Influence of Reading Speed on Pupil Size as a Measure of Perceived Relevance

Oswald Barral, Ilkka Kosunen, and Giulio Jacucci

Helsinki Institute for Information Technology HIIT,
Department of Computer Science,
University of Helsinki, Finland
{oswald.barral,ilkka.kosunen,giulio.jacucci}@helsinki.fi

Abstract. Depending on the task or the environment, we read texts at different speeds. Recently, a substantial amount of literature has risen in the field of predicting relevance of text documents through eye-derived metrics to improve personalization of information retrieval systems. Nevertheless, no academic work has yet addressed the possibility of such measures behaving differently when reading at different speeds. This study focuses on pupil size as a measure of perceived relevance, and analyses its dependence on reading speed. Our results are followed by a discussion around the need of taking into account reading speed when using eye-derived measures for implicit relevance feedback.

Keywords: pupillometry, perceived relevance, reading behavior

1 Background

1.1 Reading behavior

When using information retrieval systems to seek for information, the user adopts different reading behaviors, depending on several factors. The task to achieve, the environment or time pressure are some of them. The main component of reading behavior addressed in this study is reading speed.

Different reading speeds are usually associated to different reading tasks. Skimming can be helpful when there is a need to address a large amount of information and retain the most relevant parts of it. However, reading at fast rates involves less comprehension [1, 2]. If the goal of the reading process is to comprehensively understand the text, a normal reading speed will be adopted. On the other side, if there is a reduced available time and the amount of information is large, a faster reading speed will be more adequate, in order to focus just on the relevant parts of the text. The information seeker will therefore always adopt an optimal reading speed for every situation.

Having said that, as the amount of information available increases, the users tend to adopt faster reading rates, especially when seeking for information. Liu made an extensive survey addressing the changes of reading behavior in people ranging between 30 and 45 years old [3]. The participants in the study were asked to answer a set of questions regarding how their reading characteristics had changed over the past ten years. One of the outcomes of the survey was that 80% of the participants reported to have increased the time spent scanning and browsing, which are reading behaviors that imply high reading rates.

1.2 Pupil size as a measure of perceived relevance

Eye tracking technologies have been used in the field of information retrieval and personalized access over the past years as eye-derived metrics have proven to be useful to indicate users subjective perception of relevance [4–6]. In the goal of personalizing results, these implicit metrics are highly valuable as they provide an intrinsically individualized feedback.

Studies have shown a relationship between pupil size and user attention [7, 8]. It is well known that pupil size and cognitive load are highly correlated, different researches having approached the matter. Experiments have ranged from mathematical operations to search tasks [9]. Interestingly, Oliveira et al. [10] showed how pupil size could be of special interest when analyzing relevance in web search results. They studied both relevance of images and documents. Focusing on changes in pupil diameter, they were able to claim pupil size to be a carrier of interest-related information. Their experiments were on a very controlled level, letting the demonstration of similar conclusions in less controlled experiments as future research.

2 The present study

Given the above-mentioned reading behaviors, especially the increasing trend to read at fast reading rates, we consider highly relevant to study eye-derived implicit measures of relevance under the influence of different factors. In the present study, we focus on pupil size under the influence of reading speed. We designed an experiment in order to study whether reading speed has a direct impact on the ability of pupil size to indicate perceived relevance in documents.

2.1 Apparatus

The machine used to run the experiment was a 64bit processor Intel Core i73930k 3.20GHz 3.20GHz 16GB RAM, OS Windows 7 Enterprise SP1 with NVIDIA GEForce GTX580 GPU. The display device was a Dell 1703FPt 17" LCD Monitor at a 1280x1024 resolution. The experiment was developed using ePrime Software. The texts were displayed in an 85% window (I.e. 1088x870.4 pixels) with a 22-point font size. The subject was asked to sit 40-50 cm away from the screen approximately and to take a comfortable position. A Mirametrix S2 eye tracker operating at 60 Hz was situated under the screen and slightly moved to best fit to the subject eyes according to his natural and more comfortable position. The number of clock ticks since the booting of the operative system was used as reference for the synchronization between the Mirametrix S2 eye tracker and the ePrime software.

A first eye tracking calibration procedure was carried out at the beginning of the experiment and another one at the middle of the experiment. Each calibration procedure lasted for about 5 minutes, depending on the subject. The process was repeated up to five times to ensure optimal calibration (average error < 40 pixels). If the threshold was not reached within the first attempts, the average error margin was augmented in 10 pixels. The subject was rejected if after 5 additional attempts the average error was not fewer than 50 pixels. Two subjects out of ten were rejected due to calibration impossibility.

2.2 Participants and Procedure

Ten students (four undergraduate and six master's) participated in the experiment. Two of them were women. Eight participants reported to have advanced English reading level, and two reported a medium English reading level. None of them was a native English speaker. All of them had normal or corrected to normal vision. As already pointed out, two of the participants did not overcome the calibration procedure due to technical difficulties and their data was rejected. At the beginning of the experiment the participants were asked to sign a consent form and to indicate basic information about themselves. The data was saved anonymously in order to preserve participants privacy.

The participants were first conducted through a training session. The training consisted of two parts. The first one intended to get the users familiar with the three different speeds. As the reading speed is relative to the user's expertise or abilities, among other factors, instead of using an absolute word per minute rate for each of the speeds, an approach similar to the one by Dayson and Haselgrove was implemented [2]. The participants were first asked to read a document at a comfortable reading speed in order to be able to understand everything. They were instructed to reproduce that speed when they would be asked to read at a normal speed. They were then presented another text and asked to read it as twice as fast as the first text. If the time spent reading was higher than 70% of the previous one, they were presented a new text and asked to read faster, until they managed to spend less than 70% of the original time reading the text. They were then instructed to reproduce that speed every time they would be asked to read at a fast speed. An homologous procedure was used to train the skimming speed. Different texts were used in each of the phases in such a way that the familiarity with the text could not influence the reading speed. The participants were told explicitly to try to do their best to reproduce each of those speeds during the experiment. The second part of the training consisted of using the actual system until the participants explicitly recalled to have fully understood how they were supposed to interact with the system.

We decided to split the recording session into two parts as the participants of a pilot study reported to feel tired after having gone through the whole sequence of abstracts. Also, this allowed the recalibration of the eye-tracking device, avoiding the accumulation of systematic error [11]. Each of the two parts consisted of three topics. For each of the topics, the participants were asked to read in a given speed a sequence of abstracts. For each abstract, they were asked to assess as soon as possible using the left and right arrows whether the text was relevant to the topic (*binary-rating*). The participants were asked to keep reading until the end of the text at that given speed and to press space when done. Then, they were asked to grade, in a scale from 0 to 9, how relevant was the abstract to the topic (*scale-rating*) and how confident they felt about their answer (*confidence-rating*) .

For each of the six topics six abstracts were shown, half of them being relevant and the other half being non-relevant. The participants had to read two of the abstracts at a normal speed, two at a fast speed and two at a skimming speed. The order of the topics and the abstracts, as well as the reading speeds, was randomized. The topics were selected to be of common understanding and the participants were allowed to ask to the experimenter any question regarding the understanding of those. The topics were also selected in a way that their semantic meaning would not overlap. The relevant abstracts were selected not to be too obvious in the first lines. The non-relevant abstracts were selected to be completely non relevant to any of the topics.

3 Analysis and Results

For each abstract we took a time window of 10 seconds (i.e. five seconds before and five seconds after *binary-rating*) and averaged the values of the pupil each 500 milliseconds. We normalized the pupil data in each text by subtracting the mean of the pupil size over the entire text. Only the data of texts where the *binary-rating* and the *scale-rating* were congruent, and where *confidence-rating* was higher than 6 were taken into account (i.e. *valid-trials*). In these cases we observed a clear spike in the pupil size about 1 to 1.5 seconds after assessing the *binary-rating*. This was not surprising as the maximal pupil dilation has been reported between the event attracting attention and 1.3 seconds after [8].

In order to test for statistical significance between the spikes when assessing texts as relevant and when assessing texts as non-relevant we first took, for every abstract, the average value of the normalized pupil size in the time window of 0 to 1.3 seconds after the response time. Then, for the overall texts, as well as for each speed and each condition (the user answered relevant or answered non-relevant) we averaged the values within subjects. Finally, we performed Wilcoxon signed-rank text on the resulting paired samples.

In overall, pupil size was significantly higher when assessing texts as relevant (Mdn = 0.8) than when assessing texts as non-relevant ($Mdn = 0.66$), $z = -2.366$, $p < 0.05$, $r = -0.63$. When analyzing the texts read at normal speed, pupil size was also found to be significantly higher when assessing relevant ($Mdn = 0.93$) than when assessing non-relevant ($Mdn = 0.8$), $z = -2.197$, $p < 0.05$, $r = -0.59$. However, when analyzing the texts read at fast speed – relevant ($Mdn = 0.91$), non-relevant ($Mdn = 0.7$), $z = -1.690$, $r = -0.45$– and

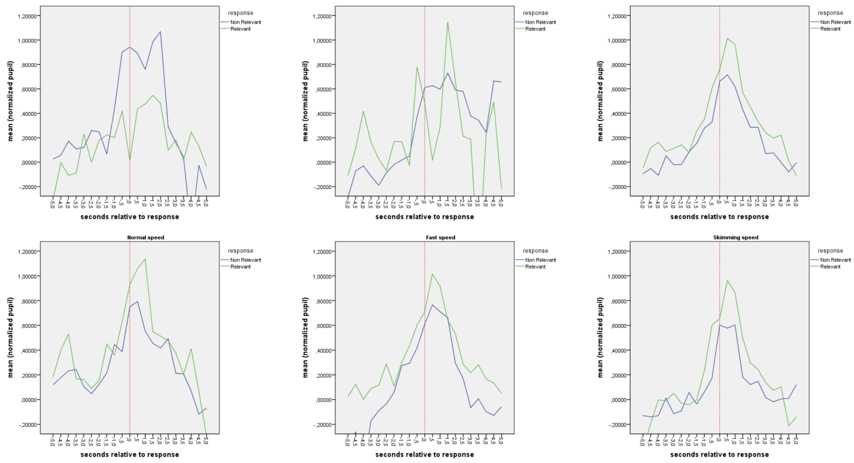

Fig. 1. Beginning from top-left: Pupillary response when *confidence-rating* is below 6; Pupillary response when *binary-rating* and *scale-rating* are not congruent; Pupillary response in the *valid-trials*. Beginning from bottom-left: Pupillary response for *valid-trials* read at normal speed, fast speed and skimming speed. The red line indicates the moment of *binary-rating*. The blue line represents the non-relevant and the green line represents the relevant texts. The plotted values are normalized within trials and averaged across participants.

skimming speed –relevant ($Mdn = 66$), non-relevant ($Mdn = 0.59$), $z = -0.676$, $r = -0.18$– no statistical significance was found.

4 Discussion

The results showed a clear relationship between the pupil dilation and the participants' subjective judgments. On top of that, the analysis of pupil size confirmed our hypothesis that its behavior would differ when reading documents at different speeds. When looking at the data without taking into account the speed in which the document was read, statistical analysis showed a significantly bigger response-related spike when the user perceived the document as relevant than when perceiving it as irrelevant. Nevertheless, when having a look at the same data but splitting the analysis by reading speed, the data showed statistical significance only when the user was reading at normal speed. That is, when the subject was given the instruction to read at faster rates than the comfortable normal reading speed, the response-related spike in the pupil size did not carry statistically relevant information regarding the judgement of the participant.

With this study we aim to raise a discussion around the fact that, when dealing with documents, different reading behaviors might have a direct impact on the reliability of our eye-derived measures. Thus, reading behaviors should be controlled and studied in order to have more accurate implicit feedback and, consequently, better personalization. As with pupil size, we believe that fixation-derived features used to infer relevance in documents will also behave differently when reading at different speeds and, therefore, need a closer look when the aim is to build realistic personalized search engines based on implicit feedback [12]. We encourage researchers to study the behavior of information seekers, and to apply such knowledge in the design of personalized information retrieval systems. We believe that a main element of the information seeking behavior that need to be understood is how the texts are addressed, studying which components have an influence on the application of implicit relevance measures. In the presented work we identified reading speed as one of these components affecting pupil size but, surely, in order to apply implicit metrics to enhance personalization in realistic systems, other measures and components of reading behavior need to be carefully studied.

References

1. Masson, M.E.: Cognitive processes in skimming stories. Journal of Experimental Psychology: Learning, Memory, and Cognition 8(5) 400 (1982)
2. Dyson, M., Haselgrove, M.: The effects of reading speed and reading patterns on the understanding of text read from screen. Journal of Research in Reading 23(2) 210–223 (2000)
3. Liu, Z.: Reading behavior in the digital environment: Changes in reading behavior over the past ten years. Journal of Documentation 61(6) 700–712 (2005)
4. Puolamäki, K., Salojärvi, J., Savia, E., Simola, J., Kaski, S.: Combining eye movements and collaborative filtering for proactive information retrieval. In: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, ACM 146–153 (2005)
5. Ajanki, A., Hardoon, D.R., Kaski, S., Puolamäki, K., Shawe-Taylor, J.: Can eyes reveal interest? implicit queries from gaze patterns. User Modeling and User-Adapted Interaction 19(4) 307–339 (2009)
6. Loboda, T.D., Brusilovsky, P., Brunstein, J.: Inferring word relevance from eye-movements of readers. In: Proceedings of the 16th international conference on Intelligent user interfaces, ACM 175–184 (2011)
7. Beatty, J., Lucero-Wagoner, B.: The pupillary system. In: Handbook of psychophysiology, New York, NY, US, Cambridge University Press 142–162 (2000)
8. Just, M.A., Carpenter, P.A.: The intensity dimension of thought: Pupillometric indices of sentence processing. Canadian Journal of Expérimental Psychology/Revue canadienne de psychologie experimentale 47(2) 310 (1993)
9. Hess, E.H., Polt, J.M.: Pupil size in relation to mental activity during simple problem-solving. Science 143(3611) 1190–1192 (1964)
10. Oliveira, F.T., Aula, A., Russell, D.M.: Discriminating the relevance of web search results with measures of pupil size. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM 2209–2212 (2009)
11. Hornof, A.J., Halverson, T.: Cleaning up systematic error in eye-tracking data by using required fixation locations. Behavior Research Methods, Instruments, & Computers 34(4) 592–604 (2002)
12. Barral, O.: Exploring the applicability of implicit relevance measures in varying reading speed for adaptive I.R. systems. Master's thesis, Universitat Politècnica de Catalunya/Helsingin Yliopisto, Spain/Finland (2013)