COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

APPLICATION OF COGNITIVE SCIENCE ON IMPROVING THE PROCESS OF USER ONBOARDING INTO A NEW SOFTWARE INTERFACE

Master's Thesis

2019

Bc. Márius RAK

COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

APPLICATION OF COGNITIVE SCIENCE ON IMPROVING THE PROCESS OF USER ONBOARDING INTO A NEW SOFTWARE INTERFACE

Master's Thesis

Study programme:	Cognitive Science
Field of study:	2503 Cognitive Science
Supervising department:	Department of Applied Informatics
Supervisor:	RNDr. Barbora Cimrová, PhD.

Bratislava 2019 **Bc. Márius RAK**

35442487

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:	Bc. Márius Rak
Study programme:	Cognitive Science (Single degree study, master II. deg., full time form)
Field of Study:	Cognitive Science
Type of Thesis:	Diploma Thesis
Language of Thesis:	English
Secondary language:	Slovak

Title: Application of cognitive science on improving the process of user onboarding into a new software interface

Annotation: For usefulness of a software, it is essential that the software is understandable to its user. However, it is not always possible to design a simple software. Therefore, it is often necessary to explain the usage of new tools to the user. This phase is in the commercial field known as onboarding. There are various techniques used during onboarding, however, verification of their effectiveness is often difficult. The aim of this thesis is to design a method for verifying the efficiency of the onboarding process using current technologies in cognitive science. The combination of event-related potentials (ERP) and eye-tracking seems promising.

Aim: 1. Application of user psychology and cognitive science on improving the user experience during human-computer interaction.
2. Design of an experiment and a method for examining the learnability of software usage.

Literature:
Marangunić, N., & Granić, A. (2015). Technology acceptance model: a literature review from 1986 to 2013. Universal Access in the Information Society, 14(1), 81-95.
Léger, P. M., Sénecal, S., Courtemanche, F., de Guinea, A. O., Titah, R., Fredette, M., & Labonte-LeMoyne, É. (2014). Precision is in the Eye of the Beholder: Application of Eye Fixation-Related Potentials to Information Systems Research. J. AIS, 15(10), 3.

Supervisor:	RNDr. Barbora Cimrová, PhD.
Department:	FMFI.KAI - Department of Applied Informatics
Head of department:	prof. Ing. Igor Farkaš, Dr.
Assigned:	25.09.2018
Approved:	28.11.2018

prof. Ing. Igor Farkaš, Dr.
Guarantor of Study Programme

.... Student

.... Supervisor

35442487

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta:	Bc. Márius Rak
Študijný program:	kognitívna veda (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor:	kognitívna veda
Typ záverečnej práce:	diplomová
Jazyk záverečnej práce:	anglický
Sekundárny jazyk:	slovenský

Názov: Application of cognitive science on improving the process of user onboarding into a new software interface
Aplikácia poznatkov kognitívnej vedy pri zlepšovaní procesu zoznamovania sa s používateľským rozhraním nového softvéru

Anotácia: Ak má byť softvér ako nástroj skutočne užitočný pre používateľa, je nutné aby bol softvér pochopiteľný. Avšak nie vždy je možné navrhnuť softvér jednoducho, preto je často potrebné nového používateľa oboznámiť s ovládaním nástroja. Táto fáza je v komerčnej oblasti označovaná ako onboarding. Existujú rôzne techniky používané počas onboardingu, no overenie, ktorá je najefektívnejšia je často náročné. Cieľom práce je navrhnuť metódu overenia efektivity procesu onboardingu pomocou aktuálnych technológií používaných v kognitívnej vede. Vhodným nástrojom sa javí kombinácia evokovaných mozgových potenciálov (ERP) a sledovania pohybu očí (eye-tracking).

Cieľ: 1. Aplikácia poznatkov používateľskej psychológie a kognitívnej vedy na vylepšenie používateľského zážitku počas interakcie človeka s počítačom.
2. Návrh experimentu a metódy na skúmanie naučiteľnosti používania softvéru.

Literatúra:
Marangunić, N., & Granić, A. (2015). Technology acceptance model: a literature review from 1986 to 2013. Universal Access in the Information Society, 14(1), 81-95.
Léger, P. M., Sénecal, S., Courtemanche, F., de Guinea, A. O., Titah, R., Fredette, M., & Labonte-LeMoyne, É. (2014). Precision is in the Eye of the Beholder: Application of Eye Fixation-Related Potentials to Information Systems Research. J. AIS, 15(10), 3.

Vedúci:	RNDr. Barbora Cimrová, PhD.
Katedra:	FMFI.KAI - Katedra aplikovanej informatiky
Vedúci katedry:	prof. Ing. Igor Farkaš, Dr.
Dátum zadania:	25.09.2018

Dátum schválenia: 28.11.2018

prof. Ing. Igor Farkaš, Dr.
garant študijného programu

.... študent

.... vedúci práce

Acknowledgements

I am thankful to my supervisor, RNDr. Barbora Cimrová, PhD. for her patience and providing directions. Many thanks also belong to my family and friends for their support.

I hereby declare that the work presented in this thesis is original and the result of my own investigations. Formulations and ideas taken from other sources are cited as such.

Abstrakt

Používateľove začiatky s novým softvérom môžu byť náročné. Používateľ môže počas procesu premeny z nového na skúseného používateľa potrebovať pomoc. Táto fáza a pomoc pri nej sa nazýva onboarding. Jednoduchosť prechodu a efektivita onboarding-u môžu rozhodnúť o úspešnom osvojení si softvéru používateľom. Pri skúmaní interakcie človeka s počítačom (HCI, z angl. Human-Computer Interaction) sa aktuálne používajú techniky, ktoré by sa zrejme dali využiť pri vyhodnocovaní úspešnosti onboardingov. Avšak, tieto techniky majú určité nevýhody, akými sú vysoká subjektivita, nízka ekologická validita a vysoké časové nároky. V tejto práci navrhujeme dva nové postupy, ktoré využívajú elektroencefalografiu (EEG), sledovanie zraku používateľa (angl. eye-tracking) a ich kombináciu. Založené sú na sledovaní oblastí záujmu (AOI, z angl. Area of Interest), potenciáloch viazaných na udalosť (ERP, z angl. Event-Related Potential) a ich fúzií – potenciáloch viazaných na fixáciu pohľadu (EFRP, z angl. eye-fixation related potential).

Kľúčové slová: Onboarding, Prijatie Softvéru, Naučiteľnosť, Eye-Fixaation Related Potential

Abstract

User's beginnings with software tools can be difficult. A user might need help during a transition phase from novice to advanced user. This process and accompanying help are known as onboarding. Ease of the transition phase and the effectivity of the onboarding can influence the acceptance of the software. Field of human-computer interaction (HCI) currently uses techniques that could be applied for evaluation of onboardings. However, these techniques have drawbacks, such as high subjectivity, lower ecological validity and higher time demands. Two new methods employing eye-tracking, electroencephalography (EEG) and their combination are proposed. The methods are based on areas of interest (AOI), event-related potentials (ERP) and their fusion eye-fixation related potential (EFRP).

Keywords: Onboarding, Software Acceptance, Learnability, Eye-Fixation Related Potential

Contents

1 Introduction

Computer tools help with various tasks and make life easier. When tools are used with ease and pleasure, they fulfill their purpose and are used often. However, beginnings with such tools can be difficult, and a tool can be rejected. Or, if the user understands everything, he uses the software happily and with ease. New user needs help in the initial phases of usage. These phases and provided help are called *onboarding* or more specifically *user onboarding*.

This is valid for a whole range of computer tools from giant enterprise level applications designated for years of use, to standard web pages, where people spend a few minutes. But what makes the difference? How is an interface perceived and understood? What causes acceptance or rejection by users? Answers can be found in various disciplines that are tied to cognitive science. Psychology, with sub-disciplines of cognitive psychology and user psychology, neurophysiology, design and art can provide valuable insights. Knowledge from the disciplines is used by professionals of field of human-computer interaction (HCI).

Aim of the thesis is to describe principles that answers these questions. The answers would be useful during design of such experiences. The thesis explores current knowledge and methodologies applied in HCI, relevant to onboarding. Limitations of current methods are addressed. To overcome the limitations. technological possibilities of eye-tracking and electroencephalography, and their combination, can be useful to enrich a portfolio of the methods. Research which could support plausibility of ideas for a new method is presented.

In a second section, the thesis offers two complementary methods, that can help software designers with designing onboardings and provide higher objectivity compared to currently used methods. The method is based on finding elements of graphical user interface that new users have trouble to understand. This is further extended to method on comparing “onboardings”.

2 Software Onboarding

2.1 Software Acceptability & Acceptance

Alexandre et al. (2018) begin their review by asking "why some tools are adopted or reject by target user" and describe terminology that can be used to explain it. Two phenomena related to user's encounter of new software are recognized. An a priori *acceptability* and a posteriori *acceptance*.

When a user encounters a new tool, he forms some mental representation of it, which might be more or less positive. This representation determines the acceptability, which is a willingness and an intent to use the tool. However, the representation of a tool cannot be yet completely formed. During software usage, the representation changes based on the user's real activity. After the representation is formed, user pragmatically evaluates the software usage and construct the acceptance.

Understanding of described processes has been studied from various points of view. Alexandre et al. (2018) review five complex approaches and some not categorized factors influencing the processes.

Tool-use Centered. Approach based on *ergonomics* or *tool-use centered* approach is focused mostly on acceptability (a priori). Alexandre et al. (2018) mention literature concerning this approach from years 1992 - 2011.

The idea is, that two main criteria - usability and accessibility determine the acceptability. Usability means, that a tool is easy and effective to use, and suitable support to complete a task.

The two criteria of usability and accessibility within the tool-used centered approach consider concepts as efficiency, effectiveness, satisfaction, ease and reliability, ease of learning, learnability, flexibility, attitude, maniability (easy to manipulate), utilisability (achieve a goal at lower cost), operability (ability to keep a tool in a reliable functioning condition), accessibility (tool is usable for everyone). These concepts were later used for the construction of ISO standard. A methodology for the evaluation of tools based on interviews, user tests, user performance and observational data was also proposed within this approach.

Social Psychology. In 1975, an approach based on *social psychology* began to emerge, with probably most recent article from 2011. The core idea is, that user is a social agent and this determines the use of a tool by the user. The approach applied theories from social psychology to human-computer interaction. The theories are *Theory of Reasoned Action (TRA)* and *Theory of Planned Behaviour (TBA)*.

The approach considers mostly ideas of motivation (engagement in tool), effective use - performance (task completion success), attitude (belief towards the tool), behavioural control (auto-evaluation of skills and resources in regard to the use of a tool), subjective norms (social pressure to use a tool or not), relative advantage, complexity, image, visibility, compatibility.

Finally, the approach recognizes *self-efficacy* as probably the most important concept. The concepts involve overall personal judgment of competency and skillfulness to use a system, belief of success after usage of a system, expectations of results, motivation and belief in fulfillment, subjective norm. All of this leads to more engagement and less auto-depreciation, which in turn directly impacting acceptability.

This approach also yields a statement that “a piece of technology must (a) have an advantage relative to traditional strategies, (b) be compatible with users’ values, experiences and needs of individuals and groups, (c) be easy to use and (d) have quick and visible positive benefits” (Alexandre et al., 2018).

Productivity-oriented. The TRA and TBP served as the basis for a proposal of the theory of *Technology Acceptance Model (TAM)* from 1989. This theory takes the productivity-oriented approach and attempts to predict the behavior of a user and acceptability of a tool by a user. Main criteria in the theory are:

1. *Perceived Usefulness (PU)* - user's belief that using a system will enhance job performance, later separated to job effectiveness, productivity and time saving, and importance to a job,

2. *Perceived Ease Of Use (PEOU)* - user's belief that using a system would be free of effort, later as physical effort, mental stress and learning difficulty

An extension to the theory was *TAM2*. This extension added social influence (subjective norm, social pressure, image, voluntarism), pertinence of the task, quality of the output, demonstrability of results, experience, attitude, behavioral intention, actual use and external variables. The TAM2 promises explanation of 60% of the variance of use.

Combining eight theories, *Unified Theory of Acceptance and Use of Technology (UTAUT)* from 2003 promise explanation of 70% of the variance of acceptability. The explanation is based on the criteria of performance expectancy, effort expectancy, social influence and facilitating conditions. Some moderators as demography or experience are also included in the explanation.

Final and a simple theory with this approach claims, that user's evaluation is based on a comparison of costs to benefits, resp. gain vs. effort. These are assessed by criteria of quality of the result and the effort (cognitive or physical cost). Probably last publication related to this approach is from 2015.

Hedonic. Satisfaction as the basis of acceptance was addressed back in 1963. The shift of paradigm distinguishes utilitarian - productivity-oriented approach and hedonic - pleasure-oriented approach. The hedonic approach, with recent literature from 2014, recognizes pleasure of use as a determinant of productivity. Satisfaction can be a result of use.

The pleasure and satisfaction are influenced by perceived enjoyment, perceived ease-of-use, system and information quality and confirmation of expectations. With the hedonic approach, the state of a "flow" is also related, which is described as an extreme concentration and involvement in a task, that also raises user's satisfaction.

User-experience. Last and most recent (with literature beginning in 2001) approach focuses on *user experience* in use-related situations and context as a holistic unit. This experience determines software acceptance. The paradigm, which is often abbreviated as *UX*, combines all previously commented approaches.

Most prominent idea is that a tool is always considered in real context and implementation. The importance of context is recognized largely by psychology. UX approach considers dynamics of interaction and also possible influences of a tool on a user. It recognizes distinction in designer's perspective and user's perspective and their influence on experience and acceptance.

Next criteria, pragmatic but also holistic, are usefulness, usability, valuability, desirability, findability, credibility, accessibility, efficiency, security, evocation, stimulation, identification, perception of quality and usefulness, emotional affect, perceived enjoyment productivity-oriented, intrinsic motivation, aesthetics, interface design, disorientation, adequacy.

Alexandre et al. (2018) combined all of these approaches, they broke them down into pieces in form of individual considered concepts, ideas and phenomena and categorized them under five meta-criteria. These are (with examples in parentheses):

1. utility (gain, instrumental qualities, needs, performance, self-efficacy, us-

ability)

2. ease of use (cognitive load, complexity, cost, self-efficacy, learnability, usability)

3. aesthetics (image, visibility),

4. contextual and social differences (attitude, beliefs and desires, motivation, self-efficacy, social acceptability, values) and

5. overall judgment (affect, beliefs, hedonic attributes, influence of the support, positive experience, trust)

Alexandre et al. (2018) conclude their review with a statement that "the subject is a psychological socially situated actor; using the tool in a given context with a specific task". They recognizes the importance of the influence of an environment and an experience (judgment, emotions, attitude, motivation, etc.) and also the dynamic and temporary character of acceptance.

The presented approaches demonstrate how difficult it is to understand and describe qualities of a tool needed to its acceptance. Variability in approaches and fields of studies involved in describing the problem shows how a truly interdisciplinary look is required when it comes to software acceptability.

2.2 Learnability

Some of the factors determining acceptation can and should be enhanced by the specific design of the tool. However, in an effort to enhance all factors even for the new users (which consider the acceptation), overall performance or qualities of the product could be impaired (Scarr et al., 2011). Also, it is sometimes not possible to redesign the whole product, therefore only the transition from new user to active user can be addressed (Grossman et al., 2009).

Factors determining acceptation are often tied to expectations, ability to achieve performance (self-efficacy), comparison of costs and benefits, and competency, these factors can and should be addressed in the transition phase. Besides the fact, that aspects related to learning were mentioned in multiple views of acceptance and acceptability, without a doubt learnability is an important factor of usability and allowing for better learnability would help in the transition phase (Alexandre et al., 2018; Grossman et al., 2009; Scarr et al., 2011).

2.2.1 Enhancing Learnability

One of the most commonly known attempt to enhance learnability of a software which has failed was "Clippy" by Microsoft (Grossman et al., 2009). But numerous other techniques can be employed to help novice users with the transition on expertise. These are sometimes referred to as *Onboarding* (Cascaes Cardoso et al., 2017; Strahm et al., 2018), mostly in commercial sphere (YesElf, 2017) and wide on-line community (Hulick, n.d.(b)).

There is no universal technique or approach with user onboarding that would fit all context, users and applications. Onboarding is always unique to a particular product and user base. What is possible, is to provide examples of techniques. Many of them are same or closely related one to another, but currently, no specification or standardization is agreed upon, so various labels are used. Some of the techniques are explained. (Hulick, n.d.(a))

Welcome Windows. One time welcome greeting screen, displayed after the first launch of an application should contain more than an ordinary greeting. The screens provide greater potential to guide a user to take a certain action. This technique is useful when some input data is required. To support the perceived importance of the step, the interface should display a minimum of visual elements and no possibly misleading content should be included. (YesElf, 2017)

Empty or Blank States. When no data is stored in a system yet, many screens or parts of screens (like so-called widgets) would be blank. This might cause negative feelings with users such as confusion. Instead of negatively informing about emptiness, these places are an important opportunity to communicate to a user what can be expected on the screen later or provide some explanation on how to input data into the system. They are also the right place to inform a user about advantages of a platform and how is currently viewed section relevant to the advantages. The blank states should be precedently used to empower a user. (Renz et al., 2014; Hulick, n.d.(a); Portman, 2017)

Hotspots, Notifications and Bubble Tips. Animated - pulsating visual elements placed next to a certain key feature of a user interface are useful for pointing the feature to the user without a noticeable warning can be considered a hotspot. After clicking the hotspot, a brief comment explaining details of feature can be shown. (YesElf, 2017)

Hulick states that they hijack user's attention and pull away from the natural sense of self-direction. However, they should be implemented in a way, that they are not noticeable on first glance so they do not provide information instantly and massively, but rather nonviolently where a user can ignore the element. This technique allows a user to learn the system while using it. (YesElf, 2017)

Bubble tips also draw users' attention, same as hotspots, with which they should be combined. They can serve to display and notify about information in a more visible and alerting way. The content of the tips should be short and concise text. Buttons might be sometimes useful. (YesElf, 2017)

Gamification, Progression Systems, Progress Bars. Gamification is the use of elements known from video games outside of gaming scope, like delivering positive reinforcement. These elements serve to improve UX and user engagement. They work because human is a competing creature. The elements can be scores, leaderboards, badges or unlocking of items. Renz et al. (2014) declare that within gamification belongs also an increase of system complexity. (Renz et al., 2014; YesElf, 2017; Portman, 2017)

Mostly progress bars and badges are considered as gamification. They display current completion state and can serve as a form of reward for a user. Progression systems lead user on his progress from novice to advanced user and this progress can be in some cases longer. Progression systems are not always progress bars - completion meters, they can also be to-do lists. They communicate already reached progress within the system and can be used also for partial tasks. (Carlèn, 2017; Morrison, n.d.; Hulick, n.d.(a); YesElf, 2017; Renz et al., 2014)

Demo Video, Tutorial. Demo video courses expect user's interest in platform functionality and possibilities. In animated fashion, they can provide information on how to navigate in a system or progress within. Renz et al. (2014) mention in their article that 24,3% of users actually finished and understood demo.

Segmented Guided Tours, Walkthroughs. Tutorials integrated into standard workflow, based on current and previous actions, might be a less distracting variant of demos, since they would appear only when relevant. They can incorporate other techniques like notifications and bubble tips, or overlays to highlight what is important and blur-out what is not. Their purpose should be to introduce key features and lead a user towards his first reached important goal. Although, these techniques pull user away from his momentum and disturb the sense of self-direction. Therefore a designer should ensure that the workflow is linear and transparent. (Renz et al., 2014; Carlèn, 2017)

Content as Tutorial. Using platform and its systems to serve content in a way, that content is a tutorial can be a neat and unobtrusive method of onboarding.

For example, in case of massive open on-line courses, Renz et al. (2014) suggest using course itself to teach how to use a platform, or just mention some features in the course on a different topic. Generally, the tutorial can fill blank space, which can be used to educate. (Renz et al., 2014; Portman, 2017)

2.2.2 Evaluating Learnability

In 2009 Grossman et al. researched learnability with findings, that the term ”learnability” is not exactly defined. Based on various published definitions, they identified *Initial Learnability* for single usage and *Extended Learnability* for a change of performance over time of consequent usage. Four criteria on user assessment are recognized:

- Level of experience with computers
- Level of experience with an interface
- Quality of domain knowledge
- Experience with similar software

Further, seven categories of learnability metrics were constructed:

- Task Metrics: Metrics based on task performance
- Command Metrics: Metrics based on command usage
- Mental Metrics: Metrics based on cognitive processes
- Subjective Metrics: Metrics based on user feedback
- Documentation Metrics: Metrics based on documentation usage
- Usability Metrics: Metrics based on a change in usability
- Rule Metrics: Metrics based on specific rules

These metrics are close to metrics of usability and are more or less just adapted for learnability construct. Therefore methodologies used for assessing usability can be used.

2.3 Human-Computer Interaction

There is a whole field of expertise formed around designing software application products, that would provide high usability. The field is called Human-Computer Interaction (HCI). As Hassenzahl et al. (2006) state, "HCI research focused almost exclusively on the achievement of behavioral goals in work settings. The task became the pivotal point of user-centered analysis and evaluation techniques (e.g. usability testing).". The concern of the HCI field is mostly user interface (UI). Nowadays mostly graphical user interfaces (GUI) are used, studied and improved, but a command line interface still provide some advantages (Scarr et al., 2011).

According to ISO 9241-11:2018 *usability* is "extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (ISO, 2018). Next, *user experience* is defined by ISO 9241-210:2010 as "person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service" (ISO, 2010).

Further User Experience (UX) formed in course of HCI. Some consider UX to be alternative to HCI (Bargas-Avila et al., 2011). However, the distinction between these two is often vague and elusive (Hassenzahl et al., 2006). It can be considered to be umbrella phrase (Bargas-Avila et al., 2011) or even whole phenomenon (Hassenzahl et al., 2006). Be it as it may, there is no need for great disambiguation among HCI, usability and UX, and for purposes of this thesis, they can be considered interchangeable.

Bargas-Avila et al. (2011) explain, that UX takes a more holistic view of a product, not only e.g. task completion time or removing usability problems. UX is focused on positive emotions, situational and dynamic aspects of using an interactive product or even stimulation or personal-growth. While preventing negative emotions such as frustration and dissatisfaction was objective of HCI, the objectives of goals of UX are emotions like joy, pride or even surprise (Hassenzahl et al., 2006). In spite of all this, usability can still be the most important value of a product and Hassenzahl et al. (2006) question whether usability is related to or independent from qualities that are proposed to be examined by UX. He closes his paper considering ideas of UX to be a responsibility of HCI.

HCI professionals have a variety of methods available for achieving good UX of a product and/or its UI. Some of them, mentioned for example by Unger et al. (2009) and Rubin et al. (2008):

- **User Interviews** - A one-on-one conversation with a participant who belongs to one of the site's primary user groups. This is usually done with 5 - 6 participants, where most are of the target group, but also one or two participants are extremely outside of target group[1].
- **Contextual Inquiry (active and passive)** - An on-site visit with participants to observe and learn about how they work in their normal, everyday environment.
- **Surveys** - A series of questions consisting of mainly closed-end answers (multiple choice) used to identify patterns among a large number of people.
- **Focus Groups** - A group discussion where a moderator leads participants through questions on a specific topic. Focuses on uncovering participants' feelings, attitudes, and ideas about the topic.

[1]Based on personal discussion with UX professional

- **Card Sorting** - Participants are given items (such as topics) on cards and are asked to sort them into groups that are meaningful to them.

- **Usability Testing** - Users try to perform typical tasks on a site or application while a facilitator observes and, in some cases, asks questions to understand users' behavior. Also called Lab test (Tullis et al., 2010).

A portal Usability Body of Knowledge by UXPA - The User Experience Professionals' Association, n.d. available at http://www.usabilitybok.org provides a categorized list of many methods to be used by UX professionals. Among them, UXPA provides a category of *Usability Evaluation Methods* divided into groups of *Formative* and *Summative* methods. With the formative methods, work with prototype or product is important and required. While summative usability tests are used at the end of development. Theofanos et al. (2005) describe formative testing as "testing with representative users and representative tasks on a representative product where the testing is designed to guide the improvement of future iterations."

2.3.1 Usability Testing

Above-mentioned technique of *Usability Testing* is important for user-centered design and is one of the most frequently used methods for testing UX of designs. It is used to evaluate a product by testing it with real users, which allows designers to obtain direct feedback on how the users work with the product. Such testing is important because users often surprise designers with unexpected actions or reactions. It is helpful to see a system being used for designers to be able to create a product that works as expected by users. (Schall et al., 2014; Unger et al., 2009; Rubin et al., 2008)

According to Theofanos et al. (2005) and UXPA - The User Experience Professionals' Association, usability testing belongs to group of formative methods.

Usability Testing should not be mistakenly interchanged for *User Acceptance Testing* which would be considered summative method. Usability testing helps designers to design software to be efficient, effective, usable and pleasurable to work with. In practice, they are both similar. (Unger et al., 2009; Rubin et al., 2008)

The technique is based on standard principles of quantitative testing and should be carried out as A/B test, where at least two versions of interfaces are compared. The testing should be controlled, with specific hypotheses prepared, and sufficient sample size of testing groups. However in practice, these rules are often overlooked and not implemented, yet tests still provide high value. Purpose of a usability test is not to provide hard - number evidence, but rather to provide grounds for informed decisions about the design of the product. Therefore the usability test tends to be more qualitative and oftentimes, only one version of the design is examined. (Rubin et al., 2008)

Rubin et al. (2008) define these to be basic elements of usability testing:

- Development of research questions or test objectives rather than hypotheses.
- Use of a representative sample of end users which may or may not be randomly chosen.
- Representation of the actual work environment.
- Observation of end users who either use or review a representation of the product.
- Controlled and sometimes extensive interviewing and probing of the participants by the test moderator.
- Collection of quantitative and qualitative performance and preference measures.
- Recommendation of improvements to the design of the product.

If the purpose of a usability test is to evaluate a single interface, the evaluation is mostly qualitative and highly subjective. Participants should cover a broader range of categories of possible users. If the purpose is to compare two various interfaces of a system, it is possible to evaluate some more objective quantitative metrics. However, the characteristics of participants (demography, skills) should be similar to avoid biases in results.

The test procedure for individual participants is then same in both cases. A moderator instructs participant about the test and its objectives. A scenario of the test, or a list of tasks is provided to the participant. It is important to precisely define the task and what does it mean to successfully finish the task. Although, free exploration of an interface without aim can be sometimes also part of a test. The participant is observed and possibly recorded by moderator and designers. (Rubin et al., 2008; Unger et al., 2009; Hertzum et al., 2009; Tullis et al., 2010)

A participant should execute the tasks with the system as naturally as possible. However many tests employ *think-aloud* technique. The technique demands that user constantly comment on performed actions and verbalize his or her thoughts. If the participant happens to forget this requirement, then the moderator reminds it and asks the participant to 'keep talking'. Experimenters also often intervene with questions for an explanation of actions. This, the requirement itself and constant reminding can be, of course, disturbing and lower ecological validity. Also, the moderator himself is a risk to ecological validity. (Rubin et al., 2008; Unger et al., 2009; Hertzum et al., 2009; Tullis et al., 2010)

Hertzum et al. (2009) researched how does thinking aloud affect results of usability testing. They mention that verbalization can lead to fewer errors and more tasks solved correctly. Further, task competition time was also affected and is considered to be susceptible to individual variation caused by thinking aloud. This can also be used as an indicator of higher task complexity, although not very accurate. However, Hertzum et al. (2009) mention studies which concluded opposite, since thinking aloud seemed to speed up tasks solving. The technique also affected hand movements and marginally raised mental workload of participants. Despite a conclusion made by Hertzum et al. (2009), that correct use of think-aloud technique can provide valid results, it was demonstrated, that the technique causes fluctuations in results.

After the test is executed, it should provide material for qualitative interpretation of each participant's behavior. Multiple sessions can provide also objective metrics which are usually based on (Tullis et al., 2010; Rubin et al., 2008; Kuniavsky et al., 2012; Schall et al., 2014):

- *task success* - how many tasks of scenario was successfully finished,
- *errors count* - how many errors happened during the session,
- *time on task* - usually, the faster is task finished, the better the experience (except for applications designed for hedonic purposes such as fun or joy; Hassenzahl et al., 2006),
- *self-reported metrics* - mostly implemented by questionnaires following test session,
- *physiology* - facial expressions, skin conductance level, pupil dilation, various EEG metrics.

Goldberg et al. (1999) state, that "The software development industry requires improved methods for the *objective* analysis and design of software interfaces." They criticize current methods because of poor objectivity and reliability, and attempted to use measurements of eye-movements to objectively evaluate interfaces. Same call was stated by Grossman et al. (2009) where they researched evaluation of learnability. As one of such techniques is offered eye-tracking (Goldberg et al., 1999; Schall et al., 2014; Guo et al., 2016; Holsanova, 2014; Bargas-Avila et al., 2011).

3 Technical Possibilities

Thanks to technical advancements two important methods of examining human behavior and cognition are available. These are *eye tracking* and *electroencephalography*. They are suitable methods for application in HCI.

3.1 Eye Tracking

Eye tracking (ET) - a technique that can be a valuable addition to usability testing methods is a non-intrusive method to monitor where and when a user looks. The technology of ET is based on infrared, or near-infrared light source directed at pupil where it causes visible reflection in a cornea - the outer-most optical element of the eye. Location of this reflection changes with eye movement. These reflections and changes are recorded by high-resolution, and ideally also high-speed, camera. This way, a computer can track a user's eye gaze.

Before the actual test is required calibration of the device in use with individual participants to achieve high accuracy. From such data, an algorithm can later interpolate the actual position of user's gaze. ET can be either head-mounted device, such as glasses or screen-based system, which is called remote eye tracking. (Tullis et al., 2010; Kuniavsky et al., 2012; iMotions, 2016; Morimoto et al., 2004; Schall et al., 2014)

Looking. Human's gaze is composed of multiple elements. Those important for eye tracking, which are most often mentioned and analyzed are:

- gaze points,
- saccades,

- minute movements (microsaccades, tremors, drifts) and
- fixations.

Human vision works on a principle of detecting changes at the retina. *Gaze point* is the instantaneous position of gaze, when information is aligned at fovea and retina sense incoming light. *Fixation* is approximate location, respectively area where multiple close gaze points and minute movements occur and are also defined by the duration of an occurrence. Duration of fixation is between cca. 50 - 600ms. *Saccades* occur between fixations, they are movements of gaze between two points and average duration is 20 - 40ms. (Tobii AB, n.d.; Tobii AB, 2018; Schall et al., 2014)

Using Eye Tracking for Usability Testing.

By exactly following user's gaze it is possible to understand better how is the work of participant going. It is possible to determine whether participant really saw an element of an interface, that designer wanted the user to see (Tullis et al., 2010; Pernice et al., n.d.; Chynał et al., 2018). On this level, it is, however, not sure whether user read and understood words, or what was he/she thinking (Kuniavsky et al., 2012). Following and reacting to user gaze during the test can be distracting, therefore it is better to watch *Gaze Replay* after the test.

Experimenter can discuss user's behavior watching record of test with user in retrospectively (Kuniavsky et al., 2012; Pernice et al., n.d.). Such consultation, however, can be misleading, since user can over-rationalize or falsely justify his/her actions. Also it is really time consuming (Pernice et al., n.d.) There are further analyses possible with recorded ET data such as *gaze plot* (*scan path*), or *heatmaps* (Tullis et al., 2010; Kuniavsky et al., 2012; Pernice et al., n.d.; Mealha et al., 2010).

Areas of Interest (AOI) (sometimes called ROI - Region of Interest) is base for some more techniques. AOIs are predefined subregions of a user interface. Later measurements such as the ratio of fixations inside respective areas to other areas or to all fixation can be applied. Some more metrics are *Time to First Fixation* or *Time Spent* in AOI. (Tullis et al., 2010; iMotions, 2015; Mealha et al., 2010; Chynał et al., 2018)

Studying cognition only using eye-tracking is difficult since measured variables have large variability and recordings can lead to incorrect and subjective conclusions (Baccino, 2011). Besides that, in practice evaluating interfaces only using eye-tracking usually brings a need for replaying usability test sessions which translates into the necessity of more of valuable time.

3.2 Electroencephalography - EEG

It is believed, that electrical signals generated by brain represent its function and state. Such a possibility can provide an advance in the field of HCI and could be employed for an evaluation of user interfaces.

3.2.1 Beginnings of Electroencephalography

The belief of representation of brain function by an electrical signal motivates scientists to apply various processing methods to the signals measured at the human scalp. History of EEG can be dated back to 1300s and neurophysiology began to form back in the 19th century. The first record of brain activity from human scalp was made in 1875. This was beginning of concept of electroencephalography - EEG, which denotes *writing* of *electrical* activity of *brain in head*.

As most influencing discovery is considered a report of so-called *alpha waves* and *alpha blocking response* by Hans Berger in 1929. He later found some correlation between mental activities and the changes in the EEG signal. During the 1950s was rapidly developed technique of applying Fourier analysis to EEG signal and during 1970s progressed investigation of evoked potentials. (Sanei et al., 2007; Luck, 2005)

3.2.2 Field Potentials

The human brain is composed of nerve cells - *neurons* which enables its functioning. At birth, approximately 10^{11} neurons are developed. A neuron consists of dendrites, a cell body with a single nucleus, and an axon.

A nerve cell receives information as stimuli from other cells. A human neuron is connected approximately to 10,000 other neurons. They are connected to the dendrites or body of a cell by a synapse. This connection serves for transmitting information between neurons in form of *action potentials*. The cell integrates received stimuli and transports the information by axon to next cells.

Under the membrane of a cell body can be recorded negative electrical potential of 60-70 mV relative to outside of the membrane. This potential changes according to the activity of the cell which is related to the activity of other connected cells. An action potential transmitted by synaptic connection can either excite or inhibit the cell, depending on a type of the synapse. If several transmissions occur together in a short time, they will be summed. If a certain threshold is reached after the summation, the cell itself will produce and send a new action potential through its axon.

The action potential is physically transmitted not only inside the cell but also around it, in extracellular space. This is directly responsible for the generation of field potentials, which are further measured on the scalp as an EEG signal. (Luck, 2005; Sanei et al., 2007; Llinas, 2008; Buzsáki et al., 2012; Léger et al., 2014)

3.2.3 Source of EEG Signal

The EEG signal obtained from a scalp measures currents of secondary electrical fields produced by magnetic fields which are generated by electrical dipoles -

extracellular electrical fields in the brain. All electrical currents flowing in extracellular space of the brain contributes to an extracellular electrical field. These currents originate in cells of the brain and their parts, e.g. dendrites, axons or axon terminals and create electrical dipoles. The extracellular field is, therefore, a superposition of currents based on the activity of cells surrounding the location of the field. Since the electrical signal is transferred almost at the speed of light, recording of these signal provides almost instantaneous information about brain state, allowing for high temporal resolution.

The activity influencing fields can be synaptic activity, which must overlap in time to induce measurable change or occurrences of action potentials. To produce measurable field, approximately 10^7 neurons with geometry perpendicular to scalp has to be simultaneously active. Also, this activity can be canceled out by groups of neurons with opposite activity.

Even though the activity of all neurons is reflected in the field, pyramidal cells contribute most. It is because of their geometry, architecture and large population. Apical dendrites of pyramidal cells lie parallel to each other and afferent inputs are perpendicular to the dendritic axis. This makes cortex, containing the pyramidal cells strongest producers of the electrical field. Therefore EEG signal recorded at scalp is mainly produced by the cortex and its pyramidal cells which is also part of the brain where most deliberate cognitive processes happen. However, this signal is still attenuated by matters between the cortex and electrode, such as cerebrospinal fluid, dura matter, bone, or the scalp itself. (Luck, 2005; Sanei et al., 2007; Woodman, 2010; Buzsáki et al., 2012; Léger et al., 2014)

3.2.4 Source Localization

Measuring an EEG and determining a distribution of observed voltage with the signal originating at various locations in the brain is relatively simple. This is called *forward problem*. On the other hand, the *inverse problem* of localizing sources of measured EEG is an ill-defined problem since an infinite number of configurations of sources can result in same observed voltage distribution. This causes a low temporal resolution of an EEG. Although some mathematical models for approximate localization exist, it is currently impossible with such a complex system as the brain is.

To help localize sources of measured brain activity, usually, blind-source separation algorithms are applied. Some of the algorithms are for example MUSIC – Multiple signal classification, FINES or LORETA – low-resolution electromagnetic tomography algorithm. The two most commonly used techniques are *PCA – Principal Component Analysis* and *ICA - Independent Component Analysis*. These two algorithms use information from all electrodes together to solve the problem. They derive components that are based on functional relationships.

PCA is used to reduce dimensionality, i.e. lowering the number of variables, leaving only those with useful information. It separates possibly correlated values and produces a same or lower number of uncorrelated variables - principal components. The components are orthogonal to each other and have the largest possible variance.

ICA attempts to find statistically independent sources of combination in measured data. To perform ICA various algorithms are applied, most commonly used is infomax, but recently an *AMICA - adaptive mixture of independent component analyzers* has been introduced. (Jung et al., 2001; Luck, 2005; Sanei et al., 2007; Urigüen et al., 2015; Kaczorowska et al., 2017; Kappenman et al., 2012)

3.2.5 Recording Devices

In the beginnings of EEG recording, ingenious devices, which used for example coils and mirrors, were used to record the signal from a scalp. Currently are for the job used tools based on standard electrical components and principles.

On the scalp are attached electrodes, which are a crucial component for acquiring high-quality measurements. The electrodes are usually Ag/AgCl disks with a diameter of cca. 3 mm. A conductive gel is applied to them to form an electrical connection. Although, various types of electrodes are used, e.g. pre-gelled disposable, *dry-contact* or *no-contact* electrodes (Chi et al., 2010).

The electrodes are further connected to amplifiers. The amplified signal is transferred to a computer using an analogue-to-digital converter (ADC) with multiple channels. To avoid measuring noise produced by static electricity of a subject's body, the measured signal is a difference of voltages from electrode measuring brain activity and a reference electrode placed somewhere on head, where no brain activity can be measured, e.g. earlobe. Voltages of both of these electrodes are measured relative to a ground electrode placed somewhere on the subject's body, e.g. chest. This technique is called *common mode rejection*. To allow proper function of this, it is important to keep impedance of electrodes low, i.e. at values less ten 5 kΩ and up-to 1 kΩ between each other.

The signal is then, according to the Nyquist criterion, recorded at frequency two times of maximum target frequency and encoded with usually 12 - 16 bits. The signal is also processed with various filters, mostly, a high-pass filter of 0.5 Hz, a low-pass filter of 50 - 100 Hz and a notch filter at 50 or 60 Hz to reject system hum. (Luck, 2005; Sanei et al., 2007; Léger et al., 2014)

3.2.6 Electrodes Placement System

Internationally accepted convention of placing and naming electrodes is called *10/20 system*. Luck (2005) describes positions of sites in following steps:

> The first step in this system is to define an equator, which passes through the nasion (the depression between the eyes at the top of the nose, labeled Nz), the inion (the bump at the back of the head, labeled Iz), and the left and right pre-auricular points (depressions just anterior to the middle of the pinnae, labeled A1 and A2). A

> longitude line is then drawn between Iz and Nz, and this line is then divided into equal sections that are each 10% of the length of the line. Additional latitude lines, concentric with the equator, are then placed at these 10% points. Most of the electrode sites can then be defined as points that are some multiple of 10% or 20% along these latitude lines.

Electrode sites are named by region and distance from the midline. Regions are abbreviated as Fp - frontal pole, F - frontal, C - central, P - parietal, O - occipital, T - temporal. Sites on left hemisphere are assigned odd numbers and those on right have even numbers, those on midline are labeled with *z* instead of a number. This system is designed for 21 electrodes, if more electrodes are used, they are placed between the designated electrodes at equidistance between them. (Luck, 2005; Sanei et al., 2007)

3.2.7 Brain Rhythms

Described electrical extracellular field, which is recorded by EEG is manifested in brain rhythms, that are further researched and examined. Properties of the field as waveform amplitude and frequency depends on changes in the activity of those nearby cells. The activity results in temporally synchronous fluctuations of potentials in large neuronal aggregates.

Impact of a contribution of a cell diminishes with a distance of a cell from the field because of interference of other cells and filtering properties of the matter which serves as a medium. The faster the signal, in means of frequency, the more attenuated it gets, while a signal with a lower frequency is less attenuated. Therefore, rhythms with higher frequency are usually of lower amplitude.

The synchrony caused by network oscillations is associated with various brain states. (Sanei et al., 2007; Buzsáki et al., 2012) Sanei et al. (2007) enumerate five major ranges of frequencies of brain rhythms:

- **Delta - 0.5 – 4 Hz.** Delta waves are primarily associated with deep sleep, but also present in the waking state. They are similar to artifact signal generated by muscles, which can be easily mistakenly considered to be delta waves.

- **Theta - 4 – 7.5 Hz.** Theta waves originates from the thalamus, which is illustrated by the similarity of name and appear with an onset of drowsiness. They are also associated with creative inspiration, deep meditation or access to unconsciousness. Theta waves seem to be associated with arousal.

- **Alpha - 8 – 13 Hz.** Alpha waves are usually sinusoidally shaped, but occasionally negative component is sharp instead of rounded. Based on the shape, they can be sometimes seen also in the range of beta waves at 20 Hz. It is the most prominent rhythm and covers a great range of activities. Some of the states that elicit alpha waves are relaxed awareness without any attention or concentration or closed eyes. On the other side, alpha waves are reduced by opening the eyes, hearing unfamiliar sound, anxiety, mental concentration or attention.

- **Beta - 14 – 26 Hz.** Beta waves are usual for wakefulness and are associated for example with thinking, active attention, focus or solving of concrete problems. High-level beta waves may indicate a panic state. Beta activity mostly occurs at the frontal and central regions.

- **Gamma - 30 – 45 Hz.** Gama waves are sometimes called *upper beta*. Oscillations at gamma frequencies are of low amplitudes are rarely encountered.

3.2.8 Evoked Potentials

Even though described oscillations may reflect brain states and brain dynamics at longer timescales, most computations in the brain happen within tens to hundreds of milliseconds. These computations result in other waveforms elicited by the brain. These waveforms are transient and evoked or induced by sensory input event and happen in physiologically relevant time window after a sensory-based trigger.

Such waveform is usually regarded as an Evoked Potential (EP) or an Event-Related Potential (ERP). In these waveforms is also manifested activity of populations of neurons. Employing appropriate techniques, this can be used to get insights on processes and neuronal computations. (Sanei et al., 2007; Buzsáki et al., 2012; Woodman, 2010; Luck, 2005; Léger et al., 2014)

3.2.9 Signal Extraction

The cognitive processes manifested in ERP are not the only contributors to the measured signal. The signal of ERP itself is usually smaller than other signal consisting of random noise resulting from either brain itself or other electrical aspects, which is not time-locked, nor related to a stimulus. This causes ERP to be typically hidden and not clearly visible in a single measurement.

To improve the signal-to-noise ratio of an ERP and hence extracting informative signal from measured data, the data are usually averaged. Averaging is done over many time-locked epochs of EEG data. It is assumed that the brain response to a stimulus is the same, every time a stimulus is processed, while noise is always different. Average of the ERP of interest would be the same as ERP at every trail, while the average of noise should be a flat line with zero amplitude. Therefore averaging both ERP and noise should provide clear sight of an ERP while none of a noise. Woodman (2010) also suggest a simple trick for enhancing signal-to-noise ratio, which is to jitter recordings with a stimulus that would be averaged out.

The signal to noise ratio increases as a function of square root of the numbers of averaged epochs. This would imply a simple maximizing number of trials to achieve clear results. However, in practice, it is usually not possible or simple, due to various human limitations (such as fatigue), resources costs or character of an experiment. Therefore it is important to find a good compromise on a number of trials.

Woodman (2010) offers simple rules for participants count. For good measures of various components, he suggests following numbers of artifact-free trials per condition: 1000 for C1, because of alpha noise, 300-1000 trials for P1 and N1, 250 for N2pc, 35-60 for P3. Basically, the later the targeted component is, the fewer trials are required. If it is not possible to achieve such numbers from one participant, it should be compensated by data averaged from more participants.

On the other hand, Boudewyn et al. (2018) experimented with amounts of trials required to achieve significant results and mentions studies where conclusions were drawn from 6 - 20 trials. The study is concluded with the statement that in within-participant design it is more useful to increase the number of trials, while in between-groups it is better to increase the number of participants, to achieve greater power.

The averaging technique itself also introduces another problem. Even though it is assumed, that brain responses with the same ERP to every stimuli presentation, in reality, each response to each stimulus differs in all ERP characteristics of interest. This variability is hidden by averaging and the average only reveals a central tendency of responses. However, this tendency may still reflect changes of various factors.

To improve studying of certain characteristics of ERP and cognitive processes, averaging of ERP epochs is time-locked to various events. It can be time-locked to any externally observable event. These are usually onset of a stimulus or behavioral response. Additional averaging techniques are *woody-filter* and area-based measures. These techniques mitigate the problem with latency variability which might reduce effect size. The woody-filter technique estimates latency of component of interest based on a template of an ERP component. This estimated peak is then used as the time-locking event of the trial. Area-based measure finds half of an area under the component of interest and time-lock the trial to this point. (Luck, 2005; Woodman, 2010; Kappenman et al., 2012; Boudewyn et al., 2018; Léger et al., 2014)

3.2.10 Noise and Artifacts

With recording informative signal, noise and artifacts are also captured. These present serious problems with the analysis. Rejecting noise of static electricity of body and various frequencies was already described. Some of the most problematic artifacts are caused by eye-blinks and eye-movements. Other artifacts can result from a change in impedance or position of an electrode, incorrect setting of amplifier gain or muscles and heart activity or residual alpha activity called "alpha ringing". (Jung et al., 2001; Luck, 2005; Urigüen et al., 2015; Kaczorowska et al., 2017)

The simplest technique to deal with eye-blinks artifacts is to reject whole measurements of frontal electrodes or to set a threshold on voltage. If this threshold is exceeded, it can be assumed, that the signal contains eye-blink artifact. This brings the issue of finding the correct value for the threshold. Low values will cause rejecting correct data, while high values will accept unwanted noise. Also throwing away valuable data is oftentimes unacceptable. Improvement of the technique is to determine the value of the threshold based on the baseline voltage. It is recommended to execute artifact rejection for each subject individually and blindly, but same or similar criteria should be applied. (Jung et al., 2001;

Luck, 2005)

A more sophisticated method is to measure electrical signal below the eye (the channel is usually called VEOG). Physiology of eye determines, that in case of an eye blink, values measured on VEOG channel will be of opposite polarity than values at affected regions. (Jung et al., 2001; Luck, 2005)

Luck (2005) recommends asking participants to wear glasses instead of contact lenses which usually cause a higher amount of eye-blinks. He further suggests shorter trials sessions.

Eye-Movements. Similar to eye-blinks, eye-movements also causes artifact in signal because of dipole effects of the eye on a signal. Movement of eyes causes changes in voltage at sites that the eyes have moved toward. This issue can be attenuated by the proper set-up of electrodes near eyes (channels marked as VEOG - vertical or HEOG - horizontal, REOG -radial). Also, since eye-movements causes a shift of visual information at the retina, it induces also ERP response, which can not be corrected as an artifact. Generally, eye-movements presents quite a challenge in obtaining artifact free ERPs with a sufficient signal-to-noise ratio. (Luck, 2005; Urigüen et al., 2015)

Aforementioned ICA and PCA are oftentimes applied for artifact corrections and provide good results (Urigüen et al., 2015; Kaczorowska et al., 2017). Urigüen et al. (2015) provides extensive list of various artifact removal approaches that were available and used in 2015.

3.2.11 ERP and Cognitive Processes

After data are technically processed, trials averaged and artifacts cleared, it is possible to extract some meaning and knowledge from the experiment. Thanks to the almost instantaneous transfer of electrical signal from its origin to the scalp, it is possible to infer conclusions on stimulus processing from the data. The

ERPs allow to observe brain processes that begun before stimulus onset and that lasted after a response was executed. The processes are manifested in peaks and troughs of a signal relative to a particular event. Basically, measured fluctuations of voltage represent the treatment of a perceived stimulus by the brain, from its sensory input, through cognitive processing to behavioral response.

ERP Component. Measured ERP is further analyzed in terms of ERP components. It might be tempting to assume, that recorded voltage fluctuations reveal exact processes. But same as with spatial distributions of dipoles, which is impossible to define based on voltage distribution on the scalp, also voltage fluctuations can cover various additions and subtractions from an overall value at the time. While some consider the underlying partial voltage changes to be components (e.g. Woodman, 2010), others regard as a component the fluctuations that are often visible in the resulting graph (e.g. Luck, 2005). Based on historical convention, the voltage is plotted reversed, i.e. with negative polarity on the upper side.

Luck (2005) defines an ERP component as "Scalp-recorded neural activity that is generated in a given neuroanatomical module when a specific computational operation is performed". Usually, a component is defined primarily by its (1) polarity, (2) timing, or latency and (3) topography or general scalp distribution. These components are sensitive to various stimuli or task manipulation. The ERP components are named by convention. First, single letter denotes polarity (P or N) and then the order of latency follows. The naming of components is not linked to underlying neural activity in any way. (Luck, 2005; Woodman, 2010; Léger et al., 2014)

3.2.12 Major Visual ERP Components

During years of research, many ERP components were described. Despite equality of names, components of different modalities are usually not linked together, especially the earlier ones.Luck (2005) and Woodman (2010) provide a summary of major ERP components. Since this thesis takes interest in graphical user interfaces, attention will be paid only to relevant, i.e. visual and semantic components.

C1. First major visual ERP component is highly sensitive to physical stimulus properties. It is not labeled with N or P because its polarity varies. It is often combined with subsequent component, which makes it difficult to examine. C1 is elicited by primary visual cortex. Its onset is around 40-60 ms and peak 50-100 ms.

P1. Next component has onset on 60-90 ms and peaks at 90 - 130 ms. It is strongest at lateral occipital sites. The latency of P1 is modulated by stimulus contrast. At least 30 distinct visual areas are activated within the first 100 ms after stimuli onset. Many of them contribute to C1 and P1 measurings. Therefore P1 is also modulated by stimulus parameters but also by subject's spatial attention and state of arousal.

N1. The component consist of several subcomponents. Earliest of them have a peak at 100-150 ms at anterior sites. Next two posterior components of parietal cortex and lateral occipital cortex peaks at 150-200 ms. They are influenced by spatial attention and appears to reflect discriminative processing.

N170 and VPP. *VPP* stands for Vertex Positive Potential. This component is sensitive to stimuli containing faces. The very top site of a head is called vertex. The VPP can be noted at 150-200 ms. Faces elicit more negative potential at lateral occipital sites, especially over right hemisphere, with peak approximately at 170 ms. It is suggested, that the vertex site is just opposite side of dipole represented at the occipital sites.

P2 Next wave is distinct and localized at anterior and central scalp sites and is modulated by target stimuli, especially the infrequent ones. However, the effect is relevant only for stimuli defined by simple features. At posterior sites, it is often difficult to distinguish P2 from overlapping N1, N2 and P3 waves. The component is not well understood.

N2. The N2 is more of a "family" of components or in other words, it covers a range of components identified in same time range suitable for N2, with a peak around 225-250 ms. A basic N2 can be deflected by repetitive nontarget stimuli. Deviant stimuli elicit larger amplitude in the range. N2 components are affected in experiments with an array of identical and one deviant item. Response at bilateral anterior sites is elicited if a deviant item is present, even if it is not a target. Later, if the deviant item is a target or resembles it, two posterior responses are elicited. First is a bilateral posterior response marked as *N2b*, which is also modulated by a probability of occurrence of a target. Second is posterior contralateral response called *N2pc* which is not probability sensitive. An N2pc, but with more parietal distribution, is observed also during visual working memory tasks.

P3. Despite extensive research of time range of P3, not much is known about underlying processes. It was proposed, that P3 is somehow related to "context updating", but even this term is not sufficiently described or explained. What is known about the wave is, that its amplitude is larger when preceded by more distractors. The amplitude is also increased by a participant's effort devoted to a task, which might reflect resource allocation. On the other side, if the subject is uncertain whether a stimulus was a target, the amplitude is smaller. Luck

(2005) concludes, that P3 can be generated only after a stimulus has been categorized (in terms of an experiment). Anything that postpones categorization of stimulus delay the P3 wave. This logical conclusion has been confirmed by many studies. The P3 is also proved to be unrelated to post-categorization processes as response selection. It can be therefore used to determine whether experimental manipulation influence categorization or response selection to a stimulus by subject.

Language-Related Components and Others. Other described components, that might be interesting, are related to language, meaning and semantic processing. *N400* is a wave with negative polarity, strongest at central and parietal sites, slightly stronger over the right hemisphere. This component is sensitive to semantic expectancies. If an incongruent word appears, N400 has a larger amplitude. However, the same effect occurs with non-word stimuli that are meaningful, e.g. line-drawings.

Syntactic violations elicit a *P600* component, but also negativity at 300-500 ms at left frontal sites. Same negativity is observed and modulated by open vs. closed questions. Further *N280* at left anterior sites is elicited for function words and *N400* for content words. Woodman (2010) mentions also *SPCN* - sustained posterior contralateral negativity and *CDA* - contralateral delayed activity associated with maintenance in visual working memory.

3.3 EFRP - Eye-Fixation Related Potential

Hutzler et al. (2007) consider determining of onset of stimuli relevant to researched cognitive process, to be critical for recording electrophysiological correlates. Current practice is based on an implicit assumption that studied cognitive process cannot start sooner than relevant stimulus is presented and hence synchronization is based on an external event of presenting stimuli to a participant. Yet, this approach limits ecological validity and does not allow to study reaction to what is attended (Hutzler et al., 2007; Léger et al., 2014). On the other side, the focus of attention on display is indicated by gaze and eye movements, which can be used for more valid methods (Goldberg et al., 1999). Eye-tracking provides insights into some cognitive processes, mainly the overt ones, but certainly doesn't reveal covert neural processes (Blair et al., 2009).

Baccino (2011) imply that method of measuring ERP complements method of eye-tracking. The term *EFRP* was introduced in 1995 and represents time-locking of EEG signal, instead of stimulus onset, to the offset of an eye-saccade, which is determined using an eye-tracking. This methods overcome presented shortcomings of standard ERP technique and can be applied in a more complex visual pattern, such as a whole displayed sentence, allowing assessing exact time of starting of a cognitive processing of certain visual information. Moreover, this allows for processing of stimuli at natural pace of participant, while remaining externally observable. (Hutzler et al., 2007; Léger et al., 2014; Baccino, 2011)

The technique of EFRP is similar to ERP in the fact, that they are obtained by averaging the EEG epochs time-locked to an event. The most notable difference is, that the epochs are shorter than those in ERP (Baccino, 2011). The advantage is natural context, without artificial restrictions. This allows studying effects, which with standard ERP cannot occur, or studying fixations themselves. Studies can be aimed at word-recognition, reading, scene perception, visual search or object identification (Rämä et al., 2010; Kaunitz et al., 2014). Numerous studies are devoted to researching reading using an EFRP (e.g. Frey et al., 2013; Baccino, 2011; Sereno et al., 2003; Dimigen et al., 2011). In their study, Sereno et al. (2003) inform, that N400 component can begin even at 200 ms and that processing speed of "visual system is more rapid than traditionally assumed" and begins in frontal cortex 80 ms post-stimulus. Some early effects occur 200 ms post-stimulus, for example, word frequency is indexed by N1 component at 132 ms (Sereno et al., 2003). Dimigen et al. (2011) also comment on sooner occurrence of N400 effect, around 250 ms. (Léger et al., 2014)

Current Status of EFRP. A brief search in database SCOPUS for `"eye fixation related potential" OR "fixation event related potentials" OR "fixation related potentials"` yielded 82 results with slightly growing numbers of articles over years, which shows that there is some interest in the topic but the technique is not widely applied. Combining electrophysiology with eye-tracking is becoming less problematic with advancements of easy-to-use eye-trackers and can be done more effectively (Hutzler et al., 2007; Sereno et al., 2003). Moreover, besides widely known EEGLAB (Delorme et al., 2004) and ERPLAB (Lopez-Calderon et al., 2014), EYE-EEG toolbox (Dimigen et al., 2011), which allows combining eye-tracker and EEG data, is available which can be used for this kind of experiments.

Lambda - Prominent Component of EFRP. At approximately 80-100 ms of saccade offset, a component called *lambda response* can be observed. The component is positive deflection at occipital sites. The lambda response is thought to consist of many neural components associated with a saccade and reflects afferent input of information to the visual cortex. It is modulated by properties of visual stimulus or attention but also by saccade magnitude. However, some studies inform that lambda response is affected by top-down processes and can index visual information processing and working memory load and/or allocation of attention. It can be even significantly attenuated by auditory working memory load. The lambda response is not yet understood well-enough and requires further attention. (Baccino, 2011; Ries et al., 2018)

Validity. Hutzler et al. (2007) tested validity of EFRP by testing an *old/new effect*. They compared the standard ERP technique with displaying single words over time with displaying the whole sentence at once and monitoring fixations at particular words. ERP waveforms were calculated only after seeing the last word. Segments of EEG data from -100 ms to 600 ms relative to fixation onsets were extracted for analysis. Their finding was, that the effect was with FERP reliable at relevant scalp regions even on regions with the smallest magnitude of the effect. Therefore, Hutzler et al. (2007), based on observation of the same effect with EFRP as with ERP, concluded validity of the EFRP approach, but do not consider it to be a substitute of the ERP approach. Even more interesting result present in the research is visible in graphs in the study. It is notable, that differences in waveforms appear approximately 50 ms - 200 ms sooner with EFRP than with ERP according to scalp location, but authors do not comment on the significance of these differences.

Limitations of EFRP. Researchers mention challenges and problems arising from this approach. Baccino (2011) enumerates some of challenges besides standard EEG artifacts. The most apparent one is overlapping of cognitive processes between successive saccades. While various effects are with ERP apparent around 300-600 ms, fixation with reading lasts only for cca. 200-300 ms. There are attempts to mitigate these problems using various algorithmic techniques (e.g. ADJAR - Adjacent Response Technique) which were originally used for experiments with a short inter-stimulus interval. Also, PCA and ICA techniques are often employed. Other techniques try to overcome the two-way data approach of the PCA and ICA. Such a multi-way technique is, for example, PARAFAC - Parallel Factor Analysis.

Baccino (2011) and Hutzler et al. (2007) also mention issue of artifacts generated by saccades and microsaccades which is inevitable. Some speculations consider microsaccades to be a source of an artifacts at gamma range. However, they are not specific to EFRP, since they are also present at ERP experiments.

To address saccades, Ries et al. (2018) mentions an algorithm by M. Plöchl, J.P. Ossandon and P. König from 2012, which objectively rejects independent components of the signal based on data from eye-tracker. Moreover, in his research Ries et al. (2018) attempt to correct EEG signal by removing artifacts from saccades and attached graphs show that artifacts are present only for -75 to 50 ms around fixation onset. Therefore saccades should not be considered an issue with studying cognition time-locked to fixation onset. On contrary EFRP can be considered as a way to avoid data corruption by saccade artifacts.

3.4 Relevant Studies and Research

Previously conducted research, studies and experiments provide some ground for the method proposed by the thesis.

Intelligent Onboarding. Gilbert et al. (2009) describe one of implementation of a tutoring system with enhancements so that it behaves intelligently. The system is built on top of a third-party application that should be explained to a new user. The "intelligence" means, that the tutor not only bluntly pops-up at a specified time but recognizes a situation and behaves accordingly. It involves a cognitive model of working of explained application, so it can guide the user through learning more relevantly and according to user's individual characteristics. The guidance involves suggestions on next actions with explanations. Also in case of incorrect action or possibly dangerous or unwanted action, it can warn a possibly confused user. Such a system allows "learning by doing". The author states that such tutorial system "reduced the time to complete a training task by an average of 14% and reduced user frustration". However, it is not certain that user completing the task understands how should be the application controlled and used.

Learnability Evaluation Using EEG. An experiment by Stickel et al. from 2007 researched possibility to asses learnability by electroencephalography. They based the assessment on alpha dominance and beta/gamma (frequency band 25-40 Hz) dominance. A software was considered learnable when alpha dominance was present, while poor learnability was stated when beta/gamma dominated. Based on this assumption they tried to distinguish good learners from poor learners. However, the effect was not confirmed. Drawbacks of this experiment were small sample size and low quality of stimuli.

Memorization Prediction by a Negativity. Using ERP technique, Fukuda et al. (2015) found neurophysiological correlate of one of mechanisms employed during memory encoding. They found out that during encoding of items that were later recognized, ERP wave showed greater positivity than with those where participants were not sure whether they have seen them. This correlate was then used to determine the probability that an item will be recognized.

This method was validated by allowing to restudy half of the items that were predicted to be not remembered. Restudying half of items predicted as poorly studied dramatically enhanced efficacy of learning against poorly studied items that were not re-studied. Fukuda et al. (2015) ordered studied items by frontal positivity and lowest 40% were marked as poorly studied. This makes possible to forecast memory in real-time.

Object Identification. Rämä et al. (2010) employed EFRP technique to study object identification. They let participants differentiate between real and chimerical objects on simple line drawings. These two categories did not cause significant changes in results.

Instead, the authors focused on the comparison of fixations categorized by order. They compared first, last and for more than two fixations on an image also second fixation (average numbers of fixations was 3,33). Their analysis was focused on the first 150 ms of fixations, considering medians of fixations durations (1st - 166, 2nd - 271, 3rd - 385 ms).

They observed a large difference in first vs. following fixations in frontal sites. However, this was probably caused only by artifacts, as an amplitude of first fixations was approximately five times larger than other fixations. Rämä et al. (2010) also observed a significant difference of P1 latency for first compared to the last fixation at posterior and occipital sites. These two sites also elicited more positivity for P1 and N1 components when comparing first to the last fixation.

Unfortunately, the study focused only on low-level sensory processing, while later components of at least the longer saccades might have provides interesting results.

Visual Search. In their experiment from 2014, Kaunitz et al. compared EFRP and ERP settings. They employed visual oddball paradigm for ERP and free-viewing visual search task for EFRP. The task was to find a target face in scene crowded with distractor faces. Before the experiment, participants were trained to avoid short fixation. This could lower ecological validity, but not significantly.

Results of experiments show, that known cognitive EEG signatures are present also in EFRP setting. Direct comparison of ERPs and EFRPs show differences in latency and topography. The neural response showed the difference between target and distractor after 250 ms in free-viewing, same as in oddball condition. The differences between ERP and EFRP was that P1 occurred 20 ms sooner. The maximum of P3 component at centro-parietal sites was for EFRP at 372 ms, while for ERP its maximum was at 440 ms.

Effect of saccade amplitude on P1 amplitude was also investigated, same as in the previous study (Rämä et al., 2010). Result confirmed the same conclusion, i.e. that longer saccades cause the larger amplitude of the P1 component. The effect can be seen even when the difference of saccade amplitudes is one visual degree.

A further contribution of the study was also a demonstration of the possibility to classify target vs. non-target stimuli (faces) based on a single trial. Results of classifier were above significantly above chance for 11 out of 12 subjects for classification based on EFRP and 10 out of 10 for ERPs. Centro-parietal component at around 450ms was the largest contribution to the classifier.

More Ecologically Valid and Extensive Visual Search. Devillez et al. (2015) executed extensive analysis of data from visual search and free exploration tasks. Stimuli were real-life scenes, such as buildings or interiors. They experimented with two types of tasks - visual search (VS) and free exploration (FE), with three experimental conditions. The first condition was to search for an item in a scene, which was present. The second was the same except the target was not present and the third was just to freely explore the scene. They were interested in fixations on a target. If a target was missing, they examined fixations on a most salient item in the scene.

In three analyses they:

1. compared the first fixation of interest for the three conditions

2. for VS task with target compared (1) fixations on target, (2) fixations on a salient object as control, (3) next fixation after a target was fixated, (4) first fixation outside of region containing a target and (5) first re-fixation of a target.

3. used same control as in 2nd analysis and compared five fixations around a target (-2nd, -1st, target, 1st, 2nd)

The study was focused on later ERP components. Contrary to the previous study (Kaunitz et al., 2014) they did not manipulate fixation duration by instructions. Rather, they included in the analysis only fixations with duration around 250 ms.

The analyses showed the significantly larger amplitude of P300 potential at centroparietal electrodes for fixation on target compared to control. This P300 response was not present for fixations outside of target, confirming, that it is a consequence of finding the target. After the target was refixated, similar P300, but with lower amplitude, was elicited. This also confirms the effect of a target on P300 and shows overlapping of components across consecutive fixations. The smaller amplitude can be caused by habituation.

Besides conclusions of authors, some remarks can be stated about the research. A relatively low number of epochs was used. In each analysis, for each of 34 participants, they averaged roughly 10 - 30 epochs. This suggests a relatively low number of required epochs to obtain valuable results in free viewing task. Target was fixated 2,56 ($\pm$0,09) times, which is similar finding to an aforementioned study (Rämä et al., 2010). Lastly, it seems that the lambda response is masking the effect of target identification manifested in P300. This can be best seen on topographic maps of consecutive fixations, where first moments of ERP are similar to control condition, but later, a clear overlap of P300 can be seen, as was also mentioned by authors.

ERP of Visual Incongruity Proverbio et al. (Proverbio et al., 2009) experimented with visual incongruity. Control group of stimuli was images of meaningful human actions (e.g. " young woman driving car, mature man praying in a pew, a 10-year-old girl practicing her flute"). The second group of stimuli pictured unexpected human behavior without any understandable goal (e.g. "businesswoman balancing on one foot in a desert, a young man playing cello with a saw, woman cutting bread with a saw in kitchen, man splashing face with pebbles").

Participants were “distracted” by a task to distinguish pictures of a natural scene with no person present from pictures with a person present. The result was a significant effect of meaningfulness over parieto-occipital cortex beginning at 170 ms and peaking at 250 ms, and over fronto-cetral region at about 400 ms. This confirms that component N400 is not sensitive only to language semantics, but also to visually processed information.

Software Interfaces Fixation Durations. Goldberg et al. (1999) measured fixation duration in their experiment. The average duration of fixations was 411 ms for ”good interfaces” and 391 ms for ”poor interfaces”, both with a standard deviation of 144. The difference is not significant despite statement mentioned in their work, that ”more difficult processing produced longer fixation durations”. This result suggests that measuring only fixation duration is not informative, but more importantly, that approximately half of the fixations could be used to analyze P300 and N400 components of ERP.

Interestingly, he excluded fixations shorter than 100 ms because of assumption, that information encoding occurs after the 100 ms. This is in line with P1 or lambda components latencies. In the article, is also mentioned another study concluding, that ”higher display densities produced 50 - 100 ms longer fixation durations than lower density displays”, which is relevant to complicated software interfaces

Employing EFRP in Context of Information Systems Use. Léger et al. (2014) experimented with EFRP technique in a multiple stimuli context of an information system use. Experimental tasks were set in a situation of reading industry report. During reading e-mail notifications poped-up on screen. The task was to decide on a relevancy of notification to the topic of the report by opening or closing the pop-up window. If participant considered the notification relevant and chose to open the e-mail, he should have read the e-mail.

Results of experiment unsurprisingly confirmed component P300 over midline parietal area - only Pz site, after notification was displayed. The onset of fixation on the notification, which lasted at least for 400 ms elicited clear N400 component also over midline parietal site and also significant positivity over frontal sites F3 and F7, which was interpreted as evidence for language-related cognitive processes. Components related to motor planning processes were also observed in association with computer mouse use. Subsequent analysis confirmed the validity and effectiveness of the EFRP method over ERP, as by ERP it was not possible to find any significant evidence for ERP components related to stimuli.

3.5 Summary of Current State

Presented studies show that designing software to be accepted for use by a target audience is not a simple task and can be addressed from multiple points of view. The most recent approach, aiming at user experience, is intertwined with the field of human-computer interaction and other disciplines, mostly psychology. This approach makes use of plenty of methods for evaluating general usability, learnability and acceptance, which leads to enhancing them. Many ideas are currently available for improving learnability and helping with the transition from novice user.

Despite, that methods are widely and successfully used in practice of UX, new methods are always required (Alexandre et al., 2018). They should either ease-up evaluation of interfaces or provide new or better insights and understandings of users views. Current technology allows examining of human behavior and reactions with high accuracy, provide information that it is not possible to verbalize or are not even available to participants and thus provide great support for new methods.

While eye-tracking is becoming more applied in the field of human-computer interaction, electroencephalography is not a widely used technique for assessing user experience. Also, they both have some drawbacks. To overcome them it is possible to fuse the two techniques. The EEG provides insight on cognitive processes, whilst eye-tracking provides information on the focus of a user's attention. Such fusion could provide a method for obtaining insights on human-computer interaction, that was not possible to obtain yet.

It was shown, that research on such combination has already been conducted and brought the result of eye-fixation related potential. However, possible or practical applications in the field of HCI is still missing. To conclude, a relevant question is "how can current knowledge and technological possibilities help solve problems of software learnability and acceptability?"

The following section will provide a proposal on methods that can solve this issue.

4 Proposal of an Experimental Method

The main goal of this thesis is to propose a new experimental paradigm. This paradigm would utilize current knowledge of research methods of eye-tracking, event-related potentials and eye-fixation related potentials, described before, to the field of human-computer interaction.

Usage of these methods could provide new findings about mechanisms involved in information processing of brain during work with a software interface. New findings might be applied to design better interfaces, namely allowing better learnability and acceptance of software products. This lowers magnitude of issues described in the first part of the thesis. It mostly addresses issues of objectivity and costs of evaluation.

Questions that this proposal should answer are:

Q1: How to discern elements of a user interface that user does not understand?

Q2: How to compare methods of explaining interface to a new user?

Control Element - Basic Block of an Interface.

Software interfaces consist of information relevant to user's real-world tasks and control elements, that could be used to input or manipulate the information. A control element is usually enclosed in a rectangular area. It often consists of a unique graphical representation.

Graphics of the area is adjusted to provide information on how to use the element and interact with it. An element can contain an icon - pictogram conveying the meaning of the element, and possibly also short textual information describing the meaning and consequences of usage. All of these parts are optional.

Purpose of such element is to allow a user to perform some action in the software and the goal of the overall graphical representation is to provide the user with complete information on what would be the consequences of the action. Examples of three such control elements can be seen in Figure 1.

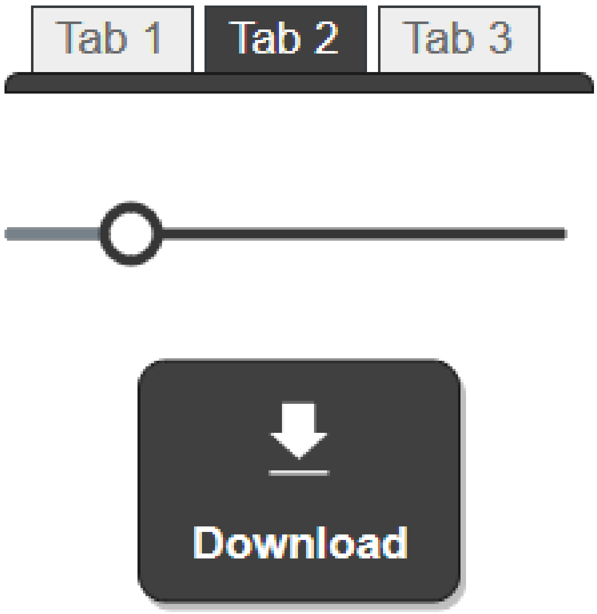

Figure 1: Examples of three control elements

A cornerstone of following proposed paradigms are control elements of a graphical software interface that user understands and knows what is their purpose and how they are used, or how do they work. Opposite to such elements would be those control elements, that does not imply understandable usage for a user. The user would not be able to infer purpose or consequences of usage of such elements and therefore those are not clear to the user. It is reasonable to expect, that elements recognized by the user and familiar to him, will activate different areas at a different time or with different power, than those that user sees for the first time or have troubles to understand. (Kaunitz et al., 2014; Devillez et al., 2015; Proverbio et al., 2009)

Event to which is ERP time-locked is in case of EFRP fixation onset or offset (Hutzler et al., 2007; Baccino, 2011; Léger et al., 2014). In this experiment, the event would be based on gaze on a graphical computer display with a software interface. Examined fixations are those with an offset in *Area of Interest* - an area of the screen where one of the examined control elements is laid (Schall et al., 2014; Pernice et al., n.d.). Epochs can be filtered based on fixation duration, dismissing fixations shorter than 100ms (Goldberg et al., 1999) or more (Devillez et al., 2015; Léger et al., 2014) to improve data reliability. This value can and should be modified according to real measured data in the phase of analysis. However, Blair et al. (2009) argue that "shifts of overt attention are not necessary for shifts of covert attention, shifts of covert attention are necessary for shifts of overt attention." This suggests, that change of fixation will not occur sooner, than covert stimulus is processed. Also, EEG signal coupled with fixations on control elements that were used immediately after fixation should be processed separately since the P300 component is modulated by relevance to task (Frey et al., 2013; Léger et al., 2014).

A user interface of a popular e-mail service contains approximately 25 - 30 unique graphical control elements. An academic information system contains approximately the same number of elements but extensively reused, which means that context can modulate usage and therefore understanding of element. A number of elements in another academic information system can be roughly estimated to 100 because of a different design approach, where elements are not reused and designed more specifically for each use-case. These numbers were obtained by brief examinations of user interfaces. These systems are designed to be as simple as possible. Although, in complex information systems, for example, an internal information system of a company, control elements are counted in hundreds[2]. With a growing complexity of software and its interface usually also grows a number of control elements. And those are usually systems that would benefit of proposed methods most.

For maximizing ecological validity of measurements, experiments should be conducted in the setting of the standard usability test, which is currently con-

[2]Based on internal data of company with focus on user onboarding.

sidered as a suitable and valid method of testing of user interfaces (Bargas-Avila et al., 2011; Rubin et al., 2008). It is important to set the usability test and tasks in a most real context and with minimum intrusion of a moderator as possible (Alexandre et al., 2018). The difference against the standard is that EEG would be employed and think-aloud would not be demanded, but suppressed or even forbidden, to minimize speech artifacts (Kaczorowska et al., 2017; Luck, 2005; Hertzum et al., 2009). That means, that participant would be instructed to accomplish various tasks, which would force the user to look and therefore fixate on various elements of the interface. Tasks should be designed to maximize the amount of tested component that are to be used in order to finish the task. This would provide more analyzable stimuli. A participant can be instructed to slow down his interaction to provide for longer fixations (Kaunitz et al., 2014) which could, however lower ecological validity (Devillez et al., 2015).

4.1 Discerning Clear and Unclear Elements

The first paradigm, offering an answer to **Q1**, is aimed at individual control elements. Using the elements as experimental stimuli, the two groups of elements - clear and unclear, should produce different EFRP waveform and it should be possible to find indexes of understanding of element (Rämä et al., 2010; Fukuda et al., 2015; Kaunitz et al., 2014; Proverbio et al., 2009). This leads to the first hypothesis.

H1: Understood, resp. clear elements of an interface produce different EFRP waveforms than unclear ones.

Individual stimuli would be assigned to two groups of *clear* and *unclear* elements based on a questionnaire filled out by each participant at the end of a session (Tullis et al., 2010; Kuniavsky et al., 2012; Fukuda et al., 2015). Each of element suitable for EFRP analysis would be displayed with a question about a purpose of the element with one correct answer, one or more incorrect answers and one “don’t know” answer. Those with correct answer selected would be assigned to a group of clear - understood elements and all others to a group of unclear - not understood elements. Example of a question in such questionnaire can be seen in Figure 2.

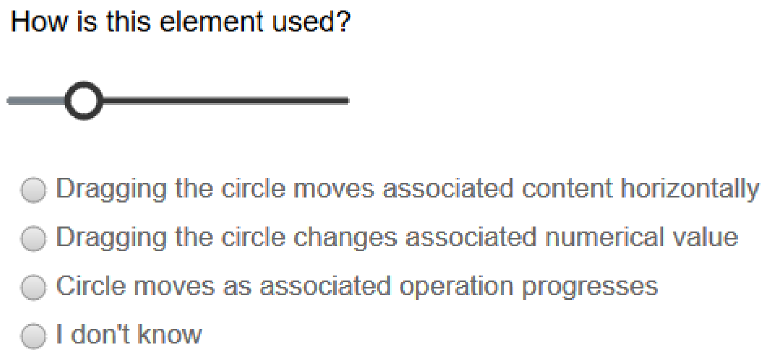

Figure 2: Examples of element in questionnaire for assigning element to one of two groups - understood or not understood elements.

To summarize, according to the first paradigm, the experiment should begin with standard usability test of a system. Free examination of the interface should be encouraged and also the scenario of the usability test should allow for that. After the usability test, the participant would fill out the questionnaire about elements automatically selected based on eye-fixations during the usability test. This procedure should provide all needed data from one participant and should be repeated with numerous participants to obtain significant results.

Dataset obtained from the experiment is described as recordings of EEG data to be averaged and time-locked according to fixation and divided into two groups as understood and not understood. For more complex interfaces number of stimuli would be in hundreds per participant, with N/2 stimuli per group (gross estimation). For more simple interfaces, the experiment would require between-group grand-average which would enhance the signal-to-noise ratio. (Woodman, 2010; Luck, 2005; Boudewyn et al., 2018)

Beside assigning elements to the two groups, further factors might modulate ERP components. An analysis should be also separately devoted to elements that were used during the interaction, i.e. user considered them relevant to the executed task since this can modulate P300 component (Luck, 2005). Components can be modulated also because of numerous element refixations, as the first sight of an element should cause different activations than the last one. Baccino (2011) states that second-pass over are of interest is because of reprocessing or verification. Next factor in the analysis should, therefore, be an order of a fixation. Fixations should be grouped as (1) first fixations, (2) second to next to last fixations and (3) last fixations (Devillez et al., 2015). EEG epochs should be therefore grouped according to these factors:

1. element understood,
2. fixation order,
3. element used.

It is expected that seeing elements that are clear and understood will produce different EFRP waveforms than those that are not clear or not understood. A difference should be visible mostly at later components that account for cognition and semantics, i.e. P300 and N400. However, these components could be visible sooner than they usually are (i.e. 300-500 ms) because of different mechanisms employed with free-viewing tasks. Understood elements should elicit the components with smaller latency and lower amplitude compared to not understood elements. The most significant difference should be visible at frontal and centro-parietal sites. Laterality of components is difficult to determine since control elements are processed as images but also by semantic meaning. (Fukuda et al., 2015; Kaunitz et al., 2014; Devillez et al., 2015; Proverbio et al., 2009; Hutzler et al., 2007)

Similar to the research of Fukuda et al. (2015) where they in real-time discriminated well-studied and poorly-studied stimuli, it could be possible to detect understood and not understood elements. Such discrimination strongly depends on discovered indexes by the experiment. Besides simple technique described by Fukuda et al. (2015), advances in single-trial ERP analysis can also provide methods for more accurate discrimination of elements. This expectation is supported also by Kaunitz et al. (2014). Obtaining of such technique would allow finding elements of a user interface that are to be redesigned or at least explained in an onboarding phase.

4.2 Comparing Learnability of Software

To address the second postulated question (**Q2**) postulated, a paradigm is focused on the entire application. However, basic blocks of evaluation are still controlled elements. According to UX paradigm, it is important to address whole application in real context of use with real task for holistic approach and also for maximizing ecological validity (Alexandre et al., 2018; Grossman et al., 2009; Hassenzahl et al., 2006; Rubin et al., 2008). The main idea is, that when user is new to application and its user interface, control element of it would elicit differently modulated ERP components than when the user is familiar with the environment. This leads to a hypothesis that:

H2: Control elements in an interface that user is familiar with, will produce different EFRP waveform than those in the interface that user is unsure how to use.

Experimental group would consist of (*G1 - new*) users that work with software for the first time and (*G2 - proficient*) users that are familiar with the software (e.g. use it on daily basis). Same as with previous experimental paradigm, users would execute tasks in a standard usability test. The tasks should be different than those executed daily by proficient users to allow for novelty even for the G2 group, which would suppress routine and habitual usage and allow for longer fixations and also require more deliberation.

To obtain EFRP waveform, measured EEG signal would be time-locked and averaged according to each fixation on any control element for each user. Further, the waveforms can be grand-averaged over an experimental group to enhance results validity. Besides factor of a new vs. proficient user, epochs should be sorted to groups according to the next two factors (1) element used and (2) fixation order, similar as in the previous paradigm.

Modulations of ERP components should be similar as in the previous paradigm since same cognitive processes, factors and mechanisms are involved. This means that amplitude and latency should be lower for P300 and N400 components at frontal and centro-parietal sites.

To only differentiate proficient from new ones using electrophysiology does not provide much of value (Alexandre et al., 2018). However, obtained differences in measurements could serve as indexes of proficiency and standardized scales can be created based on the results of such exploratory experiments. These indexes could be used to determine the best way to achieve proficiency of a user.

A new - *onboarded user*, which had not been in a contact with the software yet, should undergo an onboarding phase to learn how to work with the software. This onboarding phase can implement any methods, practices and techniques of choice of a software designer (Alexandre et al., 2018). Well-designed onboarding phase is effective and influencing and user really becomes more proficient with the software. It is reasonable to expect, that when user undergoes well-designed onboarding phase, then indexes identified in the previous step of the experiment should be more similar to users of *G2 - proficient*, than to *G1 - new.*

The method would be most useful when comparing two or more designs of onboarding processes. To do so 2+N experimental groups will be required, where

N is a number of various designs. Each participant should be subject at same usability test. As was described, *G1 - new* and *G2 - proficient* should only undergo usability test, to obtain actual indexes for the software. Participants of further groups *Gn - onboarding n* should first go through an onboarding phase and then execute usability test, where EEG would be recorded. The similarity of indexes from grand-averaged EFRP of distinct groups would determine the most effective design.

To avoid bias by a moderator and to achieve ecological validity, a moderator should not intervene during the session. This allows for conducting multiple sessions together. After the sessions are conducted, the data can be processed automatically. Hence, there is no need to replay sessions. The automatized analysis would, in the end, provide a list of control elements sorted by a rating of understanding. An analysis of the second, complementary method would provide information which condition - onboarding designs yield a better understanding of a software.

4.3 Concluding Remarks

Presented methods allow obtaining insights from usability tests which are currently difficult to gather. Besides that, the methods provide higher objectivity in evaluating the findings of usability tests. The third advantage is a possibility to obtain the insights by an automatic evaluation of the usability testings without a need for replaying individual sessions.

Since these are only proposals for experimental assessment methods, it is mandatory to validate them. To prove them useful it is required to find a minimal number of participants per group. This is subject to statistical analysis of data measured in eventual real experiments. The experiments should reveal the size of the effect of understood vs. not understood elements and users' proficiency with software on ERPs.

However, based on presented similar experiments, it is possible that around 20 usability testing sessions could be enough to decide which elements of an interface are understandable. To decide which design of onboarding phase is more effective up to 10 automatically evaluated sessions per condition could be enough to determine statistically significant results. (Boudewyn et al., 2018; Devillez et al., 2015)

Such a method could in case of widely used complex industrial applications, significantly lower personnel training costs. In the case of applications designated to wide public usage, this could mean the possibility to find the best way to improve the software acceptance.

5 Conclusion

The thesis researched reasons and determinants of acceptance of a software for use. Onboarding as an important part of an interaction between human and computer system was identified. Techniques and methodologies for achieving higher accepting and improving onboarding were introduced. Described was also the current practice of studying human-computer interaction with an emphasis on the holistic view of a tool as is recommended by a modern approach of user experience. The practice, however, contains some difficulties and deficiencies, which were identified.

The thesis further recognizes techniques of eye-tracking, electroencephalography. Their combination as a suitable basis for enrichment of currently utilized techniques is recognized. Fuse of methods of *areas of interests (AOI)* for eye-tracking and *event related potentials (ERP)* together brings *eye-fixation related potential (EFRP)* which can be essence for various new methods for field of HCI.

Two of such methods is proposed in the second section of the thesis. This section explains the principles of the methods. Advantages and possible usage of methods are explained. The methods are based on an evaluation of individual control elements of graphical user interfaces. A way to discern elements that are understood from those that are not clear to a user is provided. Also, methods describe possibility to objectively compare designs of onboardings in order improve accepting of a software. This can lead to reduced costs of evaluation and higher acceptance rate.

The methods were not validated and confirmed yet. Henceforth, the direction of research is to validate the methods and apply them to standard practice. The presented methods are a demonstration of the usage of current technical possibilities. It is believed, that many more methods could emerge from the promising technique of eye-fixation related potentials.

5.1 Interdisciplinary Character

As was stated in the work, it is necessary to involve various disciplines. Basic nature of studying human-computer interaction requires multiple disciplines as it is a combination of studying human behavior and principles of information technology and information architecture. A user experience approach itself, which is sometimes considered a separate field, but for sure close to the field of HCI, is based on principles of many fields such as psychology, art or design. Further, the involved psychology is not only users' psychology, but also cognitive psychology.

Eye-trackers, originating from psychological research are now commonly applied in the fields. Further, electroencephalography is mostly medical device but allows studying neurology, psychology, cognition. And the technique itself requires some knowledge of the discipline of electrical engineering.

The ideas in the thesis are a synthesis of many small pieces of various fields. The borders of fields which provided knowledge for the ideas are often nebulous, but the fields can be clearly identified. The thesis attempts to apply theoretical findings and knowledge of all the fields into practical usage.

References

ALEXANDRE, Boris; REYNAUD, Emanuelle; OSIURAK, François; NAVARRO, Jordan, 2018. Acceptance and acceptability criteria: a literature review. *Cognition, Technology & Work*, pp. 1–13. ISSN 1435-5558. Available from DOI: 10.1007/s10111-018-0459-1.

BACCINO, Thierry, 2011. *Eye movements and concurrent event-related potentials: Eye fixation-related potential investigations in reading*. Oxford University Press. Available from DOI: 10.1093/oxfordhb/9780199539789.013.0047.

BARGAS-AVILA, Javier A; HORNBAEK, Kasper, 2011. Old Wine in New Bottles or Novel Challenges? A Critical Analysis of Empirical Studies of User Experience.

BLAIR, Mark R.; WATSON, Marcus R.; WALSHE, R. Calen; MAJ, Fillip, 2009. Extremely selective attention: Eye-tracking studies of the dynamic allocation of attention to stimulus features in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Vol. 35, no. 5, pp. 1196–1206. Available from DOI: 10.1037/a0016272.

BOUDEWYN, Megan A.; LUCK, Steven J.; FARRENS, Jaclyn L.; KAPPENMAN, Emily S., 2018. How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology*. Vol. 55, no. 6, pp. e13049. ISSN 00485772. Available from DOI: 10.1111/psyp.13049.

BUZSÁKI, György; ANASTASSIOU, Costas A.; KOCH, Christof, 2012. The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience*. Vol. 13, no. 6, pp. 407–420. ISSN 1471-003X. Available from DOI: 10.1038/nrn3241.

CARLÈN, Filip, 2017. *User Onboarding An investigation in how to increase the activation of new customers using design*. Available also from: http://

publications.lib.chalmers.se/records/fulltext/252779/252779.pdf. PhD thesis. Chalmers University of Technology.

CASCAES CARDOSO, Marina; MARINA, 2017. The Onboarding Effect. In: *The Onboarding Effect. Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17*. New York, New York, USA: ACM Press, pp. 263–267. ISBN 9781450346566. Available from DOI: 10.1145/3027063.3027128.

CHI, Yu Mike; JUNG, Tzyy-Ping; CAUWENBERGHS, Gert, 2010. Dry-Contact and Noncontact Biopotential Electrodes: Methodological Review. *IEEE Reviews in Biomedical Engineering*. Vol. 3, pp. 106–119. ISSN 1937-3333. Available from DOI: 10.1109/RBME.2010.2084078.

CHYNAŁ, Piotr; FALKOWSKA, Julia; SOBECKI, Janusz, 2018. Web Page Graphic Design Usability Testing Enhanced with Eye-Tracking. In: KARWOWSKI, Waldemar; AHRAM, Tareq (eds.). *Intelligent Human Systems Integration: Proceedings of the 1st International Conference on Intelligent Human Systems Integration (IHSI 2018): Integrating People and Intelligent Systems, January 7-9, 2018, Dubai, United Arab Emirates*. Cham: Springer International Publishing, pp. 515–520. ISBN 978-3-319-73888-8. Available from DOI: 10.1007/978-3-319-73888-8{_}80.

DELORME, Arnaud; MAKEIG, Scott, 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*. Vol. 134, no. 1, pp. 9–21. ISSN 0165-0270. Available from DOI: 10.1016/J.JNEUMETH.2003.10.009.

DEVILLEZ, Hélène; GUYADER, Nathalie; GUÉRIN-DUGUÉ, Anne, 2015. An eye fixation–related potentials analysis of the P300 potential for fixations onto a target object when exploring natural scenes. *Journal of Vision*. Vol. 15, no. 13, pp. 20. ISSN 1534-7362. Available from DOI: 10.1167/15.13.20.

DIMIGEN, Olaf; SOMMER, Werner; HOHLFELD, Annette; JACOBS, Arthur M.; KLIEGL, Reinhold, 2011. Coregistration of eye movements and EEG in natural reading: Analyses and review. *Journal of Experimental Psychology: General*. Vol. 140, no. 4, pp. 552–572. Available from DOI: 10.1037/a0023885.

FREY, Aline; IONESCU, Gelu; LEMAIRE, Benoit; LÓPEZ-OROZCO, Francisco; BACCINO, Thierry; GUÉRIN-DUGUÉ, Anne, 2013. Decision-making in information seeking on texts: an eye-fixation-related potentials investigation. *Frontiers in Systems Neuroscience*. ISSN 1662-5137. Available from DOI: 10.3389/fnsys.2013.00039.

FUKUDA, Keisuke; WOODMAN, Geoffrey F., 2015. Predicting and Improving Recognition Memory Using Multiple Electrophysiological Signals in Real Time. *Psychological Science*. Vol. 26, no. 7, pp. 1026–1037. ISSN 0956-7976. Available from DOI: 10.1177/0956797615578122.

GILBERT, Stephen B.; BLESSING, Stephen B.; BLANKENSHIP, Liz, 2009. The accidental tutor: overlaying an intelligent tutor on an existing user interface. In: *The accidental tutor: overlaying an intelligent tutor on an existing user interface. Proceedings of the 27th international conference extended abstracts on Human factors in computing systems - CHI EA '09*. New York, New York, USA: ACM Press, p. 4603. ISBN 9781605582474. Available from DOI: 10.1145/1520340.1520707.

GOLDBERG, J.H.; KOTVAL, X.P., 1999. Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics*. Vol. 24, no. 6, pp. 631–645. Available from DOI: 10.1016/S0169-8141(98)00068-7.

GROSSMAN, Tovi; FITZMAURICE, George; ATTAR, Ramtin, 2009. A survey of software learnability: metrics, methodologies and guidelines. In: *A survey of software learnability: metrics, methodologies and guidelines. Proceedings of the*

27th international conference on Human factors in computing systems - CHI 09. New York, New York, USA: ACM Press, p. 649. ISBN 9781605582467. Available from DOI: 10.1145/1518701.1518803.

GUO, Fu; DING, Yi; LIU, Weilin; LIU, Chang; ZHANG, Xuefeng, 2016. Can eye-tracking data be measured to assess product design?: Visual attention mechanism should be considered. *International Journal of Industrial Ergonomics*. ISSN 18728219. Available from DOI: 10.1016/j.ergon.2015.12.001.

HASSENZAHL, Marc; TRACTINSKY, Noam, 2006. User experience - A research agenda. *Behaviour and Information Technology*. ISBN 0144929050033. ISSN 0144929X. Available from DOI: 10.1080/01449290500330331.

HERTZUM, Morten; HANSEN, Kristin D.; ANDERSEN, Hans H.K., 2009. Scrutinising usability evaluation: Does thinking aloud affect behaviour and mental workload? *Behaviour and Information Technology*. ISBN 0144-929X. ISSN 0144929X. Available from DOI: 10.1080/01449290701773842.

HOLSANOVA, Jana, 2014. Reception of multimodality: Applying eye-tracking methodology in multimodal research. Vol. 20, pp. 285–296.

HULICK, Samuel, (a). *(cit. 2019-01-15)*. Bulletproof User Onboarding | User Onboarding. Available also from: https://www.useronboard.com/bulletproof-user-onboarding/.

HULICK, Samuel, (b). *User Onboarding | A frequently-updated compendium of web app first-run experiences*. Available also from: https://www.useronboard.com/.

HUTZLER, Florian; BRAUN, Mario; VÕ, Melissa L H; ENGL, Verena; HOFMANN, Markus; DAMBACHER, Michael; LEDER, Helmut; JACOBS, Arthur M., 2007. Welcome to the real world: Validating fixation-related brain potentials for ecologically valid settings. *Brain Research*. Available from DOI: 10.1016/j.brainres.2007.07.025.

IMOTIONS, 2015. *7 Most Used Eye Tracking Metrics and Terms*. Available also from: https://imotions.com/blog/7-terms-metrics-eye-tracking/.

IMOTIONS, 2016. *What is eye tracking and how does it work?* Available also from: https://imotions.com/blog/eye-tracking-work/.

ISO, 2010. *ISO 9241-210:2010 - Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems*. Available also from: https://www.iso.org/standard/52075.html.

ISO, 2018. *ISO 9241-11:2018 - Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts*. Available also from: https://www.iso.org/standard/63500.html.

JUNG, Tzyy-Ping; MAKEIG, Scott; WESTERFIELD, Marissa; TOWNSEND, Jeanne; COURCHESNE, Eric; SEJNOWSKI, Terrence J., 2001. Analysis and visualization of single-trial event-related potentials. *Human Brain Mapping*. Vol. 14, no. 3, pp. 166–185. ISSN 1065-9471. Available from DOI: 10.1002/hbm.1050.

KACZOROWSKA, Monika; PLECHAWSKA-WOJCIK, Malgorzata; TOKOVAROV, Mikhail; DMYTRUK, Roman, 2017. Comparison of the ICA and PCA methods in correction of EEG signal artefacts. In: *Comparison of the ICA and PCA methods in correction of EEG signal artefacts. 2017 10th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*. IEEE, pp. 262–267. ISBN 978-1-5090-5160-1. Available from DOI: 10.1109/ATEE.2017.7905095.

KAPPENMAN, Emily S; LUCK, Steven J, 2012. ERP Components: The Ups and Downs of Brainwave Recordings. In: *ERP Components: The Ups and Downs of Brainwave Recordings. The Oxford Handbook of Event-Related Potential Components*.

KAUNITZ, Lisandro N.; KAMIENKOWSKI, Juan E.; VARATHARAJAH, Alexander; SIGMAN, Mariano; QUIROGA, Rodrigo Quian; ISON, Matias J., 2014. Looking for a face in the crowd: Fixation-related potentials in an eye-movement visual search task. *NeuroImage*. Vol. 89, pp. 297–305. ISSN 1053-8119. Available from DOI: 10.1016/J.NEUROIMAGE.2013.12.006.

KUNIAVSKY, M; GOODMAN, E; MOED, A, 2012. *Observing the User Experience: A Practitioner's Guide to User Research*. Morgan Kaufmann. Books 24x7 IT PRO. ISBN 9780123848697. Available also from: https://books.google.sk/books?id=jmj5LCAG6kwC.

LÉGER, Pierre-Majorique; SÉNECAL, Sylvain; COURTEMANCHE, François; GUINEA, Ana; TITAH, Ryad; FREDETTE, Marc; LABONTE-LEMOYNE, Élise, 2014. Precision is in the Eye of the Beholder: Application of Eye Fixation-Related Potentials to Information Systems Research. *Journal of the Association for Information Systems*. ISBN 1536-9323. ISSN 15369323. Available from DOI: 10.17705/1jais.00376.

LLINAS, Rodolfo, 2008. Neuron. *Scholarpedia*. Vol. 3, no. 8, pp. 1490. ISSN 1941-6016. Available from DOI: 10.4249/scholarpedia.1490.

LOPEZ-CALDERON, Javier; LUCK, Steven J., 2014. ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*. Vol. 8, pp. 213. ISSN 1662-5161. Available from DOI: 10.3389/fnhum.2014.00213.

LUCK, Steven J., 2005. *An Introduction to the Event-Related Potential Technique*. Cambridge, Ma: MIT Press. Available also from: http://sites.bu.edu/reinhartlab/files/2017/06/ERPbook.pdf.

MEALHA, O; VELOSO, A; ALMEIDA, S; RODRIGUES, R; ROQUE, L; MARQUES, R; MANTEIGUEIRO, C, 2010. Eye Tracking Data Representation and Visualization: on Information and Communication studies at CETAC.ME-

DIA. *Journal of Eyetracking Visual Cognition and Emotion*. Vol. 1, no. 1, pp. 1647–7677.

MORIMOTO, Carlos H; MIMICA, Marcio R M, 2004. Eye gaze tracking techniques for interactive applications. Available from DOI: 10.1016/j.cviu.2004.07.010.

MORRISON, Craig. *User Onboarding Best Practices for Retaining Your Users*. Available also from: https://usabilityhour.com/user-onboarding-best-practices/.

PERNICE, Kara; NIELSEN, Jakob. How to Conduct Eyetracking Studies. Available also from: http://www.nngroup.com/reports/how-to-conduct-eyetracking-studies/%20www.nngroup.com/reports/how-to-conduct-eyetracking-studies.

PORTMAN, Jane, 2017. *User onboarding essentials*. Available also from: https://www.invisionapp.com/blog/user-onboarding-essentials/.

PROVERBIO, Alice Mado; RIVA, Federica, 2009. RP and N400 ERP components reflect semantic violations in visual processing of human actions. *Neuroscience Letters*. Vol. 459, no. 3, pp. 142–146. ISSN 0304-3940. Available from DOI: 10.1016/J.NEULET.2009.05.012.

RÄMÄ, Pia; BACCINO, Thierry, 2010. Eye fixation-related potentials (EFRPs) during object identification. *Visual Neuroscience*. ISBN 0952-5238. ISSN 09525238. Available from DOI: 10.1017/S0952523810000283.

RENZ, Jan; STAUBITZ, Thomas; POLLACK, Jaqueline; MEINEL, Christoph, 2014. *Improving the Onboarding User Experience in MOOCs*. Germany. Available also from: https://hpi.de/fileadmin/user_upload/fachgebiete/meinel/papers/Web-University/2014_Renz_EDULEARN.pdf.

RIES, Anthony J.; SLAYBACK, David; TOURYAN, Jon, 2018. The fixation-related lambda response: Effects of saccade magnitude, spatial frequency, and ocular artifact removal. *International Journal of Psychophysiology*. Vol. 134, pp. 1–8. ISSN 0167-8760. Available from DOI: 10.1016/J.IJPSYCHO.2018.09.004.

RUBIN, Jeff; CHISNELL, Dana, 2008. *Handbook of Usability Testing Second Edition How to Plan, Design, and Conduct Effective Tests*. Second. Indianapolis, IN: Qiley Publishing, Inc. ISBN 9780470185483.

SANEI, Saeid; CHAMBERS, Ja; WILEY, John; PACIFICO, Richard J, 2007. *EEG Signal Processing*. West Sussex, England: John Wiley & Sons Ltd, ISBN 978-0-470-02581-9.

SCARR, Joey; COCKBURN, Andy; GUTWIN, Carl; QUINN, Philip, 2011. Dips and ceilings: understanding and supporting transitions to expertise in user interfaces. In: *Dips and ceilings: understanding and supporting transitions to expertise in user interfaces. Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. New York, New York, USA: ACM Press, p. 2741. ISBN 9781450302289. Available from DOI: 10.1145/1978942.1979348.

SCHALL, A.; ROMANO BERGSTROM, J., 2014. *Eye Tracking in User Experience Design*. Eye Tracking in User Experience Design. ISBN 9780124081383. Available from DOI: 10.1016/C2012-0-06867-6.

SERENO, Sara C.; RAYNER, Keith, 2003. Measuring word recognition in reading: eye movements and event-related potentials. *Trends in Cognitive Sciences*. Vol. 7, no. 11, pp. 489–493. ISSN 1364-6613. Available from DOI: 10.1016/J.TICS.2003.09.010.

STICKEL, Christian; FINK, Josef; HOLZINGER, Andreas, 2007. Enhancing Universal Access – EEG Based Learnability Assessment. In: *Enhancing Uni-*

versal Access – EEG Based Learnability Assessment. Universal Access in Human-Computer Interaction. Applications and Services. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 813–822. Available from DOI: 10.1007/978-3-540-73283-9{_}88.

STRAHM, Brendan; GRAY, Colin M.; VORVOREANU, Mihaela, 2018. Generating Mobile Application Onboarding Insights Through Minimalist Instruction. In: *Generating Mobile Application Onboarding Insights Through Minimalist Instruction. Proceedings of the 2018 on Designing Interactive Systems Conference 2018 - DIS '18*. New York, New York, USA: ACM Press, pp. 361–372. ISBN 9781450351980. Available from DOI: 10.1145/3196709.3196727.

THEOFANOS, Mary; QUESENBERY, Whitney, 2005. *Towards the Design of Effective Formative Test Reports*. Technical report.

TOBII AB. *Types of eye movement*. Available also from: https://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/types-of-eye-movements/.

TOBII AB, 2018. *Why do we move our eyes?* Available also from: https://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/why-do-our-eyes-move/.

TULLIS, Thomas; ALBERT, William, 2010. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Morgan Kaufmann.

UNGER, Russ; CHANDLER, Carolyn, 2009. *A Project Guide to UX Design: For user experience designers in the field or in the making*. Berkeley, CA. ISBN 13 978-0-321-60737-9.

URIGÜEN, Jose Antonio; GARCIA-ZAPIRAIN, Begoña, 2015. EEG artifact removal—state-of-the-art and guidelines. *Journal of Neural Engineering*. Vol. 12, no. 3, pp. 031001. ISSN 1741-2560. Available from DOI: 10.1088/1741-2560/12/3/031001.

UXPA - THE USER EXPERIENCE PROFESSIONALS' ASSOCIATION. *Usability Evaluation Methods | Usability Body of Knowledge*. Available also from: http://www.usabilitybok.org/usability-evaluation-methods.

WOODMAN, Geoffrey F., 2010. A brief introduction to the use of event-related potentials in studies of perception and attention. *Attention, Perception, & Psychophysics*. Vol. 72, no. 8, pp. 2031–2046. ISSN 1943-3921. Available from DOI: 10.3758/BF03196680.

YESELF, 2017. *5 Common Types of User Onboarding that Help Users Become Oriented with a Website and Applications*. Available also from: https://www.yeself.com/blog/5-types-of-user-onboarding-that-help-become-oriented-within-website-and-apps.