UNIVERSIDADE DE PASSO FUNDO
Graduate Program in Applied Computing

Master's Thesis

ASSESSING MOMENTARY USER EXPERIENCE WITH CONVERSATIONAL AGENTS USING AN ELECTROENCEPHALOGRAPHY DEVICE

LAÍS ANDRESSA BROCK

UPF

UNIVERSITY OF PASSO FUNDO
INSTITUTE OF EXACT SCIENCES AND GEOSCIENCES
GRADUATE PROGRAM IN APPLIED COMPUTING

ASSESSING MOMENTARY USER EXPERIENCE WITH CONVERSATIONAL AGENTS USING AN ELECTROENCEPHALOGRAPHY DEVICE

Laís Andressa Brock

Thesis submitted to the University of Passo Fundo in partial fulfillment of the requirements for the degree of Master in Applied Computing.

Advisor: Prof. Dra. Ana Carolina Bertoletti de Marchi

Passo Fundo

2020

CIP – Cataloging in Publication

B864a Brock, Laís Andressa
Assessing momentary user experience with conversational agents using an electroencephalography device [electronic resource] / Laís Andressa Brock. – 2020.
3.17 MB ; PDF.

Advisor: Profa. Dra. Ana Carolina Bertoletti de Marchi.
Thesis (Master in Applied Computing) – University of Passo Fundo, 2020.

1. User experience (UX). 2. Bioinformatics. 3. Emotiv. 4. Electroencephalography. 5. Conversational Agents. I. Marchi, Ana Carolina Bertoletti de, advisor. II. Title.

CDU: 004

Cataloging: Librarian Jucelei Rodrigues Domingues - CRB 10/1569

PPGCA
Programa de Pós-Graduação
em Computação Aplicada
Instituto de Ciências Exatas e Geociências | ICEG

ATA DE DEFESA DO
TRABALHO DE CONCLUSÃO DE CURSO DO ACADÊMICO

LAÍS ANDRESSA BROCK

Aos vinte e três dias do mês de março do ano de dois mil e vinte, às quatorze horas, realizou-se, no prédio D1 sala 01 da Universidade de Passo Fundo (UPF), a sessão pública de defesa do Trabalho de Conclusão de Curso "Assessing momentary user experience with conversational agentes using na electroencephalography device", de autoria de Laís Andressa Brock, acadêmica do Curso de Mestrado em Computação Aplicada do Programa de Pós-Graduação em Computação Aplicada – PPGCA. Segundo as informações prestadas pelo Conselho de Pós-Graduação e constantes nos arquivos da Secretaria do PPGCA, a aluna preencheu os requisitos necessários para submeter seu trabalho à avaliação. A banca examinadora foi composta pelos doutores Ana Carolina Bertoletti De Marchi, Rafael Rieder e Luciana Aparecida Martinez Zaina. Concluídos os trabalhos de apresentação e arguição, a banca examinadora considerou a candidata aprovada. Foi concedido o prazo de até quarenta e cinco (45) dias, conforme Regimento do PPGCA, para a acadêmica apresentar ao Conselho de Pós-Graduação o trabalho em sua redação definitiva, a fim de que sejam feitos os encaminhamentos necessários à emissão do Diploma de Mestre em Computação Aplicada. Para constar, foi lavrada a presente ata, que vai assinada pelos membros da banca examinadora e pela Coordenação do PPGCA.

Ana C.B. De Marchi

Prof. Dra. Ana Carolina Bertoletti De Marchi– UPF
Presidente da Banca Examinadora
(Orientadora)

Prof. Dr. Rafael Rieder – UPF
(Avaliador Interno)

Ana C.B. De Marchi

Prof. Dra. Luciana Aparecida Martinez Zaina-UFSCAR
(Avaliadora Externa)

PP

Prof. Dr. Rafael Rieder
Coordenador do PPGCA

AVALIAÇÃO DA EXPERIÊNCIA MOMENTÂNEA DO USUÁRIO COM AGENTES CONVERSACIONAIS USANDO UM DISPOSITIVO DE ELETROENCEFALOGRAFIA

RESUMO

Todos os aspectos da experiência do usuário podem ter um papel na obtenção de sentimentos sobre um determinado produto. UX momentâneo refere-se a quaisquer mudanças de sentimentos durante a interação, no momento em que ocorre. Para avaliar a UX momentânea, foram utilizadas medidas eletrofisiológicas. O objetivo deste estudo é comparar episódios momentâneos de UX enquanto os participantes interagiam com diferentes agentes conversacionais. Trinta e seis participantes de ambos os sexos foram divididos em três grupos para realizar as mesmas tarefas. O primeiro grupo foi o grupo de controle que usava o mecanismo de pesquisa do Google no navegador Chrome; o segundo grupo usou a Assistente do Google usando o teclado e o terceiro usou a Assistente do Google usando comandos de voz. O dispositivo Emotiv Insight e questionários UX foram usados para coletar dados. Os resultados entre os grupos, tanto do eletroencefalograma quanto dos questionários, foram consistentes e as diferenças foram discutidas. Houve diferença na UX momentânea do usuário em relação aos níveis de excitação coletados pelo EEG quando eles interagiam com agentes de conversação por voz. A principal contribuição é o uso da eletroencefalografia como instrumento para medir a UX momentânea e inferir que a interação pela voz mostra particularidades emocionais.

Palavras-Chave: Agentes Conversacionais, UX Momentânea, Eletroencefalografia, Emotiv.

ASSESSING MOMENTARY USER EXPERIENCE WITH CONVERSATIONAL AGENTS USING AN ELECTROENCEPHALOGRAPHY DEVICE

ABSTRACT

All User Experience aspects can play a part in eliciting feelings about a certain product. Momentary UX refers to any changes of feelings during the interaction at the moment it happens. To be able to evaluate Momentary UX, electrophysiological measures have been used. The aim of this study is to compare Momentary UX episodes while participants interacted with different conversational agents. Thirty-six participants of both genders were divided into three groups to perform the same tasks. First group was the control group using Google's search engine in Google's Chrome browser; the second group used Google's Assistant using keyboard and the third used Google's Assistant using voice commands. An Emotiv Insight device and UX questionnaires were used to collect data. The results between groups from both the electroencephalogram and questionnaires were consistent and differences were discussed. There was difference in the user's Momentary UX regarding their excitement levels collected by the EEG when they were interacting with conversational agents by voice. The main contribution is using electroencephalography as a instrument to measure Momentary UX and inferring that the interaction using voice shows emotional particularities.

Keywords: Conversational Agents, Momentary UX, Electroencephalography, Emotiv.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1. INTRODUCTION

Products like conversational agents (CAs) or dialogue systems, such as chatterbots and personal assistants are becoming popular [1]. Examples of these include personal assistants on mobile devices, from technical support help, as well as online chatbots selling products or services [2].

The use of chatbots are gradually arising in fields such as search, customer support, and entertainment. These conversational agents have captured the interest of users and companies, and have been part of headlines [3]. Most conversational agents or chatbots interact with users in text or speech and may offer different features or functions depending on their conversational platform [4]. According to Ho et al. [5], participants who interacted with CAs had emotional and relational experiences.

Some platforms also include elements that can enhance the experience of CA using pictures, sounds, buttons, quick replies, and other features [3]. Users demonstrate to have a growing interest in CAs for productivity tasks, entertainment, and communication. Since chatbots are predicted to be a new way for users to interact with services, it is crucial to understand how people experience it [6, 7].

It is also important to understand the user's feelings, state of being, sensations and emotions during the interaction with a product in order to be able to promote a good experience. Since the experience refers to many aspects, the moment and form of interaction with said product needs to be considered as part of the user experience (UX) as well [8].

UX is associated with a large range of fuzzy and dynamic concepts such as emotional, affective, experiential, hedonic, and aesthetic variables which makes it more complex to get an agreement on a universal definition [9]. Roto et al. [10] describes UX as an overall designation of how people have experienced a period of encounter with a system.

All UX aspects can play a part in eliciting feelings about a certain product. UX depends on the user's existing needs, expectations, and former experiences. Therefore, a good and satisfactory UX is subjective [11]. According to Schulze and Krömker [12], since the use of interactive applications has become an essential part of everyday life, users expect satisfying experiences, thus UX research needs approaches and methods for evaluating experience. Bargas-Avila and Hornbæk [13] states that the most frequent existing methods of UX evaluation combine during and after measurements, in which users are observed when interacting and their experience is evaluated afterwards.

When discussing UX evaluation, the time span of a specific interaction should be taken into consideration. Roto et al. [10] states that the actual experience of usage does not cover all relevant UX concerns. Other experiences extends after usage, through reflection on previous usage or, through changes in people's opinion and judgment of use. With that in mind, we must consider different time spans when discussing user experience. UX can refer to a specific feeling during the interaction, the judgment of a specific experience episode, views and consideration on a system as a whole after several uses. UX can also refer to a period before the first use, an expectation, or any moment related to the experience: before, during and after [10, 14].

There is a contrast between ‘experiencing’ and ‘an experience’, so that calls attention to what is the suitable time span when focusing on one user experience and its evaluation. Two extremes would be focusing on what a user experienced for a brief moment, for example, and on the other end, focusing on cumulative experience formed through many episodes of use and periods of non-use [10, 14].

Thus, Momentary UX refers to any changes of feelings during the interaction at the moment it happens. Anticipated UX relates to the user’s expectations before the first interaction, a thought or imagination of a interaction. Episodic UX is the thoughts and assessment of one specific episode. Cumulative UX refers to the view on the system as a whole, after having used it for a while [10, 14].

In order to measure and evaluate the different time spans of UX, different approaches are used for each type. The evaluation of a user’s experience could be a difficult task. Several studies implemented questionnaires to find out about the UX while in different environments [13]. For Momentary UX, there is a complexity to its evaluation because if the user feels or experience any interference during usage, it may change the result of the evaluation. So, if the user interrupts the experience and lose their concentration to answer a questionnaire, for example, the feelings during the interaction may suffer disturbance [10, 14].

To fulfill this need of evaluation, electrophysiological measures have been used. Measures of heart rate, sweat detection, brain activity and others have been used to measure what is happening and changing related to the users feelings while experiencing an encounter with such system [15]. Hardware and software systems have been developed to capture and decode emotions through brain electrical activity. The convenience of this can be justified in a way that brain activity has direct information about emotion, therefore can be analyzed for that purpose [16, 17].

For instance, the electroencephalogram (EEG) systems offer the possibility of gathering metrics while the user is perfoming a task during the experience, without any disturbance [15]. The brain activity is a biological signal that has unique characteristics and also features that can determine patterns which are worth recording and processing. The EEG is a record of the change of electricity caused by brain activity and has been the most extensively used signal because it measures brain electrical activity. Also, it offers the possibility of higher resolution and accuracy of data when compared to other techniques, while being non-invasive - that is, can be collected by electrodes attached to the scalp [18, 19]. Also, EEG signals could be measured at any moment and they are not dependent of other activities such as speaking or facial expressions or even answering questionnaires, which could disturb the purpose of the analysis by distraction, for example [20].

The Emotiv Insight is one EEG device that captures signals using a non-invasive method. The signal is measured and processed directly by the EmotivPRO software, developed by the same creators as the device, taking out the complexity of reading and understanding raw brain activity. The software allows to capture metrics of performance - which are divided into: interest, excitement, engagement, relaxation, focus and stress; among other features. These metrics of performance are generated by EEG measures, and could be presented in real time using brain computer interface (BCI) systems [16, 19, 21].

In this context, the aim of this study is to evaluate the Momentary UX with conversational agents' interaction episodes. The analysis used an Emotiv Insight EEG device to collect brain activity and provide performances metrics to analyze aspects of Momentary UX, as well as collected other measurements, such as questionnaires, to further investigate. The specific objectives were to compare the user's Momentary experience of interaction with a chatbot using keyboard, chatbot using voice and the control group using Google's search engine; to associate the results collected from the EEG device and questionnaires.

2. BACKGROUND

This chapter provides context and background of the terms we are using in this study. In Section 2.1 we discuss UX, in Section 2.2 we discuss the Time Spans of UX, in Section 2.3 we discuss UX Evaluation, in Section 2.4 we discuss Electrophysiologic Measures and in Section 2.5 we discuss Conversational Agents.

2.1 USER EXPERIENCE

Existing definitions for user experience range from a psychological to a business perspective. Law et al. [22] states that it is intriguing that the notion of UX has been widely disseminated and quickly accepted in the human computer interaction (HCI) community without it being clearly defined or well understood. According to Roto et al. [10], there is not one single definition that suits all perspectives of it.

The great interest around UX - both in industry and academia - might be attached to the fact that HCI experts have become aware of the limitations of the usual usability framework, which focuses mainly on user cognition and user performance [9].

The multidisciplinary nature of UX has led to many definitions and perspectives, each approaching the concept from a different viewpoint. The concept of experience is built-in to our existence as people, and so with that we could say that our experiences covers everything personally encountered, undergone, or lived through [10]. UX is related, not only to the characteristics of the designed system and the interaction, but also with the user's internal state, including his or her expectations, needs, mood and interests [23].

Law et al. [22, 9] cite there are many reasons that justify the complicated matter that is it to get an universal definition of UX. One would the that UX is associated with a large collection of hazy and dynamic concepts like aesthetic, affective and emotional variables. The exclusion or inclusion of any values or attributes can be considered arbitrary considering each author's interest. A list of human factors could be considered as characteristics in this case, for example, fun, pleasure, joy, surprise and others.

Other reason would be the fact that the unit of measurement of UX can be considered very pliable. It can range from one unique aspect of one individual users' interaction with an application to many aspects of many users' interactions. And lastly, the scenery of UX research is divided by different theoretical models that are stressing diverse ideas like emotion, value, beauty, pleasure, etc.

According to Vermeeren et al. [24], the concepts of UX and usability can be intertwined. There have been attempts to demarcate boundaries in their definitions. Usability is incorporated by UX, since it tends to focus on task performance, while UX focuses on lived experiences.

UX can be considered a consequence of an user's inner state, including their expectations, needs, and also mood. Such factors are also related to the characteristics of the designed system and the context in which the interaction occurs [23]. Promoting satisfactory UX is key to the success of an application, considering that a captive user may be interested in repeating the interaction. UX is considered dynamic due to the inner emotional state of the user, which can be altered by different contexts of use during and after interacting with a product. The aspects that can influence the experience can be split into two main parts: the pragmatic quality, related to the performance of said task, and the hedonic quality, related to the values of each user and their individual perceptions [25].

User satisfaction and contentment depends on existing needs, expectations, and experiences and, therefore is subjective. Therefore, UX understanding might depend on many aspects like length of usage time, frequency of usage, resources available and product cycle [26].

UX can be thought as a technology that achieves more than just essential needs, in a way that recognizes its use as a subjective, complex and dynamic encounter. UX is a result of a user's inner state, the characteristics of the designed system and the environment within the interaction occurs [23].

In addition to that, even thought there is a large amount of factors that might influence a user's UX with a system, Roto et al. [10] cites that the factors could be categorized into three main groups: the context (both around the user and system), the user's personal state and system properties. However, in order to analyze the experience, the factors cited above cannot be used to describe the UX, but factors and their main categories could describe the situation in which a user felt a particular experience, therefore these UX factors assist in identifying reasons behind a certain experience [13].

The former ISO standard 9241 part 210 defined UX as *"all aspects of the user's experience when interacting with the product, service, environment or facility. [. . .] It includes all aspects of usability and desirability of a product, system or service from the user's perspective".* According to Hassenzahl [27], the definition was too vague to be helpful, considering it states that "includes all aspects". And also, the term desirability could produce more questions than answers.

The current ISO standard 9241 part 210 (2019) [28] defines UX as quoted by Frison et al. [29] as *"user's perceptions and responses that result from the use and/or anticipated use of a system, product or service".* And states that experience can *"[...] occur before, during and after use"* and *"[...] is a consequence of brand image, presentation, functionality, system performance, interactive behavior, and assistive capabilities of a system, product or service. It also results from the user's internal and physical state resulting from prior experiences, attitudes, skills, abilities and personality; and from the context of use."* Thus, Frison et al. [29] states that the user, the product, system or service and the context are the main components of a user experience.

Hassenzahl [27] defines UX as a *"momentary, primarily evaluative feeling (good-bad) while interacting with a product or service"* and justifies that definition by stating that the attention should be on humans and feelings, the subjective side of product use and not only on product and materials, such as content, functionality and presentation.

2.2 TIME SPAN AND UX

Roto et al. [10] point out the dynamic and multidisciplinary nature of UX, the leads to many perspectives and definitions on it. The authors highlight the importance of thinking of time spans when analyzing UX. The actual experience does not cover all relevant characteristics of UX. Some aspects of UX do not belong within the actual experience of usage. Time spans matter in specifying UX [14].

With that in mind, we can consider different time spans when talking about the user experience. UX can refer to:

- a specific feeling during the interaction, which is referred to as Momentary UX,
- the appraisal of a specific experience episode, which is referred to as Episodic UX,
- views and consideration on a system as a whole, after several uses, known as Cumulative UX,
- a period before the first use, or any of the other time spans, an experience episode or any moment after taking a system into use, known as Anticipated UX.

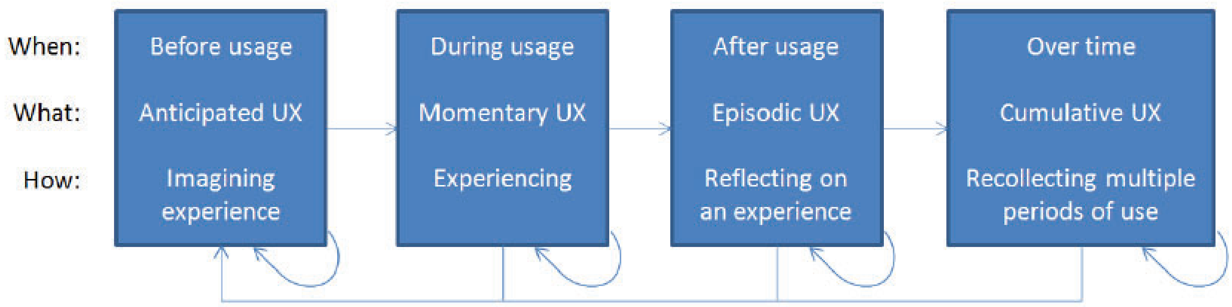

Figure 1. Figure by Roto et al. [10] representing UX time spans.

The Figure 1 by Roto et al. [10] represents graphically the time spans of UX. When they happen regarding the specific experience, how they are described and their respective titles. The subfigures also shows with directive arrows how they can be cyclic.

User expectation is one of the factors that can influence UX. Expectations can be created by other people's opinion or advertising and they can impact the way a user approaches a system. Users create expectations on a product or system even before their first encounter. Hedonic and pragmatic qualities of a said system can play an important role in determining responses related to momentary feeling during usage. Users may change their opinion of the product at the first encounter or during the actual experience [30].

According to Abro et al. [30], the notion of time is considered as a critical factor to influence UX. There is no experience without time, therefore time and experience are closely linked with each other. Time in UX can be classified into two: the physical and the psychological. The first is time that can be measured with a clock and the second reflects how the user feels time, through the person's sense of estimation.

Unique episodes of experiences result in a reflection of the experience itself. Said reflection usually represents the user's general impression of a product and many aspects are taken into consideration when thinking back on the total experience of use. The outcome of Episodic UX is not necessarily the same as a summary of all Momentary UX, because the perception of usage might change over time. When thinking back of the entire experience, people can select only some elements to remember, either positive or negative, and those will determine the overall opinion on the product [14].

Focusing on the time span of UX can help identify a user's emotional responses with accuracy. When focusing on longer periods of time, the impact of momentary experiences may reveal their impact on cumulative UX. Roto et al. [10] uses the example of how important a strong negative reaction can be. It may decrease after a successful outcome, and the overall reaction may be remembered differently.

When the focus in only on the Momentary UX, the demands on design and assessment are different than a focus on longer time spans. For long periods of usage, UX can be seen as a lifecycle or journey, from the first encounter, through episodes of use and reflection on use. Previous experiences will have an impact and influence future experiences, thinking back on one experience will create expectations or anticipations for future ones.

Abro et al. [30] stated that momentary experiences can give details on a users's emotional feedback. Therefore, it is important to clarify the time span, while assessing UX of interactive systems in a particular context.

2.3 USER EXPERIENCE EVALUATION

In order to promote good experiences, it is necessary to understand the user's feelings. Checking for efficiency and the usability of the product is not a way to completely understand how the users feel the experience. Several studies related to UX evaluation have been written using both qualitative and quantitative analysis, so that we understand what the user feels while using the product and can not only measure the experience but also improve it [31].

The abstract nature of UX makes it hard to guess what will be the outcome of design decisions on the experience. It could be very complex for the design and development teams to solve issues like social, emotional or aesthetic in a direct or explicit way. The design team would need to identify applicable methods, tools and criteria that could be used to manage the UX factors throughout the process of development [10, 22]

Given the complexity of UX, understanding how the user grades the experience is a complicated task. To evaluate the UX, there is need to collect data related to aspects of it. Some instruments used to collect these data are questionnaires or surveys, psycho-physiological measures and observation [31, 32].

There is no overall accepted measure of UX, but there are many different ways to assess it. There are evaluation methods that focus on whether an evoked emotion is positive or negative. There are methods developed for evaluating unique UX qualities such as frustration, satisfaction or fun. There are questionnaires evaluating the user's perceptions towards different aspects of the product and the experience. There are scales for effort required to finish a task, cards used to describe feelings very quickly, and the list goes on [13].

There are psycho-physiological measurements used to evaluate UX. Measurements like heart beat, skin perspiration and facial motion can contain valuable information about the emotional state of the user. The physiological reactions can be recorded by attaching sensors to the participant. This objective data can be used in combination with self-report data to find out what the user experienced [33].

Zarour and Alharbi [25] performed a research in order to synthesize findings related to UX aspects and dimensions and identify measurement methods to merge into one simplified UX theoretical framework. The authors express that their proposed framework is crucial for practical applications, the development of evaluation methods and further theoretical studies around UX. In the study, there is some evidence that an UX measurement method is a method to measure experience aspects and to gather information about the fulfillment level of a certain aspect, depending on the product evaluated.

Väänänen-Vainio-Mattila et al. [34] stated that a development centered in the user (User Centered Design or UCD) is the key to promoting good UX, because understanding user's needs and values is a good foundation for designing and evaluating possible solutions.

Isomursu [35] states that the UX is a subjective state and it does not have an objective reference, which therefore cannot be objectively measured. One user's experience cannot be experienced as such by another user. It is very complex for humans to compare experiences when they are separated by time. The human memory on experiences is unreliable, consequently also is recalling past experiences to compare with others or describe them reliably after time has passes.

When it comes to measuring UX, the author draws attention to the importance of evaluating before (the user's expectation), during (the experience) and after product usage (the user's judgment). With that remark, the changing and subjective dynamic nature of UX is clear: the expectations influence experience, experience influence judgments and these judgments will set new expectations for further expectations and so on.

These expectations can be linked to the emotional state of each user. Hole and Williams [36] proposed an evaluation method centered on emotion sampling. Users were repeatedly asked about their current emotional state while using a product by going through a number of questions. This method focused more on the experience itself. In the context of product development and measurement by this method, we would have to go further and establish a causal link between experience and product to know how the product affects the measured experience. So the focus of the evaluation is not the experience itself but the experiences cause by the product. [34]

Vermeeren et al. [24] analyzed methods for different time spans of UX. When discussing about the Momentary UX and its evaluation methods, the authors concluded that the methods represented a range of techniques that included: questionnaires, self-reporting questionnaires, thinking out loud techniques of observation and psycho-physiological measures.

The cited evaluation methods were said to have strong qualities for reasons related to validation, that is, combining at least two methods together. The study reported that physiological measures were described as being non-disruptive and optimized the results because of the lack of disturbance. The weakness informed was related to issues of practicability or feasibility, such as expertise team, specific equipment and required software and difficult data analysis, because of the evaluation instruments used [24].

2.4 ELECTROPHYSIOLOGIC MEASURES

Memory and emotion are parts of the user cognitive system, being inherent factors instinctively activated, and not perceived by the user, unlike factors that capture attention, such as interface elements or user's choices, both equally important for evaluation. Additionally, UX is a combination of aspects, and it does not happen on the system, it happens in the user's mind [27].

In recent years, the study of emotions has increased due to the growing need for applications capable of detecting the emotional state of users. An extensive part of the research in this area has been dedicated to detection of emotions. Under controlled situations, emotion detection computer systems are able to classify emotions with considerable accuracy [16].

When considering how facial and voice information are related to behavioral expression, we can conclude that those can be consciously controlled and modified, and therefore the interpretation is often subjective [16]. Nonetheless, emotions are not necessarily manifested by means of facial expressions nor voice information. So, approaches that may detect emotions have been proposed, such as skin conductance, breath, pupil dilation, heart rate and sweat detection.

Many forms of detection for emotion identification are being used. A used measure is skin conductance, that assesses the conductivity of the skin which is increased by stress. In this case, the value of skin conductance is also associated with levels of excitement [18].

One more measure is the breath. A slow breath is associated to relaxation, while a irregular rhythm, filled with variations of respiration associates to more aroused emotions such as anger or fear. A further, measures of blood pressure and heart rate are variables that relate with defensive reactions, pleasantness of a stimulus and basic emotions. Significant changes in body temperature can be observed when dealing with stress [18].

The electroencephalography is the used signal to assess brain function by a set of measurements of electrical potential differences between pairs of electrodes on the scalp. The EEG signal does not capture the activity of a single neuron, but reflects the interaction of millions of neurons in the brain [18].

According to Ramirez [16], the measured electrical activity in an EEG is distorted by the tissue and skull between the electrodes and the neurons. This induces noise and reduces the intensity of the recorded signals. In despite of this, EEG measurements still provide important insight into the activity of the brain.

EEG measurements frequency ranges from 1 to 80Hz. The signal frequencies are divided into bands, because specific frequencies are normally more notable in particular states of mind. The most important frequency waves are the alpha waves (8-12Hz) and the beta waves (12-30Hz). Alpha waves usually happen during awake relaxed mental states. Beta waves are related to an active state of mind, usually during intense focused mental activity. Alpha and beta wave activity can be used to detect emotional states of mind in humans [16, 18].

Brain Computer Interface (BCI) devices enable the connection between the human brain and a computer by capturing and analysing EEG signals to use them for controlling external devices or measuring performance metrics. Figure 2 illustrates how the BCI works. First, the brain activity is measured by the electrodes attached to the head. Later, the measured signal gets amplified and usually receives noise treatment. After that, the signal is ready to be processed with the system's algorithm of choice. And lastly, it is delivered to the BCI application.

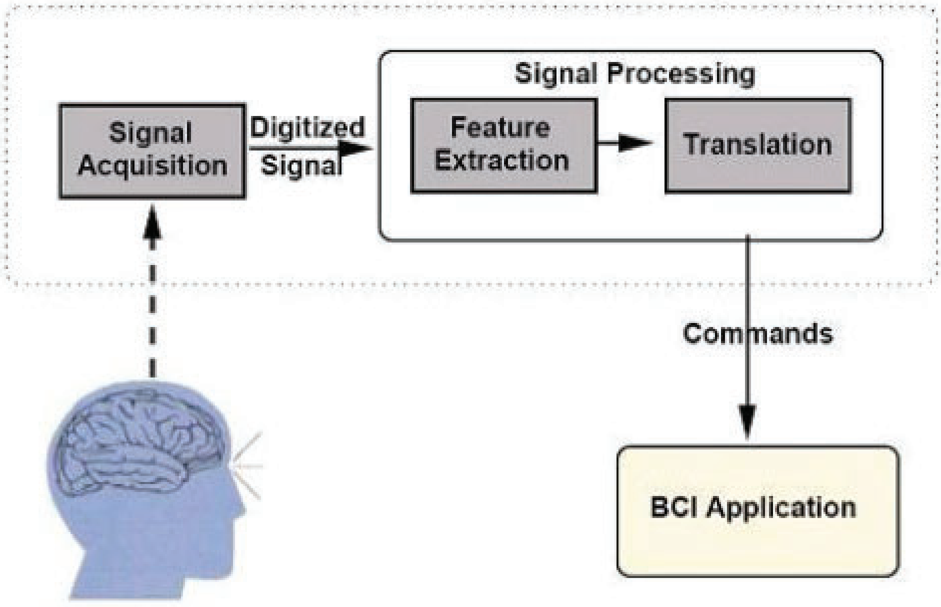

Figure 2. The workflow of a BCI system [17].

The BCI Emotiv Insight device captures signals using a non-invasive method by measuring from the scalp surface. The EEG signal is measured and processed directly by the EmotivPRO software, developed by the same creators as the device. The software allows to capture the so called metrics of performance which are definied as:

- interest,
- excitement,
- engagement,
- relaxation,
- focus, and
- stress.

These metrics of performance are obtained from the EEG brain activity and processed by the EmotivPRO algorithms for cognitive states and presented in real time [21, 19, 16].

Since portable wireless devices became more available, the recognition of the user's emotions from EEG signals is attracting more attention [37]. There are different models of recognition, they use brain waves and patterns. The valence level represents a quality of the emotion - from unpleasant to pleasant -, and the arousal level reflects a quantitative level - from not aroused to excited. Russel [38] proposed the two-dimensional Arousal-Valence model, Liu and Sourina [37] used the Valence-Arousal-Dominance, in which the valence level, arousal and dominance are combined to form a total of eight emotions.

There are an increasing number of algorithms to recognize emotions, they are composed of extraction and classification. The combination of such values results into emotions.

The EmotivPRO software uses an algorithm to create the results of the performance metrics scaled from 0 to 100. The algorithm includes 6 metrics and are explained as it follows. The stress is a measure that representes the comfort with the current challenge. The higher the stress, the higher the difficulty in completing a task. A low level of Stress could improve productivity. [21]

The engagement metric is described as the conscious direction of attention towards a task. It relates to alertness and measure the level of attention and engagement in the moment. The engagement contrasts with boredom, the higher the attention and focus, the higher the engagement metric. The interest is a level of attraction to the current moment, environment or activity. The interest can be referred to as Valence. A low level of Interest means aversion to the task, and contrariwise. [21]

The excitement level is a feeling of physiological arousal with a positive value. The greater the increase in arousal and awareness, the great the output for excitement. The measure of attention to a specific task is the measure of focus. The focus metric measures the depth of attention and how frequently the attention changes between other stimuli. A high level of switch of attention between tasks or stimuli indicates low focus and distraction. The last of six defined metrics is Relaxation and it is a level of ability to disconnect and recover from intense concentration. [21]

2.5 CONVERSATIONAL AGENTS

The act of searching is a dominant model for interacting with the web. Since the search engine's invention, we have become a searching culture. Users use search in the daily life for the biggest variance of needs. Recently, a new platform has been named to replace mobile apps and web sites: conversational agents [39].

Software applications that take part in text based conversations with users using natural languages are conversational agents. Users manifest a growing interest in CAs for productivity tasks, entertainment, and communication. The most successful applications in usage include customer assistance, small talk, search tools and virtual assistants: Amazon's Alexa, Apple's Siri, Microsoft's Cortana and the Google Assistant are some of them [3].

Wilks [40] suggested a way of differentiating characteristics to select the terms. From the author's perspective, conversational agents are distinct by their function which is to carry out tasks. Chatbots have no memory or knowledge but mimic conversation. Virtual assistants mimic conversation and are able to keep personal knowledge of the user and attempt to form a relationship with the user. Bors et al. [41] corroborates by stating that chatbots are usually used as an interface and are task oriented, while virtual assistants are user focused.

The most common interface for assistants looks like online messaging, something that the majority of users are used to daily. The interface patterns recall the usual sending and receiving messages with other people. Recently, platforms allow chatbots to access features like pictures, sounds and also access information like geographical position and further online information from specific sources [3, 39].

These conversational agents such as virtual assistants are becoming pervasive in most of devices to perform a variety of functionality [2]. The users may interact with their devices through a virtual assistant with a conversational interface. They may search for content, set a calendar appointment and other tasks [42].

According to Gruber et al. [43], if users are busy, distracted, or operating a vehicle it may be difficult for them to interact with their devices effectively, in theses cases the assistant that may be accessed with voice can be helpful. This assistant should provide an interface between the human and computer. The user may engage in a form of dialog with the assistant, either using voice or typing directly into the keyboard as in a chat, and ask or demand for tasks, information, etc.

Bors et al. [41] justifies the growing attention to CAs by stating that people are usually more comfortable when talking with each other than when operating technology, because conversation is natural and innate to humans, therefore users might opt to interact with technology in a more natural and traditional way.

Shawar [44] asked participants in a study to choose between using a chatbot or Google's search engine to perform search tasks and 47 percent chose the chatbot. Among the participants' answers, participants who chose the chatbot said that it was preferred because the chatbot had the ability to give direct answers, while Google would provide a page of links available; and Google would return more links, which would increase the browsing and searching time. The participants who preferred Google stated that their prior familiarity with Google was a main aspect of choice, since they knew exactly how to use the engine. And the other aspect was that it seemed more difficult to choose keywords to the chatbot.

Liu and Dong [45] conducted an experiment with a chatbot designed to assist students in an University. The chatbot was designed to act like an assistant and answer questions, route navigation, set appointments, etc. In the results of this experiment, the tasks with navigation showed an increase in the satisfaction in comparison to the mobile map app. However, the study suggested that users had expectations towards the chatbot and would lose patience quickly when it failed to prompt the response.

These user interfaces attempt to mimic the way humans communicate with each other in order to make the interaction flow easier and more natural. There is a variety of features in which the interface can make the user feel more like talking to another human or to a tool [46].

When users described the use of CA, they reported the use of a different language in a way that they chose to leave unnecessary words behind, choosing only keywords. Also, removing colloquial speaking, reducing the number of words used, speaking more slowly and enunciating clearly were some of the tactics described by users. Several users reported that they tried to engage a dialogue with a CA as if it were a person, and most of them said they would not do it in public [47].

3. METHOD

The study was organized according to the guidelines laid down in Wohlin et al. [48] for experiments, including the phases of scoping, planning, executing, analysing, and presenting results, along with the analysis and interpretation of the results to form conclusions.

In Section 3.1, the Research Question and plans regarding sample and variables are defined. In Section 3.2, there is a guide to how the experiment was conducted.

This study and all its procedures involving human subjects were approved by the ethics committee of the University of Passo Fundo under number 3.405.684. All participants signed an Informed Consent Form.

3.1 PLANNING

During the planning, the hypotheses are formulated, along with the definition of participants selection and the variables that will be used. Those elements together defined will result in the preparation of the experiment to be executed. We defined a research question to guide our study and be validated with the results.

- RQ1: Is there noticeable differences in the brain response of the users while interacting in different ways?

To answer our RQ, we prepared ways of measuring the signals from each user while interacting differently. Not only the EEG is going to present results of their brain waves and its emotions in each moment of the experiment, but also the participants will be able to present their sensations through questionnaires (listed and explained in Subsection 3.1.3).

3.1.1 Sample

We selected 36 participants from 23 to 55 years old up who had degrees in the field of Computer Science or Engineering as they may be more familiar with the provided tools to perform the tasks.

The participants were distributed in three groups by simple randomization, entirely by chance with no regard to the will of participants' condition nor preference [49].

Every group of participants searched the same items whilst interacting in one of three different ways as follows:

- Group 1 was the control group, using the Google platform on the Google Chrome browser

- Group 2 was using Google's Assistant using the device's keyboard
- Group 3 was using Google's Assistant using voice command

Group 1 was the control group who had no interaction with the assistant. Groups 2 and 3 interacted with the assistant, but in different ways: using keyboard and using voice.

Subjects also answered a background questionnaire characterizing the sample, detailing gender, age, and previous experiences regarding search engines and virtual assistants.

In the background questionnaire there were questions regarding participants' age, gender, last graduation and field of work. They were asked how frequently they used search engines and if they were used for every day activities or work; if they had any experience with virtual assistants or chatbots and what kind of experience they had with it.

3.1.2 Testing Environment

To perform the study, we secured a test environment previously assembled in a study room. In this room, there is a desk facing a wall that provides little interaction and close to no visual pollution. The participants were asked to seat facing the desk and the wall. The place has adequate ambient light and no sound pollution. The room temperature was preserved at 23°C (73.4°F) in all sessions.

Figure 3. Participant inside the environment created for the testing activities.

Hence the participants were in a controlled environment, they had to wear the helmet for 20 seconds to collect their initial performance metrics.

3.1.3 Assessment Frameworks

All participants filled out a background questionnaire, which was used to gather information on their previous use and experience of the tools and to characterize the sample.

We collected a variety of information while observing participants interacting with search tools in three different groups. The emotion recognition system used recorded EEG data using the Emotiv Insight wireless headset.

The helmet captures brain waves through the activity of neurons, it is a non-invasive method from the surface of the scalp. The helmet is connected to the EmotivPRO application where the waves are processed in real time and merged into performance metrics [17, 21].

We used the EmotivPRO software for acquiring the EEG data from the Emotiv headset while the participants were performing the tasks. This software is used to display real time data streams such as raw EEG and Performance Metrics, as well as contact quality and the battery level of the wireless device. The software also enables to record the metrics during a moment of time.

Figure 4. Participant wearing the Emotiv Insight helmet.

The EEG signal is measured and processed directly by the EmotivPRO software that was developed by the same creators of the helmet. The software allows us to capture the metrics of performance definied as follows: interest, excitement, engagement, relaxation, focus and relaxation [19].

We do not have direct access to the algorithm relating to the emotional parameters to raw EEG data due to intellectual property rights, the producer's algorithm is not open source. Nevertheless, studies have established the reliability of Emotiv's EEG performance and detection of emotions [16, 19, 18, 20, 50].

The Emotiv has 5 electrodes (hydrophilic semi-dry polymer) locating at channels AF3, AF4, T7, T8, Pz and two additional sensors that serve as CMS and DRL reference channels (on the left hemisphere of the head). The Emotiv Insight uses 5 data channels to collect brain data. The sampling collected rate is 128Hz, the bandwidth is 0.5-43Hz, and the digital notch filters are at 50Hz and 60Hz [21].

Figure 5. Emotiv Insight and its electrodes.

After each task, the participant had to fill out a Likert-based questionnaire, used by Mandryk et al. [51], corresponding to how they felt regarding certain emotions during the task they had just finished. This questionnaire collects information about how the participants perceived the episode of experience they just had, their Episodic UX. The term "episodic" suggests that there was a beginning and end of an interaction event and, therefore, can be used as instrument to evaluate the experience passed, when users rethink of their Momentary UX. [52]

Bargas-Avila and Hornbæk [13] looked into the the most used way to collect data in UX studies and states that the most recurrent pattern was a combination of measurements during and after the experience. Mandryk et al. [51] states that while physiological measures can be used as objective indicators to measure UX; the subjective responses may not correspond to the actual experience, since they are cognitively mediated, they may not accurately reflect what is occurring. Also, according to the authors, when being recorded, participants may sometimes answer what they think they are supposed to, not being aware of it.

That said, we can, very cautiously, expect to trace similarities between the overall answers in the questionnaire and the overall results from the EEG. To the overall perception of the words presented, we can expect that the average person would relate the emotions on the questionnaire with their feelings like below:

- The higher the Stress, the more frustrated the participant felt;
- The higher the Engagement, the less Bored the participant felt;
- The higher the Interest, the less Bored the participant felt;
- The higher the Excitement, the more Fun the participant felt;
- The higher the Focus, the more Challenged the participant felt.

Therefore, the Stress, Excitement and Focus collected by the EEG would be directly proportional to, respectively, Frustration, Fun and Challenge answered in the questionnaire. While, however, the Engagement and Interest would be inversely proportional to Boredom.

They also filled out a SUS questionnaire [53] at the end of the experiment, which was used to evaluate the overall experience regarding their opinion on the tool's usability.

3.2 EXECUTION

The following subsection 3.2.1 explains how the experiment was conducted and the collecting of data.

3.2.1 Testing Activities

The term of knowledge and acceptance and written informed consent was first handed to each of the participants so they could read and decide whether they wanted to be part of the study. After agreeing with the term, the participants had to sign to confirm the consent. After that, the test could begin. At the beginning of each test, the charge on the Emotiv Insight was checked, the participants had the headset positioned onto their heads and connected to the software.

Our goal was to create different experiences for each group while using the same tasks so that the interaction was the main change for analysis. Group 1 was the control group using Google's search engine web site, Group 2 and 3 used Google's Assistant, the first using keyboard and the second voice.

The participants received the tasks explained in one sheet of paper, along with an Apple iPad 2 unlocked, connected to the internet, and with the opened application to be used to fulfill the tasks.

The tasks were the same for all three groups. The participants were asked to search for an specific term first, and secondly read out loud a brief paragraph the participant felt summed up the term's definition.

The participants were not given instructions regarding the time of each task, they were told to feel comfortable and that they had as much time to finish the tasks as needed. The tests ranged between 5min and 19s and 11min and 6s.

The tasks were handed to the participants as it follows:

1. 1.1 Search *high blood pressure*
   1.2 Read out loud a brief phrase or paragraph when you feel you found one that best summarizes the definition you were looking for.
   1.3 Search *treatments for high blood pressure*
   1.4 Read out loud a brief phrase or paragraph when you feel you found one that best summarizes the definition you were looking for.
2. 2.1 Search *diabetes and its main types*
   2.2 Read out loud a brief phrase or paragraph when you feel you found one that best summarizes the definition you were looking for.
   2.3 Search *treatments for diabetes*
   2.4 Read out loud a brief phrase or paragraph when you feel you found one that best summarizes the definition you were looking for.
3. 3.1 Search *chatbot*
   3.2 Read out loud a brief phrase or paragraph when you feel you found one that best summarizes the definition you were looking for.

The tasks' results were previously searched to adapt the task and foresee the results. The tasks in number 2 were more complex to find a single result and a brief definition, therefore, the tasks 2.1, 2.2, 2.3 and 2.4 were supposed to take longer to be completed, because the participant had to search beyond the first result, unlike tasks inside number 1 and 3. Task number 3 was supposed to be the fastest one of the bunch.

After each group of tasks the participant had to fill out a likert scale about their Episodic UX [51], see Table 1. The participant rated Boredom, Challenge, Frustration, and

	1 Lower	2	3	4	5 Higher
Boredom					
Challenge					
Frustration					
Fun					

Table 1. Table representation of the likert scale each participant received to fulfill regarding their thoughts on the task they had just finished. Each participant received three scales, one for each bundle of tasks.

Excitement on a scale from 1 to 5, corresponding to how they felt during the task they had just finished.

After all tasks, the participants were asked to answer a SUS questionnaire [53], regarding the overall experience using the application.

The EmotivPRO software shows the EEG and calculates the performance metrics in real time as it is shown in Figure 6, a screen capture of the system during one of the participants' recording, and the black vertical markers are the time stamps of each task.

The results of each metric at each task were collected like shown in Figure 7. Time of task was crossed with the performances' metrics and the results were put into tables.

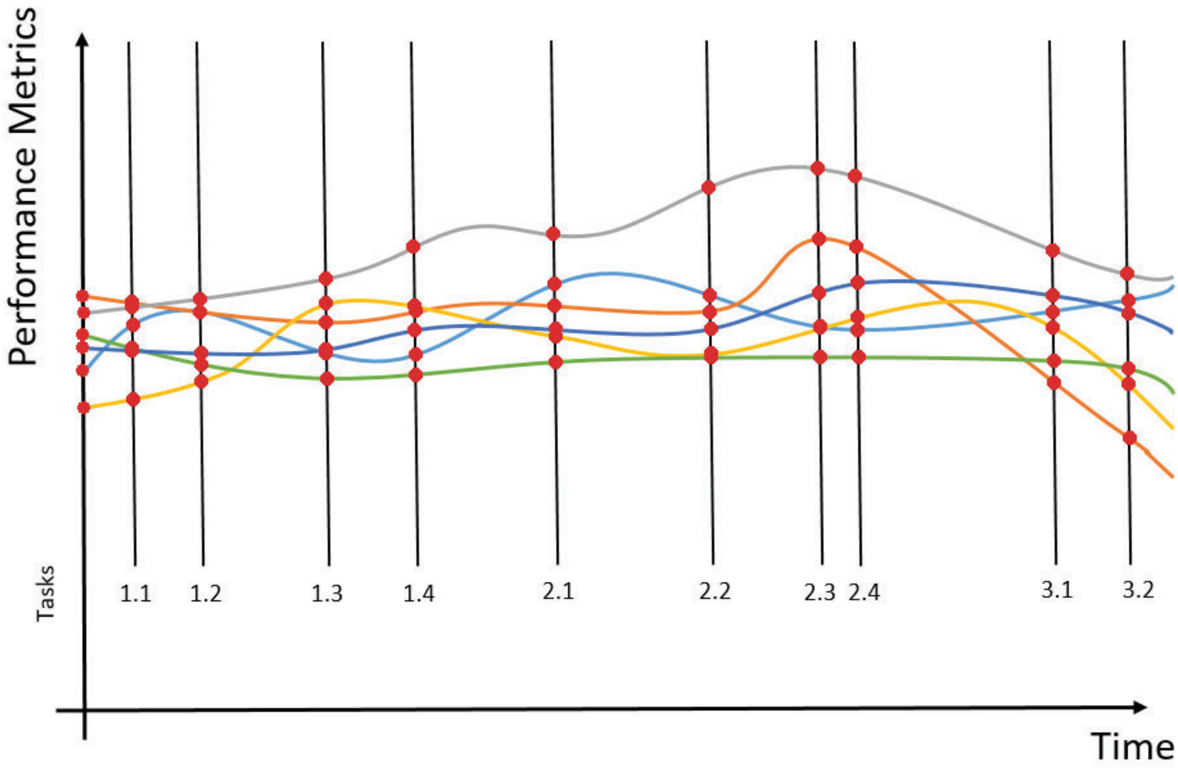

Figure 7. Figure representing how the recordings and the time stamps of data were collected. The lines are the performance metrics changing through time during the test and the red markings are when the performance metrics values were collected: at the time of each task.

During the session, there was an observer collecting timing data from the participants' tasks. The exact time of each beginning or ending of a task was taken into consideration, as well as other variables that may have occurred during the session. These timing data were later compared to the EEG data collected at the same time of the recorded session, as shown in the representation in Figure 7.

In Figure 6 we have a screen capture of one of the tests presented. The colored lines are presented by the software and they represent the variations of all six defined emotions, in order from top to bottom: Stress, Engagement, Interest, Excitement, Focus and Relaxation. The graph of emotions grows larger into the X axis as time passes by, whilst that, the observer marks the time stamps of each task, so that we can later capture the emotions felt during each one.

There were no interventions or interactions made by the observer, in order to avoid bias.

3.2.2 Statistical analysis

Each one of the 36 participants performed all 10 tasks, and each task carried its own 6 values regarding the 6 emotions of that time stamp. Also, each participant filled out 3 likert scales regarding their Episodic UX and also a SUS questionnaire about the entire experience, regarding the used tool's usability. We used descriptive statistics to summarize our data.

The Engagement variable from the EmotivPRO software measures the level of attention and concentration on the task, which is, how much of each participant was into the given task. Berka et al. [54] reported that the electrophysiological engagement measure was related with task demand, including the requirement of needed attention and the level and complexity of task processment.

McMaham et al. [15] explored results that suggested that electrophysiological engagement measure reflected information about visual processing, and attention allocation. Thus, we used engagement as the independent variable because the levels metric reflect information about attention and may be the constant representation of the user's immersion to the experience of the task [54, 15].

For the descriptive statistical analysis, we paired the variables to Spearman's correlation. The selected independent variable chosen was Engagement, to be paired with all the other five metrics, at the same time stamp and divided by Groups. To rank the correlation between EEG variables, we used Engagement as the independent variable to associate with all the other variables, while considering their values for the same task at the same time stamp. We generated box plots and scatter plots to analyze the variables.

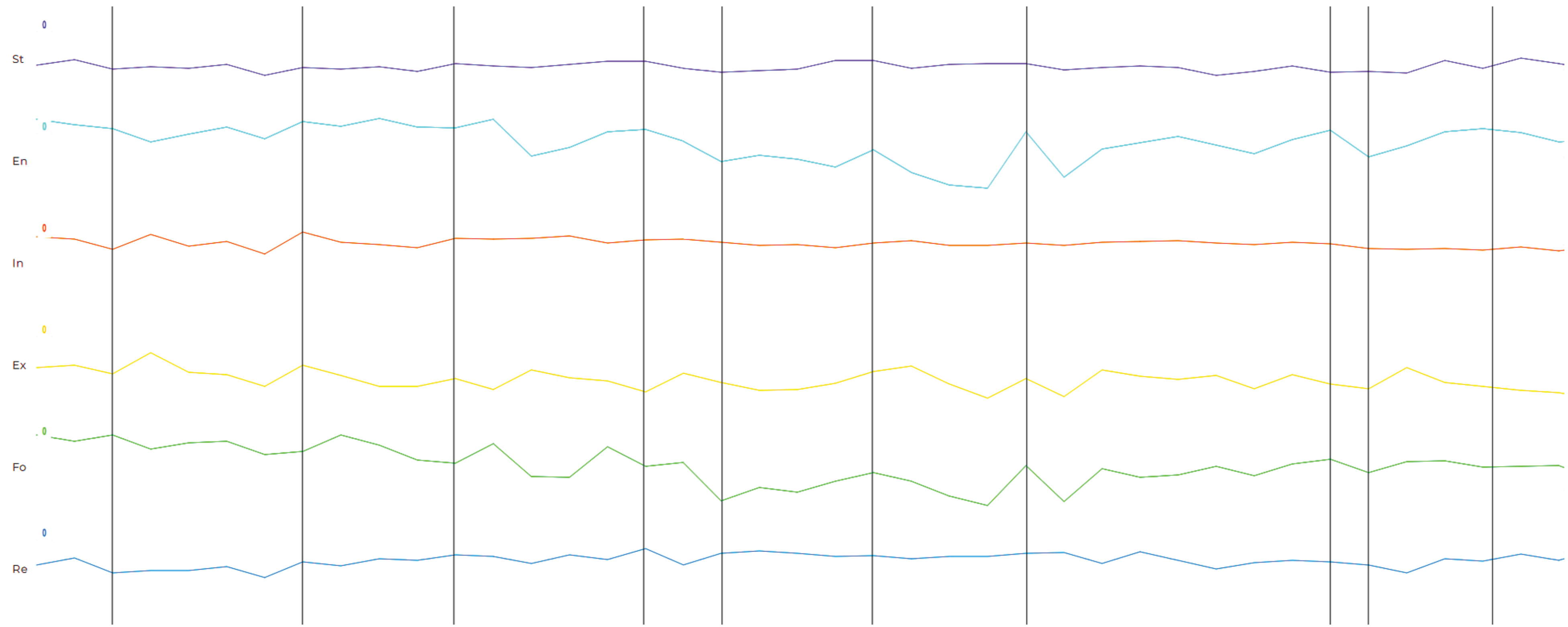

Figure 6. Figure of a screen capture of the EmotivPRO during one of the participants' recording with the time stamp as a black marker on.
From top to bottom the metrics: stress, engagement, interest, excitement, focus and relaxation

4. RESULTS

EEG data was collected from 36 participants, 12 for each group, ranging from ages of 23 to 53 (mean age = 30.83, σ = 7.18), see Figure 8. Groups 1 and 3 had equally 50% of men and women, while Group 2 had 66% of women and 33% of men. All participants reported they had no brain or cognitive dysfunctions.

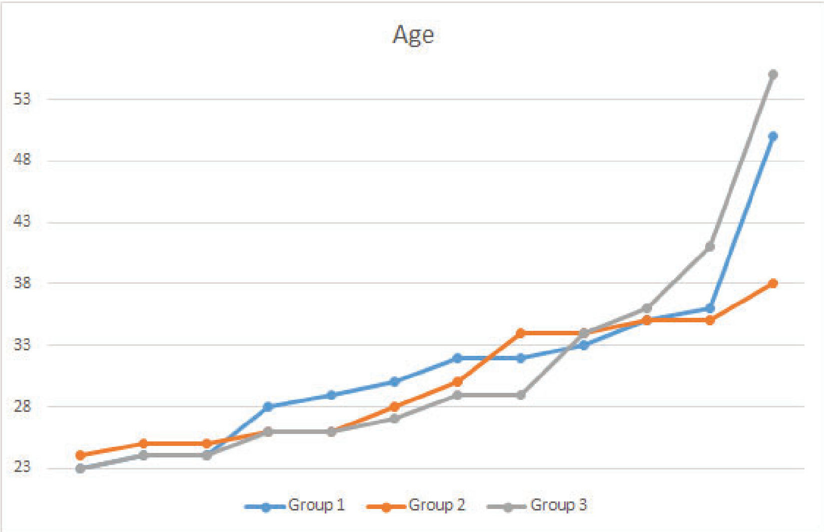

Figure 8. Graph representation of the sample's age divided by Groups.

Participants were asked how frequently they used search engines, all of Group 1 answered that they used it daily; 91% of Group 2 said that they used it daily and the other 9% chose frequently; in Group 3, 83% said daily and the other 17% chose frequently.

When asked about the use of search engines, 91% of Group 1 said they used it both in their every day lives and work, whilst 9% said only work; all of the participants in Group 2 said they used it both in their every day lives and work; and in Group 3 there were 33% saying they used it for work, while the other 66% chose every day life and work.

When asked about if they had any experience with chatbots, 50% of the participants in Group 1 answered Yes and the other half said No. In Group 2, 58% answered Yes versus the 42% that said No. In Group 3, 50% of the participants had had some kind of experience versus the other half.

The participants that said they had previous experience with chatbots were asked about what kind of experience they had. From the participants that said they had experience in Group 1, 16% said they had experiences both as user and developer, and the other 84% said they had experiences only as a user. In Group 2, 8% of the participants said they had experiences as a developer while 92% had experiences as a user. In Group 3, 25% of the participants had developed chatbots while 75% had only been users.

The participants answered a likert scale regarding how they felt after each bundle of tasks and a System Usability Scale, regarding the system they had just used. Therefore Group 1 answered the SUS about the Google Search engine, Group 2 about using the Assistant with keyboard and Group 3 about using the Assistant with voice interaction. The mean of the results are on Table 2.

	Group 1			Group 2			Group 3		
	Task 1	Task 2	Task 3	Task 1	Task 2	Task 3	Task 1	Task 2	Task 3
Boredom	2.083	2.333	2	1.833	1.583	1.333	1.667	1.833	2
Challenge	1.417	2.25	1.5	1.917	2.5	2.167	2.167	2.75	2.667
Frustration	1.5	1.833	1.667	1.833	1.833	1.833	1.667	2.75	2.333
Fun	3.167	2.417	2.917	2.167	2.25	2.667	2.583	2.5	2.833
SUS		83			85			76	

Table 2. Table containing the sample's Episodic UX likert scale results divided by tasks and SUS questionnaire results, all divided by Groups.

For each participant, there were 66 collected values, representing every one of the 6 metrics provided by the software, in 10 different time of tasks plus the initial markings collected. For all results, the mean, standard deviation and percentiles was calculated, the results of the means are represented in Table 3.

For each performance metric, we generated box plot charts to represent the minimum (the lowest data excluding outliers), the maximum (the highest data excluding outliers), the sample median, and the first quartile (the middle value between the smallest and the median of the dataset) and third quartile (the middle value between the largest and the median of the dataset). Any data not included between the above said was plotted as an outlier with a dot. The groups were represented side by side by different colors and the metrics are shown how they changed through time in Figures 9, 10, 11, 12, 13 and 14.

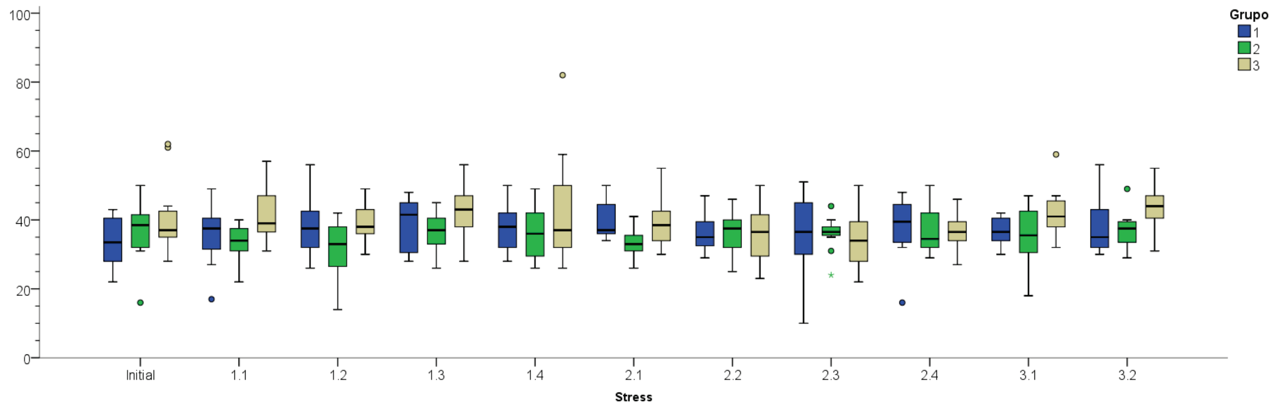

Figure 9. Figure representing the Stress levels collected in each task, divided by Groups, in a box plot chart.

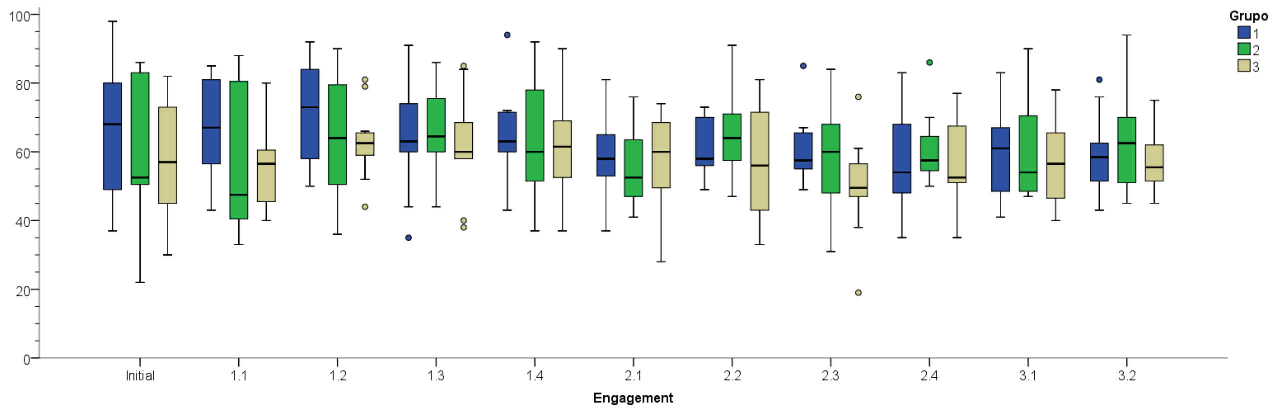

Figure 10. Figure representing the Engagement levels collected in each task, divided by Groups, in a box plot chart.

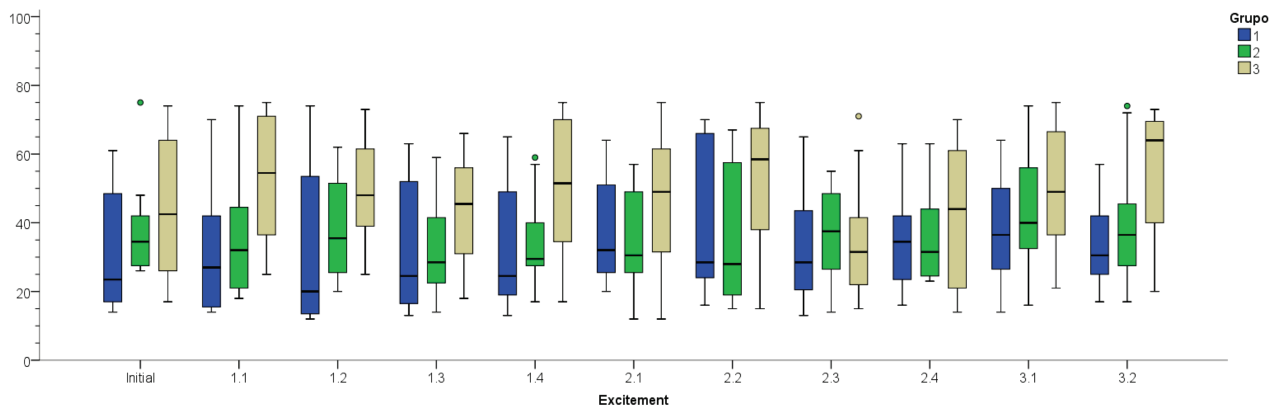

Figure 11. Figure representing the Excitement levels collected in each task, divided by Groups, in a box plot chart.

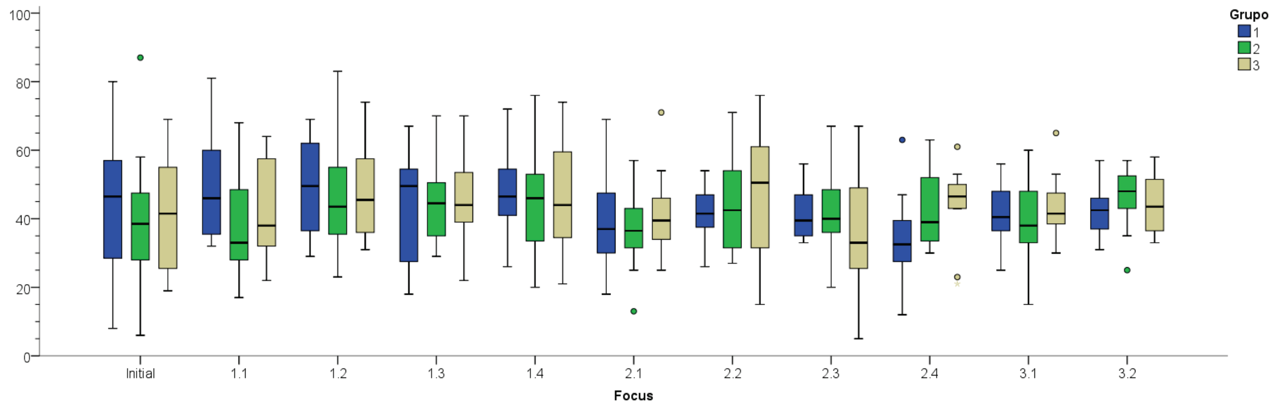

Figure 12. Figure representing the Focus levels collected in each task, divided by Groups, in a box plot chart.

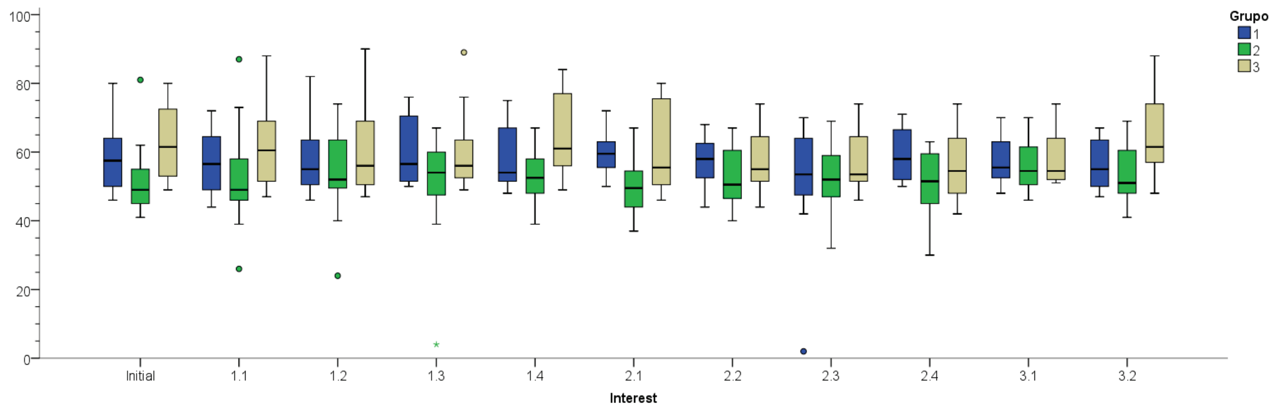

Figure 13. Figure representing the Interest levels collected in each task, divided by Groups, in a box plot chart.

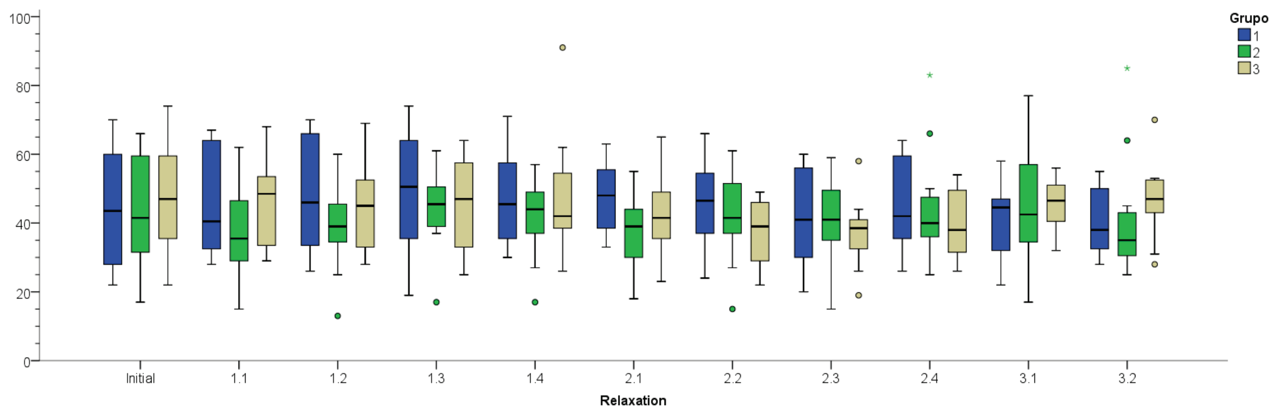

Figure 14. Figure representing the Relaxation levels collected in each task, divided by Groups, in a box plot chart.

Statistical homogeneity was found between groups in tasks 1, 2 and 3, therefore, tasks 1.1 and 1.2 were selected to represent the correlation analysis.

The scatterplots in Figures 15, 16 and 17 represent the five correlations at the first three time stamps, respectively: Initial, task 1.1 and task 1.2. Respectively, in correlation to Engagement: Stress, Interest, Excitement, Focus and Relaxation.

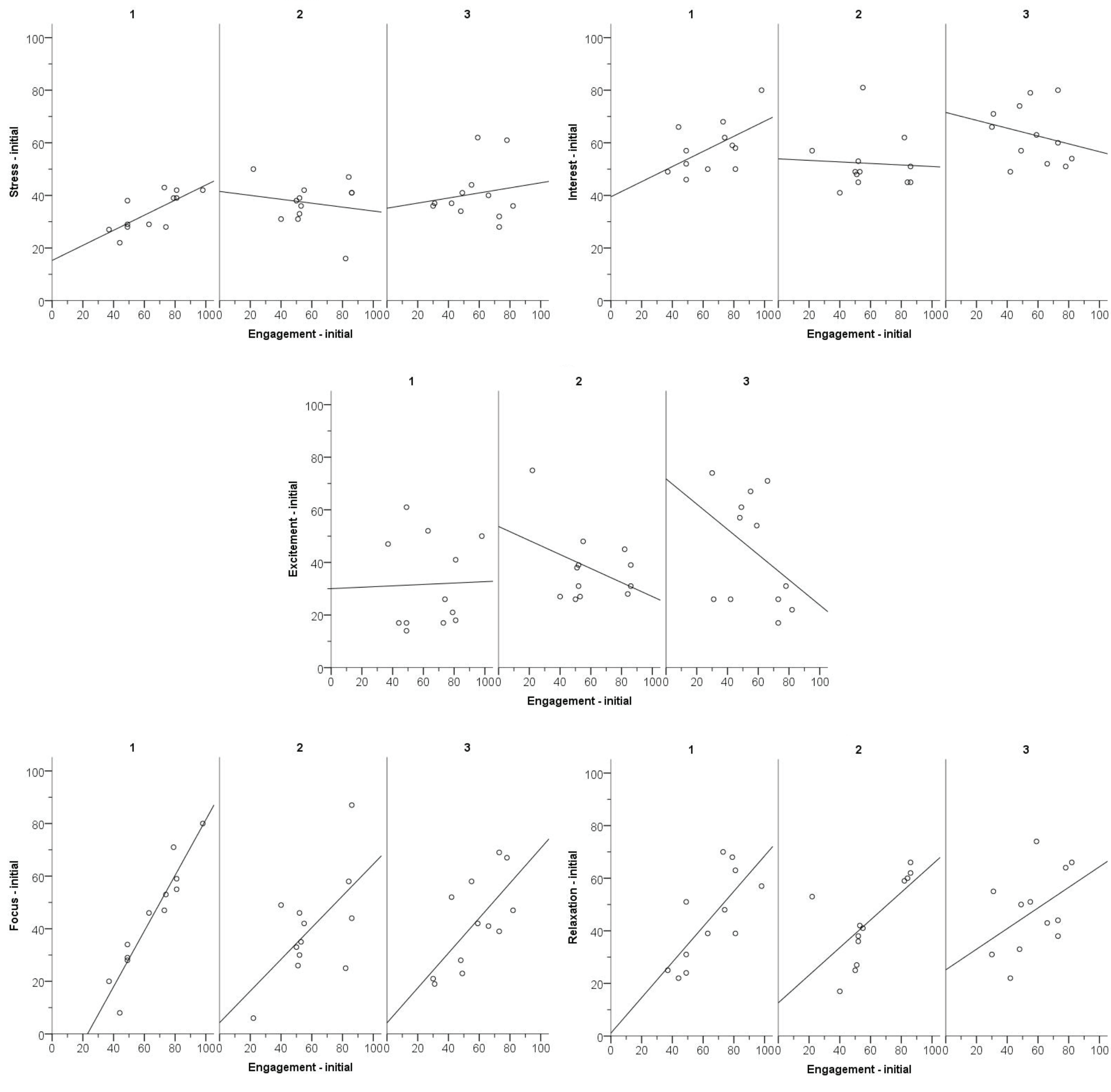

Figure 15. Figure with the bundle of correlations of variables of the time stamp Initial.

In the Initial correlation in Figure 15, all groups have homogeneity in the correlation of variables, as they were similar between the three groups.

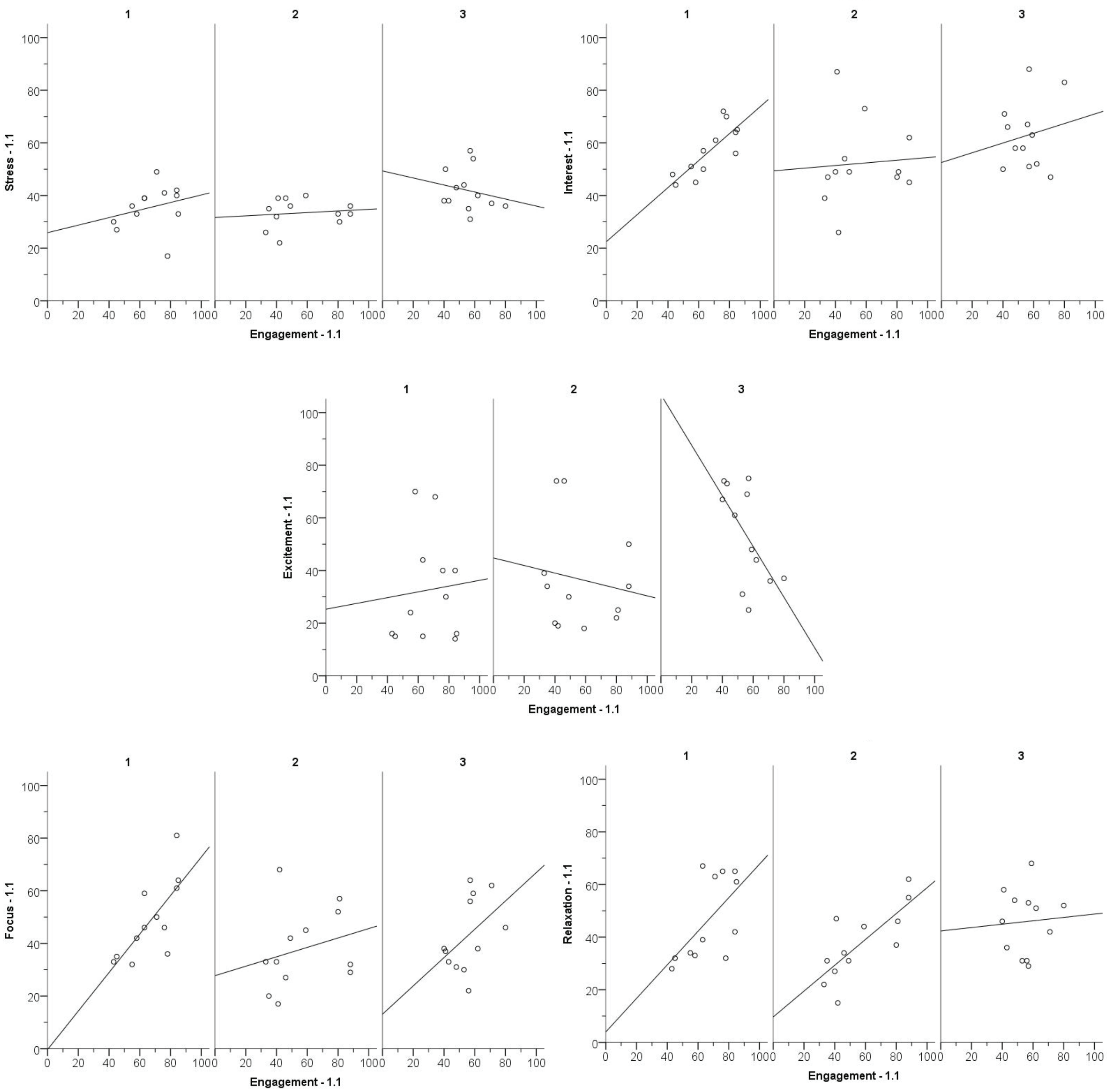

Figure 16. Figure with the bundle of correlations of variables of the time stamp 1.1.

In the Task 1.1 (Figure 16), the correlations in Stress, Interest, Focus and Relaxation remain the same for all groups. In the Excitement metrics in Group 3, the correlation became negative, that is, as the Excitement in Group 3 got higher, the Engagement became lower. The rest of the metrics remained similar as the Initial.

During the Task 1.2, there was still homogeneity for Stress, Focus and Relaxation. The correlation for excitement in Group 3 changed, and the Excitement and Engagement became more proportional. The Interest metric in Group 3 changed in this task, and it got higher as the Engagement level got higher as well.

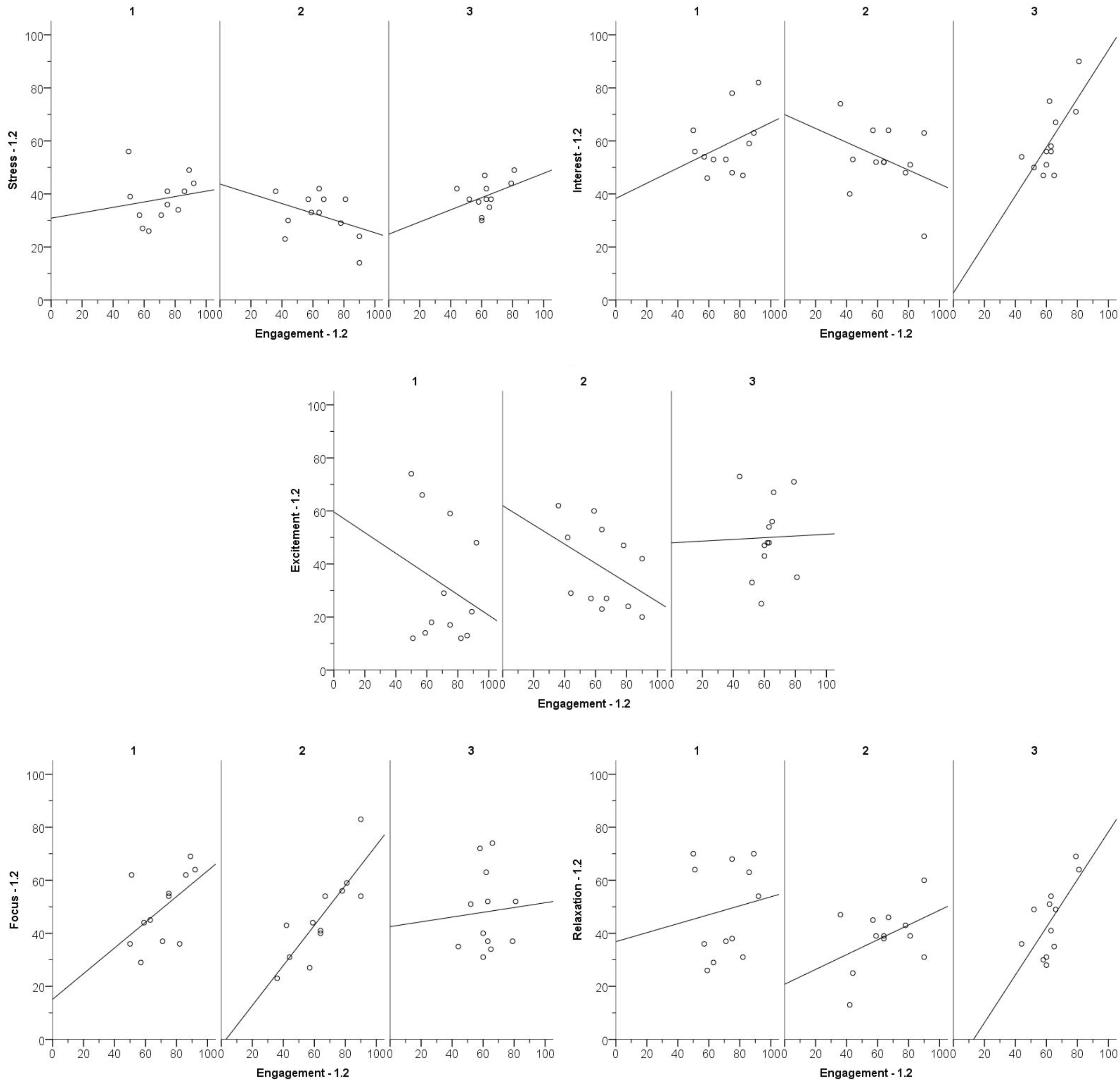

Figure 17. Figure with the bundle of correlations of variables of Task 1.2.

4.1 VALIDITY THREATS

In order to reduce the possibility of threats to the experiment, some strategies were applied considering levels as cited by Wohlin et al. [48]. Four types were considered: Conclusion validity, Internal validity, External validity, Confirmalibity and Credibility.

For Conclusion validity, the collected data was both collected by the software used and also by the observer, to be later checked in together with the observations notes resulting in a higher level of trust. For Confirmalibity and Credibility validity, several studies regarding the used device and its software were analyzed in order to certify their source and validation.

For Internal validity, the participants were handed the instructions after the helmet was set and ready to begin recording. During pilot tests, the users were able to rehearse the tasks if the instructions were handed before the full set up, and that could be an influence factor to the results.

For External Validity, as mentioned in the Section 3.1.1, the selected participants were all from areas of technology in which they were used to similar technologies and would not be disturbed nor feel unprepared when using the device. The more natural the way they acted around the device, the less the perception of it would influence the results.

Group 1	Initial	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	3.1	3.2
Stress	33.833	35.500	38.083	38.500	37.833	39.583	36.417	35.750	38.083	36.917	37.750
Engagement	64.750	67.083	70.833	64.583	65.417	58.750	60.833	60.750	56.917	59.333	59.167
Interest	58.083	56.917	58.583	59.917	58.917	59.917	57.583	51.750	59.167	57.583	56.083
Excitement	31.750	32.667	32.000	32.333	32.333	36.500	39.417	33.833	34.667	38.250	33.833
Focus	44.167	48.750	49.417	42.583	47.083	38.833	42.083	41.333	33.583	41.500	42.417
Relaxation	44.750	46.750	48.833	48.750	47.333	47.333	45.917	42.167	45.417	41.667	40.667

Group 2	Initial	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	3.1	3.2
Stress	37.083	33.417	31.917	36.917	36.250	33.500	35.917	36.083	37.000	35.583	37.083
Engagement	59.417	56.833	64.333	66.500	64.250	55.750	65.250	58.083	60.667	59.583	63.250
Interest	52.167	52.250	53.083	50.333	52.583	50.333	52.333	51.833	51.000	56.083	53.167
Excitement	37.833	36.583	38.667	32.333	33.917	34.750	37.000	36.917	35.750	43.417	40.000
Focus	40.083	37.917	46.250	45.667	45.333	36.250	44.833	40.917	42.500	39.250	46.000
Relaxation	43.833	37.583	38.750	43.833	41.917	37.333	41.750	40.583	44.583	45.250	40.583

Group 3	Initial	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	3.1	3.2
Stress	40.667	41.917	39.250	42.083	42.417	39.583	35.917	34.417	36.417	42.083	43.333
Engagement	57.167	55.583	62.750	61.667	61.583	58.000	57.333	50.000	57.083	56.083	57.167
Interest	63.000	62.833	60.167	60.417	65.333	60.667	57.417	56.833	56.250	58.667	65.583
Excitement	44.333	53.333	50.000	43.917	50.333	46.250	52.667	34.917	41.167	50.250	55.083
Focus	42.167	43.000	48.167	45.417	46.083	41.500	47.083	35.667	44.083	43.250	43.917
Relaxation	47.583	45.917	44.750	45.750	47.167	42.417	37.500	37.250	40.000	45.583	46.833

Table 3. Table representing the average data of 36 participants in the EEG collected, divided by groups.

5. DISCUSSION

There was no significant difference between the groups interacting differently in the results from the EEG. The control group using Google's search engine had an outcome that was similar to the other groups. Interacting with search engines is a predominant mean of navigation for most users [39]. This might be related to the fact that conversational agents are ubiquitous in everyday life so the experience around them is quotidian. Also, literature states that people engage psychologically when interacting with conversational agents much like they do with other people, resulting in a close to conventional interaction [5, 47].

The stress level collected was higher in Group 3 with the assistant using voice commands and that can be explained in the relation between the used tool and the stress reaction. According to Kostov [55] when users are interacting with systems, as the realism increases and makes the system act more like a human, the users tend to become less tolerant to imperfections. That could play a part into elevating the stress levels, due to user's prior higher expectations. Also, people tend to assume that conversational agents are worse at tasks than humans, and people tend to put extra effort into the choice of words so that the CA understands. This could play a part into elevating the stress level as well, as in the fear of failing to complete the task [5, 21].

Considering the stress as the challenge of the task, we should consider it as a part of the UX but not essential, since it is an emotion very close to each individual's personality [56]. All of the participants were selected from the field of technology, so that they could be more familiar with the technologies we selected, and they should not be a challenge to use nor fulfill the tasks [47].

There was no significantly difference of engagement levels throughout the test between groups. There was a slight decrease in engagement in Group 3. As mentioned before, Group 3 had the highest level of stress, which can be related to the lower engagement, considering the more stressed the user, the less immersed. The engagement level, as suggested, has a direct connection to the level of attention and concentration given to the task. According to the researchers that developed the software used, it mostly measures the level of immersion in the moment, being a mixture of attention and concentration. [21].

Theses results of small variance in the engagement corroborate with the results obtained from Zabčíková [57]. When the author investigated the usability of Emotiv while the participants were responding to visual and auditory stimuli, the engagement level had the least variation. The reason for the small variance was that the participants were committed to the task, and they did not lack a lot of concentration.

McMaham et al. [15] used the Emotiv during video game play to establish levels of task engagement and found slight increase of the engagement levels during death events.

The authors found that the higher levels of engagement may not suggest that the user was more engaged nor aroused when the character died in the game, but that it reflected that the user was entering a state of higher vigilance and concentration. Thus, it also endorses the findings in this study, when we relate the engagement levels to levels of concentration.

In addition, Luger and Sellen [47] reported in a study with users interacting with conversational agents that all users engaged a certain level of effort when speaking to ensure a successful interaction, and that appears to have a direct relation to the maintenance of the level of engagement and concentration.

In comparison to this, excitement had the biggest differences between groups. Group 1 had the lowest mean and the biggest variance, while Group 2 had the lowest variance and its mean was above Group 1. Group 3 had the highest levels of Excitement. The excitement measure is described as a physiological arousal for a positive value. It is characterized by activation in the sympathetic nervous system [21].

Žabčíková [57] found that when the participants when receiving visual or auditory stimuli that related to them personally, their excitement levels increased. Luger and Sellen [47] wanted to understand interactional factors by interviewing users of conversational agents. The findings of interviews described that several users said they had attempted to speak to the agent thinking of it as though it were a person and, as one might expect, that may relate to the significant increase in the excitement level throughout the tasks.

A low level of interest performance metric would be a participant feeling reluctant to a task, while the opposite would be how interested in or attracted to the task the user feels. Considering the nature of the tasks being very user-friendly, these levels could show how the participant felt towards each task [21, 47]. Not only, Luger and Sellen [47] stated that a conversational interaction should be, ideally, immersive, and should result in spontaneous involvement with “hypnotic effect”, which could explain the higher level of the interest metric in Group 3, that was the group within a conversational interaction.

When we discuss the focus metrics, it is close to measuring the ability of attention of the given moment. The focus was consistent among the 3 groups and had got lower towards the end of the entire test. That could relate to the participants losing their attention in the moment as the test became repetitive. The Relaxation metric tries to define the ability of a participant to win back their equilibrium after a moment of loss of attention [21].

Considering all the data collected were emotions happening at the same moment and derived from the same set of brain waves, we decided to analyze if any emotions had an effect on the other, being the cause of increase or decrease of the similar.

In the initial time span, there was no significant difference between the groups, the slight changes were in the interest and engagement correlation and, excitement and engagement. However, there is no evidence of meaning for these correlations, since the users were all in the same controlled environment waiting for the test to begin.

In the first task, while the users were performing a search in all Groups, there was homogeneity between all groups for stress, focus and relaxation. The excitement correlation in Group 3 had a change in results. We could infer that the more excitement, the less immersed and concentrated the user is, and consequently the engagement levels were lower, given the correlation between the two variables.

Group 3 was the only one where users were experiencing listening to results, instead of reading them from the tablet. According to Hiyoshi-Taniguchi et al. [58], this kind of interaction can be linked to the excitement arousal, since human attention tends to depend on information such as goals, objective assignments, and interpretation. When users are performing a task with emotional qualities, they are linked to excitement levels, which are directly related to a range of physiological responses like pupil dilation, eye widening, sweat, increase of heart rate and muscle tension, and digestive inhibition. Those actions can decrease the attention and focus, decreasing the detection of engagement [21].

During the second task, there was still homogeneity for stress, focus and relaxation. The correlation for excitement became a lot more similar between the groups in comparison to the first task, because now the interaction between groups were more alike, all the users were in time to read out loud a brief definition found before.

The interest correlation had a significant increase in relation to the first task in Group 3. That could relate to the emotional feel to the fact that users in Group 3 could feel they were in a conversation, so both their engagement and interest increased.

When attempting to establish a correlation between electrophysiological measures and questionnaires, the EEG collected measures are based on patterns of brain activity and rely on physiological information that users cannot fully control. While the questionnaires, not only are based on the evaluation of the participant about the experience episode, but also can be answered inconsistently. [51, 52]

By tracing boundaries for the mean of the results, we can carefully associate results from the EEG and questionnaires, if they are above average, below, etc. The results for Group 1 were very low in stress, which corroborates the questionnaire having low rate of frustration. The engagement and interest were the highest points of the EEG results, which again, corroborates the answers for boredom staying low in the questionnaire. The excitement in Group 1 was low, which does not exactly match the answers in the questionnaire stating a higher amount of fun. However, as mentioned, the excitement level has a quality of relating to positive physiological arousal such as increase of heart rate and pupil dilation, and these aspects do not always translate as fun for most people, so that could lead to a mistaken interpretation of the results.

In Group 2, the EEG results of engagement and interest were the highest and the stress was the lowest collected metric, confirming the low frustration stated in questionnaire by the same group.

In Group 3, the EEG metrics of Group 3 were low in Stress, confirming the questionnaire stating the low Frustration. Engagement and Interest were the highest metrics collected in the EEG, in agreement with the low metric of boredom collected by the questionnaire.

The questionnaire of usability applied at the end of each test, had similar results between Group 1 and Group 2, even though these groups were using different applications. However they were both interacting using the keyboard, which can lead to an easement of use, and a good level of reliability. Also, it corroborates with the fact that conversations tools are quotidian, therefore their results for usability are almost identical to the results of the most used model of interaction with the web [39].

Whislt that, the average of the SUS questionnaire in Group 3 was lower, while using the same application in Group 2, except that In Group 3 they were interacting using voice. So that leads us to understand that the voice command application, demanded higher effort to use. This also endorses the fact that users reported that they had to chose more carefully for keywords when speaking to the agent and enunciating well to be understood. [47]

6. LIMITATIONS

The electrophysiological methods can have limitations in terms of the accuracy of their outcome, considering the feeling of being part of a study in an observed environment can lead to significant changes in the results produced by the tests. We can not prove them, because the only way to test them is through these said environments.

Another limitation, as previously cited by Vermeeren et al [24], is the difficulty of data analysis. The amount of data produced by the software can be overwhelming to analyze, added to the fact that the algorithm to process said data is not open sourced. There are several studies that have used these equipment and software for scientifically purposes, so we can infer through sources that they are worthy of trust [16, 19, 18, 20, 50].

Therefore, one more limitation if the use of the EEG provided by the Emotiv company. One future work could include comparision with different technologies and EEG helmets. Using only one conversational agent is also a limitation for this study. For future studies, it would be interesting to compare different conversational agents and validate if their different algorithms influence in the Momentary User Experience.

Furthermore, we cannot make a general statement after a study involving a limited number (36) of subjects. We generated a large amount of data, which was not be possible to discuss all aspects collected in one study, so there should be further more investigation with the sample provided from these tests. And also, there should be more studies increasing the sample of the study to provide further knowledge.

7. CONCLUSIONS

This study investigated the measurement of the Momentary UX and the differences perceived between groups interacting differently. One of our contributions was the use of the Emotiv Insight electroencephalogram to collect brain activity and analyze differences in groups. There was difference in the user's Momentary UX regarding their excitement levels collected by the EEG when they were interacting with conversational agents by voice. One of the reasons for this is justified by the emotional qualities brought by the similarity with a human conversation, since the users were speaking and listening to the agent.

The main contribution is the comparison of using different forms to measure Momentary UX and inferring that the interaction using voice shows emotional differences. The self-report questionnaires can be associated to the results of the EEG. More research should be conducted to expand the understanding of Momentary UX and its perceptions while using conversational agents. This study aims to offer a contribution that we hope would encourage further studies on this subject.

REFERENCES

[1] ENDRES, D.; ALAM, M.; ŞOTROPA, D. *Graph-Based Representation and Reasoning: 24th International Conference on Conceptual Structures, ICCS 2019, Marburg, Germany, July 1–4, 2019, Proceedings*. Springer International Publishing, 2019. (Lecture Notes in Computer Science). ISBN 9783030231828. Available: <https://books.google.com.br/books?id=wAGfDwAAQBAJ>.

[2] SERBAN, I. et al. A deep reinforcement learning chatbot. 09 2017.

[3] KLOPFENSTEIN, L. C.; DELPRIORI, S.; RICCI, A. Adapting a conversational text generator for online chatbot messaging. In: *Internet Science*. [S.l.]: Springer International Publishing, 2019. p. 87–99.

[4] KARWOWSKI, W.; AHRAM, T. *Intelligent Human Systems Integration 2019: Proceedings of the 2nd International Conference on Intelligent Human Systems Integration (IHSI 2019): Integrating People and Intelligent Systems, February 7-10, 2019, San Diego, California, USA*. Springer International Publishing, 2019. (Advances in Intelligent Systems and Computing). ISBN 9783030110512. Available: <https://books.google.com.br/books?id=X5uCDwAAQBAJ>.

[5] HO, A.; HANCOCK, J.; MINER, A. S. Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication*, Oxford University Press (OUP), v. 68, n. 4, p. 712–733, may 2018. Available: <https://doi.org/10.1093/joc/jqy026>.

[6] SKJUVE, M. et al. Help! is my chatbot falling into the uncanny valley? an empirical study of user experience in human-chatbot interaction. *Human Technology*, Jyvaskyla University Library, p. 30–54, feb. 2019. Available: <https://doi.org/10.17011/ht/urn.201902201607>.

[7] BIDUSKI, D. et al. Assessing long-term user experience on a mobile health application through an in-app embedded conversation-based questionnaire. *Computers in Human Behavior*, Elsevier BV, v. 104, p. 106169, mar. 2020. Available: <https://doi.org/10.1016/j.chb.2019.106169>.

[8] ROGERS, Y.; SHARP, H.; PREECE, J. *Interaction design: Beyond human-computer interaction.* 3rd. ed. [S.l.]: John Wiley and Sons, 2013.

[9] LAW, E. L.-C. et al. Understanding, scoping and defining user experience. In: *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*. ACM Press, 2009. Available: <https://doi.org/10.1145/1518701.1518813>.

[10] ROTO, V. et al. User experience white paper: Bringing clarity to the concept of user experience. In: *Dagstuhl Seminar on Demarcating User Experience*. [S.l.: s.n.], 2011. p. 12.

[11] ZAROUR, M.; ALHARBI, M. User experience aspects and dimensions: Systematic literature review. *International Journal of Knowledge Engineering*, v. 3, p. 52–59, 12 2017.

[12] SCHULZE, K.; KRöMKER, H. A framework to measure user experience of interactive online products. In: *Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research*. ACM Press, 2010. Available: <https://doi.org/10.1145/1931344.1931358>.

[13] BARGAS-AVILA, J. A.; HORNBæK, K. Old wine in new bottles or novel challenges. In: *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI*. ACM Press, 2011. Available: <https://doi.org/10.1145/1978942.1979336>.

[14] MARTI, P.; IACONO, I. Anticipated, momentary, episodic, remembered: the many facets of user experience. In: . [S.l.: s.n.], 2016.

[15] MCMAHAN, T.; PARBERRY, I.; PARSONS, T. D. Evaluating player task engagement and arousal using electroencephalography. *Procedia Manufacturing*, Elsevier BV, v. 3, p. 2303–2310, 2015. Available: <https://doi.org/10.1016/j.promfg.2015.07.376>.

[16] RAMIREZ, R.; VAMVAKOUSIS, Z. Detecting emotion from EEG signals using the emotive epoc device. In: *Brain Informatics*. [S.l.]: Springer Berlin Heidelberg, 2012. p. 175–184.

[17] STRMISKA, M.; KOUDELKOVA, Z. Analysis of performance metrics UsingEmotiv EPOC+. *MATEC Web of Conferences*, EDP Sciences, v. 210, p. 04046, 2018. Available: <https://doi.org/10.1051/matecconf/201821004046>.

[18] POMER-ESCHER, A. G.; SOUZA, M. D. P. de; FILHO, T. F. B. Methodology for analysis of stress level based on asymmetry patterns of alpha rhythms in EEG signals. In: *5th ISSNIP-IEEE Biosignals and Biorobotics Conference (2014): Biosignals and Robotics for Better and Safer Living (BRC)*. IEEE, 2014. Available: <https://doi.org/10.1109/brc.2014.6880963>.

[19] STRMISKA, M.; KOUDELKOVA, Z.; ZABCIKOVA, M. Measuring brain signals using emotiv devices. *WSEAS Transactions on Systems and Control*, World Scientific and Engineering Academy and Society (WSEAS), 2018.

[20] PHAM, T. D.; TRAN, D. Emotion recognition using the emotiv EPOC device. In: *Neural Information Processing*. [S.l.]: Springer Berlin Heidelberg, 2012. p. 394–399.

[21] EMOTIV | Brain Data Measuring Hardware and Software Solutions. EMOTIV, 2019. Available: <https://www.emotiv.com/>.

[22] LAW, E. et al. Towards a shared definition of user experience. In: *CHI '08 Extended Abstracts on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2008. (CHI EA '08), p. 2395–2398. ISBN 978-1-60558-012-8. Available: <http://doi.acm.org/10.1145/1358628.1358693>.

[23] HASSENZAHL, M.; TRACTINSKY, N. User experience - a research agenda. *Behaviour & Information Technology*, Informa UK Limited, v. 25, n. 2, p. 91–97, mar. 2006. Available: <https://doi.org/10.1080/01449290500330331>.

[24] VERMEEREN, A. P. O. S. et al. User experience evaluation methods: Current state and development needs. In: *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*. New York, NY, USA: ACM, 2010. (NordiCHI '10), p. 521–530. ISBN 978-1-60558-934-3. Available: <http://doi.acm.org/10.1145/1868914.1868973>.

[25] ZAROUR, M.; ALHARBI, M. User experience framework that combines aspects, dimensions, and measurement methods. *Cogent Engineering*, Informa UK Limited, v. 4, n. 1, dec. 2017. Available: <https://doi.org/10.1080/23311916.2017.1421006>.

[26] KARAPANOS, E.; HASSENZAHL, M.; MARTENS, J.-B. User experience over time. In: *Proceeding of the twenty-sixth annual CHI conference extended abstracts on Human factors in computing systems*. ACM Press, 2008. Available: <https://doi.org/10.1145/1358628.1358891>.

[27] HASSENZAHL, M. User experience (UX) - towards an experiential perspective on product quality. In: *Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine on - IHM '08*. ACM Press, 2008. Available: <https://doi.org/10.1145/1512714.1512717>.

[28] INTERNATIONAL Organization for Standardization. ISO 9241-210:2019(en) - Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems. 2019. Geneva, CH.

[29] FRISON, A.-K.; WINTERSBERGER, P.; RIENER, A. Resurrecting the ghost in the shell: A need-centered development approach for optimizing user experience in highly automated vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour*, Elsevier BV, v. 65, p. 439–456, aug. 2019. Available: <https://doi.org/10.1016/j.trf.2019.08.001>.

[30] ABRO, A. et al. Understanding factors influencing user experience of interactive systems: A literature review. v. 10, p. 18175–18185, 01 2015.

[31] MAIA, C. L. B.; FURTADO, E. S. A systematic review about user experience evaluation. In: *Design, User Experience, and Usability: Design Thinking and Methods*. Springer International Publishing, 2016. p. 445–455. Available: <https://doi.org/10.1007/978-3-319-40409-7-42>.

[32] MASHAPA, J.; GREUNEN, D. van. User experience evaluation metrics for usable accounting tools. In: *Proceedings of the 2010 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on - SAICSIT 2010*. ACM Press, 2010. Available: <https://doi.org/10.1145/1899503.1899522>.

[33] ROTO MING LEE, K. P. B. C. A. V. E. L. K. V.-V.-M. J. H. M. O. V. *All About UX*. 2011. Available: <https://www.allaboutux.org/>.

[34] VääNäNEN, K.; ROTO, V.; HASSENZAHL, M. Towards practical user experience evaluation methods. 01 2008.

[35] ISOMURSU, M. User experience evaluation with experimental pilots. In: . [S.l.: s.n.], 2008.

[36] HOLE, L.; WILLIAMS, O. M. Emotion sampling and the product development life cycle. In: . [S.l.: s.n.], 2008.

[37] LIU, Y.; SOURINA, O. EEG-based valence level recognition for real-time applications. In: *2012 International Conference on Cyberworlds*. IEEE, 2012. Available: <https://doi.org/10.1109/cw.2012.15>.

[38] RUSSELL, J. A. Affective space is bipolar. *Journal of personality and social psychology*, American Psychological Association, v. 37, n. 3, p. 345, 1979.

[39] CLARK, J. A. Anticipatory design: Improving search UX using query analysis and machine cues. *Weave: Journal of Library User Experience*, University of Michigan Library, v. 1, n. 4, mar. 2016. Available: <https://doi.org/10.3998/weave.12535642.0001.402>.

[40] WILKS, Y. Is a companion a distinctive kind of relationship with a machine? In: . [S.l.: s.n.], 2010.

[41] BORS, L.; SAMAJDWER, A.; OOSTERHOUT, M. van. *Oracle Digital Assistant*. Apress, 2020. Available: <https://doi.org/10.1007/978-1-4842-5422-6>.

[42] BROWN, F. A.; MILLER, T. M.; MORRIS, R. *Virtual assistant focused user interfaces*. 2014.

[43] GRUBER, T. R. et al. *Using context information to facilitate processing of commands in a virtual assistant*. 2018.

[44] SHAWAR, B.; ATWELL, E. Different measurements metrics to evaluate a chatbot system. In: . [S.l.: s.n.], 2007. p. 89–96.

[45] LIU, R.; DONG, Z. A study of user experience in knowledge-based qa chatbot design. In: *IHSI*. [S.l.: s.n.], 2019.

[46] QVARFORDT, P.; JöNSSON, A.; DAHLBäCK, N. The role of spoken feedback in experiencing multimodal interfaces as human-like. In: . [S.l.: s.n.], 2003. p. 250–257.

[47] LUGER, E.; SELLEN, A. "like having a really bad pa": The gulf between user expectation and experience of conversational agents. In: . [S.l.: s.n.], 2016. p. 5286–5297.

[48] WOHLIN, C. et al. *Experimentation in software engineering*. [S.l.]: Springer Science and Business Media, 2012.

[49] SAURO, J.; LEWIS, J. R. Quantifying the user experience - practical statistics for user research. In: SAURO, J.; LEWIS, J. R. (Ed.). Boston: Morgan Kaufmann, 2012. p. 269 – 272. ISBN 978-0-12-384968-7. Available: <http://www.sciencedirect.com/science/article/pii/B9780123849687000102>.

[50] ASPINALL, P. et al. The urban brain: analysing outdoor physical activity with mobile EEG. *British Journal of Sports Medicine*, BMJ, v. 49, n. 4, p. 272–276, mar. 2013. Available: <https://doi.org/10.1136/bjsports-2012-091877>.

[51] MANDRYK, R. L.; INKPEN, K. M.; CALVERT, T. W. Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour & Information Technology*, Informa UK Limited, v. 25, n. 2, p. 141–158, mar. 2006. Available: <https://doi.org/10.1080/01449290500331156>.

[52] SOKKAR, A. A.; LAW, E. Validating an episodic UX model on online shopping decision making. In: *Proceedings of the 5th ACM SIGCHI symposium on Engineering interactive computing systems - EICS 2013*. ACM Press, 2013. Available: <https://doi.org/10.1145/2494603.2480301>.

[53] BROOKE, J. *SUS: A quick and dirty usability scale*. 1996.

[54] BERKA, C. et al. Eeg correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine*, v. 78, p. B231–44, 06 2007.

[55] KOSTOV, V.; FUKUDA, S. Emotion in user interface, voice interaction system. In: *SMC 2000 Conference Proceedings. 2000 IEEE International Conference on Systems, Man and Cybernetics. Cybernetics Evolving to Systems, Humans, Organizations, and their Complex Interactions*. IEEE, 2000. Available: <https://doi.org/10.1109/icsmc.2000.885947>.

[56] OSWALD, L. M. et al. Relationship between cortisol responses to stress and personality. *Neuropsychopharmacology*, Springer Science and Business Media LLC, v. 31, n. 7, p. 1583–1591, jan. 2006. Available: <https://doi.org/10.1038/sj.npp.1301012>.

[57] ŽABčíKOVá, M. Measurement of visual and auditory stimuli using eeg headset emotiv epoc+. *MATEC Web of Conferences*, v. 292, p. 01023, 01 2019.

[58] HIYOSHI-TANIGUCHI, K. et al. EEG correlates of voice and face emotional judgments in the human brain. *Cognitive Computation*, Springer Science and Business Media LLC, v. 7, n. 1, p. 11–19, jul. 2013. Available: <https://doi.org/10.1007/s12559-013-9225-0>.

UPF
UNIVERSIDADE
DE PASSO FUNDO

UPF Campus I - BR 285, São José
Passo Fundo - RS - CEP: 99052-900
(54) 3316 7000 - www.upf.br