Fixation-Related EEG Frequency Band Power Analysis: A Promising Neuro-Cognitive Methodology to Evaluate the Matching-Quality of Web Search Results?

Christian Scharinger[(✉)], Yvonne Kammerer, and Peter Gerjets

Leibniz-Institut für Wissensmedien, Schleichstr. 6, 72076 Tübingen, Germany
{c.scharinger,y.kammerer,p.gerjets}@iwm-tuebingen.de

Abstract. We used fixation-related EEG frequency band power analyses to study the evaluation of web search results. Participants were presented thirteen search queries, each followed by a Google-like list of six search results. One of the search results was a complete match (i.e., having strong lexical and semantical overlap with the search query), one a complete mismatch (i.e., having neither lexical nor semantical overlap with the search query), and four were partial matches, having either only semantical or only lexical overlap with the search query. We analyzed the EEG alpha frequency band power for individuals' fixations during the initial viewing of each search result. We observed that on average the EEG alpha frequency band power at parietal electrodes was significantly lower for the complete match than for the other search results. Thus, this decreased alpha power might reflect participants' detection of the hit in a list of search results.

Keywords: EEG · Alpha · Eye-tracking · Web search

1 Introduction

Using web search engines like Google has nowadays become the typical starting point for any information gathering process. Despite the importance and prevalence of using web search engines in our daily life, research to date has only rarely addressed the cognitive aspects of conducting web searches by means of the electroencephalogram (EEG). In the current study we were interested in one of these cognitive aspects, namely cognitive load (CL, cf. [1]). We assessed CL during participants' evaluation of web search results by means of EEG alpha frequency band power.

Previous research showed that for increasing CL (i.e., increasing demands on executive functions and working memory), the EEG alpha frequency band power at parietal electrodes decreases [2, 3]. Especially, a decrease of alpha frequency band power in the range between 10 to 13 Hz (i.e., the upper alpha) has been associated with semantic processing demands [4, 5]. Importantly, EEG alpha power might be used as a reliable measure of CL in the context of hypertext reading and link selection [6, 7].

The use of fixation-related EEG frequency band power analysis (see Sect. 2.4) in the current study provided participants a normal reading situation during the web

search evaluation. We manipulated the manner of overlap between different search results and the search queries, thus creating search results of different matching-quality [7]. We used four different matching conditions: A complete match, showing sematic and lexical overlap with the search query (the HIT), partial matches, showing either semantic or lexical overlap (SEM and LEX), and a no-match (MISS), showing neither sematic nor lexical overlap with the search query. We expected that depending on the amount of semantical and lexical overlap between search query and search result the induced CL would vary. The MISS should be easily identified as irrelevant for the current search query, thus resulting in lower CL as compared to the CL during reading of search results that had semantic and/or lexical overlap with the search query (i.e., potential matches that therefore had to be more thoroughly processed).

## 2   Methods

2.1   Participants

Twenty-two healthy university students (age mean = 22.41, *sd* = 3.43, 17 f/5 m) participated in the study and received a payment of 8 €/h. They were all native speakers of German, right-handed, and had normal or corrected-to-normal visual acuity. The study was approved by the local ethic committee. Participants gave their written informed consent at the beginning of the study. Due to technical problems, one participant had to be excluded from further data analysis.

2.2   Materials and Procedure

We designed the task material and procedure in analogy to a study described in [8]. We used thirteen different fact-finding search queries addressing a variety of topics. Each query was formulated as a whole-sentence question presented at the center of the screen (see Fig. 1). Each search query was followed by a search engine result page (SERP), i.e., a Google-like list of six search results. One of the search results was a HIT, one a MISS, and four were partial matches (SEM, LEX). While the SEM search results – like the HIT – would provide an answer to the search query, the LEX search results – like the MISS – would not. The order of the different search result categories varied between SERPs to avoid list sequence effects (e.g., [9]). Furthermore, the order of the search queries was varied between participants.

Each search result consisted of one line of a blue-colored heading (Arial, 26 pts) and a content-summary of two to three lines (Arial, 16 pts, black color). The search results consisted of only textual content information, and no source information was provided. This was because we were specifically interested in participants' evaluation of search results based on content information only.

Participants were instructed to read the search results and then to evaluate the matching between the search query and each search result by mouse-clicking first on the best matching search result and then in descending relevance-order on the other search results. The presentation speed of search queries and SERPs was self-paced. A button labeled "To the search results" was positioned below the search query. By
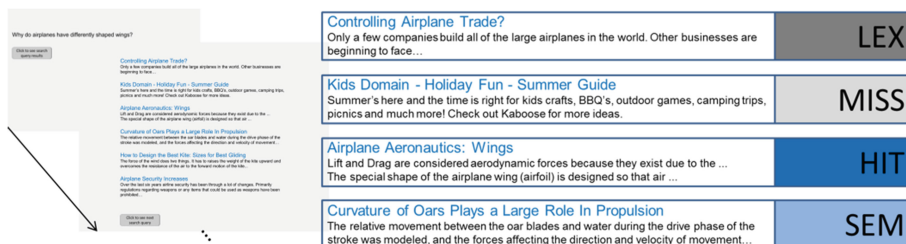
Fig. 1. Exemplary sequence of a search query ("why do airplanes have differently shaped wings?") followed by a corresponding SERP (showing four out of six search results). The task materials were originally presented in German.

clicking on this button, participants reached the corresponding SERP. Here, after a mouse-click on a search result, the blue-colored heading changed to a dark red to indicate that the mouse-click had been registered. Participants always stayed on the SERP until they clicked on a button labeled "Next search query" at the bottom of the SERP.

The task procedure was identical for all participants. At the beginning of the web search evaluation, after the calibration of the eye-tracker, written task instructions were presented as the first page on the screen, which was then followed by the first search query (see Fig. 1 for exemplary task materials).

2.3 Apparatus

The experiment was run in a quiet, dimly lit room. Participants sat in a comfortable chair in front of a 22-in. Dell monitor (1680 × 1050 pixels screen resolution) while their EEG and eye-tracking data were recorded. We used a light-gray background-color on all displayed web pages to provide a constant and pleasant brightness value to minimize eye-strain.

Eye-tracking data were recorded using a 250 Hz SMI (SensoMotoric Instruments) infrared remote eye-tracking system that was positioned below the monitor. A chin rest was used to avoid head movements during data recording and to guarantee a fixed distance of about 70 cm between the eyes and the eye-tracking device. The eye-tracking data were recorded at a sampling rate of 250 Hz (SMI iView X 2.7.13). The eye-tracker was calibrated using the built-in calibration routines (SMI Experiment Center, 9-point calibration) before the written task instruction appeared on the screen.

EEG data were recorded from 27 electrode sites (Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, O2) positioned according to the international 10/20 system. The right mastoid served as reference during recording. Ground electrode was positioned at AFz. Three additional electrodes were placed around the eyes for recording of the electro-occulogram (EOG). EEG data were recorded (PyCorder 1.0.2) at 500 Hz sampling rate (ActiCHamp, Brainproducts, Inc.) using active electrodes (ActiCap, Brainproducts, Inc.). Impedances were kept below 5 kOhm.

2.4   Analysis Procedure

Eye-tracking and EEG data were preprocessed and synchronized as described in [7]. For each SERP six equal-sized rectangular areas of interest (AOIs) were defined, one for each of the six search results (SMI BeGaze 3.41). The AOIs were used to analyze the EEG data eye-fixation related. We collapsed the data of the two semantic and the two lexical search results, resulting in four AOI categories that reflected the four different categories of search results (i.e., HIT, SEM, LEX, and MISS). Using these four AOI categories, portions of the EEG data were selected for further analyses that were related to fixations of these AOIs. Specifically, we analyzed the EEG data for individuals' fixations during the initial viewing of each search result. The initial viewing was defined as the first sequence of fixations of an AOI that in sum lasted longer than 500 ms. The EEG data was epoched in one-second-long data epochs, time-locked to the onset time of the first viewing of the AOI and labeled with the corresponding AOI category. In total, for each participant and each category thirteen epochs (HIT, MISS) respectively 26 epochs (LEX, SEM) were created.

Mean EEG frequency band power was then calculated for the one-second EEG data epochs using fast-fourier transforms (FFTs) for the upper alpha frequency band spectrum (10 Hz to 13 Hz). The alpha power was then averaged for each participant over all epochs of each of the four AOI categories (i.e., the matching-conditions).

3   Results

Participants' average reading times for the SERPs were 39.15 s (sd = 11.91). The average initial viewing durations were: HIT, 2.84 s (sd = 1.44), SEM, 2.64 s (sd = 1.36), LEX, 1.92 (sd = 0.97), and MISS, 1.42 s (sd = 0.63). A one-factorial repeated measures ANOVA revealed a main effect of matching-condition, $F(3,60) = 24.43$, $p < .001$, $\eta_p^2 = .55$. Paired-sample t-tests (two-way, Bonferroni-corrected) confirmed a significant decrease in viewing durations from HIT to LEX to MISS as well as from SEM to LEX to MISS (all p < .001), whereas there was no significant difference in viewing durations between SEM and HIT (p > .99).

Interestingly, the EEG alpha frequency band power data (see Fig. 2) showed a different result pattern. A one-factorial repeated measures ANOVA also revealed a main effect of matching-condition, $F(3,60) = 7.91$, $p < .001$, $\eta_p^2 = .28$. The mean upper alpha power at electrode Pz for the conditions SEM (8.50[1], sd = 0.50), LEX (8.43, sd = 0.62), and MISS (8.22, sd = 0.48) did not differ (p > .88). However, the HIT showed significantly lower upper alpha power than the other three conditions (HIT vs. MISS, $p = .036$, HIT vs. SEM, $p = .001$, HIT vs. LEX, $p = .028$). Topoplots (see Fig. 2, right part) showing the upper alpha frequency band power expressed as ERD/ERS%-values [6] at all electrode sites over the scalp underlined the parietal-central localization of the decreased alpha for the HIT. This is in line with literature on CL [3, 6] and justifies our selection of Pz as indicative electrode.

---

[1]   EEG alpha power is expressed in $10*\log_{10} \mu V^2/Hz$. For reasons of readability we omitted writing the unit in the text.
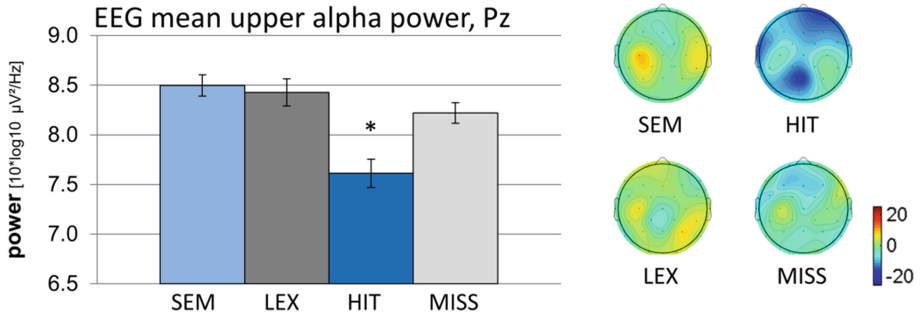
Fig. 2. Left: bar plots showing the mean alpha frequency band power (10–13 Hz) at parietal electrode (Pz) for the partial matches (SEM, LEX), the complete match (HIT), and the mismatch (MISS). Note. * = p < .05. Right: Topoplots showing alpha ERD/ERS% values [6]. As baseline for the ERD/ERS% served the mean alpha frequency band power of all conditions.

4   Discussion

We used a methodology of fixation-related EEG frequency band power analysis to study users' cognitive load (CL) during the evaluation of web search results. With respect to CL, our initial hypothesis was not supported by the EEG data. EEG alpha frequency band power did not differ between the no-match and partial matching search results. Instead, we observed a significantly decreased upper alpha frequency band power – indicating increased CL – for the search result that had semantical and lexical overlap with the search query (i.e., the HIT), as compared to the other search result categories. Importantly, this outcome could not be simply attributed to confounding factors like different viewing times of the four AOIs, as the result pattern of mean AOI fixation durations was different from the EEG outcomes. The observed EEG pattern indicates that when readers encounter a hit for their current search query, they might process this search result more thoroughly. This might lead to increased CL and hence decreased EEG alpha frequency band power. This interpretation is corroborated by studies showing EEG alpha to reflect semantic processing demands [4] and also by studies showing decreased alpha frequency band power as indicative for complex decision making [10]. However, it is an open question why the processing of the search results with only semantic overlap (SEM) did not cause a similar or even higher CL.

Clearly, the current study might only serve as a starting-point for a more thorough examination of cognitive aspects of web search evaluation in future research. Nonetheless, it might indicate that the methodology of fixation-related EEG frequency band power analyses can provide important insights into these cognitive aspects that are not covered by other measures. Especially, alpha frequency band power might be a good measure to cognitively evaluate the matching-quality of search engine results and thus add an important factor to traditional methods of web search evaluations.

References

1. Gwizdka, J.: Distribution of cognitive load in web search. J. Am. Soc. Inf. Sci. Technol. **11**, 2167–2187 (2010)
2. Scharinger, C., Soutschek, A., Schubert, T., Gerjets, P.: When flanker meets the N-back: what EEG and pupil dilation data reveal about the interplay between the two central-executive working memory functions inhibition and updating. Psychophysiology **52**, 1293–1304 (2015)
3. Gevins, A., Smith, M.E., McEvoy, C.L., Yu, D.: High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. Cereb. Cortex **7**, 374–385 (1997)
4. Klimesch, W., Doppelmayr, M., Pachinger, T., Russegger, H.: Event-related desynchronization in the alpha band and the processing of semantic information. Brain Res. Cogn. Brain Res. **6**, 83–94 (1997)
5. Klimesch, W.: Alpha-band oscillations, attention, and controlled access to stored information. Trends Cogn. Sci. **16**, 606–617 (2012)
6. Antonenko, P., Paas, F., Grabner, R., van Gog, T.: Using Electroencephalography to measure cognitive load. Educ. Psychol. Rev. **22**, 425–438 (2010)
7. Scharinger, C., Kammerer, Y., Gerjets, P.: Pupil dilation and EEG alpha frequency band power reveal load on executive functions for link-selection processes during text reading. PLoS ONE **10**, e0130608 (2015)
8. Keil, F.C., Kominsky, J.F.: Missing links in middle school: developing use of disciplinary relatedness in evaluating internet search results. PLoS ONE **8**, e67777 (2013)
9. Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., Granka, L.: In Google we trust: users' decisions on rank, position, and relevance. J. Comput. Commun. **12**, 801–823 (2007)
10. Davis, C.E., Hauf, J.D., Wu, D.Q., Everhart, D.E.: Brain function with complex decision making using electroencephalography. Int. J. Psychophysiol. **79**, 175–183 (2011)