The Effect of Search Engine, Search Term and Occasion on Brain-Computer Interface Metrics for Emotions When Ambiguous Search Queries Are Used

Wynand Nel[1][a], Lizette de Wet[1][b] and Robert Schall[2][c]
[1]Department of Computer Science and Informatics, University of the Free State, Nelson Mandela Drive, Bloemfontein, South Africa
[2]Department of Mathematical Statistics and Actuarial Science, University of the Free State, Nelson Mandela Drive, Bloemfontein, South Africa

Keywords: Human-Computer Interaction, Brain-Computer Interface, Emotions, Ambiguous Search Queries, Long-Term Excitement, Short-Term Excitement, Engagement, Meditation, Frustration.

Abstract: World Wide Web (WWW) searches are the primary source of information for many people for which different search engines are available. Depending on the search query, which might be ambiguous, search engines can return thousands of results to the user potentially causing frustration and a dislike towards the search engine. In this study, using a Brain-Computer Interface (BCI) we investigated the Long-Term Excitement, Short-Term Excitement, Engagement, Meditation and Frustration of study participants while they were performing ambiguous searches using Google, Yahoo! and Bing. The captured emotional data as well as pre-test and post-test questionnaire data suggest that the different search engines and search terms had an influence on the emotions of a participant during searches with ambiguous search queries.

1 INTRODUCTION

In the information age we search daily for answers to specific questions. Such searches can be done non-electronically (using, for example, printed material), or electronically (using technology to perform digital searches). Information Retrieval can be defined as a process and technique of searching, recovering and interpreting of information that is stored inside a file, catalogue or computer system (Dictionary.com, 2018; Merriam-Webster, 2018; TheFreeDictionary by Farlex, 2015).

Technology practitioners, commentators and researchers must consider the implications and measurement of the usability and associated User Experience (UX) of technological products. This is particularly true for the UX when a user does digital searches.

Digital/World Wide Web (WWW) searches are part of the everyday lives of people searching for answers in an era where the Internet is the primary source of information for many people. For this reason, we studied the effect of *search engine (Google, Yahoo! and Bing), search term (shoot, divide and seal)* and *occasion (first, second and third)* on various brain-computer interface (BCI) metrics, for different emotions, when ambiguous search queries are used during an Internet search.

2 BACKGROUND

The Internet consists of millions of linked computers worldwide. The connectivity allows those computers to communicate, carry data and exchange information (Mouton, 2001).

According to Edosomwan and Edosomwan (2010) the ultimate goal of a website is to share information, but the millions of pages which are added to the WWW daily, provide various types of data and information, which in turn, provide a challenge for information retrieval. In order to find

[a] https://orcid.org/0000-0001-5579-6411
[b] https://orcid.org/0000-0001-6819-6984
[c] https://orcid.org/0000-0002-4145-3685

Nel, W., de Wet, L. and Schall, R.
The Effect of Search Engine, Search Term and Occasion on Brain-Computer Interface Metrics for Emotions When Ambiguous Search Queries Are Used.
DOI: 10.5220/0008164900280039
In *Proceedings of the 3rd International Conference on Computer-Human Interaction Research and Applications (CHIRA 2019)*, pages 28-39
ISBN: 978-989-758-376-6

specific information on the WWW the user needs to know the address (URL) of that information on the WWW, but this URL is rarely available to the user beforehand. Thus, searching with search engines has become the only viable navigational construct (Goodman and Cramer, 2010).

Search engines are among the most accessed web sites (Edosomwan and Edosomwan, 2010; Oberoi and Chopra, 2010; *Oxford English Dictionary*, 2010) and rely on users to supply them with a search string upon which the search engine returns the results (documents or web pages) matching the word(s).

Search engines can return thousands of results to the user depending on the search query that is entered. The user then needs to work through these results until he/she has found the information required. A potential problem is the type of search query that the user enters during the search process. According to Teevan, Dumais and Horvitz (2007), users prefer short search queries which may result in ambiguity and the search engine returning more results than needed. For example, a user can enter the word "apple" as search query, which could imply an interest in fruit, but the search engine may instead return more results on Apple Incorporated. The large number of irrelevant results, in turn, may cause user frustration and other negative emotions, as he/she now needs to work through the large amount of search results, many of which are irrelevant.

The frustration and negative emotions caused by a large amount of results returned by a search engine when an ambiguous search query is used can create dislike in a user towards the specific search engine; when asked to do so, that user might rate the search engine with a bad UX score. This process can thus, for example, influence the UX measurements on search engines.

3 SEARCH QUERIES

Search engines rely on the users to supply a search query in order to retrieve the relevant information. These search queries can be either *ambiguous* (the search query can have more than one meaning, for example, *bat*); *broad* (the search query can have many sub-categories, for example, *sport*); *proper nouns* (the search query can be names or locations, for example, *Babcock*); or a *clear query* (the search query is very specific with a narrow topic, for example, *University of Chicago*) (Azzopardi, 2007; Dou et al., 2007; Elbassuoni et al., 2007; Sanderson, 2008; Song et al., 2007).

3.1 Ambiguous Queries

Ying et al. (2007) argue that word ambiguity is a severe problem in keyword-based search methods. Furthermore, Sanderson (2008) and Song et al. (2009) report that roughly 7%–23% of the queries presented to search engines are ambiguous, with the average length of queries being one word. The ranking algorithm of search engines struggles to give high quality results when such short queries are submitted, as it does not offer enough information to the search engine. This results in the search engine providing a diversified set of results (Luo et al., 2014).

An example of an ambiguous search query is the word *ruby*. When a person searches for the word *ruby*, what should the search engine return? Is the person in search of information on the Ruby gemstone or rather on the Ruby computer programming language? If the person is indeed in search of the Ruby programming language, specifically what is he/she searching for? Is he/she looking for Ruby documentation, regarding how the programming language works, or rather a brief explanation of what Ruby is?

Different methods exist and are in use by many search engines to assist the user in mitigating ambiguous queries. These methods include *word sense disambiguation* (Gale et al., 1992; Voorhees, 1993; Ying et al., 2007), *personalisation of web pages* ("E-business solutions", 2011; Linden, 2007), *query expansion* (Carpineto et al., 2001; Mitra et al., 1998), *suggesting corrections to misspelled queries* (Loukides, 2010), *click-through data* (Leung et al., 2008) and *clustering* (Baeza-Yates et al., 2007).

4 USABILITY AND UX

Various definitions exist for usability and every person working in the field might have his/her own definition (Tullis and Albert, 2013). Steve Krug (2006, p. 5) defines usability as, *"... making sure that something works well: that a person of average (or even below average) ability and experience can use the thing—whether it's a Web site, a fighter jet, or a revolving door—for its intended purpose without getting hopelessly frustrated*".

UX, however, as defined by Tullis and Albert (2013), the User Experience Professionals Association (2014), and Nielsen and Norman (2015), involves a broader view on every aspect of the interaction or anticipated interaction (International Standards Organization – ISO FDIS 9241-210, 2009) that a user has with a company, its services and products, taking into account the perceptions, thoughts and feelings of the user.

Thus, UX can be interpreted as being more comprehensive than usability. Tullis and Albert (2013) emphasise that usability and UX are two separate concepts, where UX include additional aspects, such as the feelings, thoughts and perceptions of the users as they interact with the product.

To differentiate between usability and UX, one could list and compare usability and UX goals (Preece et al., 2015; Rogers et al., 2011). Usability goals have traditionally been viewed as being concerned with meeting specific usability criteria, such as effectiveness and efficiency, whereas UX goals have been concerned with clarifying the nature of the UX, such as to be aesthetically pleasing (GFK, 2015; Preece et al., 2015; Rogers et al., 2011). GFK (2015) developed a tool to measure UX, called the UX Score. They mention that it is important to move beyond only looking at usability measurements. They argue that in order to understand the total UX, it is important to consider product fit and engagement, learnability and look and feel as well.

5 BCI

BCI, also known as Brain-Machine Interface (BMI), is an augmented technique that translates intentions into operational commands through a functional interface without requiring any motor action, allowing individuals to communicate and control external devices like a computer (Brandman et al. 2018; Hammer et al., 2018; Shah, 2018; Waldert, 2016). Wolpaw et al. (2002) define a BCI as a communication system where the messages sent by a person to the external world, does not pass through the brain's normal output pathways of nerves and muscles (e.g. speech and gestures). Instead, a BCI device harnesses bio-potentials, which are electric signals originating from the brain and nervous system (Colman and Gnanayutham, 2013), which are under the conscious control of the user (Wolpaw et al., 2002).

The BCI establishes a direct, non-muscular connection between the brain and the electronic device by measuring Electroencephalography (EEG) signals on the outside of the skull before decoding it into computer-understandable commands (Colman and Gnanayutham, 2013; Nicolas-Alonso and Gomez-Gil, 2012; Thorpe et al., 2005; Wolpaw et al., 2002). An overview of the components of a BCI system is shown in Figure 1.

BCIs can be classified according to their invasiveness and can either be invasive (e.g. intracortical – signal acquisition is recorded within the brain – or non-invasive – EEG signal acquisition is recorded from the scalp (Kameswara Rao et al., 2012; Nicolas-Alonso and Gomez-Gil, 2012; Shah, 2018; Wolpaw et al., 2002).

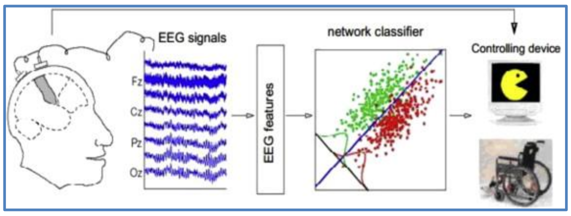

Figure 1: Schematic View of a BCI System (Karlovskiy and Konyshev, 2007).

5.1 Emotiv EPOC Neuroheadset

The Emotiv EPOC Neuroheadset, a consumer-grade EEG device, was the chosen BCI for this research study, as it is a non-invasive, low-cost BCI that did not pose a risk to the participants (Shah, 2018). This headset, developed by Emotiv (a neuro engineering company), was designed for human-computer interaction and is a high-fidelity, high-resolution 14-channel wireless neuroheadset. It can detect, user-trained mental commands (Cognitiv™ suite), subconscious emotional states (Affectiv™ suite), and facial expressions (Expressiv™ suite), which allow the computer to react to a user's moods and deliberate commands in a more natural way (Emotiv, 2012, 2014a). According to Maskeliunas et al. (2016), the Emotiv EPOC was originally developed as an input device for video games, but is becoming increasingly popular as a research tool, due to its usability and flexibility.

The Affectiv™ suite (Figure 2) deemed to be the most appropriate detection suite for this research study, as it monitors the user's emotional state (engagement, boredom, excitement, frustration and meditation level) in real-time, and enabling an extra dimension in interaction, which allows the computer to respond to the emotions.

As there are no international recognised units for emotions, the Affectiv™ suite produces numbers based on each user's historical range. An excitement level of 1 is the maximum excitement for that specific user, where 0 indicates a catatonic state (Gmac, 2014). Three distinct Affectiv™ detections, namely *instantaneous excitement*, *engagement* and *long-term excitement*, are available through this suite (Emotiv, 2012, n.d.).

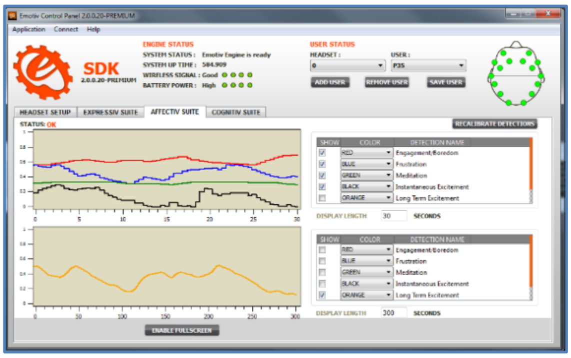

Figure 2: Emotiv Control Panel – AffectivTM Suite.

Engagement is experienced as attentiveness and the conscious direction of attention towards task-relevant stimuli. This is characterised by alpha and beta waves, as well as an increase in physiological arousal. Boredom is the opposite pole of this detection but it does not always correspond with a subjective emotional experience that all users describe as boredom. The user's engagement score will increase when he/she writes something on paper or types on a keyboard, and will decrease rapidly when he/she closes his/her eyes. The related emotions to engagement are vigilance, alertness, stimulation, concentration and interest (Emotiv, 2012, n.d.).

The user experience *instantaneous excitement* is an awareness or feeling of physiological arousal with a positive value. Excitement is characterised by the activation of the sympathetic nervous system. This results in a range of physiological responses in the user: sweat gland stimulation, pupil dilation, eye widening, blood diversion, heart rate and muscle tension increases and digestive inhibition. The related emotions to instantaneous excitement are titillation, nervousness and agitation. Instantaneous excitement is measured over time periods as short as several seconds (Emotiv, 2012, n.d.).

Long-term excitement is experienced and defined in the same way as instantaneous excitement. However, long-term excitement is measured over longer time periods, typically measured in minutes (Emotiv, 2012, n.d.).

The Affectiv^{TM} detection of the Emotiv EPOC Neuroheadset is capable of detecting emotions at a rate of four detections per second. These detections, the automatically scaled SDK (Software Development Kit) values, are derived from a trailing sample of two to nine seconds for the different emotions (long and short-term excitement, meditation, frustration and boredom) (Emotiv, 2014b).

Short-term excitement has the fastest response and uses a trailing buffer of two seconds to analyse for events. According to Emotiv (Gmac, 2012), it is possible to see a response in the short-term excitement within half a second if there was a significant event. For engagement, a one second trailing buffer is used to analyse for events, but as the data are based mostly on high frequency signals, a response can be noticed between a half and one second. For meditation, comparisons over several seconds are used to derive the information and it takes one to three seconds to respond to an event. Frustration uses a 10-second trailing buffer to analyse for events resulting in four to six seconds' delay to respond to a significant change (Gmac, 2012).

6 METHODOLOGY

The following research question was formulated, *"What are the effects of Search Engine, Search Term and Occasion on the BCI Metrics (Minimum, Maximum, Average and Fluctuation) for the different emotions (Long-Term Excitement, Short-Term Excitement, Engagement, Meditation and Frustration) when ambiguous search queries are used?"*

In order to answer the research question, 36 participants (19 males and 17 females) were recruited. They were all first-year students (20 to 25 years of age) at the University of the Free State (UFS), enrolled in the computer literacy course. Each participant completed a pre-test questionnaire, performed three ambiguous searches and completed a post-test questionnaire.

6.1 Pre-test Questionnaire

Before the participants were instructed to complete the pre-test questionnaire, a unique user profile was created on the Emotiv control panel before fitting the Emotiv EPOC Neuroheadset on the participants' heads. The Emotiv EPOC Neuroheadset learns and adapts to the user's range and scale of response (brainwaves) over time. The scale is rapidly adjusted by the Emotiv Affectiv^{TM} suite over the first few minutes of use and stabilises very well over about 40 minutes. A recommendation from one of the Emotiv forum administrators was to involve participants with questionnaires or other activities for several minutes at the start of the testing session and then, after approximately 10 minutes, the detections will be well stabilised (Gmac, 2012).

Completing the pre-test questionnaire took the participants a few minutes, allowing for the stabilisation of the brainwave detections of the headset. The questionnaire consisted of five sections and asked each participant to answer questions related to personal information, computer, WWW and search engine experience, searching and search engines, usability testing and BCI usage.

The pre-test questionnaire data show that the majority of the participants rated their technological experience as Expert Frequent Users when evaluating their computer experience, WWW usage and experience and search engine experience. The data show that Google is the most popular search engine, being nominated the favourite search engine amongst all 36 participants.

6.2 Tasks

The tasks consisted of three pre-selected ambiguous search queries to be used in WWW searches with three search engines (Google, Yahoo! and Bing). Each participant had to carry out three searches on three occasions (cross-over design), using each of the three different search engines and three different search terms. Participants were randomized to the six unique search engine/search term combinations using a Graeco-Latin Square design (Kempthorne, 1983 p. 187). This design allowed for the statistically efficient assessment of the effects of both search engine, search term, and occasion (an effect of occasion could be due to learning tiring effects on the users, for example).

Twenty-seven ambiguous search terms were considered (Table 1) for inclusion in the study. In order for a search term to be selected, it had to exhibit the same characteristics across all thee search engines. As the user was not allowed to change the search string via the keyboard, it was imperative that the "Searches related to *[original search string]*" in Google ("Google Search Engine", 2018), "Also Try" in Yahoo! ("Yahoo! Search Engine", 2018), and "Related searches" in Bing ("Bing search engine", 2018), displayed the same results across the three search engines once the search engine had completed the search. These suggestions are provided by the specific search engine to assist users in using more relevant search queries. Google and Yahoo! display these items towards the bottom of the web page, whereas Bing displays them at the top right, as well as at the bottom of the web page.

Table 1: Ambiguous search terms.

Apple	History	Racquet
Ball	Hook	Ruby
Bat	Java	Science
Canon	Match	<u>Seal</u>
Develop	Math	<u>Shoot</u>
Drink	News	Sport
<u>Divide</u>	Number	Star
Fight	<u>Power</u>	Study
Game	Python	Tree

The initial ambiguous search terms were *Power*, *Divide* and *Seal*, but a day before the formal data capturing started, Yahoo! and Bing no longer returned the same "Also Try"/"Related searches" items for the term *Power*. The ambiguous search term *Shoot* replaced *Power*.

The participants completed each search task while wearing the BCI headset. The headset recorded the participants' emotional data in real time while they were busy with the tasks.

The emotional data gathered from the 36 participants included their *Long-Term Excitement*, *Short*-Term *Excitement*, *Engagement*, *Meditation* and *Frustration*. The recorded emotional data were cleaned and normalised before four summary metrics were calculated per emotion, from the individual profiles, recorded over time. The metrics included:

- Minimum (*Min*): Minimum value measured during the recording period.
- Maximum (*Max*): Maximum value measured during the recording period.
- Average (*Avg*): Arithmetic mean of all values measured during the recording period.
- Fluctuation: Calculated as the normalized peak-trough fluctuation, namely (max–min)/average.

6.3 Post-test Questionnaire

Participants completed a post-test questionnaire answering questions related to BCI, WWW, their physical and personal experience, System Usability Scale (SUS) questionnaire (one for each search engine), emotions and general searching experience.

The purpose of the post-test questionnaire was to obtain data on the participants' perceptions, emotions and feelings after completing the three search tasks, while wearing the BCI.

The majority of the participants indicated that they were confident using Internet Explorer and Google, but not so much Yahoo! and Bing. The average SUS score has also confirmed this.

7 STATISTICAL ANALYSIS

During this study it was of interest to determine whether there were statistically significant differences in the mean BCI metrics, regarding the five emotions (*Long-Term Excitement*, *Short-Term Excitement*, *Engagement*, *Meditation* and *Frustration*), between the three *Search Engines* (*Google*, *Yahoo!* and *Bing*), *Search Terms* (*Shoot*, *Divide* and *Seal*) and *Occasions* (*First*, *Second* and *Third*).

In order to statistically assess the effect of *Search Engine*, *Search Term* and *Occasion*, respectively, on the BCI metrics calculated from five emotions, the following null hypotheses were formulated:

$H_{0,1}$: There are no differences between the mean BCI metrics (Minimum, Maximum, Average and Fluctuation) calculated for the different emotions (Long-Term Excitement, Short-Term Excitement, Engagement, Meditation and Frustration) with regard to the factors Search Engine (Google, Yahoo! and Bing), Search Term (Shoot, Divide and Seal) and Occasion (First, Second and Third).

Thus, in effect, 60 hypotheses were tested (5 emotions each with 4 metrics and 3 factors; 5 x 4 x 3 = 60).

The BCI data, that is, the four summary metrics *Minimum*, *Maximum*, *Average* and *Fluctuation* for each emotion, were analysed using ANOVA fitting the factors *Participant*, *Occasion*, *Search Term* and *Search Engine*. From this ANOVA, F-statistics and P-values associated with testing of the significance of the factors *Occasion*, *Search Term* and *Search Engine* were reported. Furthermore, mean values of each metric and emotion, for each level of the factors, *Occasion*, *Search Term* and *Search Engine*, were reported.

With regard to each metric and emotion, the three search engines were compared by calculating point estimates for the pairwise differences in mean values between search engines, as well as 95% confidence intervals for the mean difference and the associated P-values.

The mean values and overall F-tests for the effect of *Search Engine* (*Google*, *Yahoo!* and *Bing*), *Search Term* (*Shoot*, *Divide* and *Seal*) and *Occasion* (*First*, *Second* and *Third*) on the four summary metrics (*Minimum*, *Maximum*, *Average* and *Fluctuation*) of the BCI data will be discussed, per emotion (Long-Term Excitement – Section 7.1, Short-Term Excitement – Section 7.2, Engagement – Section 7.3, Meditation – Section 7.4 and Frustration – Section 7.5), in the sections to follow.

7.1 Long-Term Excitement

The effect of search engine, search term and occasion, for the emotion *Long-Term Excitement*, will be discussed below.

7.1.1 Effect of Search Engine

For the emotion *Long-Term Excitement*, the *Minimum* metric showed statistically significant differences between search engines ($P = 0.0075$), with *Yahoo!* having the lowest mean minimum (0.20) followed by *Bing* (0.24) and *Google* (0.29). In contrast, the *Maximum* and *Average* metrics did not show statistically significant differences. The *Fluctuation* metric showed statistically significant differences between search engines ($P = 0.0053$) with *Yahoo!* showing the greatest mean fluctuation (1.27), followed by *Bing* (1.00) and *Google* (0.87). The fact that *Yahoo!* had the greatest mean *Fluctuation* metric can probably be explained by the fact that this search engine had the smallest mean *Minimum* metric, while the mean *Maximum* and mean *Average* metrics did not differ significantly between search engines.

7.1.2 Effect of Search Term

Regarding *Long-Term Excitement*, the *Minimum* metric showed statistically significant differences between search terms ($P = 0.0036$), with *Divide* having the lowest mean minimum (0.19) followed by *Shoot* (0.26) and *Seal* (0.28). The *Maximum* and *Average* metrics did not show statistically significant differences. The *Fluctuation* metric showed statistically significant differences between search terms ($P = 0.0008$) with *Divide* showing the greatest mean *Fluctuation* (1.32) followed by *Shoot* (0.88) and *Seal* (0.94). The fact that *Divide* had the greatest mean *Fluctuation* metric can probably be explained by the fact that this *Search Term* had the smallest mean *Minimum* metric, while the mean *Maximum* and mean *Average* metrics did not differ significantly between *Search Terms.*

7.1.3 Effect of Occasion

Finally, for *Long-Term Excitement* none of the metrics showed statistically significant differences between occasions. Thus, the order in which the participants used the three search engines and three search terms did not have an effect on *Long-Term Excitement*. This suggests that learning or tiring effects did not occur, or at any rate did not affect *Long-Term Excitement* so that measurements taken on the different occasions were comparable.

7.2 Short-Term Excitement

The effect of search engine, search term and occasion, for the emotion *Short-Term Excitement*, will be discussed below.

7.2.1 Effect of Search Engine

For the emotion *Short-Term Excitement*, the *Minimum* metric showed statistically significant differences between search engines (P = 0.0013), with *Yahoo!* having the lowest mean minimum (0.04) followed by *Bing* (0.05) and *Google* (0.1). In contrast, the *Maximum* and *Average* metrics did not show statistically significant differences. The *Fluctuation* metric showed statistically significant differences between search engines (P = 0.0064), with *Yahoo!* showing the greatest mean fluctuation (2.53) followed by *Bing* (2.15) and *Google* (2.10). The fact that *Yahoo!* had the greatest mean *Fluctuation* metric can probably be explained by the fact that this search engine had the smallest mean *Minimum* metric, while the mean *Maximum* and mean *Average* metrics did not differ significantly between search engines.

7.2.2 Effect of Search Term

For the emotion *Short-Term Excitement*, the *Minimum* metric showed statistically significant differences between search terms (P = 0.0418), with *Divide* having the lowest mean minimum (0.04) followed by *Shoot* (0.07) and *Seal* (0.08). In contrast, the *Maximum* and *Average* metrics did not show statistically significant differences. The *Fluctuation* metric showed statistically significant differences between search engines (P = 0.0496) with *Divide* showing the greatest mean fluctuation (2.46), followed by *Shoot* (2.20) and *Seal* (2.12). The fact that *Divide* had the greatest mean *Fluctuation* metric can probably be explained by the fact that this search term had the smallest mean *Minimum* metric, while the mean *Maximum* and mean *Average* metrics did not differ significantly between search terms.

7.2.3 Effect of Occasion

For the emotion *Short-Term Excitement*, none of the metrics *Minimum*, *Maximum*, *Average* and *Fluctuation*, showed statistically significant differences between occasions. Thus, the order in which the participants used the three search engines and three search terms did not have an effect on the *Short-Term Excitement*. This suggests that learning or tiring effects did not affect the *Short-Term Excitement*, similar to *Long-Term Excitement*, and measurements taken on the different occasions were comparable.

7.3 Engagement

The effect of search engine, search term and occasion, for the emotion *Engagement*, will be discussed below.

7.3.1 Effect of Search Engine

For the emotion *Engagement*, none of the metrics showed statistically significant differences between search engines. Inspecting the individual mean values per BCI metric, it can be seen that the differences were small, indicating that this BCI emotion does not discriminate between the three search engines.

7.3.2 Effect of Search Term

For the emotion *Engagement*, the *Minimum* metric showed statistically significant differences between search terms (P = 0.0004), with *Divide* having the lowest mean minimum (0.40) followed by *Seal* (0.46) and *Shoot* (0.46). The *Maximum* metric also indicated statistically significant differences between the search terms (P = 0.0025). *Divide* has the highest mean maximum (0.83), followed by *Seal* (0.78) and *Shoot* (0.77). The *Average* metric is the only metric not showing statistically significant differences. The *Fluctuation* metric showed statistically significant differences between search engines (P < 0.0001) with *Divide* showing the greatest mean fluctuation (0.74) followed by *Seal* (0.55) and *Shoot* (0.52). The fact that *Divide* had the greatest mean *Fluctuation* metric, can probably be explained by the fact that this search term had the smallest mean *Minimum* and highest mean *Maximum* metric. This search term was also the only search term that required the participant to view three web pages before finding the correct answer. The findings of the SUS scores indicated that the participants experienced the task using the search term *Divide*, to be more difficult than terms *Shoot* and *Seal*.

7.3.3 Effect of Occasion

For the emotion *Engagement*, the metrics *Minimum*, *Maximum* and *Fluctuation* did not show any statistically significant differences between occasions. The *Average* metric showed statistically significant differences (P = 0.0176), with the *First Occasion* having the highest average mean (0.61), followed by the *Second* (0.59) and *Third Occasions* (0.58).

7.4 Meditation

The effect of search engine, search term and occasion, for the emotion *Meditation*, will be discussed below.

7.4.1 Effect of Search Engine

For the emotion *Meditation*, the *Minimum* (P = 0.0038), *Maximum* (P = 0.0126) and *Fluctuation* (P = 0.002) metrics showed statistically significant differences between search engines. The *Average* metric did not show statistically significant differences. *Yahoo!* showed the greatest mean *Fluctuation* (0.70), followed by *Bing* (0.58) and *Google* (0.55). The fact that *Yahoo!* had the greatest mean *Fluctuation* metric can probably be explained by the fact that this search engine had the greatest mean *Maximum* metric, and mean *Average* metrics did not differ significantly between search engines.

7.4.2 Effect of Search Term

For the emotion *Meditation*, the *Minimum* metric showed statistically significant differences between search terms (P < 0.0001), with *Divide* having the lowest mean minimum (0.24) followed by *Shoot* (0.26) and *Seal* (0.27). The *Maximum* metric did not indicate statistically significant differences between the *Search Terms*. The *Average* metric indicated statistically significant differences with P = 0.0024. The *Fluctuation* metric also showed statistically significant differences between search terms (P < 0.0001), with *Divide* showing the greatest mean fluctuation (0.74) followed by *Seal* (0.55) and *Shoot* (0.54). The fact that *Divide* had the greatest mean *Fluctuation* metric can probably be explained by the fact that this search term had the smallest mean *Minimum* metric.

7.4.3 Effect of Occasion

For the emotion *Meditation*, none of the metrics showed statistically significant differences between occasions. Thus, the order in which the participants used the three search engines and three search terms did not have an effect on *Meditation*. This suggests that learning or tiring effects did not affect the *Meditation*, similar to *Short-* and *Long-Term Excitement*, and measurements taken on the different occasions were comparable.

7.5 Frustration

The effect of search engine, search term and occasion, for the emotion *Frustration*, will be discussed below.

7.5.1 Effect of Search Engine

For the emotion *Frustration*, the *Minimum* (P = 0.0229) and *Fluctuation* (P = 0.0081) metrics showed statistically significant differences between search engines. The *Maximum* and *Average* metrics did not show statistically significant differences. *Yahoo!* showed the greatest mean *Fluctuation* (1.53), followed by *Bing* (1.30) and *Google* (1.28). The fact that *Yahoo!* had the greatest mean *Fluctuation* metric can probably be explained by the fact that this search engine had the lowest mean *Minimum* metric, while the *Maximum* and *Average* metrics did not show statistically significant differences between search engines.

7.5.2 Effect of Search Term

For the emotion *Frustration*, all of the metrics, *Minimum* (P = 0.0017), *Maximum* (P = 0.0102), *Average* (P = 0.0072) and *Fluctuation* (P < 0.0001), showed statistically significant differences between search terms. The *Fluctuation* metric for the search term *Divide*, showed the greatest mean *Fluctuation* (1.62), followed by *Shoot* (1.27) and *Seal* (1.12). This search term was also the only search term that required the participant to view four web pages before finding the correct answer. The findings of the SUS scores indicated that the participants experienced the task using the search term *Divide* to be more difficult than terms *Shoot* and *Seal*.

7.5.3 Effect of Occasion

For the emotion *Frustration*, none of the metrics showed statistically significant differences between occasions. Similar to *Meditation*, *Short-*, and *Long-Term Excitement*, it can be deduced that the effects of learning or tiring did not affect the participants *Frustration* levels.

8 DISCUSSION

The answer to the research question, *"What are the effects of Search Engine, Search Term and Occasion on the BCI Metrics (Minimum, Maximum, Average and Fluctuation) for the different emotions (Long-Term Excitement, Short-Term Excitement, Engagement, Meditation and Frustration)?"* can be summarised as follows.

The metrics *Minimum* and *Fluctuation* showed a statistically significant effect of *Search Engine* with regard to all emotions except *Engagement*, and a statistically significant effect of *Search Term* with regard to all emotions. It was also found that the *Average* metric only had a statistically significant effect on the emotions *Meditation* and *Frustration* with regard to *Search Term*. The *Minimum* metric seemed to be the most sensitive of all the metrics that were investigated. *Occasion* generally had no statistically significant effect with regard to any metric or emotion.

For the emotions *Long-* and *Short-Term Excitement*, two metrics (*Minimum* and *Fluctuation*) showed statistical significant differences for the effect of *Search Engine* and *Search Term*. None of the metrics showed any statistically significant differences for the effect of *Occasion* suggesting that learning or tiring effects did not affect *Short-* and *Long-Term Excitement*, and measurements taken on the different occasions were comparable.

For the emotion *Engagement*, none of the metrics showed statistically significant differences for the effect of *Search Engine*. The effect of *Search Term*, on the other hand, did show statistically significant differences, indicating that the participants had to concentrate while searching for the correct answers. There was a statistically significant difference between *Occasions* regarding the *Average* metric for this emotion, but after investigating the individual mean values it was found that the differences in mean values for the *First*, *Second* and *Third* occasions were very small and thus can be ignored.

Three metrics (*Minimum*, *Maximum* and *Fluctuation*) of the emotion *Meditation* showed statistically significant differences for the effect of *Search Engine* and *Search Term*. None of the metrics showed any statistically significant differences for the effect of *Occasion* suggesting that learning or tiring effects did not affect *Meditation*, similar to *Short-* and *Long-Term Excitement*, and measurements taken on the different occasions were comparable.

For the emotion *Frustration*, two metrics (*Minimum* and *Fluctuation*) showed statistically significant differences for the effect of *Search Engine*, and all the metrics showed statistically significant differences for the effect of *Search Term*, confirming the results gathered in the post-test questionnaire where approximately a quarter of the participants indicated that they felt frustration and did not know wat to do while being confronted with the tasks. None of the metrics showed any statistically significant differences for the effect of *Occasion*.

In summary, it seems clear that *Occasion* did not have any statistically significant effect regarding any emotion. This finding suggests that learning or tiring effects did not affect any of the emotions and that measurements taken on the different occasions were comparable in this respect. In turn, this finding suggests that a study design where participants complete several tasks according to some form of cross-over design, as was done in the current study; measurements of the kind taken in the present study are not affected by learning or tiring effects.

Table 2 summarises the significant factors on the BCI metrics for the different emotions. *Search Engine* was a significant factor for the emotions *Long-Term Excitement*, *Short-Term Excitement*, *Meditation* and *Frustration*, while *Search Term* was a significant factor for all five of the emotions. In contrast, *Occasion* was only a significant factor for *Engagement* but as mentioned above, this can be ignored.

The following null hypotheses can thus be rejected:

$H_{0,1a}$*: There are no differences between the mean BCI metrics (Minimum, Maximum and Fluctuation) calculated for the different emotions (Long-Term Excitement, Short-Term Excitement, Meditation and Frustration) with regard to the factors Search Engine (Google, Yahoo! and Bing).*

$H_{0,1b}$*: There are no differences between the mean BCI metrics (Minimum, Maximum, Average and Fluctuation) calculated for the different emotions (Long-Term Excitement, Short-Term Excitement, Engagement, Meditation and Frustration) with regard to the factors Search Term (Shoot, Divide and Seal).*

9 CONCLUSION

The statistical analysis showed that *Search Engine* (*Google*, *Yahoo!* and *Bing*) was indeed a significant factor for the emotions *Long-Term Excitement*, *Short-Term Excitement*, *Meditation* and *Frustration*, while the *Search Term* was a significant factor for all five of the emotions. This indicated that the different search engines and search terms had an influence on the different emotions of a participant when ambiguous search queries were used. The different occasions did not show any statistically significant differences, indicating that learning or tiring effects did not affect any of the emotions and that measurements taken on the different occasions were comparable in this respect.

The post-test questionnaire revealed that the majority of the participants found the usability test exciting, with low levels of frustration, while being engaged in the tasks. These findings contradict the BCI data, which clearly indicated that *Search Engine* and *Search Term* affected frustration. This phenomenon might be explained by the fact that the participant responses were captured after the completion of the tasks and that they felt more relaxed at that time, not remembering how they felt before (or as the literature indicated, not willing to share their true emotions (Schall, 2015; Tullis and Albert, 2013)). Another explanation might be that the emotions detected by the Emotiv EPOC Neuroheadset, specifically *Frustration*, are more sensitive than what users experienced. As mentioned earlier, the *Frustration* emotion used a 10-second trailing buffer to analyse for events, resulting in 4 to 6 seconds' delay to respond to a significant change. However, this was kept in mind when the BCI data were cleaned up and prepared for statistical analysis. Another possible reason might be the fact that during the tests, the computer froze three times and the recording software crashed four times. In each case, the test had to be restarted and participants might have experienced some level of frustration. The post-test questionnaire also documented that some participants indicated that they were glad that the test session was over, which might also indicate that frustration was present. However, the majority of the participants indicated that they felt relaxed and that they were not bored. The results show that the participants were positive towards the overall usability test, which suggests that their emotions did not negatively affect the reliability of the data that were captured.

In the light of the above findings, the answer to this research question is that factors *Search Engine* and *Search Term* do have an effect on the BCI metrics for the different emotions mentioned. Furthermore, it was found that factor *Occasion*, did not have an impact on the results and can thus be ignored.

10 FUTURE RESEARCH

Future research can include the following:

- Different ambiguous search queries to the ones used in this paper could be used. The researcher should also ensure that the number of steps/web pages needed to complete each task is the same. It will also be valuable to see how the search engines have adapted over time to accommodate ambiguous search terms.
- Multiple ambiguous search terms should be identified and different search terms should be used at different phases of the study in order to perform a longitudinal study. However, future researchers should keep in mind that the data collection per phase should be completed as soon as possible, as search engines adapt without warning.
- Different BCI devices should be compared. The limitation of the Emotiv EPOC Neuroheadset is that its sensors need to make contact with the scalp. This limits the participants to those not wearing wigs, or having weaved or braided hair.

Table 2: Summary of Significant Factors on BCI Metrics for the Different Emotions.

Factors	**Metric**	**Long-Term Excitement**	**Short-Term Excitement**	**Engagement**	**Meditation**	**Frustration**
Search Engine	Min	✓	✓		✓	✓
	Max				✓	
	Avg					
	Fluctuation	✓	✓		✓	✓
Search Term	Min	✓	✓	✓	✓	✓
	Max			✓		✓
	Avg				✓	✓
	Fluctuation	✓	✓	✓	✓	✓
Occasion	Min					
	Max					
	Avg			✓		
	Fluctuation					

REFERENCES

Azzopardi, L. 2007. Position paper: Towards evaluating the user experience of interactive information access systems. *Proceedings of SIGIR 2007 Workshop: Web Information Seeking and Interaction*. 60–64.

Baeza-Yates, R., Hurtado, C. and Mendoza, M. 2007. Improving search engines by query clustering. *Journal of the American Society for Information Science and Technology*. 58(12).

Bing search engine. 2018. [Online], Available: www.bing.com [2018, November 26].

Brandman, D.M., Hosman, T., Saab, J., Burkhart, M.C., Shanahan, B.E., Ciancibello, J.G., Sarma, A.A., Milstein, D.J., et al. 2018. Rapid calibration of an intracortical brain–computer interface for people with tetraplegia. *Journal of Neural Engineering*. 15(2). [Online], Available: http://stacks.iop.org/1741-2552/15/i=2/a=026007.

Carpineto, C., de Mori, R., Romano, G. and Bigi, B. 2001. An information-theoretic approach to automatic query expansion. In Vol. 19. ACM Transactions on Information Systems *ACM Transactions on Information Systems (TOIS)*. 1–27. [Online], Available: http://www.accessmylibrary.com/coms2/summary_0286-10728134_ITM.

Colman, J. and Gnanayutham, P. 2013. Assistive technologies for brain-injured gamers. In G. Kouroupetroglou (ed.). Hershey, PA: IGI Global *Assistive Technologies and Computer Access for Motor Disabilities*. 28–56.

Dictionary.com. 2018. *Information retrieval*. [Online], Available: http://dictionary.reference.com/browse/information+retrieval [2018, November 21].

Dou, Z., Song, R. and Wen, J. 2007. A large-scale evaluation and analysis of personalized search strategies. *Proceedings of the 16th international conference on World Wide Web - WWW '07*. 581.

E-business solutions. 2011. [Online], Available: http://www.it-architects.co.uk/a_-_z_glossary_index/E-Business_Solutions_Glossary/e-business_solutions_glossary.html [2011, March 08].

Edosomwan, O.J. and Edosomwan, T. 2010. Comparative analysis of some search engines. *South African Journal of Science*. 106(11/12). [Online], Available: http://www.sajs.co.za/index.php/SAJS/article/view/169/439.

Elbassuoni, S., Luxenburger, J. and Weikum, G. 2007. Adaptive personalization of web search. *Proceedings of SIGIR 2007 Workshop: Web Information Seeking and Interaction*. 1–4.

Emotiv. 2012. *Emotiv EPOC specifications. Brain-computer interface technology*. [Online], Available: http://www.emotiv.com/upload/manual/sdk/EPOCSpecifications.pdf [2014, February 06].

Emotiv. 2014a. Emotiv EPOC and TestBench spesifications. *Brain Computer Interface and Scientific Contextual EEG*.

Emotiv. (2014b). Emotiv media. Retrieved 6 February 2014, from http://www.emotiv.com/ about/media/

Emotiv. (n.d.). Emotiv software development kit. User manual for release 2.0.0.20. Emotiv.

Gale, W.A., Church, K.W. and Yarowsky, D. 1992. Using bilingual materials to develop word sense disambiguation methods. In Montréal, Canada *Fourth International Conference on Theoretical and Methodological Issues in Machine Translation*. 101–112.

GFK. 2015. *UX Score*. [Online], Available: http://www.gfk.com/solutions/ux/our-products/ux-measurement/pages/ux-score.aspx [2015, November 18].

Gmac. 2012. *Affective suite baseline data / delay on data / 3D map*. [Online], Available: http://emotiv.com/forum/messages/forum12/topic2047/message11705/?sphrase_id=1442#message11705 [2014, August 15].

Gmac. 2014. *Affectiv suite*. [Online], Available: http://emotiv.com/forum/messages/forum12/topic4084/message18512/?sphrase_id=1442#message18512 [2014, August 15].

Goodman, E. and Cramer, M. 2010. [Online], Available: http://www.comscore.com/Press_Events/Presentations_Whitepapers/2010/The_Future_of_Search_The_Emerging_Power_of_Real-time_Personalized_Search.

Google Search Engine. 2018. [Online], Available: www.google.co.za [2018, November 26].

Hammer, E.M., Halder, S., Kleih, S.C. and Kübler, A. 2018. Psychological predictors of visual and auditory P300 brain-computer interface performance. *Frontiers in Neuroscience*. 12:1–12.

International Standards Organization - ISO FDIS 9241-210. 2009. *Ergonomics of human system interaction - Part 210: Human-centered design for interactive systems.* Switzerland: International Organization for Standardization (ISO).

Kameswara Rao, T., Rajyalakshmi, M. and Prasad, T.V. 2012. An exploration on brain-computer interface and its recent trends. *International Journal of Advanced Research in Artificial Intelligence*. 1(8):17–22.

Karlovskiy, D.V. and Konyshev, V.A. 2007. Visualmind framework for brain-computer interface development. In *Proceedings of the 3rd Russian-Bavarian Conference on Bio- Medical Engineering*. 15–18. [Online], Available: http://www5.informatik.uni-erlangen.de/Forschung/Konferenzen/Biomed07/sessions/Article_(Karlovskiy)_eng.pdf.

Kempthorne, O. 1983. *Design and Analysis of Experiments*. Revised ed. Florida: Robert E Krieger Publishing Company.

Krug, S. 2006. *Don't Make Me Think! A Common Sense Approach to Web Usability*. 2nd ed. Vol. 277. Berkeley, California USA: New Riders Publishing.

Leung, K.W., Ng, W. and Lee, D.L. 2008. Personalized concept-based clustering of search engine queries. *IEEE Transactions on Knowledge and Data Engineering*. 20(11):1505–1518.

Linden, G. 2007. *Personalized search primer and Google's approach*. [Online], Available: http://www.readwriteweb.com/archives/personalized_search_primer.php [2011, March 08].

Loukides, M. 2010. *What is data science? Analysis: The future belongs to the companies and people that turn data into products.* [Online], Available: http://radar.oreilly.com/2010/06/what is data science.html [2010, July 01].

Luo, C., Liu, Y., Zhang, M. and Ma, S. 2014. Query ambiguity identification based on user behavior information. *Information Retrieval Technology*. (863):36–47.

Maskeliunas, R., Damasevicius, R., Martisius, I. and Vasiljevas, M. 2016. Consumer-grade EEG devices: Are they usable for control tasks? *PeerJ*. 4(e1746).

Merriam-Webster, I. 2018. *Dictionary: Information retrieval*. [Online], Available: http://www.merriam-webster.com/dictionary/information retrieval [2018, March 17].

Mitra, M., Singhal, A. and Buckley, C. 1998. Improving automatic query expansion. In New York: ACM New York *SIGIR '98 Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Mouton, J. 2001. *How to succeed in your Master's and Doctoral Studies. A South African guide and resource book.* Van Schaik publishers.

Nicolas-Alonso, L.F. and Gomez-Gil, J. 2012. Brain-computer interfaces, a review. *Sensors*. 12(2):1211–1279.

Nielsen, J. and Norman, D. 2015. *The definition of user experience*. [Online], Available: http://www.nngroup.com/articles/definition-user-experience/ [2018, March 12].

Oberoi, I.S. and Chopra, M. 2010. *Web Search Engines – A Comparative Study*. [Online], Available: http://www.idt.mdh.se/kurser/ct3340/ht09/ADMINISTRATION/IRCSE09-submissions/ircse09_submission_14.pdf.

Oxford English Dictionary. 2010. 2nd ed. Oxford University Press.

Preece, J., Rogers, Y. and Sharp, H. 2015. *Interaction Design - Beyond Human-Computer Interaction*. 4th ed. West Sussex, United Kingdom: John Wiley and Sons Ltd.

Rogers, Y., Sharp, H. and Preece, J. 2011. *Interaction Design - Beyond Human-Computer Interaction*. 3rd ed. West Sussex, United Kingdom: John Wiley and Sons Ltd.

Sanderson, M. 2008. Ambiguous queries: Test collections need more sense. *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 499–506.

Schall, A. 2015. The future of UX research: Uncovering the true emotions of our users. *User Experience Magazine*. [Online], Available: http://uxpamagazine.org/the-future-of-ux-research/.

Shah, J.J. 2018. *A Closed Loop Brain Computer Interface System: Hardware Implementation*. San Diego State University.

Song, R., Luo, Z., Wen, J., Yu, Y. and Hon, H. 2007. Identifying ambiguous queries in web search. *Proceedings of the 16th international conference on World Wide Web - WWW '07*. 1169.

Song, R., Luo, Z., Nie, J.Y., Yu, Y. and Hon, H.W. 2009. Identification of ambiguous queries in web search. *Information Processing and Management*. 45(2):216–229.

Teevan, J., Dumais, S.T. and Horvitz, E. 2007. Characterizing the value of personalizing search. *SIGIR'07*.

TheFreeDictionary by Farlex. 2015. *Information retrieval*. [Online], Available: http://encyclopedia2. thefreedictionary.com/Information+Retrieval [2015, March 17].

Thorpe, J., van Oorschot, P.C. and Somayaji, A. 2005. Pass-thoughts: Authenticating with our minds. 121.

Tullis, T. and Albert, B. 2013. *Measuring the User Experience - Collecting, Analyzing, and Presenting Usability Metrics*. 2nd ed. Amsterdam: Elsevier/Morgan Kaufmann.

User Experience Professionals Association. 2014. *Definitions of user experience and usability*. [Online], Available: https://uxpa.org/resources/definitions-user-experience-and-usability [2018, March 12].

Voorhees, E. 1993. Using WordNet to disambiguate word senses for text retrieval. In Pitsburgh, PA, USA *ACM-SIGIR '93*. 171–180.

Waldert, S. 2016. Invasive vs. non-invasive neuronal signals for brain-machine interfaces: Will one prevail? *Frontiers in Neuroscience*. 10(JUN):1–4.

Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G. and Vaughan, T.M. 2002. Brain–computer interfaces for communication and control. *Clinical Neurophysiology*. 113(6):767–791.

Yahoo! Search Engine. 2018. [Online], Available: https://za.yahoo.com/ [2018, November 26].

Ying, L., Scheuermann, P., Li, X. and Zhu, X. 2007. Using WordNet to disambiguate word senses for text classification. In Vol. Part III. Berlin Heidelberg: Springer-Verlag *ICCS*. 781–789.