one interesting finding that could add to the value of such methods was that there were cases where distorted sounds not reported in subjective responses were identified on the ERPs. This is evident by the fact that ERP signals have the same trend as in the case of a subjectively

reported distorted signal. This could potentially indicate a detection of emotions that do not penetrate up to the subjective behavior level. From the spectral asymmetry perspective, like the ERP cases, reduction in quality showed an increase in alpha and theta levels representing fatigue and drowsiness. For validation purposes, in one of the studies, eye blink was also monitored, longer blink duration was recorded in cases of lower quality videos than higher quality ones. The longer duration eye blinks also indicate a high level of fatigue. In this study where eye blink was monitored SCL and ECG were also measured showing no sign of variation towards a change in quality. In a study that used a 2D and 3D video as a stimulus with varying levels of quality, EEG recordings showed a high frontal asymmetry in the alpha power band, which reflects emotional affect towards the two different quality levels.

Overall the review showcases that the approach of measuring ERPs and spectral waves using EEG for quality of experience purposes is valid. Changes were observed on both the P300 and the alpha band power variations due to quality impairments, although applications of the two approaches may vary. Quality reduction and its effect on frequency components, represented by a reduction in cognitive states changes, was also visible. ERP observation seems more suitable when using short stimuli. A relatively high number of repetitions is needed to obtain a smooth averaged ERP. The authors note that solving such issues will make the potentials of EEG, as a measure for quality of experience, immense. (Arndt et al., 2016). This study is valuable in a sense that it includes a detailed review of EEG studies with respect to the quality of experience and includes references that have guidelines in how to conduct experiments using EEG. The reviewed papers themselves could be used as references for those who would want to conduct an experiment using EEG, depending on the context of use.

6.   Application Areas

Psychophysiological measurements can potentially be applied to a range of disciplines. The possibilities of continuously monitoring human emotion and possibly using that data to make applications and software adhere to user's need either as an interaction interface or as a user experience evaluation technique are interesting. As a result, various studies are being conducted to make such potential a reality. EMG and EEG tools on usability and user experience studies have been experimented with in media research, the gaming world, assisted living studies, ambulatory services, e-commerce and other areas where creating a positive experience is an essential aspect.

In this research, to a varying extent, we have covered studies across different environments. The application contexts were divided in two distinctions of tools (EEG and EMG) being implemented as a part of an interaction design and when used as an objective user experience evaluation tools. Both formats intend to create a better user experience through creating more intelligent and responsive to emotion systems or assisting the development environment with a better understanding of the human experience.

Facial EMG is mostly used to study a valence of emotion via the activities around the corrugator major, zygomatic supercilii, and orbicularis oculi muscles. The corrugator and zygomatic are the most studied from the three and are affiliated to negative and positive representations of an emotion, respectively. Measuring user frustration levels (Hazlett, 2003), evaluating boredom, challenge, excitement, frustration and fun human emotions in a play environment (Mandryk., 2006), measuring user's perceived task difficulty (Branco et al., 2005) are few of the studies that use facial EMG as a tool to measure human emotion in the context of interacting with a computer. However, it should be noted that EMG's applications come with its own limitations and use context issues. It is very important to understand the context of use, experiment environment, and the software applications or systems when trying to draw a pattern between emotion and such tools.

A thorough analysis of previous studies and the issues faced comes very handily when trying to make use of psychophysiological tools. For example, facial EMG had issues of giving out false positives in gaming, and it is recommended that it be used in combination with other tools to validate the actual emotion (Kivikangas et al., 2011). The result, in this case, raises the issue of social masking, in a sense that people may not show their true feelings in particular social context. Bearing in mind the issues related to social masking, Ravaja, Cowely and Torniainen (2016) suggest that best form of EMG's use seems to be when used with participants reacting to scenes on screens. When implementing facial EMG in emotion-related studies, it is important to take the context of use and the environment into consideration. An emotion of challenge also showed a very minimal intensity than expected in a play environment (Branco et al., 2005). Players responded that they tend to relax to increase their performance when faced with a challenge, in a play environment. Frustration level also shows low intensity than subjectively reported as well. Hence, such an application demands a careful understanding and caution. Experiments should be conducted keeping theses varying possibilities in mind.

Other types of EMG are also studied, more in user experience studies through human-computer interaction. Before such studies, researchers note that surface EMG has application in sports training, rehabilitation, machine and computer control, occupational health and safety, identifying posture disorders, medical diagnosis, ergonomics (Naik et al., 2006; Kaneko, 2008; Ravaja, Cowely & Torniainen 2016). Detecting the direction of listening for music applications, to develop a responsive system to users' needs (Kaneko, 2008), a wearable interface for mobile devices (Constanza, Inverso and Allen, 2005), exploring the possibility of hand gesture recognition systems for human-computer interaction (Naik et al., 2006), are few of the studies that experiment with EMG as a communication method. However, EMG tools are susceptible to noise and other interferences making its application

on a day to day activity a bit more difficult. Also, when EMG or any other tools is implemented as an interaction device, in most cases, there is a need for users' training on how they can use their muscle activities. We are talking about a change in the whole dynamics of human-computer interaction in a sense.

In the context of biofeedback systems, such as the one implemented by Valve, that makes use of the skin conductance level (SCL) to understand the arousal of players and adjust the game difficulty level (Kivikangas et al., 2011), there is no need for training the users. Such systems create a loop without the necessity of the user's conscious input. In the context of systems where EMG or EEG are used to automatically get the user emotion data and use that in creating a more exciting or enjoyable experience through a biofeedback mechanism, user training is not the concern. Applying facial EMG and EEG in a real-world environment is, at least at the current level, hindered by the presence of electrodes around the face. In this context, the main challenge here is to develop systems that are accurate and deployable in a day to day use context.

The potentials of EEG as both as a user experience evaluation tool and an interaction interface is also explored by various studies. Its application areas include persuasive advertising and e-commerce (Saari, Ravaja, Laarni, Turpeinen, Kallinen, 2004; Kivikangas et al., 2011), as a direct input mechanism for games and toys, as a design component in adaptive systems to users' affective states (Kivikangas et al., 2011), to investigating physiological correlation between perceptual and attentional processes (Arndt et al., 2016), clinical and psychiatric studies (Wand, Gwizdka, Chaovalitwongse, 2016), brain-computer interface (Culpepper & Keller, 2003), social behavior studies, urban design (Skroumpelou et al., 2015), reading news (Rajanen, Salminen, Ravaja, 2015), and in measuring subjectivity in user experience studies (e.g., Cernea, Olech, Ebert, 2012).

Obviously, the application areas for a device that can directly measure activities in the human brain is immense. EEG tools come with their own limitation in their spatial resolution, price, and complexity. The potentials of its application could also be dictated by the reduction of its limitations. As mentioned previous in section 4.2, limitations of price are partially being addressed. The complexity in terms of its physical presence is also improved with the introduction of EEG headsets such as Emotiv Epoc and NeuroSky. Such devices also face their own challenge in improving their precision and accuracy. With improved functionality, an application on a day to day activity could still be limited due to the presence of electrodes around the head. There are also steps to take in how the electrodes are preserved and applied. Overall, the signs of improvement and the wide range of studies conducted using the tools is a clear sign that their application is not too far ahead. The applications of both EEG and EMG, in the user experience domain, will potentially supply the field with lots of knowledge regarding the human emotion, which in turn could be used to create a lot more intelligent systems that create a pleasant experience.

7.   Future Directions

The usage of psychophysiological tools in the context of user experience studies is evidently on the rise. Various limitations slowed down the rate at which such application are researched and implemented. Limitations include price, expertise, high cost, time consumption in design and analysis, and device format. Solving the limitations could be the determinant factor in paving the way for the potentials of PM tools in UX to be fulfilled. There are steps taken in addressing some of the limitations thus far. For example, more affordable and less complex EEG tools such as the Emotiv Epoc devices have occurred in the market.

The increase in number of studies that make use of EMG and EEG, will surge the level of knowledge in the area and facilitate the development of more algorithms, standardizations and deployment methods. Development of more efficient algorithms that help analyze psychophysiological signals, will pave the way for future application developments and more profound understanding of event related emotions (Kivikangas et al., 2011). Especially in the EEG context, the problem is mostly on the ability to distinguish between the desired and irrelevant signals. Methods such as Independent Component Analysis (ICA) and time locking event-related potentials facilitate the deployment EEG tools (Kivikangas et al., 2011; Arndt et al., 2016). Tools such as the Emotiv Epoc have their own classifications of signals into categories of emotion, expression, and commands. The tool has detections enabled for detection of a set of facial expressions, emotional states and cognitive commands (Cernea, Olech, Ebert & Kerren, 2012). The precision and accuracy of such tools could be limited and the detection and categorizations need to be studied for their efficiency. However, the possibilities of such devices that may facilitate such emotion detections and classifications is a very good start for future research and application developments using the tools.

Steps to develop standardizations of using psychophysiological tools in the context of quality of experience is also noted in (Arndt et al., 2016). They also state that in the video quality experts group, one section has taken the initiative to study the topic thoroughly. The idea is to study the method's robustness, with various labs using different types of PM tools being involved. Continuous development of best approaches to get the best out of these tools is necessary. One future approach could be combining different methodologies to have a more reliable and generalizable data (Kivikangas et al., 2011). Not only combining the different PM tools but also with traditional tools could be one approach (Mandryk et al., 2006). They suggest combining EMG with video observation tools, where researchers can identify interesting features such as increase or decrease in an emotional state, and investigating those time frames instead of the whole video could save a lot of time. In addition, in the context of false positives, such as the case with facial EMG false smiles and frowns, using a combination of tools might help distinguish the real emotion masked by the subject. One idea could be to monitor EEG together with EMG and compare EMG valence states with corresponding EEG approach/withdrawal readings.

8.   Discussion

The study intends to give a knowledge base about psychophysiological tools, EEG and EMG in specific, and their application in user experience research. The study reviews limitations that exist in the area, measures to solve such issues, advantages these tools bring compared to traditional methods, the context of use, emotion indicators that are measured, and application areas in HCI and UX research. Psychophysiological tools are the umbrella of the study, but main the focus is on EEG and EMG in specific.

The format of the research is designed in such a way that aims to create an understanding about *how can EEG and EMG psychophysiological measurements be used in user experience research*? For answering the main research question, five sub-questions have been derived and answered. Thus, chapter 4 deals with advantages and limitations (RQ1&2), chapter 5 deals with EEG and EMG's use context on user experience research (RQ3), chapter 6 deals with application areas (RQ4) and chapter 7 deals with future directions (RQ5). In this chapter, the intention is to give a summed-up view of the general idea and implication of the review. The order will also follow a similar trend in a sense that it starts with explaining limitations and advantages, then discuss use context, followed by a brief overview of application areas, and finally, make few suggestions regarding future directions.

8.1 Advantages and limitations

Advantages of psychophysiological tools originate from two aspects. One is in the form of addressing the limitations with traditional user experience evaluation methods. In this sense, the use of psychophysiological tools could be as a replacement to conventional methods and as a supplement to give more insight into the user's state of emotion during an experience. By traditional methods, we mean surveys and questionnaires during and post usage, observational analysis using video recordings, and heuristic evaluations.

In comparison to questionnaires and surveys during or after usage of an application or a product, psychophysiological methods help avoid social bias. In most cases, such methods are prone to errors caused due to user's answering styles, social masking, wording limitation in the questionnaire, and participant's conscious knowledge of experience, observer bias, participant's memory, and interruption of the process of experience. In addition, and user's responses might be influenced by the novelty of applications. (Potters & Bolls, 2012; Kivikangas et al., 2011; Constanza, Inversion & Allen, 2005; Mandryk et al., 2006). Video recordings face problems of illumination, view point variations, in addition to being very rigorous and time-consuming. (Partala et al., 2005; Mandryk et al., 2006). Heuristic evaluations are mostly done with experts, and when it comes to novel applications, there are no experts. Also, the latter method is more useful to uncover usability issues in usage rather than aspects of user experience such as playability (Mandryk et al., 2006).

The advantages of psychophysiological tools over subjective methods such as questionnaires during and post usage was shown on some of the studies reviewed. Cases of emotions in the users' subconscious level, where subjective methods did not manage to indicate were seen on (Arndt et al., 2016; Mandryk et al., 2006). Arndt et al. (2016) used EEG for their experiment while Mandryk et al. (2006) conducted theirs using facial EMG. The advantages of EEG and EMG is not only in detecting emotions that the users may not be able to explain or notice for that matter, but also in the ability of monitoring experience continuously. The ability to continuously monitor user's emotion supplies researchers with a dynamic understanding of user experience which in turn can be temporally related to interface events, navigation logs and other user behavior (Hazlett, 2003). The ability to time lock events to user emotions could also be helpful in the context of PM tools as supplementary to tradi-

tional methods. For example, with video observation, in which case researchers can identify interesting features such as increase or decrease in an emotional state, and then investigate the corresponding time frames in a video recording instead of examining the entire video. Such possibilities will drastically reduce the time required to qualitatively analyze user interaction with entertainment technologies (Mandryk et al., 2006).

Another advantage or need for PM tools originate from their added potential in measuring the user experience directly and their potential use as interaction tools. Advancement of mobile devices and the multi-disciplinary nature of human-computer interaction are other factors that drive the need for studies of PM tools in the area. The ability to communicate with a computer and other machines with various gestures has been a potential area of application for EEG and EMG. Such advancements, in the long run, will supply a computer with the ability to be well versed in the complex human communication mechanisms. Tan Le (2014) presented, on TED talks, an EEG (Emotiv Epoc) based application that is used as a command input interface to guide the movement of a wheelchair. Eye blinks dictates left and right turn instructions, and a smile translates as a command for moving straight in a forward direction. The application examples also include smart home applications where a process of thought measured using EEG, acting as an instructor for various tasks such as opening/closing a curtain.

Psychophysiological tools come with their limitations, which could be the reason behind the lack of its typical applications. Price, complexity, device format, time consumption, shortage of sufficient knowledge, and data dependency on interpretation are the major limitations (Kivikangas et al., 2011; Potter & Bolls, 2012; Ganglebauer et al., 2009; Pentus et al., 2014; Ravaja, 2004; Park, 2009). Price limitation was the biggest factor especially on the usage of EEG. Previous tools cost thousands of euros limiting their user experience studies. However, the emergence of small headsets such as Emotiv Epoc, have paved the way for cheaper tools that for UX research. However, these devices come with their limitation since their simplicity and price reduction comes at the cost of their precision and interference tolerance. Such limitations may add more work in data analysis to distinguish the desired signal and avoid unwanted ones. Nonetheless, their simplicity to be worn as headsets and their wireless capability, together with the reduced price, makes them valuable in addressing IOT challenges and trends of real time EEG signaling (Konstantinidis et al., 2015).

There are various studies conducted using EEG and EMG with promising results. Emotiv Epoc itself supplies researchers with built-in software, which contains categories of emotions. Some studies make use of these built-in types for user experience studies (Cernea et al., 2012). The availability of such functionalities might also address issues that originate from data analysis complexity. Most user experience studies experts could find it hard to use various sophisticated algorithms such as Independent Component Analysis (ICA), one of the algorithms that can help in distinguishing and mapping desired signals. Monitoring Event-Related Potentials (ERPs), indicated by pick voltage signals in response to a stimulus might be simple, but signals need to be smoothed to remove unwanted signal through averaging several trials of a single recording. However, the categories themselves need to be evaluated extensively before they can be considered fully reliable. Cernea et al. (2012) used some Emotiv Epoc EEG tool's emotion categories, in their study of subjectivity. In their study they state that most of the attributes managed to accurately relate to the user's actual emotion with few exceptions. Further evaluations of the classifications and their validation could lead to improvements in complexity issues in this regard. Another option studying human emotion on EEG tools is analyzing frontal asymmetry. (Ravaja et al., 2016; Kivikangas et al., 2011.)

Facial EMG is relatively simpler to record and analyze and are commonly used to infer emotional valence in response to a stimulus (Ravaja et al., 2016). In the case of surface EMG, algorithms such as ICA are mandatory to distinguish between the desired and unwanted muscle activities. The problem of the devices' complexity and the need for expertise some of the limiting factors in the number

of studies conducted using these tools. However, as mentioned in previous paragraphs, although the accuracy and precision are not efficient, development of systems such as the built-in software in Emotiv Epoc, which entails emotion and cognition categories, is a step towards an easier implementation of such tools. However, despite the improvements in the analysis and deployment side, the device format with electrodes attached to the face and scalp areas, could also be a limiting factor when we think of their real-world application.

Other issues include the fact that a controlled set up is needed since most of the devices are susceptible to noise and also generate noise (Arndt et al., 2016). Actually, Ganglbauer et al. (2009) even go as far as claiming that PM tools are not ideal in situations that include people in motion, due their susceptibility to interference that originate from activities such as movement and also external noises. On the other hand there are other studies that claim such tools can be applied in subtle environments as interaction devices (Constanza et al., 2006). However, in general, it is very important to keep in mind the context of use and the limitations of the devices when conducting experiments.

Implementation of experiments using PM tools is also one aspect of the drawbacks. Depending on the tool in use, there is a need for procedures and caution in implementation of experiments. In addition, the format of the devices, with electrodes attached to one's face and scalp could hinder a day to day application of the tools (Constanza, Inverso & Allen, 2005; Kivikangas, Chanel, Cowley & Salminen, 2011; Potters & Bolls 2012; Ravaja, Cowley, & Torniainen, 2016). Time consumption is another issue that arises from the implementation and analysis limitations. In comparison to other methods such as surveys and questionnaires, using PM tools could be a lot of work. As a result, researchers might opt to the traditional methods, if the advantages PM tools bring does not outweigh the limitations.

Facial EMG faces signal interpretation complexity as well. In some cases, distinguishing between a real and a fake smile/frown is not straight forward. There have been reported cases where people hide their real emotions and show a fake smile, in competitive scenarios or when interacting with people of higher order (Ravaja et al., 2016). Such cases alert researchers into paying attention to the context of their study. For example, one of the cases where such a false positive occur was in a gaming environment where players might want to hide their real emotions when playing against a co-located person. A positive valence in the zygomatic muscles at a person's own death, in a game, was recorded. Ravaja et al. (2016) suggest that such cases could be part of opposite emotion expression by players for emotion coping. They state that facial EMG does not always represent a true internal emotion and is susceptible to display rules. In a slightly different context, Branco et al. (2005) reported a case where zygomatic muscle activities showed a stronger relationship to difficulty levels, contrary to the notion that the muscle activity is always occurring in response to positive stimuli.

False positives and issues with social masking and display rules also raise the question of the improvements PM tools bring to the traditional methods. It will be intriguing to know the extent at which social masking and display rules may have on facial-EMG compared to subjective methods such as questionnaires. Of course, the context of display rules varies in the two contexts. In the case of the false positive in facial-EMG recorded, was in the gaming world where there is competitiveness between people, whereas in traditional user experience measures the social masking could be that the subject may not be entirely honest in their response.

In using PM tools, it is important to be aware of the context and make sure that our interpretations take different aspects into consideration. One way to walk around issues of false positives could be using the tools in combination with each other and other methods. For example, it would be interesting to implement facial EMG together with EEG and replicate the same gaming experiment. The false smiles may be contradicted by the frontal asymmetry of a user corresponding to approach/withdrawal towards a stimulus. Built in emotion categories such as excitement, engagement, fun could also be

used to validate if real positive feelings occur in facial EMG. One problem here could be that facial EMG is susceptible to noise and that EEG generates a lot of noise (Potter & Bolls 2012, Arndt et al., 2016). Such issues should also be taken into consideration when implementing psychophysiological tools in combination.

On the other hand, false positives could have values in themselves. For example, in studies of social behavior and display rules affect people, it could be intriguing to measure facial EMG in combination with other tools and analyze the context in which social masking takes effect. When we think about a computer-mediated communication and sufficiently intelligent systems, it could be important to teach the computers themselves about social behavior. Other issues with EEG and EMG implementation are device format and time consumption in signal analysis (Constanza et al., 2005; Kivikangs et al., 2011; Potter and Bolls, 2012; Ravaja et al., 2016). Careful planning of experiments, mostly controlled, data analysis, proper implementation of the experiment such as; placement of electrodes, skin abrasion, gelling electrodes are aspects of PM tools that need caution. Proper planning and implementation are determinant to the outcome of an experiment.

8.2 Design context

In this study, EEG and EMG are reviewed in two contexts; as part of a design facilitating human-computer interaction and as evaluation tools that measure user experience of products. The general goal of both approaches is to enhance user experience and facilitate functional, exciting, pleasant, enjoyable, and useful products.

8.2.1 Psychophysiological measurements as UX design components

As design components, EEG and EMG tools serve as command interfaces, biofeedback mechanisms, usability logs and other contexts within the media, gaming and software development domain. Most studies were experimental applications, prototypes and implementation models instead of a full-blown application. There are some commercial scale applications of SCL as a biofeedback mechanism, one example is in a gaming sector by Valve (Kivikangas et al., 2011). The company designed a game that adjusts difficulty level based on the player's skin conductance level, which increases while sweating. However, most of the studies covered in this research only go as far as validating the potentials of EEG and EMG tools to be interaction interfaces. Most work focuses on validating signal detection and recognition rates and their potential integration into various systems (Culpepper & Keller, 2003; Partala et al., 2005; Naik et al., 2006; Kaneko 2008).

Some of the studies in this review that present the potentials of EEG and EMG as an input/output interface tools (Constanza et al., 2005; Culpepper & Keller, 2003; Naik et al., 2006). Types of EMG equipment, mostly presented in the context of interaction devices are surface EMG. Facial EMG mainly measures the valence of emotion through the smile and frown muscles. EEG applies in both contexts, and there is no distinction as surface and facial since there is only one type, which measures activities from the surface of the scalp. Devices in this context might vary, in the number of electrodes and their capacity of measurement. Laufer & Nimeth (2007) used SCL to study participant's stress level in a play environment and use the data to predict user action in a game. The possibility to predict user actions, in the long run, could potentially help design systems that trigger those actions.

Applications of PM tools in the design component context applies to various important fields such as the health sector. For example there have been studies where EEG headsets is used to monitor brain activities of epileptic patients while they carry out their normal activities (Gyselinckx, 2010). The application is part of a design framework abnormal activities in the brain can help predict epilepsy pattern signaling for ambulatory services. In addition to being indicators of valence of emotion, facial EMG could also have a potential in helping computers learn emotional intelligence. Partala et al.

(2005) studied facial EMG as a potential tool for communicating emotion related information with a computer. They claim the computer can have a real time understanding of the human emotion, which in turn can help it make user experience decisions, making it a more emotionally intelligent system.

Constanza et al. (2005) studies the potentials of EMG as an input and output mechanism. They developed an algorithm that managed to detect subtle gestures to a varying extent. Although the accuracy was not 100%, it was enough to show that such systems could be developed. Naik et al. (2006) also studied the use of EMG as an interaction mechanism. Their study focused on hand gestures that could potentially be used to pass commands to a computed. The experiment was controlled and the participants carried out minimal activities, which raises the question of reliability in the context of people conducting different activities. However, this are baby steps that lead to bigger and better things. With a compilation of such works and continuous development on the basis of corresponding results, the realization of complete systems that includes PM tools as a design component is not too far ahead.

8.2.2 Psychophysiological measurements as UX evaluation method

Various studies that implemented EEG and EMG tools as user experience evaluation methods were studied in this review. The studies were divided into two aspects with some that evaluate aspects of emotion and other's that evaluate the tools' usability capabilities. The distinction in this case is top level and does not contain significant variation. As a result, there were some commonalities between studies across the two categories. In this section the intention is to give a general idea of EEG and EMG's use in the user experience context.

Various aspects of user experience have been experimented with, using EEG and EMG tools. A product's learnability, difficulty level, and quality were studied using the tools (Stickel et al., 2007; Branco et al., 2005; Hazlett, 2003). Emotion aspects such as frustration, boredom, excitement, fun, engagement, satisfaction, and challenge are studied with their representation and detection rates with extent. The state of attention, effects of workload, vigilance, fatigue, error recognition, vigilance, and cognitive alertness are also aspects that can be measured with such tools. (Frey et al., 2013)

The extent of success in measuring aspects of emotion, cognitive states, performance and others, depends on the context of the study and the product and software in question. For example, the aspect of challenge studied by Mandryk et al. (2006) indicated an inverse relation to the assumption made by the researchers. The assumption was that the player's arousal would increase when the challenge increases. However, participant's comments revealed that when facing a challenge, they tend to relax to improve their performance. The study was in a gaming environment, which showcases that context must be taken into consideration when making an assumption regarding people's emotional responses. In the same study, frustration and boredom levels were also found to be lower in intensity than what participants rated in questionnaires. On subjective responses when asked their frustration levels, participants, for example, may pick "very frustrated" from a Likert scale. However, being frustrated in a gaming environment may be relatively less intense than having a car broke down on a highway. The measure emotion, in this case, shows a less intense frustration potentially because of the context of interaction since the overall experience in a gaming environment is fun. Even though the assumption may not be conclusive, it is very good indication that context of what we are measuring is critical when making assumptions and analyzing the data we collect using PM tools. We cannot take the information as it is and interpret it, without taking the whole picture into consideration.

Hazlett (2003) also studied frustration levels but in the context of website usage in this case. Data gathered, in this case, was not only evaluated on people's responses but also assessing the expected difficulty levels of the websites via attributes such as some pages, steps it takes to complete a task, frequency of accurately completing a task and others. This could be a good example of taking various

aspects of evaluation into context when measuring aspects of emotion in different context. The study suggests that the data gathered through facial EMG, namely corrugator activity, was in line with their expected values of frustration.

Various components of EEG and EMG are measured for user experience studies. When it comes to EEG, there are two fundamental ways data can be assessed; one is data can be analyzed concerning a short and distinct event that elicits an event-related potential (ERP). Here the amplitude of the ERP's component can vary with the level of quality perceived by the user. This method is monitoring components such as P300, increase in positive voltage occurring following a stimulus onset. The second way is, data can be studied using an analysis of the frequency band power. This is especially interesting when drawing a conclusion about the mental state of participants, or to describe the change in the mental state between conditions. The latter one is where the alpha, beta, theta and gamma bands of frequency come into play (Arndt et al., 2016; Potter & Bolls, 2012. A relatively high number of repetitions is needed to obtain a smooth averaged ERP. Accordingly, ERP observation seems more suitable when using short stimuli.

According to Kivikangas et al. (2011), frontal alpha asymmetry is mostly used when studying emotion and motivation using EEG. Frontal asymmetry dose not directly measure valence but it represents a broader motivation tendency towards approach and withdrawal related behavior and emotions. Greater left frontal activity indicates a propensity to approach a stimulus and the greater right activity indicates an inclination to withdrawal from a stimulus. Rajanen et al. (2015) have used frontal asymmetry to assess the approach motivation towards print or digital media when reading a newspaper. Other studies also explore a relative dominance of different frequency bands as indicators of user experience issues. For example Stickel et al. (2007) state that software is considered learnable in the case of an alpha wave dominance, whereas the high dominance of beta and gamma shows a software as barely learnable. Arndt et al. (2016) also state high frontal asymmetry in the alpha band power to be indicative of emotional affects.

Overall, there are links between the frequency bands and aspects of emotion, which facilitate different emotional interpretations. Alpha activity is related to deep relaxation, beta with alertness/conscious focus, the delta is associated with sleep in healthy humans, while theta relates to creativity/relaxation/depression (Potter & Bolls, 2012). Occurrence and intensity of one or more of this waves have its indication of emotion. For example in the study conducted by Arndt et al. (2016) reduction in quality of a stimulus showed an increase in both alpha and theta levels which represents fatigue and drowsiness. The result was also supported by monitoring eye blinks that were in line with indications of fatigue, represented through longer blink durations.

Facial EMG measures corrugator supercilii (associated with frown muscles), zygomatic major (related to smiling) and the frontalis (responsible for moving the eyebrows upwards) muscle activities (Partala et al., 2005; Potters & Bolls, 2012, p. 125; Branco et al., 2005). However, in most studies, the focus is on two of the three muscle activities, which are the zygomatic major and corrugator supercilii. In one of the studies reviewed (Branco et al., 2005) frontalis data was dropped due to interferences on trials. Such issues might explain why there is more focus on the other two muscle activities. The focus on zygomatic and corrugator muscles could be because they are direct representations of a frown/smile, supplying researchers with a platform to have a positive to negative valence of emotion platform. Depending on the focus of study, researchers focus on the activity of their interest. Hazlett (2003), for example only focused on data collected from the corrugator muscles, since the study focuses on user frustration levels. Partala et al. (2005) used a combination of the two muscle activities. The subtracted the corrugator supercilia activity value from the zygomatic major activity values. If the score is above zero, then the experience is positive, and below zero as a negative emotion. Their finding suggested that this approach had the highest accuracy level in detecting the valence of emotion.

8.3 Application areas and future directions

EEG and facial EMG as user experience evaluation techniques apply in multiple application areas. Applications include smart homes, clinical studies, urban planning, gaming, media research, behavioral studies, user experience studies and any research that involve usage of products and the resulting human emotion (Naik et al., 2006; Kaneko, 2008; Ravaja, Cowely & Torniainen 2016; Branco et al., 2005). Various studies across different domains have been reviewed in this project. Two major distinctions were designed with respect to application context 1) PM tools as UX design components and 2) PM tools as UX evaluation tools. The focus in both cases was on EEG and EMG measurement tools, which are the focus of the study.

From the design component perspective, studies that either study the tools as potential input/output mechanisms and other that consider the tools as biofeedback mechanisms were reviewed. As an input/output interaction mechanism, the application areas could be any interaction with a computer and mobile devices. Potentially all the functionalities that we carry out with our mouse and keyboard could be replaced or supplemented with such tools. On top of the activities we carry out, there could potentially also be systems that we can instruct with the process of our thought and emotion expressions represented by muscle activities. For example, there has been a case of a wheelchair being instructed by recordings from Emotiv Epoc headset (Tan Le, 2014). In the demo, a person was moving his wheelchair with instructions such as eye blinks and head movements.

In the context of biofeedback mechanisms, the computer or system records expression or emotion signals and reacts to them in different contexts. For example the game company Valve, used SCL to monitor people's stress level while playing a game (Kivikangas et al., 2011). The task level of the game is increased and decreased based on the values of the SCL. Gyselinckx et al. (2010) also studies ambulatory monitoring of brain activity for epileptic patients while they carry out their normal day to day tasks with an EEG headset is experimented with in a prototype level (Gyselinckx et al., 2010). The application helps to study abnormal brain activities that may help predict epilepsy patterns for ambulatory services. This is also another case where PM tools create a design loop with biofeedback mechanisms that in turn can initiate a response from a system. The difference in the application areas in these two example also tells the wide range of application areas such tools can potentially be deployed.

As a user experience evaluation tool, PM tools can be implemented across different disciplines such as persuasive advertising and e-commerce (Saari, Ravaja, Laarni, Turpeinen, Kallinen, 2004; Kivikangas et al., 2011), as a direct input mechanism for games and toys, as a design component in adaptive systems to users' affective states (Kivikangas et al., 2011), to investigating physiological correlation between perceptual and attentional processes (Arndt et al., 2016), clinical and psychiatric studies (Wand, Gwizdka, Chaovalitwongse, 2016), brain-computer interface (Culpepper & Keller, 2003), social behavior studies, urban design (Skroumpelou et al., 2015) and user experience studies (Rajanen, Salminen, Ravaja, 2016; Cernea, Olech, Ebert, 2012).

Ideally the potentials of such tools that can record signals and activities that represent human emotion, have a wide range of potential applications. The deployment and day to day application of this tools is to some extent slowed down by its limitations. However, there have been continuous development in the number of studies that make use of such tools and wide range application of the tools in the user experience domain is only a matter of time. Improvements in signal analysis, device complexity, price, and corresponding expertise could hasten the realization of such tools in every day applications.

The continuous increase in the number of studies conducted using EEG and EMG tools in UX, will result in an increase in knowledge leading to development of more algorithms, standardizations and

deployment methods. Such developments will pave the way for more usage of the tools and a realization of their full potential in the future (Kivikangas et al., 2011). In this review, there were studies that showed that algorithms such as ICA are shown to be efficient in distinguishing EEG and surface EMG signals from unwanted recordings. Studies along this lines and further improvement of the approaches will help solve complexity limitations in the future (Kivikangas et al., 2011; Arndt et al., 2016). One idea for the future could be to design a method that can automate such data analysis tools so that the tools themselves will be equipped with mapping signals to emotions. Of course, as stated in previous sections, tools such as Emotiv Epoc has their own classification of signals into categories of emotion, expression and commands (Cernea et al., 2012).

Constant work on finding the best approaches and implementation formats for PM tools is important. In the future, works maybe conducted using PM tools in different combinations, in order to narrow down to accurate interpretation of data. Combining different PM tools could help acquire a more reliable and generalizable data (Kivikangas et al., 2011). There were sample studies is this review that developed a model using multiple PM tools (Mandryk et al., 2006). In fact, the latter study is one of the few studies that developed an actual model of user experience evaluation. Work along similar lines is important since as the field matures, models and tools that can be used for user experience evaluation studies should be developed. The challenge in this sense is that emotions themselves are context dependent and a single model across different application areas could be difficult to implement. Continuous experiments and collective knowledge is very important in addressing issues of such context dependencies and designing experiments that cater for them.

9.   Conclusion

The purpose of this study was to give a solid foundation about the use of psychophysiological tools in user experience research. The intention was to give a reference point where a compiled understanding of previous implementations and studies of psychophysiological tools, EEG and EMG in specific, can be found. We intend to address the phenomenon of increased research of such tools in the area of user experience studies. The compiled reviews of previous studies and their analysis can help construct a body of knowledge

To help assist reach our goals of answering *how EEG and EMG psychophysiological measurements can be used in user experience research*? Five main research questions were formulated. Questions include: 1) what limitations exist and how they are overcome? 2) What are the advantages in comparison with traditional methods? 3) How are psychophysiological tools used in user experience studies; a. as part of UX design or b. as a UX evaluation tool? 4) What applications areas are researched? And 5) what are the possible future research directions?

A narrative literature review was used to help answer the research question and reach our goal of laying a solid knowledge foundation in the topic of psychophysiological measurement tools in UX research. With the aim of developing a comprehensive understanding, topics are studied, categorized summarized and synthesized. Previous studies were included and excluded based on relevance on the basis of insight into the topic, which developed through the process of reading. Accordingly the study made distinctions of previous work regarding psychophysiological tools in user experience studies; one is their implementation as design component and the second is their use as user experience evaluation tools. The second category was also further divided into two concepts of measurement; emotion attribute evaluation and usability construct evaluation. The classification and categorization of the review could potentially give other researchers a clear view of the phenomenon depending on what they want to focus on.

It is apparent that EEG and EMG tools have a big potential in evaluating and improving user experience of products. The study covers valuable works done with the tools being used as design component as well as evaluation method tools. A compiled review of studies distributed over wide range of application areas and use contexts, all of which have commonalities in what they measure and want to achieve, was needed. In all use contexts and different application areas, the bottom line is supplying users with systems and services that elicit an improved and pleasant user experience.

Facial EMG is proved to measure activities that represent the valence of emotion. Various aspects such as task difficulty level, perception, and focus were measured in studies reviewed. Signal and expression interpretation was seen to be context dependent in itself, and caution needs to be taken when using such tools. In some cases, the facial experiences did not index the true inner emotional states due to display rules and context of use. However, those false positives may also have value in studies of social behavior regarding how people are influenced by display rules. Such knowledge can also assist improve the emotional intelligence of machines. In the context of interaction mechanism, other versions of EMG such as surface EMG were used. With electrodes attached on parts of muscles that can be voluntarily activated/deactivated, recording those activities and using them as command inputs is a possibility. This context faces a challenge since there will be multiple sources of signals that can be recorded at the same time. Algorithms such as ICA, come in handy in such cases. Facial EMG could also be a design component as well, both as an interaction mechanism and as a bio feedback system. Voluntarily frowning and smiling as an interaction mechanism could be uncomfortable in a way. Biofeedback mechanism to design systems that react to user's emotion is, potentially, a solid prospect for facial EMG is UX study.

EEG had wide range of application both as affect evaluation and as an interaction mechanism. The capability of directly monitoring brain activities supplies researchers and other industries with wide range of application potentials. As is the case with other PM tools, EEG's implementation was limited by its price, complexity, format, and lack of expertise. As an interaction device EEG has been experimented with, in wide range of application area such as IOT, gaming, and health. From the user experience evaluation perspective, it has been used to measure different aspects of emotion and usability constructs such as learnability, subjectivity, difficulty levels, attention, workload, stress, frustration, excitement, boredom, fun and engagement. However, EEG faces issues of price, complexity, expertise, and device format. Some of the issues are gradually being addressed indicating a continuous development towards a more common applicability.

The implications of this study are mainly on the user experience and human computer interaction research domain. Since the area of study is relatively new, a research that presents a concise description of the work done thus far, making classifications and analysis is vital. Results of this study supply researchers with a compiled review of what work has been done and categorizes them in clear and simple manner. Researchers and practitioners can get a clear overview on the possibilities, strengths and limitations of using EEG and EMG psychophysiological measurement in UX research. Industries in the gaming, medical, and software development world can benefit from understanding the way PM tools are manifesting themselves in user experience and usability studies. User experience and HCI are multidisciplinary studies and everyone that is engaged in product development of any kind would benefit from such studies.

One limitation could be the extent to which the reviewed studies were critically analyzed. However, the main goal of this research was not a critical analysis of existing research, instead, we intended to give an insight of what is being done in the study of psychophysiology as a component of the user experience domain. Weaknesses of the studies that might bring differences in results and implications are of course catered for. Questionable results and approaches are either discarded or included with the concern put in perspective. We believe that this study managed to review and assess the phenomenon, which is psychophysiological measurements in UX studies. The study can supply readers with a concise knowledge of what is meant by EEG and EMG in the context of UX studies, what has been done in this regards and what could be the possibilities in the future. It could be a starting point for those interested in going into detailed studies of either tools in UX context. Researchers can continue on this study and study EEG and EMG in more details. Experiments could be conducted in combination of the two tools and others, in order to get a more reliable data. Through further literature reviews, narrative and SLR, further classifications and analysis of EEG and EMG can be made. Such advancements are necessary in order for the field to mature and achieve the realization of emotionally intelligent interactive systems.

References

Al-Husain, L., Kanjo, E., & Chamberlain, A. (2013). Sense of space: mapping physiological emotion response in urban space. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication* (pp. 1321-1324). ACM.

Anttonen, J., & Surakka, V. (2005). Emotions and heart rate while sitting on a chair. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 491-499). ACM.

Boell, S. K., & Cecez-Kecmanovic, D. (2015). On being 'systematic'in literature reviews in IS. *Journal of Information Technology*, *30*(2), 161-173.

Boivie, I., Gulliksen, J., & Göransson, B. (2006). The lonesome cowboy: A study of the usability designer role in systems development. *Interacting with computers*, *18*(4), 601-634.

Branco, P., Firth, P., Encarnação, L. M., & Bonato, P. (2005. Faces of emotion in human-computer interaction. In *CHI'05 Extended Abstracts on Human factors in computing systems* (pp. 1236-1239). ACM.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2007). Psychophysiological science: Interdisciplinary approaches to classic questions about the mind. *Handbook of psychophysiology*, 1-16.

Cernea, D., Olech, P. S., Ebert, A., & Kerren, A. (2012). Measuring subjectivity. *KI-Künstliche Intelligenz*, *26*(2), 177-182.

Costanza, E., Inverso, S. A., & Allen, R. (2005, April). Toward subtle intimate interfaces for mobile devices using an EMG controller. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 481-489). ACM.

Culpepper, B. J., & Keller, R. M. (2003). Enabling computer decisions based on EEG input. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *11*(4), 354-360.

Cunningham, W. A., Dunfield, K. A., & Stillman, P. E. (2013). Emotional states from affective dynamics. *Emotion Review*, *5*(4), 344-355.

Dalgleish, T. (2004). The emotional brain. *Nature Reviews Neuroscience*, *5*(7), 583-589.

Darwin, C. (1965). *The expression of the emotions in man and animals* (Vol. 526). University of Chicago press.

De Smedt, T., & Menschaert, L. (2012). VALENCE: affective visualization using EEG. *Digital Creativity*, *23*(3-4), 272-277.

Dysinger, W. S., & Ruckmick, C. A. (1933). *The emotional responses of children to the motion picture situation* (Vol. 3). Macmillan.

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic? *Emotion Review*, *3*(4), 364-370.

Frey, J., Mühl, C., Lotte, F., & Hachet, M. (2013). Review of the use of electroencephalography as an evaluation method for human-computer interaction. *arXiv preprint arXiv:1311.2222*.

Ganglbauer, E., Schrammel, J., Deutsch, S., & Tscheligi, M. (2009, August). Applying psychophysiological methods for measuring user experience: possibilities, challenges and feasibility. In *Workshop on user experience evaluation methods in product development.*

Gruber J., Yale University (2013 Producer/online lecture). Human Emotions, available from https://www.youtube.com/watch?v=slwdzpg4Aqk&index=3&list=PLh9mgdi4rNewieO9Dsj-OhNBC9bF4FoRp

Gyselinckx, B., Van Hoof, C., Ryckaert, J., Yazicioglu, R. F., Fiorini, P., & Leonov, V. (2005, September). Human++: autonomous wireless sensors for body area networks. In *Custom Integrated Circuits Conference, 2005. Proceedings of the IEEE 2005* (pp. 13-19). IEEE.

Gupta, R., Khomami Abadi, M., Cárdenes Cabré, J. A., Morreale, F., Falk, T. H., & Sebs, N. (2016, June). A Quality Adaptive Multimodal Affect Recognition System for User-Centric Multimedia Indexing. In *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval* (pp. 317-320). ACM.

Hassenzahl, M., & Tractinsky, N. (2006). User experience-a research agenda. *Behaviour & information technology*, *25*(2), 91-97.

Hazlett, R. (2003, April). Measurement of user frustration: a biologic approach. In *CHI'03 extended abstracts on Human factors in computing systems* (pp. 734-735). ACM.

Hedberg, H., Iivari, N., Rajanen, M., & Harjumaa, L. (2007, May). Assuring quality and usability in open source software development. In *Emerging Trends in FLOSS Research and Development, 2007. FLOSS'07. First International Workshop on* (pp. 2-2). IEEE.

He, T., Bi, L., Lian, J., & Sun, H. (2016, June). A brain signals-based interface between drivers and in-vehicle devices. In *Intelligent Vehicles Symposium (IV), 2016 IEEE* (pp. 1333-1337). IEEE.

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International journal of human-computer studies*, *64*(2), 79-102.

Iivari, N., Hedberg, H., & Kirves, T. (2008, September). Usability in company open source software context-Initial findings from an empirical case study. In *IFIP International Conference on Open Source Systems* (pp. 359-365). Springer US.

ISO 9241-11 (1998). Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability. Geneva, CH: International Standard Organization.

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

Kaneko, N. (2008, April). Detecting the direction of listening with the EMG signals measured behind ears. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 535-538). ACM.

Kim Ann Zimmermann, Live Science Contributor | March 11, 2016 06:05pm ET. Nervous System: Facts, Function & Diseases. Accessed on 18th October 2016. URL: http://www.livescience.com/22665-nervous-system.html

Kivikangas, J. Matias. (2015). *Emotion and Social Context in Digital Game Experience* (Doctoral dissertation). University of Helsinki database, HELDA. (URN: ISBN: 978-951-51-1818-9).

Kivikangas, J. M., Chanel, G., Cowley, B., Ekman, I., Salminen, M., Järvelä, S., & Ravaja, N. (2011). A review of the use of psychophysiological methods in game research. *Journal of gaming & virtual worlds*, *3*(3), 181-199.

Konstantinidis, E., Conci, N., Bamparopoulos, G., Sidiropoulos, E., De Natale, F., & Bamidis, P. (2015, December). Introducing Neuroberry, a platform for pervasive EEG signaling in the IoT domain. In *Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare* (pp. 166-169). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1998). Emotion and motivation: measuring affective perception. *Journal of Clinical Neurophysiology*, *15*(5), 397-408.

Lang, P. J. (2014). Emotion's response patterns: The brain and the autonomic nervous system. *Emotion Review*, *6*(2), 93-99.

Laufer, L., & Németh, B. (2007, August). Your skin knows when you will jump. In *Proceedings of the 14th European conference on Cognitive ergonomics: invent! Explore!* (pp. 277-280). ACM.

Lottridge, D. (2008, April). Emotional response as a measure of human performance. In *CHI'08 Extended Abstracts on Human Factors in Computing Systems* (pp. 2617-2620). ACM.

MacLure*, M. (2005). 'Clarity bordering on stupidity': where's the quality in systematic review? *Journal of Education Policy*, *20*(4), 393–416.

Mandryk, R. L., Atkins, M. S., & Inkpen, K. M. (2006). A continuous and objective evaluation of emotional experience with interactive play environments. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 1027-1036). ACM.

Miller, N. E. (1959). Liberalization of basic SR concepts: Extensions to conflict behavior, motivation and social learning. *Psychology: A study of a science, Study*, *1*, 196-292.

Naik, G. R., Kumar, D. K., Singh, V. P., & Palaniswami, M. (2006). Hand gestures for HCI using ICA of EMG. In *Proceedings of the HCSNet workshop on Use of vision in human-computer interaction-Volume 56* (pp. 67-72). Australian Computer Society, Inc.

Nielsen, J. (1993) Usability Engineering, Morgan Kaufmann Publishers Inc.

Park, B. (2009, July). Psychophysiology as a tool for HCI research: promises and pitfalls. In *International Conference on Human-Computer Interaction* (pp. 141-148). Springer Berlin Heidelberg.

Partala, T., Surakka, V., & Vanhala, T. (2005). Person-independent estimation of emotional experiences from facial expressions. In *Proceedings of the 10th international conference on Intelligent user interfaces* (pp. 246-248). ACM.

Pentus, K., Mehine, T., & Kuusik, A. (2014). Considering Emotions in Product Package Design through Combining Conjoint Analysis with Psycho Physiological Measurements. *Procedia-Social and Behavioral Sciences*, *148*, 280-290.

Potter, R. F., & Bolls, P. (2012). *Psychophysiological measurement and meaning: Cognitive and emotional processing of media*. Routledge.

Rajanen, D., Salminen, M., & Ravaja, N. (2015). Psychophysiological responses to digital media: frontal EEG alpha asymmetry during newspaper reading on a tablet versus print. In *Proceedings of the 19th International Academic Mindtrek Conference* (pp. 155-162). ACM.

Rajanen, D., Salminen, M., & Ravaja, N. (2016). Reading a Newspaper on Print versus Screen: A Motivational Perspective. In *System Sciences (HICSS), 2016 49th Hawaii International Conference on* (pp. 630-637). IEEE.

Ravaja, N. (2004). Contributions of psychophysiology to media research: Review and recommendations. *Media Psychology*, *6*(2), 193-235.

Ravaja, N., Cowley, B., & Torniainen, J. (2016). A short review and primer on electromyography in human computer interaction applications. *arXiv preprint arXiv:1608.08041*.

Saari, T., Ravaja, N., Laarni, J., Turpeinen, M., & Kallinen, K. (2004). Psychologically targeted persuasive advertising and product information in e-commerce. In *Proceedings of the 6th international conference on Electronic commerce* (pp. 245-254). ACM.

Schnittker, R., Schmettow, M., Verhoeven, F., & Schraagen, J. M. C. (2016). Combining situated Cognitive Engineering with a novel testing method in a case study comparing two infusion pump interfaces. *Applied ergonomics*, *55*, 16-26.

Styliadis, C., Konstantinidis, E., Billis, A., & Bamidis, P. (2014, May). Employing affection in elderly healthcare serious games interventions. In *Proceedings of the 7th international conference on PErvasive technologies related to assistive environments* (p. 32). ACM.

Skroumpelou, K., Mavros, P., & Smith, A. H. (2015). Are we there yet? Exploring distance perception in urban environments with mobile Electroencephalography. In GIS Research UK 2015.

Stickel, C., Fink, J., & Holzinger, A. (2007). Enhancing universal access–EEG based learnability assessment. *Universal Access in Human-Computer Interaction. Applications and Services*, 813-822.

Tan Le, (2014, April 14). A headset that reads your brainwaves TEDTalks [Video file]. YouTube, retrieved from: https://youtu.be/fs2GDSYYCoA

Wang, S., Gwizdka, J., & Chaovalitwongse, W. A. (2016). Using Wireless EEG Signals to Assess Memory Workload in the $ n $-Back Task. *IEEE Transactions on Human-Machine Systems*, *46*(3), 424-435.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly*, xiii-xxiii.

Zhao, L., & Deek, F. P. (2006, April). Exploratory inspection: a learning model for improving open source software usability. In *CHI'06 extended abstracts on Human factors in computing systems* (pp. 1589-1594). ACM.