When Relevance Judgement is Happening?
An EEG-based Study

Marco Allegretti
Mitech Srl
Mantova, IT
marco.allegretti@
mitechsrl.it

Yashar Moshfeghi
School of Computing Science
University of Glasgow
Glasgow, UK
Yashar.Moshfeghi@
glasgow.ac.uk

Maria Hadjigeorgieva
School of Psychology
University of Glasgow
Glasgow, UK
maria.hadjigeorgieva@
glasgow.ac.uk

Frank E. Pollick
School of Psychology
University of Glasgow
Glasgow, UK
Frank.Pollick@
glasgow.ac.uk

Joemon M. Jose
School of Computing Science
University of Glasgow
Glasgow, UK
Joemon.Jose@
glasgow.ac.uk

Gabriella Pasi
DISCo
University of Milano-Bicocca
Milan, IT
pasi@disco.unimib.it

ABSTRACT

Relevance is a central notion in Information Retrieval, but it is considered to be a difficult concept to define. We analyse brain signals for the first 800 milliseconds (ms) of a relevance assessment process to answer the question "when relevance is happening in the brain?" with the belief that it will lead to better operational definitions of relevance. For this purpose, we devised a user study in which we captured the brain response of 20 participants. Using a 64-channel EEG device, we measured the electrophysiological activity of the brain while the subjects were in the phase of giving an explicit judgement about the relevance of presented images according to a given topic. Analyses were then performed over different time windows of the recorded EEG signals using repeated measures ANOVA. Data reveal significant variation between relevance and non-relevance within the EEG signals from the presentation of the image to 800 milliseconds afterwards. At an early stage these differences were located at frontal and posterior electrode sites. However, at later stages these differences were located in central, centro-parietal and centro-frontal areas. Our findings are an important step towards (i) a better understanding of the concept of relevance and (ii) a more effective implicit feedback systems.

Categories and Subject Descriptors: H.3.3 Information Storage and Retrieval - *Information Search and Retrieval - Search Process*

General Terms: Performance, Experimentation

Keywords: Relevance, EEG

1. INTRODUCTION

In this paper we focus on the process of relevance assessment performed by humans in an image retrieval task. The main aim of this study is to identify the time frame in which the brain has the highest activities with regard to the process of relevance assessment from the moment of observing a stimulus. Identification of the period of time in which there is a clear association of the activation of brain regions with the task of relevance assessment can be the basis to detect implicit signals of the relevance of images. This may lead to robust definitions and implementation of the concept of relevance.

Relevance is a core notion in Information Retrieval. It has been advocated that the relevance of an information object to the information need of a specific user is a subjective and multidimensional concept, which encompasses various properties and characteristics of the sought information objects [3]. In Information Retrieval systems searchers express their information needs via a set of query words, a process considered to be uncertain and noisy [13]. A progressive disambiguation of the user information need can be achieved through an interactive and iterative process known as the relevance feedback cycle. This process may rely either on explicit or implicit user's indications of relevant objects that the system processes to construct a better representation of the user's needs. To account for the highly interactive nature of the IR task, a number of relevance feedback techniques have been introduced in the literature, which vary from explicit to implicit.

Techniques that rely on explicit feedback are cognitively difficult to use, and hence implicit feedback based systems have been proposed, which, though noisy, allow us to collect several distinct signals of the user's interests through the analysis of both the user's actions and user generated contents [14]. Over the last few years affective and physiological features have been considered as valid ground to define implicit feedback techniques [1, 10, 8]. For example, Arapakis et al., studied the role of emotions in information retrieval, and introduced a number of models [1, 2]. In their subsequent work, they showed that emotional features can

be effectively included in building implicit feedback systems [2], and they can also be used to personalise searching [1]. Another example is the work by Moshfeghi et al. in which they demonstrated that, in addition to emotional features, physiological features can also be used to model relevance [8], and they can also be used to predict task types [9].

More recently, a new wave of research was initiated to better understand relevance by analysing brain activity. Moshfeghi et al. [10] show that it is possible to identify brain regions activated during the process of relevance judgement. Another example is the study conducted by Eugster et al. [4] using EEG to show that the frequency content of the EEG signal over extended time windows, along with Event Related Potentials (ERPs) can be used as an effective feature set for decoding relevance of text. Likewise, Kauppi et al. [6] using magnetoencephalography (MEG) show that the frequency content of the MEG signal along with eye movement data can be used for decoding relevance of images. Finally, Zhang et al, [15] show that brain signals monitored via EEG may be used to predict term relevance. Despite the great progress made over this period in terms of better understanding relevance and how to use it more unobtrusively, these studies fail to answer an important and complementary question which is *when* the relevance assessment is happening in the brain. In this paper we aim towards answering this research question.

We conjecture that EEG signals can be exploited to detect significant brain activities in the process of assessing relevance of images. In particular, we aim to identify the time frame in which that significative activity appears in a user's brain while they are assessing pictures. In this spirit, this paper reports the results of an EEG-based user study on image search. We therefore devised an experiment consisting of collecting data via a 64-node EEG in a lab-based user study, and analyzed the collected data off-line. In contrast to the approaches [4, 6], we examined the time variation of the amplitude of the EEG response to relevant and non-relevant items so as to better understand how the difference in relevance evolves over time. In fact, our hypothesis is that there is a time frame in which the EEG signals would be different for relevant documents from non-relevant documents. The rest of the paper is organised as follows: the experiment methodology is described in Section 2, results and discussion are presented in Section 3, and finally the paper is concluded in Section 4.

2. EXPERIMENTAL METHODOLOGY

In this section, we report detailed settings of our experimental methodology.

Experiment Design: This study used a within-subject design. The independent variable was the relevance (with two levels: relevant, and non-relevant), The dependent variables were the EEG signal of the brain.

Tasks: During the experiment, subjects were asked to judge the relevance of images according to a topic. The images used were selected from the ImageCLEF 2009 - Photo Retrieval Task[1]. To this end, 100 images (stimuli), 50 relevant and 50 non-relevant, were shown to each subject. In order not to tire them, stimuli were presented in two separate tasks of 50 images each. For both tasks users could choose between two different topics from a total of four topics, one for each ImageCLEF used in the experiment (11, 49, 02 and 46).

Apparatus: The experiment was run in a purpose-designed room, with a soft light in order to avoid distractions. Stimuli were presented with E-Prime 2.0[2] installed on a PC with a video resolution of 1024 x 768 pixels. Subjects were able to interact with the computer via a keyboard.

EEG signals were recorded with a 64 Ag/AgCl electrodes BrainProducts cap (BrainCapMR)[3]. Eye movements and the electrical activity of the heart were monitored by 2 electrodes (EOG, which was the combination of VEOG and HEOG, and ECG), while the other 62 channels were placed on the scalp at the locations based on the International 10-20 system[5]. All the 64 channels were referred to the reference electrode placed in the middle of the scalp. Data was digitally sampled for analysis at 1024 Hz sampling rate with a 0.016 to 250 Hz analogue bandpass filter and a resolution of 0.5 μV per bit.

Participants: Data were collected from 20 participants (11 females and 9 males) with European origins (10 from UK and 10 from other EU countries). They were recruited from the University of Glasgow population (12 BSc, 5 MSc and 3 PhD) and the average age was 22.35 ($\pm$ 1.66) years (from 20 to 25). All of them had a normal or corrected-to-normal vision, no history of neurological or psychiatric disorders and declared not to be on any drug or medication which could affect their EEG.

Procedure: The user study was carried out in the following manner. The formal meeting with the participants took place in the laboratory setting. At the beginning of the session the participants were given an information sheet which explained the conditions of the experiment. They were then asked to sign a Consent Form and were notified about their right to withdraw at any point during the study, without affecting their legal rights or benefits. Then, they were given an Entry Questionnaire to fill in. The session proceeded with a brief tutorial on the use of the interface with a short training task.

Figure 1 shows how a stimulus was presented to participants. At the beginning, a fixation cross was shown for 1000 ms and it was used to capture the attention of the participant in the middle of the screen. Then, an image, which was randomly chosen among the 50, was presented for 1000 ms. During this period of time, from the appearance of the fixation cross on the screen to the disappearance of the image, the participant was asked to stay as relaxed as possible, and to try not to blink or move[4]. After the presentation of the stimulus, participants were asked to provide a feedback about relevance by pressing a key of the numeric keypad on the right-hand side of the keyboard. The subjects were instructed to press key 1 with the first finger of their right hand for positive feedback (which means the image was relevant), or press key 2 with the second finger for negative feedback. In order not to put any pressure on the participants, and reduce the probability of errors in pressing buttons, there were no time limits to answer the question. In the end, a blank screen was shown for 1500 ms to let them rest before the next picture.

---

[1]More information about the ImageCLEF 2009 - Photo Retrieval Task at: http://www.imageclef.org/2009/Photo. Relevant images were randomly taken from ImageCLEF 2009 tasks 11, 49, 03 and 46.

[2]E-Prime, Psychology Software Tools©. https://www.pstnet.com/

[3]Brain Products©. http://www.brainproducts.com

[4]These requirements are important in order to ensure good quality of the collected EEG signals.
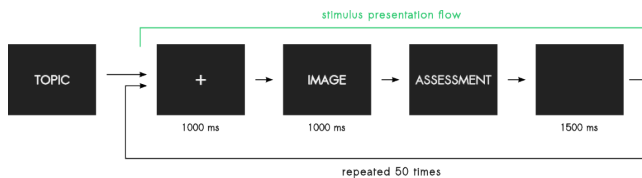
Figure 1: The figure shows the structure of a task. From the left, the first screen was dedicated to show the topic. Once ready, a participant could press any key on the keyboard to start. Firstly, a fixation cross was presented for 1000 ms. Then, a random image from the collection was shown for 1000 ms. This was followed by the question about relevance, which had no time limit. At the end, a black screen was shown for 1500 ms.

Finally, an Exit Questionnaire was administered at the end of the session. Participants were then asked to sign a payment form, prior to receiving the payment. Each study took approximately 100 minutes to complete. Users could only participate once. Prior to running the actual user study, a pilot study was performed using 5 participants to confirm that the process worked correctly and smoothly. A number of changes were made to the system based on feedback from the pilot study.

Preprocessing Steps: The acquired EEG signals were pre-processed using the tools provided by EEGLAB[5]. The pre-processing steps started by referencing the data to the common average (CAR). Then, a band-pass filter, from 0.3 to 40 Hz, was applied to remove power line noise. Epochs were then extracted from 500 ms before stimulus presentation to 1000 ms afterwards. Bad epochs containing only noise were manually removed. Finally, a spatial filter was run using Independent Component Analysis (ICA) in order to remove interference caused by eye blinking and other artifacts (e.g., muscle movement).

Time Frame Based Analysis: We decided to examine the behavior of the EEG signals across all the first 800 ms from the time when the stimuli was presented to the participants. In order do so, we defined four consecutive windows of time as follows: from 0 to 180 ms, from 180 to 300 ms, from 300 to 500 ms, and from 500 to 800 ms. A similar analysis strategy was adopted by Koelstra et al. [7] and included a time window of 300-500 ms. A decision for the other values used in these time boundaries was based on preliminary examination of the data, consideration of the task and consideration of the literature on the timing of EEG events in relation to perceptual and cognitive processing. For example, since approximately 150 milliseconds is the shortest duration of Event Related Potential (ERP) related to visual object categorization [12] we chose 0-180ms as our first epoch to include any such ultra-rapid potentials. Following Koelstra et al. we also chose 300-500 ms, a range that would include activity such as the P300, which has been related to attentional and memory processing [11].

Statistical Analysis of Button Responses: The collected responses were compared to the relevance assessment set (Qrel) provided with the ImageCLEF 2009 tasks. In general, the overall accuracy was 0.89 (± 0.10). In particular, the accuracy over relevant stimuli was 0.90 (± 0.12), while over non-relevant stimuli it was 0.87 (± 0.14). Regarding the response time, the overall mean response time was 676 (± 506) ms, where an average of 669 (± 480) ms was taken

by subjects to judge relevant images, while 683 (± 530) ms was taken to judge non-relevant ones.

3. RESULTS & DISCUSSIONS

This section presents the analysis of our EEG results. We firstly calculated the mean value for each channel obtained by averaging over the 20 participants' recorded signals and over the target window of time for two conditions: relevance and non-relevance. The Mean Significant Difference (MSD) was then calculated between the two conditions in each time range. This was calculated by subtracting the mean signal for the non-relevant condition from the mean signal for the relevant one. We finally obtained an F-value for each of the 62 electrodes by performing an ANOVA test.

Figure 2 shows topographical plots of the significance in difference (F-value) between the two conditions in the different time ranges. In these plots, the blue color indicates that there is no difference between relevance and non-relevance. Conversely, colors from yellow to red indicate a significant difference between two conditions, and the more the color is red the higher the difference. Moreover, the electrodes that carry significant differences ($p \leq 0.001$) are highlighted with a white dot.

0 - 180 ms This is the first window of time which goes from the initial presentation of the image to 180 ms afterwards. Looking at the first topographical plots on the left in Figure 2, it is clear that there are no apparent differences between the two conditions. This was also confirmed by an ANOVA test that did not reveal any significant effects. One possible reason for this lack of a difference between conditions is that this early time window is dominated by encoding of the stimulus features. Since both the relevant and non-relevant images are rich in detail it is possible that the coding of these stimulus features might mask any potential differences in processing relevance.

180 - 300 ms The window of time 180 - 300 ms shows an early process of implicit judgements of relevance. This process takes place in different areas of the scalp, as shown in the second topographical plot in Figure 2. The most relevant ones are located in the frontal area and in particular in F1 ($F(1, 19) = 22.063$, $MSD = 1.123\mu V$) and AF4 ($F(1, 19) =$, $MSD = 1.126\mu V$) locations. Since in both electrodes the MSD value is positive, the potential associated with relevance is greater than non-relevance.

300 - 500 ms This time-window shows (third topographical plot in Figure 2) that the positive value of MSD has moved to central areas. In particular, the most significant difference between the two conditions can be found in the electrodes located at C2 ($F(1, 19) = 29.031$, $MSD = 1.074\mu V$) and CP2 ($F(1, 19) = 27.655$, $MSD = 1.273\mu V$) positions on the scalp.

This window of time is also the same analysed by Koelstra et al. in their study. Looking at both our and their results, it is possible to state that they tally. In fact, the electrodes which carry the most significant differences between the two studies are very similar. In both experiments the area surrounding electrode CP2 is the one which carries the greatest difference between the two conditions[6].

500 - 800 ms The last window of time taken into account is the one which shows the most significant difference

---

[6]In our experiment the electrode which shows the greatest significant difference is C2. Since Koelstra et al. used a 32-channels EEG cap instead of a 64-channel one, they did not have this channel.
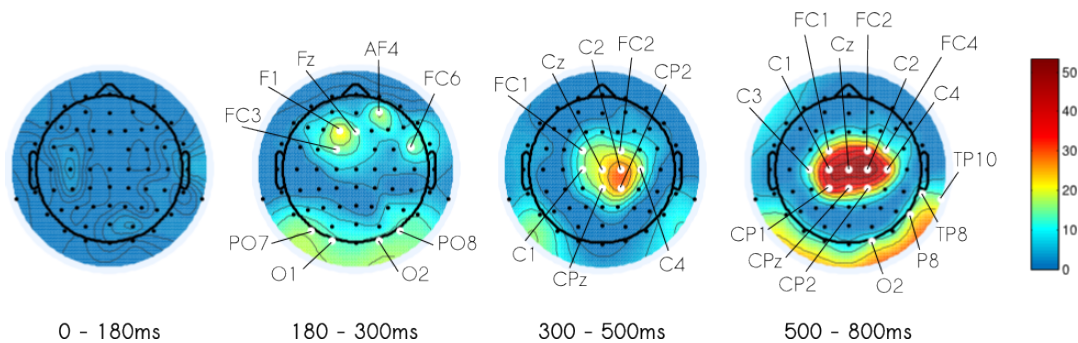
Figure 2: Topographical plots wihch show the significance of difference (F-value) between the two conditions in each window of time.

between relevance and non-relevance condition. As shown in the topographical plots in Figure 2, in the fourth plot the difference between the two conditions is more accentuated than in the third one. Nonetheless, it is possible to notice that the region of interest is still located in the centre of the scalp. In this last time range, the electrodes which carry the greatest difference between the two conditions are C2 ($F(1, 19) = 53.369$, $MSD = 1.781\mu V$), Cz ($F(1, 19) = 48.154$, $MSD = 1.800\mu V$) and C1 ($F(1, 19) = 43.257$, $MSD = 1.848\mu V$).

Discussion Timing and topography of our results suggest a timecourse of processing relevance that involves an early stage in the time window of 180 - 300 ms and a later stage that begins around 300 ms and builds through 800 ms. This first stage was evident in the larger potentials for relevant images found at frontal and posterior electrode sites. The timing of this difference suggests a role for late perceptual encoding of relevance and might explain the fMRI activity reported by Moshfeghi et al [10] in the inferior temporal cortex. Moreover, this early difference in amplitude might be related to establishing differences in the EEG frequency content used as features in recent research [6, 4]. From 300 to 800 ms after the stimulus was presented, the mean difference between EEG signals elicited by relevant and non-relevant images became greater in the electrodes centered around Cz at the central area of the scalp. Similar results have been obtained by Eugster et al. [4]. They found a peak difference at 757 ms between relevant and non-relevant stimuli at the CZ and C4 electrode sites, arguing that this potential was due to processing of memory. Despite the different nature of the two studies (Eugster et al. focused on textual terms, while we focused on images), it is possible to assert that this later stage might be common across the visual modalities of text and image, and related to cognitive encoding of visual information. Hence, regardless of context, our brain needs around 800 ms to judge the relevance of a visually presented stimulus.

4. CONCLUSION

In this paper we aimed to answer a fundamental question which is *when* the relevance assessment is happening in the brain. Focusing on the process of assessing relevance of images, we tried to identify the time frame in which users' brain EEG signals are the strongest and most distinguishable from the time that the images are shown to them. Our hypothesis was that there is a time frame in which the difference between EEG signals elicited by relevant and non-relevant im-

ages would be significant. To investigate that, we analysed how relevance assessment evolves over different windows of time within the first 800 millisecond of relevance judgment. Our findings revealed time frames as well as regions which carry the most significant difference between relevance and non-relevance conditions within this period of time. Despite the fact that the process of relevance assessment is very complex, this novel study provides an important step towards unravelling the nature of relevance and operationalising it for IR processes.

5. REFERENCES

[1] I. Arapakis, K. Athanasakos, and J. M. Jose. A comparison of general vs personalised affective models for the prediction of topical relevance. In *SIGIR*, pages 371–378, 2010.

[2] I. Arapakis, Y. Moshfeghi, H. Joho, R. Ren, D. Hannah, and J. M. Jose. Enriching user profiling with affective features for the improvement of a multimodal recommender system. In *CIVR*, 2009.

[3] P. Borlund. The concept of relevance in ir. *Journal of the American Society for information Science and Technology*, 54:913–925, 2003.

[4] M. J. Eugster, T. Ruotsalo, M. M. Spapé, I. Kosunen, O. Barral, N. Ravaja, G. Jacucci, and S. Kaski. Predicting term-relevance from brain signals. In *SIGIR'14*, pages 425–434. ACM, 2014.

[5] H. Jasper. The ten-twenty electrode system of the international federation. In *Electroencephalogr Clin Neurophysiol*, volume 10, pages 371–375, 1958.

[6] J. P. Kauppi, M. Kandemir, V. M. Saarinen, L. Hirvenkari, L. Parkkonen, A. Klami, R. Hari, and S. Kaski. Towards brain-activity-controlled information retrieval: Decoding image relevance from MEG signals. *NeuroImage, in press*, 2015.

[7] S. Koelstra, C. Mühl, and I. Patras. Eeg analysis for implicit tagging of video data. In *ACII*, pages 1–6, 2009.

[8] Y. Moshfeghi and J. M. Jose. An effective implicit relevance feedback technique using affective, physiological and behavioural features. In *SIGIR*, pages 133–142, 2013.

[9] Y. Moshfeghi and J. M. Jose. On cognition, emotion, and interaction aspects of search tasks with different search intentions. In *WWW*, pages 931–942, 2013.

[10] Y. Moshfeghi, L. R. Pinto, F. E. Pollick, and J. M. Jose. Understanding relevance: An fmri study. In *ECIR*, pages 14–25, 2013.

[11] J. Polich. Updating p300: An integrative theory of p3a and p3b. volume 118, 2007.

[12] C. M. S. Thorpe, D. Fize. Speed of processing in the human visual system. volume 381, pages 520–522, 1996.

[13] P. Ingwersen, and Kalervo Järvelin. *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer, 2005. xiv, 448 S. ISBN 1-4020-3850-X, 2006.

[14] R. W. White. Implicit feedback for interactive information retrieval, 2005.

[15] Y. Zhang, J. Zhang, M. Lease, and J. Gwizdka. Multidimensional relevance modeling via psychometrics and crowdsourcing. In *SIGIR'14*, pages 435–444. ACM, 2014.