Eye tracking and EEG features for salient Web object identification

Gino Slanzi, Claudio Aracena, and Juan D. Velásquez

Department of Industrial Engineering, Universidad de Chile
Av. República 701- P.O. Box: 8370439
Santiago, Chile
gslanzi@ing.uchile.cl, caracena@ing.uchile.cl, jvelasqu@dii.uchile.cl

Abstract. We propose a biological-based feature comparison for identifying salient Web objects. We compare several features extracted from eye tracking and EEG data with a baseline given by mean fixation impact introduced by Buscher. For this, we performed an experiment with 20 healthy subjects in which gaze position, pupil size and brain activity were recorded while browsing in a Web site adaptation. Our results show that there are EEG features that could be related to Web user attention in objects. In particular the Gamma Band RMS and the EEG Variance indicate that the longer subjects view a web object (more attention), the less brain signal disturbance appears. We also discarded pupil size features due to low correlation with baseline. These results suggest that EEG features could be used to identify salient objects without using the time users spent on them as done in previous methodologies.

1 Introduction

The penetration of the Web has changed people's behaviour over time. For example, when any sort of information or product is required, people usually, and almost naturally, check the Web. Thus, companies and organizations have wanted to get presence in this network and increase their sales and market position. To achieve effectiveness in their goals it is necessary to design web sites that can attract more customers than competitors' web sites.

However, designing and implementing attractive websites require knowledge about customer behavior and preferences. For that purpose there are several techniques for discovering customer experience while browsing a website, including polls, surveys, weblog analysis, etc. In addition to those techniques several modern methodologies such as mouse tracking have been developed in order to extract more objective patterns from web user behavior.

In this study we compare different types of data analysis from web user behavior, including eye tracking, pupil size and electroencephalography. The aim is to find out what the most relevant objects on a web site are depending on the different data analyses. In addition, we discuss the salient web object identification differences between each analysis.

This comparison is interesting due to the fact that each biological response can explain different human behaviors. For instance, eye fixations have been related with attention in the focus area [1], pupil size has been related with cognitive load or mental activity [4] and EEG signals have been related with many phenomena, in particular, emotional or cognitive states [7].

To achieve our objective, we performed an experiment where the gaze movements, pupil size and EEG signals were recorded for 20 subjects. The task consisted of browsing 32 web pages of the MBA program of the University of Chile's web site.

The paper is organized as follows, first we present some related research, and then describe our approach for identifying salient objects and features from biological signals used as comparison measures. After specifying the experimental setup, we will attempt to answer the data processing and research questions. Next, the results are shown along with their pertinent discussion, and finally we conclude our study and propose future work.

2 Related Work

One remarkable line of salient web object identification was developed by Buscher et al. The main motivation comes from the need to understand how people allocate visual attention on Web pages, taking into account its relevance for both web developers and advertisers. In 2009, authors implemented an eye-tracking-based analysis in which 20 users were shown 361 pages while performing information foraging and inspection tasks [1]. The main assumption was that gaze data could represent a proxy of attention. From that, an analysis framework was developed by first generating a tool that allows DOM elements to be characterized and a mapping performed between gaze data and the DOM elements. The second part involves the use of the extracted web features in a machine-learning setting to predict salient elements on a web page.

Another relevant contribution by Buscher et al. is the introduction of the concept of *fixation impact*. It allows the identification of the set of elements that are under the gaze of the user at a certain time. It follows empirical studies that show that human vision is characterized by a narrow window of high acuity along with the standard gaze area. Thus, when visualizing an element, it means that other elements in the surroundings are also being considered. Therefore, given a fixation point, a DOM area is selected in order to identify every element under it. A distance score is given to each element based on its coverage, assuming a Gaussian distribution. The fixation impact is computed using this distance and also incorporating a time dimension, which means the fixation duration.

A methodology to extract salient web objects was developed by Velásquez et al. This methodology started with the analysis of plain text for identifying the *Website Keywords*, defined as "*word or possibly set of words that is used by visitors in their search process and characterizes the content of a given web page or web site*" [10]. Afterwards, the methodology was extended, defining a *web object* as "*any structured group of words or multimedia resource within a* *web page that has metadata to describe its content*" and a *Website Keyobject* as "*the web object or group of web objects that attracts web users' attention and that characterizes the content of a given web page or web site.*" Thus the main objective of the methodology turned into identifying *Website Keyobjects* instead of *Website Keywords* [9]. One problem presented in the methodology was the application of a survey for collecting information about user preferences, thus acquiring subjective results. To solve this problem, eye-tracking technology was incorporated, with the result that the time spent on each object was able to be extracted in a more precise and objective way [8].

Dimpfel and Morys in [2] used quantitative features from EEG to analyze five websites and an eye-tracking device was added, mainly for tracking gaze movements. These features first tried to measure attention and activation and then results were compared with a typical survey. The results show that using EEG features can be helpful in website analysis, but more studies are needed to confirm if this kind of research could be helpful in other scenarios, such as advertising.

Khushaba et al. [5] have been researching consumer neuroscience, in particular, user preferences using EEG and ET data. Their studies aim to find interdependencies among the EEG signals from cortical regions in a decision-making environment, and also a way to quantify the importance of different product features such as shape, color, texture, etc., in these decisions. Results showed there was a clear and significant change in the EEG power spectral activities taking place mainly in the frontal, temporal, and occipital regions when participants indicated their preferences.

3 Proposed Approach

The main goal of this study is to investigate salient Web object identification using different biological features that we describe in this section. In particular, we propose a comparison of eye gaze, pupil dilation and electroencephalogram features for identifying these relevant Web objects.

3.1 Web Object Identification

The initial element of analysis is the *web object*, which is defined as any combination of DOM elements that comprises an idea or a concept. For example, the combination between an image and an adjacent text paragraph could represent a defined block on which the user focuses his attention. The decision of using an aggregated representation and not the original DOM elements resides in our observation that *(a)* DOM objects are usually too small to satisfy the level of granularity the visual attention provides. Since fixation usually encompasses a set of elements, *(b)* using an aggregated representation provides a better understanding of the user interest, as the level of information that can be extracted from the semantic combination is richer.

The task of grouping DOM elements into web objects is not a trivial process, since several criteria could be used, leading to different sets. In our case, the selection task consists of presenting all the Web pages in the Web site to the expert in a sequential order. For each page, the expert is asked to arrange the elements into groups, and write down a unique identifier. The criteria for grouping is left to the expert, therefore no specific requirement is requested. Finally, through a DOM manipulation process, the web object is fully identified and its characteristics, such as size and position, are extracted.

3.2 Comparison Measures

There are several ways to estimate and rank which elements capture the attention of users, such as post-navigation surveys and questionnaires. Although these approaches are easy to perform, they do not provide a robust response because each user has different perceptions for each web element, hence, these kinds of responses are not representative of the real degree of relevance a user gave to each Web object.

An experiment was performed where an eye-tracking device was used to capture subjects' gaze movements and measure pupil size. In parallel we included the brain signal recording with an electroencephalogram device. Thus, it allowed us to have a triple data stream of biological signals.

As we collected the raw data, we extracted relevant features included in the state of the art for providing an objective comparison. For each data type it is possible to obtain several features that we discuss below.

Eye Gaze Measures. We used a type of eye tracking that allowed us to obtain a reliable spectrum of the visual activity for each Web page and for each Web element. Then, having this recording as a stream of data with an associated time component, we identified interest points, formally called *fixations*, which consist of periods of time in which the user was focused on a defined point. Each fixation is preceded and followed by a *saccade*, which is a transition between elements.

As a quantitative metric for fixation time, we followed the approach by Buscher et al. [1], namely the *mean fixation impact* (MFI). In that sense, we take into account that as human vision has a narrow window of high acuity, called *fovea*, when fixating on a specific area, the user is also gathering information from its surroundings. If the user is focusing on a Web object, attention resources are also being distributed to other elements as well. To capture that phenomenon, it is assumed that the attention allocation follows a Gaussian distribution of volume 1. This forms a circle around the fixation point and for all Web elements that intersect, an attention score is computed.

One particular element can receive attention from several fixations during a Web session, so the score is defined as the addition of all the contributions. This represents the attention based on an information foraging task on a specific page. Formally, given F_{pu} the set of fixations the user u produced on page p, the aggregated fixation impact for the Web object o is

$$I(o,p) = \sum_{u \in U} \sum_{f \in F_{pu}} \Phi(f,o) \tag{1}$$

where $\Phi(f,o)$ computes the proportion of attention based on the Gaussian distribution for a given fixation f on the element o.

Pupil Dilation Measures. Pupil size has been related with different cognitive processes since it is closely linked with the sympathetic and parasympathetic systems. For instance, Hess et al. [4] studied the relation of pupil size and mental activity in simple problem solving and Goldinger et al. [3] studied this response related with memory retrieval, among others.

Since it is not a straightforward task to define a measure that can express all the underlying patterns present in pupil size, we used two measures based on studies of pupil size and its relationships.

- **mean pupil size**: This metric is calculated as the mean of pupil size while subjects are fixating their gaze on a particular web object. Then, for each object a grand average considering all subjects is calculated. This measure indicates a level of pupil size for each object and can lead to different interpretations according to object characteristics, for example, it would be different if the object corresponds to plain text or an image.
- **mean delta indicator**: A pupil size versus time wave is generated when users browse the Web page. This signal can be described as a smooth continuous curve. In most cases, this curve has a unique maximum and minimum and the difference between them, called *delta indicator*, can show a strong or soft biological reaction depending on its value. A possible interpretation of this measure can be the arousal level, which means that if there is a large *delta indicator* it could indicate that an object provokes a higher level of attention in users. As well as mean pupil size, the *delta indicator* is calculated for each object considering a grand average among all users.

Electroencephalogram Measures. We include brain activity analysis by means of an EEG recording. For each object, a mean EEG signal is calculated and transformed into different parameters. In order to use brain activity as a measure for identifying user interest of each Web object, we propose the following features:

- **Frequency Band Features**: A useful way to analyze EEG waves is to separate them into different frequency bands. A possible way is to use *Wavelet Transform*, in which each of these bands coincide with the standard EEG frequency bands (*Delta* $0-4$ Hz, *Theta* $4-8$ Hz, *Alpha* $8-16$ Hz, *Beta* $16-32$ Hz, *Gamma* $32+$ Hz). Then for each band the energy, RMS and power were computed.
- **Typical Statistics**: Mean, variance and standard deviation was calculated for each object signal, in order to analyze if these values could be useful as a quantifier of users' attention, focus or interest in objects.

3.3 Research Questions

In order to compare the previously-defined features, we will consider the MFI as a baseline for comparison, due to its relevance as a proxy of user attention [1]. As we anticipate, for establishing an objective comparison between our baseline and the other features, we propose the following research questions:

1. **Which features can be discarded due to low correlation with the baseline feature, and which ones could be considered the most similar?**
2. **Among the most similar features, what are the differences regarding Web object type and structure?**

4 Experimental Setting

In order to obtain the data, an experiment was performed considering different aspects that allowed the reproduction of user Web browsing, while monitoring and recording eye gaze position, pupil dilation and EEG. This experimental stage took place at the Neurosystems Laboratory of the Faculty of Medicine of the University of Chile.

Participants. We used 20 healthy subjects, 3 females and 17 males aged between 22 and 25 years old (mean = 24.2, variance = 1.64). All subjects declared having normal or corrected-to-normal vision and did not have any neurological or psychiatric illness. All participants had to sign an informed consent approved by the *Ethical Committee of the Faculty of Medicine of the University of Chile.*

Design. We used a website adaptation based on an MBA program website offered by the University of Chile. Thirty-two web pages containing 359 objects were transformed to images. Objects were extracted according to the process described in subsection 3.1. Each image was divided depending on its length to generate sub-pictures of 1600 x 900 pixels. The experiment consisted of a website simulation made up of images, where subjects could move below, above and forward at will. The instruction that was given was **Browse the site freely, with no time limits (minimum nor maximum) on each page. Use the keyboard up and down arrows for scrolling and right arrow to show the next page**.

Instrumentation. Image presentation was controlled by the *Experiment Builder* software by *SR Research.* Pictures were displayed on a 32″ LG screen located in the experimental room, at a distance of 80 cms from the subject. Pupil size and gaze position was recorded using an *EyeLink 1000* eye tracker by *SR Research*, this device recorded both eyes at a rate of 500 Hz during all the experiment. Subjects' heads were adjusted by a chin support that helped to keep the head steady. For brain activity data, the *BioSemi Active 2* EEG system was used at a 2000 Hz sampling rate. 32 scalp electrodes were placed according to the $10-20$ international system, in addition to 8 external electrodes placed in the ocular and mastoid zone. The experimental room had no light on during recording sessions.

5 Results and Discussion

Once the data was fully acquired, some treatments were performed in order to answer the research questions proposed in 3.3. This section describes this treatment and gives a solution to those questions and the respective discussion.

Data Pre-processing and Transformation. The eye tracking and EEG data were preprocessed separately. For the eye-tracking data, the pupil dilation signal was taken and preprocessed by linearly interpolating blinks and fixing saccade offsets. Additionally, a low-pass filter of 2 Hz was used to remove noise. Ocular positioning was used to determine what object was being seen and the time spent on each one.

For the EEG data, first of all, the sampling rate was reduced from 2000 to 500 Hz, for synchronization with the eye-tracking data. Then the data was filtered with a $1 - 60$ Hz bandpass filter. Blinks, saccades and other irregularities were removed using the Matlab toolbox Eeglab.

Considering a time window of 300 ms for the minimum fixation, a mean signal was calculated for each object. This was done for both pupil dilation and EEG signal (for the average of the 32 scalp electrode signals). As mentioned in 3.2 a series of features were obtained for these signals: Merging the Web object identification and the eye tracking and EEG data allowed us to characterize each object according to the eye gaze, pupil dilation and EEG features. Particularly, for EEG frequency band features, the Daubechies-5 wavelet function in addition to a 6 level decomposition were used, in order to have 6 scales of details (d1-d6) and a final approximation (a6), since the sampling rate was set at 500 Hz.

RQ1 - Feature Comparison. The set of Web elements defined for the study Web site was arranged in decreasing order according to their MFI values. Buscher et al. used this feature as a proxy for users' visual attention, thus based on this evidence, we can have a list of ordered Web elements as a comparison baseline. In this sense, it is important to say that every object within the complete Web site was considered unique. That is, if object $ID1$ was present on pages 1, 2, and 3, in the baseline list this object would appear three times, together with its respective impact value and page. In the end the baseline list consisted of 1014 elements.

In the same way, for each object we computed the proposed features and proceeded to order them without considering objects with respective EEG feature values equal to 0. We also noticed that sorting elements decreasingly, yielded no relationship at all, whereas the increasing sorting had *high* levels of similitude with the baseline for some variables. Then, to obtain a comparison measure, we took the first 100 objects of both baseline and EEG features, and counted the repeated objects.

Accordingly, we found EEG-based features that allow the identification of approximate users' attention in a similar way to the MFI. There were also variables that did not have any relation with the baseline, which were then discarded for further analysis, for example the *mean*.

We performed an analogous procedure for the pupil dilation features, *mean pupil size* and *mean delta* of each object, finding that these variables could not represent users' attention in the same way as the baseline, having no correlation with it at all. Thus, pupil size variables were discarded as well. In Table 1 is an example set of features compared with the baseline, where variables can be seen with both high and low correlation to it.

Table 1. Example of EEG and ET feature comparison performance

Feature	γ Power	γ Energy	γ RMS	MeanEEG	Var	DeltaInd	MeanPupil
Matching	74	79	80	2	79	32	5

From not discarded features, we chose those that could represent in a better way relations with the MFI. In the previous analysis, *Energy*, *RMS* and *Power* features for all bands have analogous numbers of salient web objects compared with our baseline. Nevertheless, it is possible to observe that the *RMS of Gamma band* have a slightly better performance in most of the cases. Another relevant feature is the *EEG Variance* which shows good performance too.

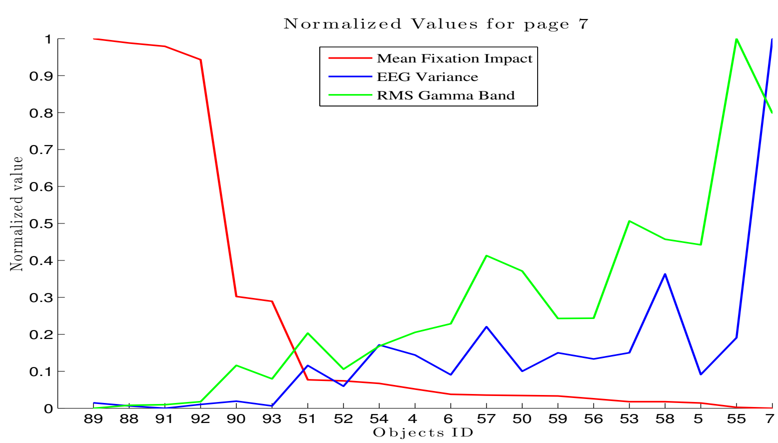

Fig. 1. Normalized values of *Mean fixation impact*, *EEG Variance* and *RMS of Gamma band* for page 7.

For comparing, graphs were constructed to show how the features behave versus our baseline. These graphs depict feature values per ID object on a particular page, while for display purposes we just included 20 objects as a maximum and range normalized values in [0,1]. Figure 1 shows the graph for MFI, *EEG Variance*, and *RMS of Gamma band* for page 7. As seen, objects are in descending order according to MFI values. A negative correlation exists between baseline features and the others ($R = -0.41$ for *EEG Variance* and $R = -0.61$ for *RMS of Gamma band*). This finding could suggest that the longer subjects view a web object, that is, the more attention given, the less brain signal disturbance appears. This fact can be supported with evidence from [6], where it was found that a lower signal amplitude level is related to an attention/learning state.

RQ2 - Web Objects Difference Analysis. We constructed attention maps for each feature under study, thus the differences between them can be graphically observed. The way to build the attention map is to depict a red circle on each web object, where the bigger the circle is, the more attention was attracted to it. Since *EEG Variance* and *RMS of Gamma band* has a negative correlation with MFI a conversion was made, in which small values in *EEG Variance* and *RMS of Gamma band* correspond to large circles on maps.

Figure 2 shows attention maps. The MFI map presents more attention on web objects with text as we expected, because it is necessary to spend more time to get an idea about them, unlike with pictures. However, if we look at the *EEG Variance* and *RMS of Gamma band* maps, we can observe that pictures and text objects have a similar amount of attention. This evidence could suggest that these features are a proxy of attention independent of the amount of time spent on web objects.

(a) Mean fixation impact (b) EEG Variance (c) RMS Gamma band

Fig. 2. Attention maps for selected features for page 7.

6 Conclusion and Future Work

In this work we studied the relationship between EEG and eye-tracking features in order to improve and obtain a more objective method of salient web object identification. We conducted an experiment where EEG and eye-tracking data were recorded while subjects were surfing on a web site. Then, features were extracted and compared in order to determine which was more closely related to our baseline feature, *mean fixation impact*. The results suggest that *Gamma band RMS* and the *EEG variance* features can help to identify salient objects without considering the time that subjects spend on each object as eye-tracking features do.

As future work we want to use EEG features for salient web object identification and compare the results with similar methodologies such as those proposed by Buscher et al. in [1]. From this comparison we expect to improve the results obtained by that author.

Acknowledgements

This work was partially supported by the Millennium Institute on Complex Engineering Systems (ICM: P-05-004-F, CONICYT: FBO16).

References

1. Buscher, G., Cutrell, E., Morris, M.R.: What do you see when you're surfing?: using eye tracking to predict salient regions of web pages. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp. 21–30. ACM (2009)
2. Dimpfel, W., Morys, A.: Quantitative objective assessment of websites by neurocode-tracking in combination with eye-tracking. Journal of Behavioral and Brain Science 4, 384–395 (2014)
3. Goldinger, S.D., Papesh, M.H.: Pupil dilation reflects the creation and retrieval of memories. Current Directions in Psychological Science 21(2), 90–95 (2012)
4. Hess, E.H., Polt, J.M.: Pupil size in relation to mental activity during simple problem-solving. Science 143(3611), 1190–1192 (1964)
5. Khushaba, R.N., Wise, C., Kodagoda, S., Louviere, J., Kahn, B.E., Townsend, C.: Consumer neuroscience: Assessing the brain response to marketing stimuli using electroencephalogram (eeg) and eye tracking. Expert Systems with Applications 40(9), 3803–3812 (2013)
6. Li, X., Hu, B., Zhu, T., Yan, J., Zheng, F.: Towards affective learning with an eeg feedback approach. In: Proceedings of the first ACM international workshop on Multimedia technologies for distance learning. pp. 33–38. ACM (2009)
7. Pastor, M.C., Bradley, M.M., Löw, A., Versace, F., Moltó, J., Lang, P.J.: Affective picture perception: emotion, context, and the late positive potential. Brain research 1189, 145–151 (2008)
8. Velásquez, J.D.: Combining eye-tracking technologies with web usage mining for identifying website keyobjects. Engineering Applications of Artificial Intelligence 26(5), 1469–1478 (2013)
9. Velásquez, J.D., Dujovne, L.E., LHuillier, G.: Extracting significant website key objects: A semantic web mining approach. Engineering Applications of Artificial Intelligence 24(8), 1532–1541 (2011)
10. Velásquez, J.D., Weber, R., Yasuda, H., Aoki, T.: A methodology to find web site keywords. In: e-Technology, e-Commerce and e-Service, 2004. EEE'04. 2004 IEEE International Conference on. pp. 285–292. IEEE (2004)