Date of acceptance Grade

Instructor

Exploring the applicability of implicit relevance measures in varying reading speed for adaptive I.R. systems

Oswald Barral

Helsinki September 23, 2013

M. Sc. Thesis

UNIVERSITY OF HELSINKI
Department of Computer Science

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta — Fakultet — Faculty	Laitos — Institution — Department
Faculty of Science	Department of Computer Science

Tekijä — Författare — Author

Oswald Barral

Työn nimi — Arbetets titel — Title

Exploring the applicability of implicit relevance measures in varying reading speed for adaptive I.R. systems

Oppiaine — Läroämne — Subject

Computer Science

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
M. Sc. Thesis	September 23, 2013	69 pages + 20 appendices

Tiivistelmä — Referat — Abstract

This thesis goes further in the study of implicit indicators used to infer interest in documents for information retrieval tasks. We study the behavior of two different categories of implicit indicators: fixation-derived features (number of fixations, average time of fixations, regression ratio, length of forward saccades), and physiology (pupil dilation, electrodermal activity). Based on the limited number of participants at our disposal we study how these measures react when addressing documents at three different reading rates. Most of the fixation-derived features are reported to differ significantly when reading at different speeds. Furthermore, the ability of pupil size and electrodermal activity to indicate perceived relevance is found intrinsically dependent on speed of reading. That is, when users read at comfortable reading speed, these measures are found to be able to correctly discriminate relevance judgments, but fail when increasing the addressed speed of reading. Therefore, the outcomes of this thesis strongly suggest to take into account reading speed when designing highly adaptive information retrieval systems.

ACM Computing Classification System (CCS):
H.3.3 [Information Search and Retrieval],
H.1.2 [User/Machine Systems],
J.4 [Social and behavioral Sciences]

Avainsanat — Nyckelord — Keywords

perceived relevance, reading speed, eye tracking, pupil size, electrodermal activity, implicit feedback

Säilytyspaikka — Förvaringsställe — Where deposited

Kumpula Science Library, serial number C-

Muita tietoja — övriga uppgifter — Additional information

Contents

Acknowledgements

I would like to express my gratitude to Professor Giulio Jacucci for supervising this thesis as well as giving me the opportunity to work in the best entourage and conditions one could ask for. I would also like to share the credit of my work with Ilkka Kosunen who gave me everyday mentorship and advice as well as helped me with data analysis related issues. I would never have been able to write this thesis without his guidance.

I would like to thank every member of the Ubiquitous Interaction group for the warm welcome they gave to me while doing this thesis. Everyday scientific dialogue enriched myself and consequently improved the outcomes of this thesis.

Last, but not least, I would like to thank the participants in my experiment, who have willingly shared their precious time to take part in the study.

1 Introduction

1.1 Background

Eyes behave in different ways according to the task to perform. When reading, eye movements, patterns and characteristics have well been studied by psychologists [RM76] [RP81] [RP89] [Ray98] [LF00]. Different purposes of reading require different reading behaviors. The way in which we read the newspaper while having breakfast differs drastically from the way we read a scientific article while seeking for related literature. The speed of reading is a main component of reading behavior, every reader having a comfortable reading rate in which comprehension is optimal [Mas82] [DH00]. Speed of reading is also influenced, sometimes, by external circumstances such as deadline pressure or fatigue. Additionally, while entering the digital era, the overall reading behavior changed, as a consequence of the increasing amount of documents and formats available. The largest amount of text is read nowadays from web-based formats on screens, in detriment of classic books or printed articles. Moreover, the most reported reading behavior while using screens is skimming, which is essentially reading at extremely fast rates [HTIH96].

Analysis of eye movements has been raising a big interest among computer scientists. In the field of human-computer interaction, the use of eye movements has widely been spreading in areas such as usability research or interaction in virtual environments [TJ00] [Duc02] [MR02] [PB05] [ZRZ08]. The understanding of user interest through the analysis of gaze data is one of the hot spots within the community [CLWB01] [AHK+09]. The aim is to be able to infer user interest or preferences based on the movement of the eyes, a specific application being found in information retrieval systems [MJL07] [BDBE12] [PSS+05]. The systems where such an application is useful are called *adaptive* or *personalized* information retrieval systems, and their aim is to filter the retrieved information based on the user preferences and characteristics. A wide range of methods has already been explored in this goal [ABD06] [KT03] [HKTR04]. The use of eye movements and physiology to predict perceived relevance is of special interest as the user is then giving a highly valuable feedback to the system, without even being aware of doing so [BDvE08].

1.2 Methodology

Given that eye movements behave in different manners when addressing texts at different speeds, eye-based measures studied until the moment to infer interest in information retrieval systems may behave in different ways as well. Few researches, even none, have been conducted in the field of inferring relevance from eye movements according to the reading speed. Furthermore, the ability of physiology to indicate perception of relevance has not been exhaustively explored yet. Consequently, the following research questions are addressed: 1) Are eye-based features used in the literature to infer relevance of documents behaving in different ways when the reading speed changes? If so, which are these ways? 2) Is physiology able to indicate user's perception of relevance and is such ability dependent on the speed of reading? 3) What is the applicability to modern adaptive information retrieval systems? An experiment will be addressed in order to find an answer to the above-mentioned questions.

Most of the studies done in the field of inferring relevance of documents from user's eyes have been analyzing results of highly controlled experiments. Researchers knew the word positions and even their semantic meaning and relevance [SKSK03] [SPK04] [AHK$^+$09] [LBB11]. This thesis' analysis is going to focus our analysis on a higher level. Neither the position of the words nor their semantic value is going to be controlled as, in our view, in the domain of highly dynamic and adaptive systems these factors are hardly controllable. Therefore the following fixation-derived features will be analyzed: number of fixations, average time of fixations, regression ratio and length of forward saccades. This thesis will try to replicate earlier results in perception of relevance inference, as well as study their behavior when addressing texts at different reading speeds. Additionally, an in-depth study of pupil size behavior together with electrodermal activity will be carried out, in order to study to what extend physiology is able to indicate perception of relevance in varying reading speeds. The direct implications for information retrieval and adaptive information retrieval systems will be discussed as well.

1.3 Structure of the thesis

This thesis will first review the background regarding patterns and speed of reading, eye tracking state-of-the-art methodologies, and implicit measures used in the literature to enrich information retrieval systems. Then, a detailed experiment procedure will be exposed and the analysis and results will be discussed. Below are explained the motivation of each section as well as a brief examination of their content:

Section 2 will give the reader an insight into reading behaviors and speeds. The main behavioral changes related to the overcrowding of digital documents will first be addressed. Next, an overview of the behavior of the user while seeking for information, namely while using information retrieval systems is going to be addressed. The section will pay special attention to skimming speed as is seen as one of the most characteristic behaviors when seeking for information in digital environments.

After briefly reviewing eye tracking history, section 3 will give an outline of the main computer science fields where eye tracking has been of interest. The section will then focus on the specific methodology and device used for recording the eye-gaze data in the thesis' experiment, as well as presents a detailed description of the fixation-recognition algorithm implemented.

Different techniques have been explored to enhance information retrieval systems. Section 4 will study in detail the wide and heterogeneous range of input signals used in such a goal. Special attention will be given to the specific field of eye-gaze analysis to infer perceived relevance. Additionally, a review of the state-of-the-art papers in this specific field will also be carried out.

The experimental design, methodology, procedure and extracted features will be exposed in Section 5.

Section 6 will first give an overview of the collected data. Then it will focus on the analysis of fixation-derived features when addressing texts perceived as relevant and non relevant, in varying reading speeds. In parallel, the section will deepen in the understanding of the relationship of such features with the speed of reading, independently of relevance prediction.

Section 7 will analyze the effect of the response event in the behavior of pupil diameter as well as of electrodermal activity. Whether those physiological signals are able to predict relevancy and in which way speed of reading influences such predictions will rigorously be discussed.

Section 8 will give a summary of what has been done, what has been accomplished, which had been the limitations of the present study and what kind of potential is seen for future work.

Few appendices are attached at the end of the document containing python and SPSS code snippets as well as relevant SPSS outputs. A sample comma-separated file used to compute the eye-derived features is included as well.

2 Reading patterns and speed

The fact of people not always reading at the same speed is well known. Reading speed depends on different factors like the task to achieve or the environment. The change in reading speed is a direct consequence of the change in reading behavior. Addressing different tasks might require different behaviors, and reading is not an exception. For instance reading a novel or reading an advertisement in the street will require different approaches from the reader.

The specific changes of reading behavior due to the switching to the digital era have raised a big interest in the academic community. In the early 80's the benefits and drawbacks of reading from paper or from a screen were addressed [MLTB82] [GAF+87].

This section is going to have a quick overview on that early research. We are going to stress as well modern reading and information-seeking behaviors, derived from the explosion of information availability and accessibility. To conclude, the section will focus on the skimming versus reading paradigm, directly related to modern reading behaviors.

2.1 Switching from paper to screen.

When the digitalization of the society was becoming an immediate reality, questions regarding how the digitalization of documents would affect the behavior of readers and the society as a whole were raising. A doubt regarding the extinction of printed documents was generalized.

One of the main fields of study was whether reading from screens could affect the comprehension of readers. That is, if reading from a screen was improving or decreasing the comprehension or performance of readers. The speed of reading or the reader comfort was other aspects sensitive to variation according to the reading format.

One of the first studies addressing those questions was the one led by Muter [MLTB82]. Their early studies did not show any relationship between comprehension and reading format. Nevertheless reading speed while reading from a screen proved to be about 28.5% slower than when reading form printed format. Those results were confirmed by later studies [Dil92] [MM91]. For instance Dillon found similar conclusions, stressing that reading from a screen was about 20% to 30% slower than from paper [Dil92].

Still in the 80's John D. Gould [GAF+87] and his team identified the low resolution of displaying technologies as the main cause explaining the reported slower reading speed from screens. The bad quality of those screens worsen the user comfort, hence deteriorated the reading speed.

It is interesting as well to recall the work done by Muter and Maurutto [MM91]. While they found reading from screens slower than from paper, they reported their results as being less pronounced than the earlier research done in that focus. The divergence could be attributed to the progressive adaptation of the society to digital environments as well as to the progress of displaying technologies. Those findings were corroborating Gould's explanations and gave a hint towards the possibility of having, in a near future, no handicap regarding the reading speed in digital environments. However, nowadays, it is still found of better comfort to read from paper than from the more contemporary digital displays such as e-readers [SZM13].

2.2 Reading behavior while seeking for information in digital environments

Despite the possible lack of comfort while reading from screens, it is undeniable that the amount of non-printed read text is growing. Nowadays more text is read from screen than from paper [Liu05] [NHJW06] [RNW+08]. It is noteworthy to recall the reading behavior associated to the digital environment.

The way we seek for information is changing radically, due to the explosion of the amount of information available, and the easiness of its accessibility. Other factors like its organization, as well as its presentation, are playing an important role in readers' change of behavior.

Scientists are reported to read now more articles and from a broader range of journals. The third part of the scientific articles is read from digital sources and half of the scientists use digital libraries [NHJW06]. A notorious amount of studies have addressed that matter, social sciences academics being reported to read less from digital sources than academics from science faculties. Scholars in the business schools and economics are the ones found to use more actively digital publications

[Nel01] [Smi03]. There are academic disciplines, such as history, where the use of electronic journals is significantly lower. The main reasons are the small amount of publications available in that format and the lack of awareness of the academics [NHJW06].

Research carried out by the CIBER [RNW+08] analyzed the logs of the London University library in order to study the behavior of scholars while seeking for information with the aim of improving and adapting digital libraries. Users were described as to spend a relatively short amount of time in each page, smaller than the minimum coherent time required fully reading and understanding an article. Even if they were reported to download the articles when they had the choice, there was no evidence of posterior reading. Due to the huge amount of articles available, the characteristic general behavior is to spend a big amount of time browsing the different documents, while spending little time in each paper. The large amount of abstracts and titles prompt the reader to browse from a document to another, in order to seek for relevant information. They do not spend long time in each article, but they jump rapidly between documents. These guidelines were reported as a general behavior, the diversity of information seeking patterns being influenced by factors such as the field of interest or the expertise [RNW+08].

Liu made an extensive survey addressing the changes of reading behavior in people from 30 to 45 years old [Liu05]. 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. Even if both the familiar and labor environments need to be taken into account, the overall results showed an increment in the time spent reading. 67% of the participants reported to read more than they did ten years before. Liu explained this increment as a result of the increase in the quantity and availability of information. Links between related documents induce the readers to switch from documents in a fast and dynamic way, exposing them to a large volume of information.

Liu described the *screen-based reading* behavior having the following characteristics: increasing time spent browsing, documents are only read once, the *keyword spotting* phenomena, non-linear reading, selective reading, lower in-depth reading and lower concentrated reading. An explanatory table with the metrics used to evaluate those characteristics can be found in the study [Liu05]. The *keyword spotting* phenomenon occurs when dealing with a big amount of text. User seeks for relevant keywords in order to spend the available time reading the text located near that region of text, skipping the non relevant information. The habit of just reading the first page of text, without even scrolling before jumping to another document was also one of those characteristics reported. Regarding in-depth and concentrated reading though, the users reported to prefer the printed format since it is easier to take annotations and highlight relevant parts of text. We would like to stress the fact that 80% of the participants reported to have increased the time spent scanning and browsing.

Those findings are the results of self-reported questionnaires from US citizens between 30 and 45 years old. Further studies should be addressed in order to extract more general conclusions including different age, cultural and geographical population segments.

2.3 Reading vs. skimming

As discussed in the above section, one of the main changes in the reading behavior is the increase of scanning and browsing through documents. The characteristic of scanning or browsing is the fact of reading different parts of text, at a high speed, in order to find relevant information to the given task, as fast as possible. Skimming, as defined by the Oxford dictionary is "*read (something) quickly so as to note only the important points.*", which is a main component of browsing or scanning for information. In this section we are going to analyze the characteristics of the different reading speeds, giving special attention to skimming.

Reading speed is not always a matter of voluntary actions. It has been proved that font size as well as line spacing has a direct influence on the reading speed [DK98]. It has also been found that shorter lines cause slower reading. Documents presented in columns are read slower than documents with a larger line length. In these cases reading speed decreases when there is an increasing need of scrolling [DH01] or because of the shorter length of saccades [RP89]. Still, it is said that after 100 characters per line, increasing the line length does not improve the reading speed [DH01].

Nevertheless, 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 keep the most interesting parts of it. On the other hand, it has been shown that fastest speeds involve less comprehension of text [Mas82] [DH00]. If the goal of the reading process is to be aware of every single part of the text, then a normal reading speed should be adopted. If there is a reduced available time and the amount of information is large, then a faster reading speed will help to focus just on the more relevant parts of the text.

The reason why skimming strategy is adopted when facing a huge amount of information was studied [Mas83]. Skimming technique requires jumping parts of text, some information not being processed ergo lost. A method where the time of fixations in words would decrease in order to have shorter fixations but in more words was suggested. Obviously no experiments can be designed controlling these parameters since they are completely dependent of human physiology. Additionally, a minimum fixation time is required to process the information. Whichever the reason is, the oculomotor system adopts skimming as a reading behavior when the reading speed is highly pushed forward [Mas83].

Early studies regarding reading speeds traits were carried out by Masson's team [Mas82] [Mas83]. Reading speed is highly dependent on the reader expertise and abilities but in overall, normal and skimming reading speeds were reported to be of an average rate of 225 words/minute and 375 words/minute accordingly [Mas82]. Later studies also found similar rates being 244 words/minute and 460 words/minute for normal and skimming speeds accordingly [DH00].

One of the main interests while analyzing different reading speeds was to find out whether the level of comprehension was affected when the speed changed [Mas82] [DH00] [DP09]. The type of information remembered has been proved not to be dependent on the reading speed. When asking about details or general questions regarding the comprehension of the document, no relationship has been found between the reading speed and the level of detail correctly answered. When the user reads at faster rates though, the amount of overall remembered information decreases, both regarding concrete and general details [Mas82] [DH00]. It has also been reported that when asking comprehension related questions, the response time increases along with reading speed [Mas82].

The finding of an optimal reading speed that would maximize comprehension, if there was such, was addressed by researchers such as Walczyk [WKMB99]. Their assumption was that under no time pressure, the subjects would read in a relaxed way, while under high time pressure, stress and lack of time would not allow an optimal task execution. They formulated the hypothesis that an optimal time pressure, or *mild time pressure* in which the user would perform the task in an optimal way should exist. Subsequently a reading speed should be associated to that optimal time pressure. The findings indicated a better performance under a certain time pressure, but no explicit *mild time pressure* was derived.

Other effects have been found to be suitable to affect comprehension [DH01]. One of those parameters being line length, 55 characters per line found to be optimal both for normal and skimming speeds.

Duggan and Payne focused their study on understanding the behavior of normal reading and skimming in long documents [DP09]. The analyzed texts were both unstructured documents and hierarchically structured texts. The research motivation was promoted by the increasing amount of structured text available in the World Wide Web (WWW). They found out that when reading really long texts, skimming behavior was reporting better comprehension of the macro-structure of the text than normal reading. That is, when the different parts of text were well structured and linked, the skimmers were reported as to have a better *structure map* of the text. Nevertheless, if the structure of the text was not clear, skimming did not show any significant improvement regarding normal reading.

3 Eye tracking

One main focus of this thesis is to analyze gaze data while reading under certain circumstances. In order to record such information, eye tracking techniques have been widely explored during the last century. This section will first give a quick review of eye tracking history and techniques. Then, a brief overview of the different applications of eye tracking technology in computer science will be presented. Finally, a detailed inspection of the specific eye tracking technique and device used for the data collection of this thesis as well as a description of the fixation-detection algorithm implemented will be carried out.

3.1 Eye tracking history

It is commonly said that eyes are the reflection of the mind or the soul. Scientists and social scientists have always been interested in eye's reactions, being interpreted as an outside gate for the mind. At the early beginning of the 20th century concrete measurement of the eyes was becoming a reality. An enormous variety of techniques have already been tested and used for different purposes. Nowadays eye tracking is getting available to the general public, as devices are getting cheaper and simpler.

The first publication introducing an accurate device and method to measure eye movements was reported in 1901 by Dodge and Cline, which is said to be the first objective eye tracker in modern times [Rob68]. Based on this first approach, other methods and techniques were implemented and tested along the 1920's and 1930's in an increasing and heterogeneous demonstration of technical creativity.

During the 1950's techniques using contact lens flourished, from the use of accelerometers [Tho65] to the use of mounted mirrors on the lens [RR50], the last one being the most commonly used. These proved to be very precise methods, since the tracking was made directly touching the eye itself. Nevertheless these techniques were also found to be tremendously annoying as the fact of wearing lenses and the setting up made them highly intrusive. Additionally head fixation was a usual requirement, as those systems were not compatible with free head movements, increasing their intrusiveness. One of the main restrictions that intrusive systems present is the difficulty to use such methodologies outside the research domain, outside the laboratories. Those systems can hardly reach large audiences which, in some cases, can be a strong inconvenient, especially when the goal is to commercialize the final product.

Non-intrusive techniques were later on explored, from cornea-retinal potential measurement to video images and computer vision based methods. Non-intrusive techniques are usually less accurate than intrusive methods, but they allow smoother and better overall human-system interaction. Nowadays the researcher can choose between a wide range of techniques, giving priority to usability or to accuracy. Non-intrusive techniques are reaching a more than sufficient grade of accuracy, being viable to use them in research fields such as, for instance, human-computer interaction. The pupil-corneal reflection method will be discussed in more details in section 3.3 as it is the one that will be used in the study of this thesis. Early reviews of eye tracking techniques can be found in Robinson [Rob68] or Young [You63] surveys. Additionally, a complete and more contemporary review can be found in [RS13].

3.2 Eye tracking in Computer Science

Even if the use of eye tracking begun within the social sciences community, when trying to better understand user behavior and physiology, it has also proven to be of great value in the field of computer science. Duchowski has made a review of the different areas that use eye gaze measurement for research purposes, from psychology to industrial engineering or neuroscience [Duc02]. Study of eye movements and eye tracking techniques has widely been explored within the computer science community too. Eye tracking has been used in a wide and heterogeneous range of fields in computer science, mostly in human-computer interaction related areas.

The first area that we want to describe is the one named by Duchowski as *eye-based interaction*. This area is extremely useful for making computers accessible to people with specific physical disabilities, most of the research being done within that focus. In this field, two main applications have specially been explored. The first one is the one known as eye tipping. The usual scenario involves a keyboard displayed on the screen while the user types looking at the letters for a fixed period of time (known as *dwell* time). Other techniques have been explored to trigger events (in this case selecting a letter) such as blinking, for instance. For the specific case of eye typing, Majaranta and Räihä have reviewed the state-of-the-art, discussing in detail the different approachable selection mechanisms [MR02].

The other main focus within *eye-based interaction* category is the use of eyes as a pointer on a screen. A large variety of studies have explored that possibility [ZMI99] [ZRZ08]. It has been shown that for the moment, mainly due to eyes' physiology, one can hardly use eyes as a cursor on a conventional operating system. Nevertheless it has been proved to be of a noteworthy usefulness when selecting areas or regions -instead of specific locations- in virtual reality systems, given that eyes are much faster than gestures, for example [TJ00]. Other problems might arise when dealing with eyes as an input signal. Lack of precision due to the devices, drift (explained in details in section 3.3) or to the *Midas Touch problem* are some examples [ADS05]. It is of notorious importance to take into account all these factors when designing a gaze-based interaction system.

Jacob first reported the *Midas Touch problem* in 1990. In his own words: "*At first, it is empowering simply to look at what you want and have it happen. Before long, though, it becomes like the Midas Touch. Everywhere you look, another command is activated; you cannot look anywhere without issuing a command.*". After exploring different techniques for triggering a selection event, *dwell* time was found to be the best solution [Jac90].

A good example of eye-based interaction system is the Eye-Bed [SBSL02]. The system was developed with the idea of attending the needs of people with tetraplegia who would permanently be laying on a bed. A big screen was placed in the ceiling and an eye tracker was placed on a lamp, mounted to the headboard. Simple eye gestures were used for the user to interact with the system such as very fast versus very slow blinks, staring at an object, closing the eyes (interpreted as the user sleeping), etc. For this kind of scenarios, eye-based interaction has proven a great usefulness, since the eyes are barely the only movements that the user can afford.

Jacob made an extensive analysis of whether *non-command* systems using as only input eye movements were feasible [Jac93]. *Non-command systems* were defined as the ones that would not require explicit commands or interaction from the user, in such a way that the user would have the impression of the system acting according to his needs. We have just described eye-movement based interaction systems. These systems work in a way that specific eye movements are programmed to trigger events, usually requiring specific user training. Contrarily, Jacob introduced the concept of taking advantage of natural eye behavior in order to interact with the system, no artificial or learned eye movement being required. The conclusion was that a fully eye-controlled *non-command* interface was hardly going to be a reality. The need of buttons for selection confirmation in menu commands being one of the main issues reported [Jac93].

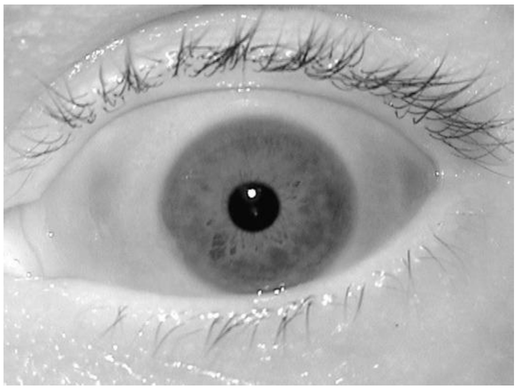
Interface usability analysis is another area of human-computer interaction where eye movements have proven a noteworthy utility. In Duchowski ‘s review, different papers are analyzed, considering, for instance, studies that compare mouse and eye gaze interaction, evaluation of Web pages’ structure and layout or click-down menus [Duc02]. Eye movements’ analysis concedes to interface designers the ability to understand in a different and even deeper way user interaction behavior, compared with classical explicit evaluation questionnaires. An increasing number of studies have been conducted in this area, pointing to the fact that eye gaze analysis might become a standard in the usability evaluation process: “*Eye-tracking studies in HCI are beginning to burgeon, and the technique seems set to become an established addition to the current battery of usability-testing methods employed by commercial and academic HCI researchers. This continued growth in the use of the method in HCI studies looks likely to continue as the technology becomes increasingly more affordable, less invasive, and easier to use. The future seems rich for eye tracking and HCI.*” [PB05]. A classified summary of all kind of metrics can be found in Pool and Ball state of the art review made out from the results of a broad range of eye tracking centered usability research studies [PB05].

Last but not least, we would like to pay special attention to the field of adaptive information retrieval systems. Computer scientists have found in the use of eye gaze data a way to get into their systems very valuable input regarding the behavior of users. Sometimes the user is not even aware of giving such input, which is usually seen as an advantage. Such input might even not be possible to be given in an explicit way, just because humans are not in control of the entirety of their behavior. Modern information retrieval systems try to be personalized, fitting to the user's personal needs and characteristics, a goal that eye tracking techniques and eye gaze analysis are helping to accomplish. Section 4.5 will focus on the state-of-the-art of eye tracking applications in the field of information retrieval relevance inference.

3.3 Device specifications and functioning

In the study carried out for the purposes of this thesis, we use the Mirametrix S2 eye tracker [Mir13] in order to track the user eye gaze data. The device implements the bright pupil tracking method, one of the two eye tracking methods generally used in human computer interaction research. Those methods are known as *corneal-reflection/pupil center* methods and can be observed in figures 1 and 2 [HJM+89] [MKAF00]. Both methods use an infrared light source and a camera sensitive to the infrared range. The infrared light is used because of its invisibility for the human eyes, in order to not annoy nor modify user normal behavior. The difference between the two methods lies in the location of the light source. In the dark pupil method, the infrared source is not located on the vision axis and the pupil appears in the camera as a black ellipse (see figure 1). In the bright pupil method, the light source is located in the same axis of the vision, the pupil being displayed as a bright ellipse (see figure 2).

(a) source: [HPI12]



(b) source: [BBH+09]

Figure 1: Pupil and first Purkinje image using dark pupil method

In order not to be exposed to a major amount of infrared noisy sources the bright pupil method has to be used in indoor spaces, having a better performance in dark environments. The camera captures the light reflected in the pupil, which appears in the image as a bright ellipse, easy to computationally detect by means of computer vision techniques since the iris appears as a dark area. This effect is comparable to the one that natural light does in cats' eyes during the night [HJM+89].

In order to track gaze direction and user point-of-regard, is not enough to track the pupil. The glint, or first Purkinje image, is also needed for such a goal [HJM+89] [CC73]. The first Purkinje image is a reflection point off the corneal surface that appears in the camera as a bright area. It has been proved that the relative position between this two reflection points is not affected by translation but it is modified under rotation [CC73]. This means that even if the user moves the head (translation) the relative position of the two points will not be affected. If the user changes the point-of-regard (rotation), this will directly modify their relative position. Therefore, the system can be aware of user's point-of-regard by tracking the relative position of glint and pupil [CC73]. The point-of-regard can also be approximately determined just by tracking the glint but then, the measurement will not be independent of head movements thus, fixation of the head will be required [PB05].

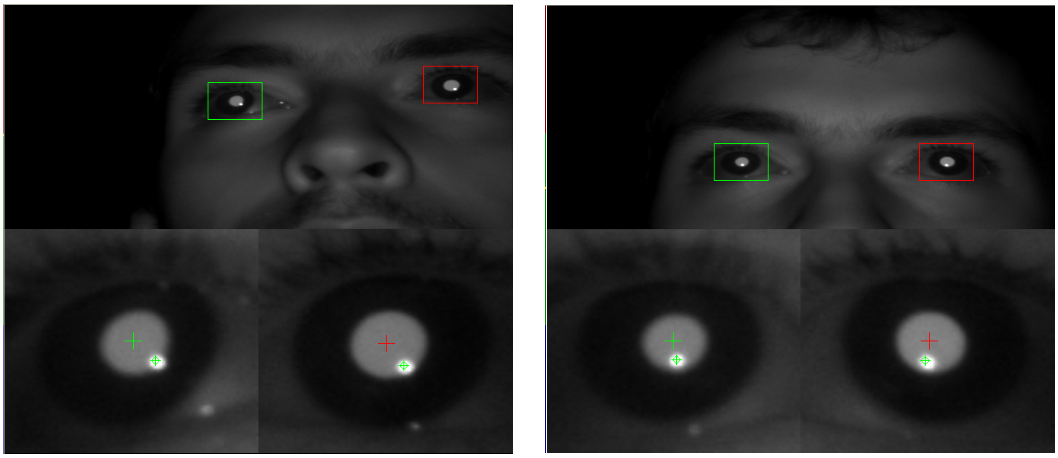
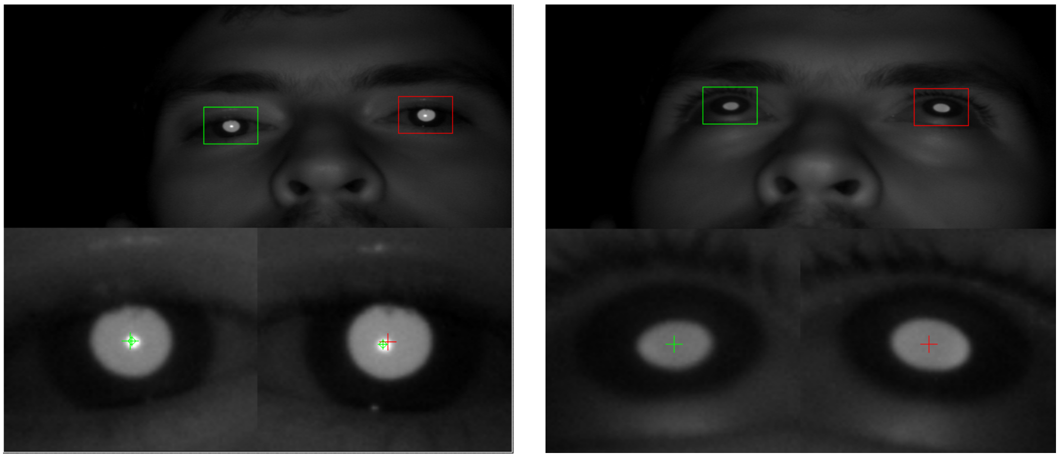


Figure 2: Pupil and first Purkinje image using bright pupil method

Nevertheless there are two critical eye positions where the gaze location cannot be tracked. The first one occurs when the user is looking directly to the infrared light source. In that case, the center of the pupil and the center of the glint overlap, converging in the same point, being impossible to extract a vector to determine the gaze position. This undesired effect could easily be avoided by placing the device under the screen so that the glint and pupil images are never overlapped unless the user point-of-regard is outside the display [HJM+89]. The second critical eye position occurs when the user is looking for instance to a corner of a very large screen, making a very pronounced eye rotation in such a way that the first Purkinje image is not displayed (there is no reflection). This effect can also be avoided by adjusting correctly the relative distance between the user and the screen, even though the accuracy might be affected. Sometimes, mainly because of devices' limitations, one cannot avoid this issue. Figure 3 shows the described critical eye positions.

Combining both bright and dark pupil methods Miramoto et al. got an extremely accurate tracking system capable even to track the gaze of multiple users [MKAF00].



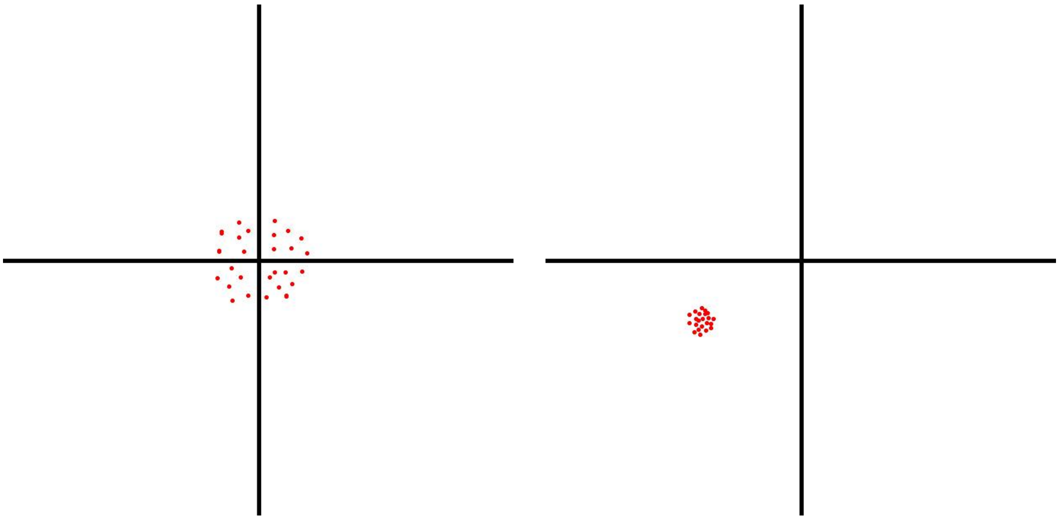
(a) The glint reflection and the pupil center overlap.

(b) No glint reflection because of excessive eye rotation.

Figure 3: Critical situations for eye gaze tracking

Eye tracking lack of precision and accuracy are determined by the so called *variable error* and *systematic error* respectively (see figure 4) [HH02]. The first one is highly dependent on the device technology and can be corrected using some specific eye tracking data post-processing. The second one is user dependent and is one of the main reasons why eye tracking devices require user-specific calibration. Furthermore, systematic error can appear after calibrating, as a result of user fatigue or elapsed time since the calibration process. This phenomenon is also known as drift. Mechanisms to on-line correct the systematic error and even automatically launch the calibration process have been reported [HH02].

The mirametrix system works in a way that first nine calibration points are recorded



(a) High variable error and low systematic error

(b) High systematic error and low variable error

Figure 4: Variable error influences precision and Systematic error influences accuracy

[Mir13]. The user is required to look at these coordinates, in such a manner that the system associates to each of these a specific relative position of both the glint and pupil centers. Once these nine points are successfully recorded (about 15 seconds), the system is able to track the point-of-regard in every position of the screen, by means of computer vision techniques and trigonometric calculations. The mirametrix S2 device specifies that there will never be a drift over 0.3 degrees. Furthermore the device takes less than 16ms to reacquire the eyes image in case of need. Following the official device' specifications, its accuracy is in the range of 0.5-1 degrees of visual angle, meaning that with the user staying at 50 cm from the device, the error in the screen is going to be in the range of 0.44 cm to 0.87cm approximately. The data is logged at 60 Hz and the user has a head moving freedom of 25cm x 11cm x 30cm (width, height, depth).

3.4 Fixation-recognition algorithm

In the process of analyzing raw gaze data in the goal of extracting gaze-based features, a first fixation-detection layer must be executed. The problem of extracting fixations is not trivial and, in fact, a full dissertation could be written addressing the topic. A large selection of methods has been already explored [SG00], but no standard algorithm or method has been adopted by the researchers.

The API of the Mirametrix device [Mir13] provides the researcher with fixation related information. As no further information is available regarding the algorithm used, the decision of implementing and using an own fixation detection algorithm was taken, in order to best fit the experimental setup. The algorithm that we implemented was similar to the one described by Buscher [BDvE08], with a few modifications, to adapt it to our scenario. Within Salvucci classification [SG00], our algorithmic approach would be in the section of *dispersion-based algorithms*. Those algorithms base their fixation recognition in the spatial proximity of gaze data points, assuming that close consecutive data points belong to the same fixation. Our algorithm proceeds as follows:

- We consider that a fixation is made of, at least, five consecutive data points. That is -at a rate of 60Hz- around 83 milliseconds. As we will discus in section 4.4 there is some divergence in the literature regarding which is the minimum fixation duration (ranging from 60 to 100ms [Jac93] [LF00]). We considered a middle point in a way to ignore meaningless fixations but, in the same time, avoiding being too restrictive. As we are not much interested in the short fixations, the minimum length threshold in this case is not really relevant.

- A fixation is initialized when five consecutive data points fit into a 30x30 pixels square. That is, the same window used as the one in Buscher algorithm [BDvE08]. According to our setup, 30 pixels are, at a distance of 0.45cm, approximately 1 degree of visual angle, which is reasonable, according to the analysis made by Gale [GJ84] and to the Mirametrix device specifications regarding accuracy (see section 3.3).

- Once the threshold of five points is reached, the pixel window is increased by 20 pixels and all the consecutive data points fitting in the window are considered belonging to the fixation. The fixation ends when five consecutive data points fall outside of the fixation window. That is, similarly as Bucher's algorithm, ours is tolerant to eye tracking noise and technology inaccuracy by excluding from the fixation a maximum of four consecutive outlier data points.

- The position of the resulting fixation is computed as the average of all the data points, and the duration as the timestamps subtraction of the last and the first data points forming the fixation. The implementation of the algorithm can be found in the appendices.

4 Implicit signals in information retrieval systems

It has been shown that relevant information can be extracted from the natural behavior of users interacting with computers, adding extra value to the system. Claypool et al. claimed "*Every user interaction with the system can contribute to an implicit rating*", speaking in terms of information retrieval systems [CLWB01]. A large amount of implicit information can be logged during a single user-system interaction deriving in a large amount of data. Each of the points has a low feedback value by itself, but between-points combination and analysis can potentially reveal highly valuable feedback [Nic98] [ABD06] [GPS99]. Furthermore, early studies claimed to have gained the same accuracy predicting relevance using implicit indicators than explicitly asking the user for feedback [MS94] [KMM+97]. In the following sections we are going to review the state-of-the-art of implicit signals and discuss their applicability in adaptive information retrieval systems.

4.1 Implicit and explicit signals

Explicit indicators have already been deeply explored and are nowadays used in a wide range of systems, with special incidence in recommendation software. Explicit feedback is also present in everyday life situations such as statistical polls, movie ratings, user satisfaction forms, etc. Explicit signals are those given by the user *explicitly*, meaning that he is aware of the action of giving feedback to the system. By contrast, the user is not aware of giving implicit feedback, and there lies the enormous potential of those signals. On the other hand, a computational processing and storing cost arises when dealing with implicit signals due to the large amount of data that can be recorded in a single user-system interaction [Nic98].

A wide range of studies have compared and even combined explicit and implicit feedback [Nic98] [KMM+97] [PSS+05] [BDBE12] [KT03]. In terms of information retrieval systems, the better accuracy of explicit in front of implicit feedback at the expense of the user experience has been discussed [KT03]. Classic explicit feedback systems ask the user for a specific ranking for each of the explored documents, in order to use that feedback to improve future searches. Nevertheless, users are most of the times too busy or just lazy to think about the ranking, even being aware that doing so would improve their user experience, as the system would better adapt to their specific needs [KMM+97] [Kel05]. This was one of the main arguments that launched Konstan et al. [KMM+97] into the exploration of implicit feedback, in order to make the interaction with the system more smooth and natural. If the goal is to have enriched information systems without the need of bothering the user, a solution has been found in asking experts for explicit feedback. Nevertheless, economic cost is then suitable to arise: "*Expert annotations require effort and have economic value, so the marketplace will undoubtedly assign them a price*" [OM96].

Nichols was the first to think about combining implicit and explicit feedback in information retrieval systems. That is, to use implicit indicators not only to have a larger amount of signals to infer user interest in documents or terms but to use them as a tool to give value to specific explicit ratings. The idea was that implicit indicators could tell about the real behavior of the user, identifying, for instance, mistakes while explicitly rating [Nic98]. Likewise, the combination of two implicit signals has also been pointed as a tool to better understand user behavior while performing a task. Claypool et al. described the following scenario: "*If a user does not read a document for very long, but they do bookmark it, the short time might suggest that they do not like the page, while the bookmark might suggest that they do. In this case, they probably bookmarked it for later reading and we do not yet know if they like it or not.*" [CLWB01].

The benefits of relating both kinds of relevance feedback in information retrieval systems are still nowadays a hot spot within the community. Additionally, encouraging positive results have been found by combining both kinds of signals, maximizing the impact in the system personalization of the two implicit and explicit feedbacks [KT03] [AHK+09] [FKM+05] [ZZ06].

4.2 *Traditional* implicit signals

It is difficult to classify implicit signals. As mentioned earlier, these encompass all the information that the system can get from the user in a non-obtrusive way. The collected raw information needs to be processed for it to be used as implicit feedback. Then, the system may use such information in order to understand the user behavior and consequently, user attention, interest or perception of relevance.

We name *traditional* implicit signals those that have been more intensely studied until the moment. These are the ones obtainable from the digital environment, namely from the system framework itself. The more exhaustively studied in the field of information retrieval, in the goal of revealing user interest in a specific document, have been: saving or printing a document, selecting or bookmarking a document and reading time. Other commonly analyzed indicators are, for instance, the amount of scrolling, the number of mouse clicks, the links visited or the user annotations [KT03]. A large amount of implicit indicators have been studied and a classification of part of them can be found in Oard and Kim study [OK01]. This classification has been enlarged later on by Kelly and Teevan [KT03]. In the aim of classifying interest indicators, Claypool et al. made another interesting classification [CLWB01]. Their arrangement was on a much higher level, taking into account both the explicit as well as the whole variety of implicit interest indicators.

Reading time has definitely been one of the most analyzed interest indicators in information retrieval systems. Still, implicit signals are highly context dependent meaning that no absolute correlation between reading time and interest has been derived, inasmuch as some studies have found very high correlation, and others did not found any. Morita and Shinoda found very high correspondence between reading time and explicit relevance feedback from the users of their information retrieval system [MS94]. They concluded that using 20 seconds as a reading time threshold for inferring interest, 30% of the relevant articles were retrievable with a 70% of precision. Another interesting finding was the independence between the length of the article and the elapsed time spent by the user to decide whether the article was relevant or not. That is, the decision was usually made within the first lines, length and reading time variables being found independent. Research made by Konstan et al. [KMM+97] and Claypool et al. [CLWB01] derived similar conclusions.

Kelly and Belkin [KB01] tried to find correlation between explicit feedback and three implicit measures: amount of scrolling, reading time and amount of user-system interaction. Contrarily to the studies mentioned above, none of the three indicators analyzed showed correlation with interest. They reported the divergence between the results to be due to the dissimilarity between their experiment conditions and the one used in Morita and Shinoda study [MS94]. These divergent results proved again the high context dependence of implicit indicators. Dependence on the type of read text has been reported by Kim, Oard and Romanik [KOR00] as reading time and other implicit signals like printing or saving, proved to behave in a different way when reading full scientific articles or when only addressing their abstracts.

Negative interest indicators are reported to be the signals that show non-interest [CLWB01]. These indicators are rarely used due to their complexity and difficulty to analyze. For instance, Claypool and his team discussed the fact of considering an absence of indicator as a negative indicator. Not doing something can be the consequence of a wide range of factors, which made the non-interest inference quite doubtful. Genuinely, inferring non-interest has been found to be much more complicated than inferring interest. When making assumptions, not only when dealing with *negative interest indicators* but with the entire domain of implicit signals, one needs to be aware of all the incident factors. The high context dependence makes the derivation of absolute conclusions extremely difficult. Additionally the user behaves in different ways when performing different tasks: the behavior while reading is completely different from the one while browsing (see section 2). A system using implicit indicators must always be aware, at any moment, of the system context.

Last, we would like to highlight an interest inference technique in information retrieval and recommender systems that can be considered half in the path of implicit and explicit methods. This technique is the collaborative filtering, and has already rigorously been studied [SM95] [HKTR04]. The basic principle of the mechanism is to group similar users based on their shared preferences and behaviors in order to use that information to infer interest. In order to classify users, the system uses the information regarding which are the documents visited by each user, commonly in combination with explicit feedback assessed by the users themselves or by experts in the domain. It is not uncommon that studies combine collaborative filtering with other implicit feedback sources and techniques, trying to make the system as more unobtrusive as possible. The motivation is to get a system able, based on inferred interest, to group users and get personalized recommendations autonomously, while minimizing the need of explicit feedback [KMM+97] [PSS+05].

4.3 *External* implicit signals

Claypool et al. [CLWB01] used the term *external interest indicators* in their classification of implicit interest signals to describe all the indicators coming from outside the system context. Those are mainly derived from the user's physiology. A wide battery of psychophysiological signals is suitable to reveal usefulness for our computer systems. Electrocardiogram (ECG), eye gaze direction, pupil diameter, electrodermal activity (EDA), electromyography (EMG), electroencephalogram (EEG), blood volume pulse (BVP), respiration rate and temperature are some examples [KLT+10] [SBL+09] [SPK04]. The main difference with the implicit indicators we have been describing up to this point is that, in order to retrieve physiological signals, an external specific device is required for each of the indicators. Additionally, it is usual that the devices need specific calibration or setup process, to the detriment of a smooth and natural user-system interaction. Nevertheless the calibration process, if needed, is usually done prior to the interaction with the system, so that the user does not need to modify the natural behavioral patterns during the actual interaction with the system, contrarily as when giving explicit feedback. The devices are a potential encumbrance for the user even if, nowadays, are getting less and less intrusive. Nevertheless in many cases where calibration is needed, once the setup process is done, the user is not aware of using an external device, to the point of forgetting about it. An example of completely non-intrusive system is the eye tracker that we will use in this study, which has already been described in details in section 3.3.

Psychophysiological signals have been studied in HCI mainly for inferring user emotions. A wide range of areas has been approached, from computer games [KLT$^+$10], to affective interfaces [LN02]. Nevertheless, there is not much research done in the scope of inferring interest through psychophysiological indicators. As interest has been described as a special kind of emotion [Sil08] and psychophysiological signals have already been successful to indicate emotions, evidence arises in the possibility of inferring interest through these implicit signals. Furthermore, the special case of eye depending external indicators have strongly been studied in the field of attention, interest and perceived relevance inference, especially in the information retrieval field [SKSK03] [OAR09]. A review of the state-of-the-art is discussed in section 4.5.

Electrodermal activity (EDA) has been one of the most widely used psychophysiological signals. Psychologists' interest for electrodermal activity began in 1888, in the laboratory of the French neurologist Jean Charcot [DSF07]. Electrodermal activity is a measure of the skin conductance, which can be measured in a relatively simple way. A small current is passed through the skin and by means of Ohm's law computations, the skin resistance, therefore the skin conductance, can be measured. Eccrine sweating glands are the ones responsible of the measured variance in skin conductance when analyzing EDA. Human body has two different kind of sweating glands, the apocrines (their main function being the cooling of the body) and the eccrines, which are activated mainly as a sign of arousal or specific phasic reactions to stimulus such as novelty, intensity, emotional content and significance. When measuring EDA we are interested in the eccrine glands, which can be found in the palm of the hands, the phalanges or the soles of the feet [DSF07]. When dealing with this kind of psychophysiological signal, environmental factors like the time in the day (morning, evening), humidity, temperature of the room as well as subject-dependent physiological factors need to be taken into account [DSF07]. Computer scientists found in this measure a high potential to be used as an *external* implicit signal for computer systems, some applications in computer games already been explored [BMA+01] [KLT+10].

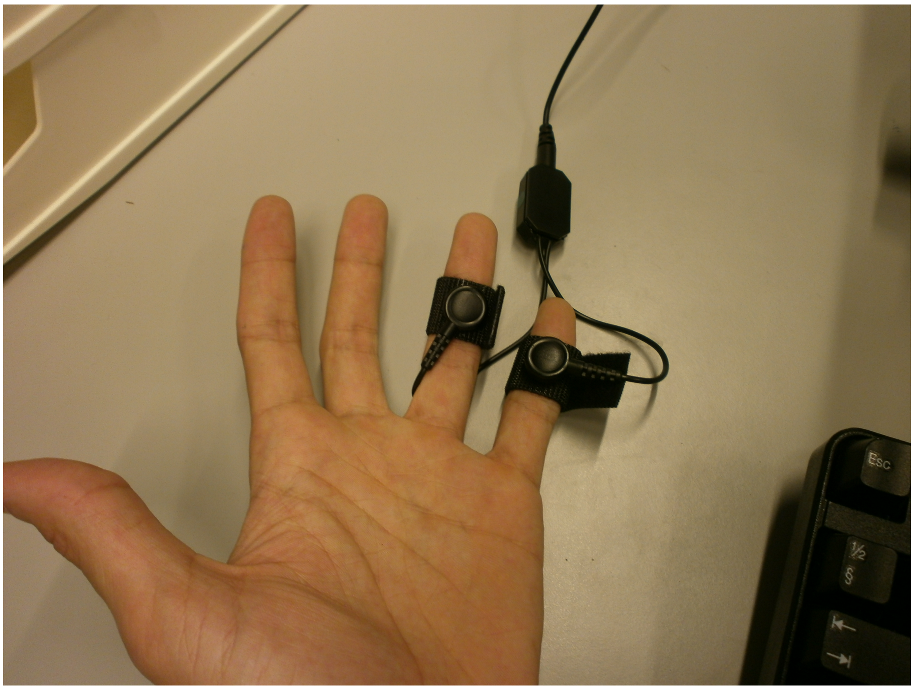


Figure 5: Recording of electrodermal activity (EDA)

4.4 Eye gaze properties while reading: Features to infer user interest

One of the purposes of this thesis is to study how eye movements are able to indicate reader's perception of relevance. That is, what the subject thinks is relevant. Whether the user' subjective judgment is correct, is outside the control of the researcher designing the study. Then, relevancy is a subjective perception and in order to make systems adapt to the user needs, perception of relevance is what we are interested in. For the purposes of this thesis, interest and perceived relevance are assumed to be really close, sometimes even used as synonyms. Additionally, in order to use eye gaze data to infer interest, some principles of eye movements' behavior while reading need to be understood.

The user eye gaze behavior while reading has well been studied [RM76] [RP81]

[DMR88] [RP89] [Ray98] [LF00]. We are going to focus our attention in the two basic eye movements: saccades and fixations. Other types of eye movements have been reported [Jac93] but are not relevant for the purposes of this survey. In order to change the point-of-regard, the eyes make fast, ballistic movements named saccades. Between saccades, the eye remains relatively stable: the fixations. It has been shown that while reading, fixations are ranging from 60ms to 500ms long, with an average of 250ms [LF00]. No consensus exists in the literature regarding the minimum time of fixations though. It is sometimes reported of 60 milliseconds, but 80ms or 100ms are even more commonly reported [Ray98]. Fixation average length is dependent on the specific task carried out by the reader, varying for instance between silent and oral reading [Ray98].

Saccades are approximately 1 to 15 characters long, depending on multiple characteristics of the reader [RP89]. Rayner and Pollastek concluded that, while reading English documents, 85% of the saccades between two consecutive fixations were progressive saccades (from left to right) [RP89]. Another type of saccade was defined, the regressive saccades or just regressions. These were defined as movements from the actual position to positions in already visited parts of the text (movements from right to left or to an upper line of the text) [RP89].

The more interesting eye-related studied features in the goal of inferring readers' interest are reviewed below. We are going to let word-level features outside the review as in this thesis we are not interested in the term-based inference of relevance. Instead, we will review the eye-gaze properties that have shown to indicate user preferences or interests, without the need of being aware of further word-level semantic or syntactic content. Nevertheless some of the outstanding surveys based on word-level features are going to be discussed in section 4.5.

- The relationship between **fixation duration** in a specific region of text and user interest has exhaustively been analyzed [RM76] [RP81] [DMR88] [LBB11]. Despite being proved to be a good indicator of user interest [LBB11], other factors are also the cause of such behavior. We know, for instance, that the ambiguity of a word within the context is suitable to evoke longer fixation duration [DMR88]. Other word-related factors such as the size or the frequency within the document can also influence the fixation length [RM76] [RP81]. In these situations, longer fixation demonstrates user attention, but not user interest. Therefore, we can assume fixation time as an indicator of user attention but the assumption of fixation time in a specific word being an absolute indicator of user interest can not easily be done.

- Recently Buscher et al. [BDBE12] studied the relationship between eye movements while reading and user interest. The research was focused on the application of eye-derived implicit feedback in the domain of web search scenarios. They found the **amount of text read line by line** as well as the **mean forward saccades length** to be good discriminants of user perceived relevance. These metrics are a measure of the amount of text read without skipping information, a quite “intuitive” interest indicator. When perceiving relevance in a document, the first measure was found to increase while the second was found to decrease.

- Few studies on the field reported **regression ratio** as being a nice indicator of interest [BPB$^+$06] [MJL07]. This metric measures the proportion of saccades that are regressions. Still and all Buscher [BDBE12] reported regressions being attention indicators. As already discussed in terms of fixation time, attention is also the result of other factors than interest such as difficulty in reading or understanding a word or a sentence.

- A measure named **thorough reading ratio** has also been studied when trying to infer users perception of relevance in documents [MJL07] [BDBE12]. The overall goal of the metric is to measure reading intensity, while being implemented in slightly different ways. Results have been quite positive when trying to discriminate relevant and non relevant documents. Moe et al. claimed "*About 2.5 as much time was spent reading thoroughly in relevant elements compared to irrelevant elements.*" [MJ07].

- Another interesting feature is the **pupil size**. It is not as commonly used as fixation-derived metrics, but studies found relationship between pupil size and user attention [BLW00] [JC93]. A delay of about 1.3s between the event that attracts user attention and the pupil maximal dilation has been reported [JC93]. It is well known that pupil size and cognitive load are highly correlated. Different kinds of experiments have addressed the matter, ranging from mathematical operations to searching tasks [HP64]. There might be a relationship between cognitive load and interest or perception of relevance, but no direct implication has yet been found. Furthermore, other parameters such as luminosity can also cause pupil dilation. Just and Carpenter [JC93] warned of the possibility of data misinterpretation derived from those external incident factors, stressing the high sensitivity of this measure.

- Measurement of **blink rate** have also been addressed in some studies, and similar issues have been found regarding the high sensitivity to external factors such as light exposition, fatigue or workload [GW03]

4.5 Eye tracking as implicit input for inferring perceived relevance of documents

One of the main motivations of this thesis is, while focusing on implicit feedback for tasks of information retrieval, to try to better understand the relationship between eye gaze and user perception of relevance in documents. The problem gets simplified if the assumption of perception of relevance being strictly related with user interest is made. The issues concerning the relationship between interest and attention have already been introduced. We need to be aware of the fact that attention might not always be an indicator of interest but a physiologic response to other factors [DMR88]. In this section we are going to focus on revising the state-of-the-art of eye tracking in the benefits of modern information retrieval systems.

An interesting selection of word-level metrics that can be used to retrieve user interest from raw eye gaze data, in the goal of extracting relevance of documents for personalized information retrieval tasks, can be found in the appendix of Salojärvi et al. study [SKSK03]. A classifier was built using these features, being able to classify relevant documents to a given topic as such, simply based on the user eye gaze data. The system was actually able to classify according with user perceived relevance, finding out that whenever there was a misclassified data point (an irrelevant document classified as relevant) it was due to the fact that the user itself was considering that document as relevant, when in fact it was not. The study recalled that it is easier to extract user interest from a paragraph or a text than from just a phrase or word.

Jarkko Sälojärvi et al. designed a controlled experiment were the relevancy of words was already known and used as base line [SPK04]. They reported a comparative study between different mathematical models in the goal of inferring user interest from eye movements. Hidden Markov models, support vector machines and linear discriminant analysis were confronted. Their analysis was on the word level, matching each fixation to the nearest word in order to extract eye-word related features (I.e. fixation duration on a word, number of fixations on a word, etc.). Accuracies higher than 75% predicting relevance from words while reading where reached using hidden Markov models.

Ajanki and his team built a system using the same word-level metrics [AHK+09] finding meaningful positive conclusions regarding the relationship between gaze patterns and user interest. They studied the possibility of generating queries using eye movements both as only input and combined with explicit relevance feedback. They reported their system to be able, to some extent, to anticipate user actions. That is, the system was capable to autonomously generate queries entirely based on eye gaze data.

Oliveira et al. [OAR09] showed how pupil size could be of special interest when analyzing relevance in web search results. They studied both relevancy 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. They also stressed the potential issue of pupil size being a delayed measure of interest, as the relevant changes were reported to occur about 400 to 500 milliseconds after the event. The stated fact improves the challenge of using pupil size as an on-line interest indicator, specially regarding fast adaptive dynamic systems.

Moe et al. [MJL07] designed a low controlled experiment, in a way to extract conclusions as realistic as possible. They let the participants choose one of the predefined tasks, and browse any available document they wish. The system interface designed was similar to a web browser, for the interaction to be as natural as possible. The behavior of three reading features described in section 4.4 (regressions, total viewing time and thorough reading) was analyzed. Some of the measures, such as thorough reading, proved to be useful when inferring interest. Their findings showed that low controlled experiments are able to present similar results than previous, more controlled experiments. Nevertheless, due to the limited size of their experiment, the results cannot be taken for granted.

Recently Loboda and his team made a research focused on sentence-terminal words [LBB11]. Their aim was to infer relevance of sentences only looking at sentence-terminal words. The data analysis showed that words ending relevant sentences attracted more fixations and for longer time than those ending non relevant sentences. However, relevance of terminal-words was not controlled, relevance control being on the sentence level. Additionally their analysis was based on a small amount of subjects, which made the reliability of their results low.

Recently Buscher et al. [BDBE12] highlighted the fact that few work has yet been done in the segment scope, meaning that the relevancy of documents is extracted from the eye gaze data of the full document length. The other approach would be to infer relevance form the segment scope, i.e. a paragraph or a section of the text. This thesis will analyze the reading behavior in abstracts of scientific papers.

5 Experiment

In chapter 2 we have seen that reading behavior is highly related with different parameters. The task to achieve, the environment, the support for the reading or the time pressure are some of the incident factors discussed. The main component of reading behavior that we have reviewed is the speed of reading. We have seen that the change in reading speed can either be altered voluntarily (e.g. as a way to achieve a given task optimally), or non voluntarily (e.g. when the format or the support of reading changes). Additionally, the fact of comprehension being affected with the change of reading speed has also been discussed. As comprehension is determinant when perceiving interest for documents, we understand that it is of main importance to study the behavior of eye-derived features to infer interest when reading at different speeds. The fact of the increasing “scanning-skimming” behavior adopted by the users when seeking for information in modern systems, mainly due to the ridiculously big amount of information available, is an extra motivation.

We designed an experiment in order to study the effects of varying reading speed in eye-derived measures as well as part of human physiology. The participants were asked to read abstracts from scientific papers in a normal, fast or skimming speed. They were asked to assess as soon as they could, whether the abstract was related or not to a given topic. They were also asked, after reading the whole abstract, to grade in a 0 to 9 scale both the relevance and the certainty of their answer. Eye movement, electrodermal activity as well as reading times, responses and other metrics were recorded.

5.1 Apparatus

The machine used to run the experiment was a 64bit processor Intel Core i7-3930k 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. The Mirametrix S2 eye tracker was situated under the screen and slightly moved to best fit to the subject eyes according to his natural and more comfortable position. For the recording of the electrodermal activity (EDA) the proComp infinity encoder was used and the electrodes were placed on the participant's non dominant hand. 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, the proComp infinity and the ePrime software.

Figure 6: Experiment setup

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.

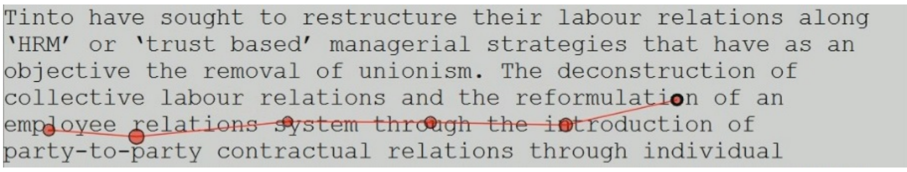
No specific calibration process was required for the recording of the electrodermal activity. As the electrodes were placed on the hand of the subject, the participants were asked, as only requirement, to clean carefully their hands with a non abrasive soap before the experiment in order to homogenize the measurements since the last cleaning of the hands might influence the skin conductance [DSF07].

5.2 Procedure

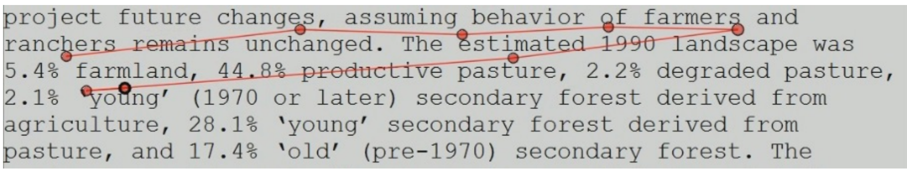
Ten students (four undergraduate and six Master's) participated in the experiment. Two of them were women. Eight participants reported themselves as having advanced English reading level, and two reported a medium English reading level. None of them was a native English speaker. Seven different mother tongues were reported for the overall of the ten subjects. 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 (see section 2), instead of using an absolute word per minute rate for each of the speeds, an approach similar to the one made by Dayson and Haselgrove was implemented [DH00]. The participants were first asked to read a document at their comfortable speed in order 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 new speed was more than 70% of the previous one, they were asked to read faster until they reached the goal. They were then instructed to reproduce that speed every time they would be asked to read at a “fast” speed. An homologue procedure was used to train the skimming speed, i.e. they were asked to read at a twice rate and asked to repeat the procedure until their reached at most a rate of 70% of the “fast” speed. Different texts were used in each of those 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. This approach not only takes into account the different subjects particularities when reading, adapting the reading speed accordingly. It also ensures that the reader will entirely read the whole amount of text. One of the biggest problems of having a fixed-time limit for each of the speeds is the fact that the participant could read the text at a lower reading rate and simply do not manage to finish the reading of the whole text in time.

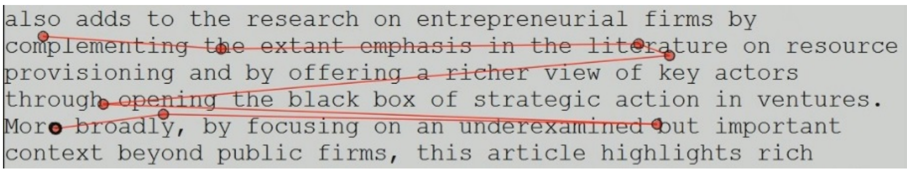
The second part of the training consisted of using the actual system until the participants explicitly recalled to have fully understood how they were supposed to interact with the system.



(a) Normal reading speed



(b) Fast reading speed



(c) Skim reading speed

Figure 7: Eye movements for a period of around 2 seconds. The dots’ size indicates fixation duration. Lines indicate saccades.

We decided to split the recording session into two parts as the participants of a pilot study reported to feel very tired after having gone through the whole sequence of abstracts. The splitting, besides lowing the load for the user, allowed the recalibration of the eye tracking device, ensuring the data to be more reliable as the system was prone to accumulate variable error (see section 3.3).

Each of the two parts consisted of three topics. For each of the topics, the user was asked to read in a given speed a sequence of abstracts (figure 8). For each abstract, the user was asked to assess as soon as possible using the left and right arrows whether the text was relevant according to the topic. The participant was asked to read until the end of the text on that given speed and press space at the end (figure 9). Then, when the text was fully read, the participants had to grade, in a scale from 0 to 9, how relevant was the abstract to the topic (figure 10), and then asked to rate the certainty of their answer (also in a scale from 0 to 9, figure 11).

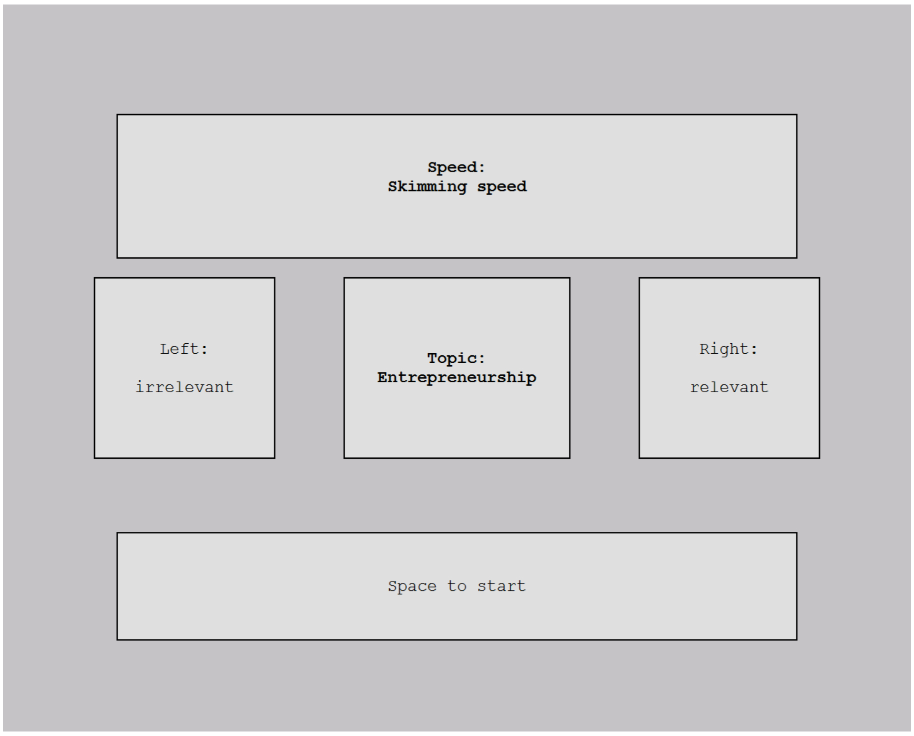


Figure 8: The participant is indicated the speed and topic

For each of the six topics six abstracts were shown half of them being relevant and the other half being non-relevant. The participant 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 could

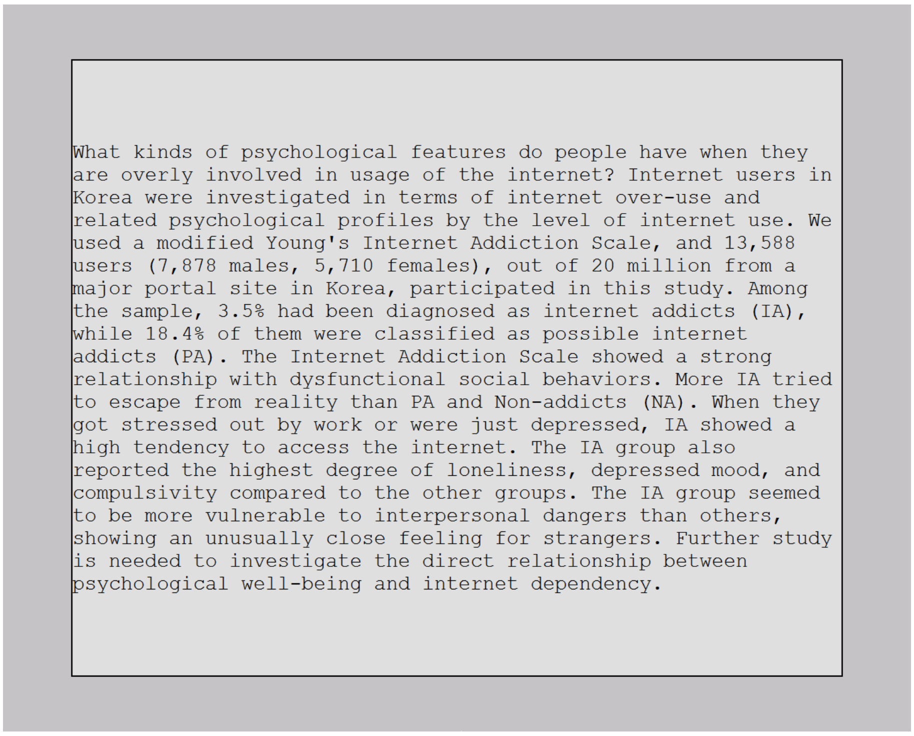


Figure 9: The abstract is displayed to the participant

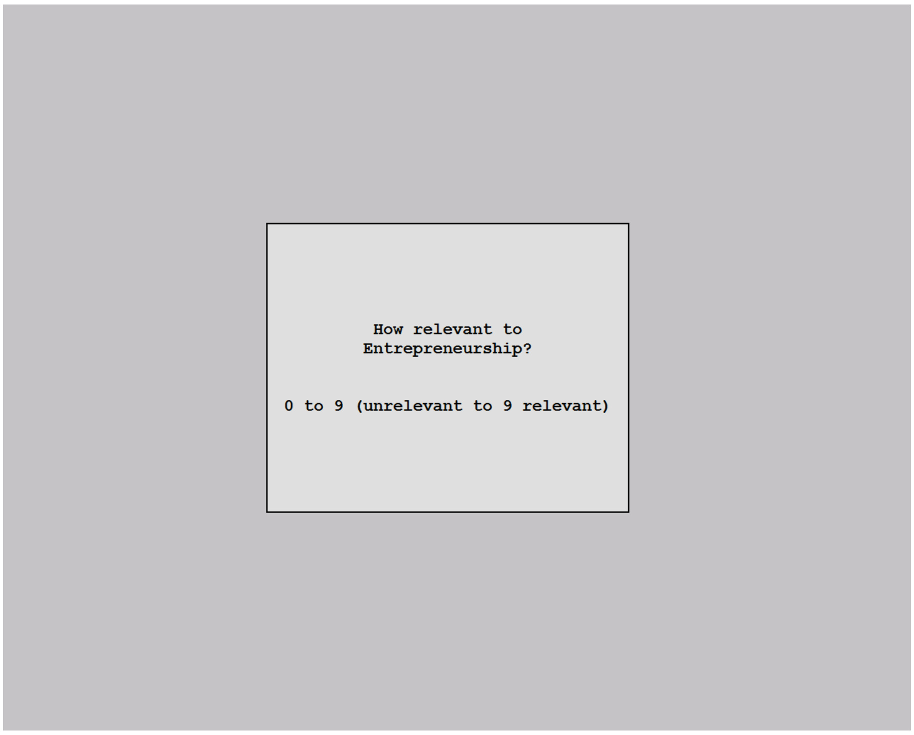


Figure 10: The participant is asked to grade the relevance.

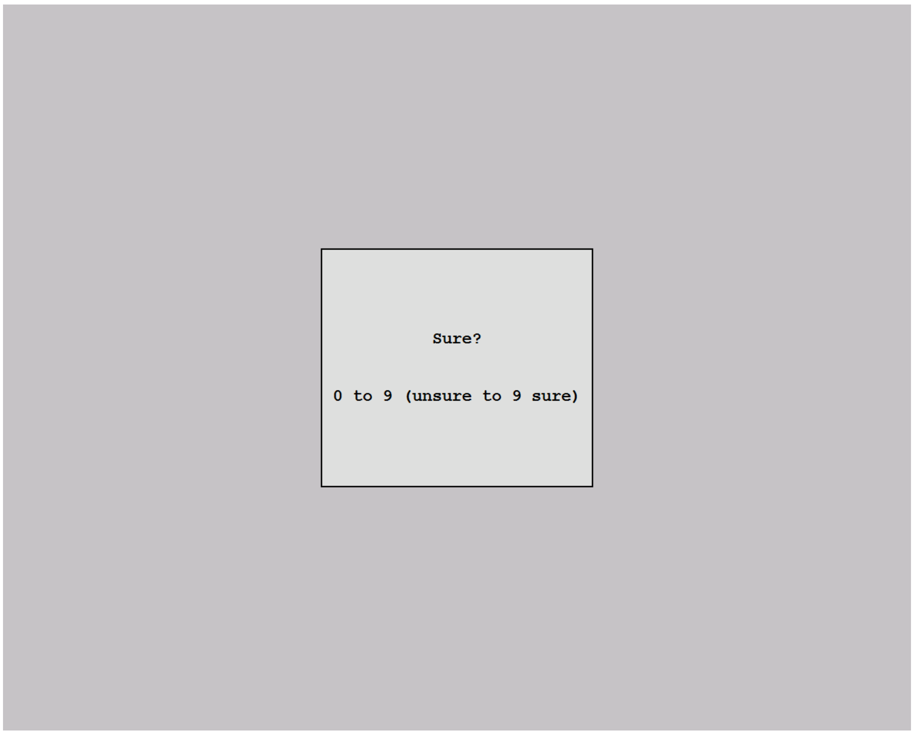


Figure 11: The participant is asked to assess the certainty of the grading.

ask the experimenter any question regarding the understanding of those. They 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 (even if that was not always possible given that a good abstract will always show in the first lines which topic it is about). The non relevant abstracts were selected to be completely non relevant to any of the topics. These requisites were specified to reach, to the extent possible, a one to one relationship between articles and topics and to avoid both user confusion and data misinterpretation.

5.3 Feature extraction

In order to extract eye-gaze derived features, first a preprocessing of the data was carried out. The logs of the ePrime software were transformed into coma-separated files, enclosing all the relevant environment related information (response-time, response answer, number of words in the abstract, etc.). A sample file can be found in the appendices. Eye tracking logs were then split according to the time frame when the stimulus were shown, in order just to take into account the eye movements during the reading of the abstracts.

For each of the read texts, a sequence of features was extracted from the raw data. The values of electrodermal activity were also included in the features vector. Table 1 shows the parameters used in the analysis.

Name	**Description**
subjectNmbr	The subject number
sessionNmbr	The session number within a subject
abstractNmbr	The abstract number inside the session
numWords	The number of words of the abstract
relevantClass	Whether the given abstract is relevant to the topic
speedClass	The instructed speed: normal, fast or skim
topic	The topic
response	The value of the binary explicit relevance response
elapsReadTime	The total amount of time spent reading
elapsRespTime	The time spent to assess whether the abstract was relevant or not
gradeAnsw	The assessed value of relevance when finishing reading
sureAnsw	The assessed certainty of gradeAnsw
coherent	Whether the subject binary response and numeric feedback are coherent
speed	The real speed (number of words / time reading)
numFix	The total number of fixations
avgTimeFix	The average fixation time
numSaccades	The number of saccades
numRegressions	The number of regressions
numFwdSaccades	The number of forward saccades
sumLengthFwdSaccade	The sum of the forward saccades length
avgPupilSizeR	The average pupil size of the right eye
avgPupilSizeL	The average pupil size of the left eye
EDA	The average value of Electrodermal activity

Table 1: Parameters for the analysis

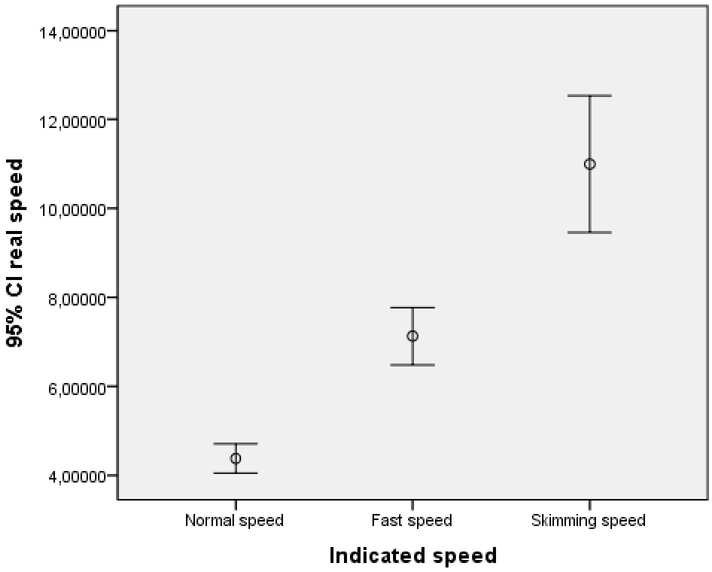
6 Analysis of fixation-derived features

Before any specific analysis, we first made an overall overview of the data to see whether the behavior of subjects was according to the specified when designing the experiment. As the dataset consisted of a limited amount of subjects, we were able to first look at their behavior one by one, using plots and basic descriptive statistics. We had to discard one of the subjects as we found out that, for some reason, he did not read the texts according to the indicated speeds. Figure 12 shows the behavior of a normal subject and the behavior of the bad subject. The data from the mentioned subject was discarded, that is, any of the following reported results or descriptive statistics have taken into account the data from the above mentioned user.

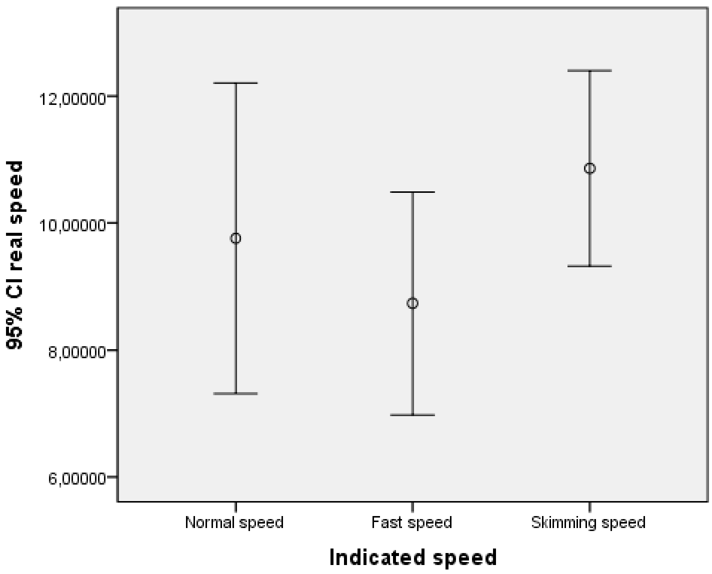
We found that for normal speed, the subjects where reading at a mean rate of 4.55 words per second (equivalent to 273 words per minute; s.d. 1.23 words per second). When they were asked to read at a fast speed, the mean rate was 6.18 words per second (equivalent to 371 words per minute; s.d. 1.79 words per second). When asked to read at skimming speed the mean rate was of 9.15 words per second (equivalent to 549 words per minute; s.d. 3.28 words per second). For normal and fast speeds, the rates are similar to the ones found by Masson [Mas82] or Dyson and Haselgrove [DH00]. The extremely fast reading speed, which we named skimming speed was read at rates comparable to the ones reported by Muter and Maurutto study [MM91].

While designing the experiment, we tried to choose topics of common understanding, relevant and not relevant articles being fairly obvious. As a result the mean of assessed certainty was 8.42 (s.d. 1.4), 8.38 (s.d. 1) and 8.11 (s.d. 1.4) for normal, fast and skimming speed accordingly, in a scale from 0 to 9. The participants assessed to be fully sure of their answer (pressed 9), thus to have fully understood the relationship between the text and the topic, in more than 60% of the read texts (75%, 64.3% and 63.1% for normal, fast and skimming speed). There was no appreciable difference was found in the understanding of the different topics, the assessed certainty mean ranged between 8.12 and 8.45 within the six different topics.

After this first overview of the data, we decided to take into account only the data of texts in which the subjects where congruent in their answers and the assessed certainty was over six. A feedback was considered congruent in one of the two following cases: a) the subject first assessed non relevant and at the end graded lower than 5; b) the subject first assessed relevant and at the end graded higher or



(a) Randomly choosen participant showing appropriate behavior



(b) Participant whose data was rejected

Figure 12: Error graph of the real speed according to the instructed speed

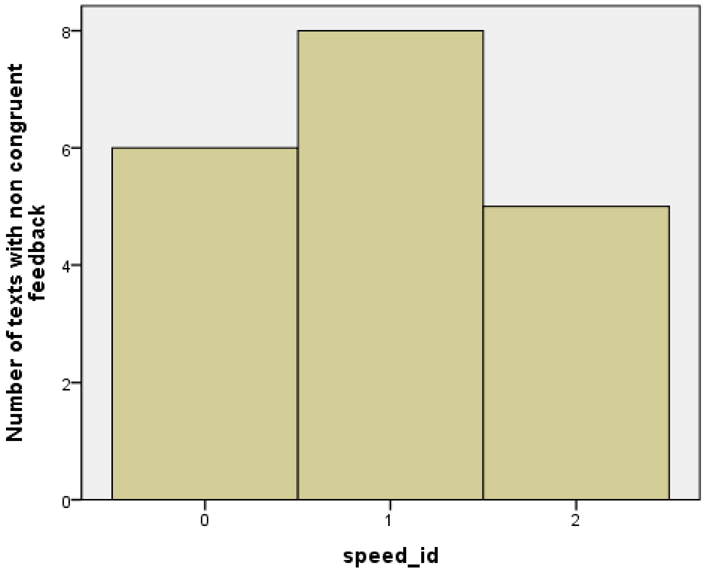
equal to 5. Our aim was to discard from the dataset all the data points referring to ambiguous relevance judgments, which could derive into data misinterpretation.

People were not found to be less congruent when reading at higher speed rates, as one could assume. For texts read at normal speed 7.1% of the given feedback was not congruent (that is 6 out of 84), 9.5% of the feedback given for texts read at fast speed was not congruent (that is 8 out of 84) and regarding texts read at skimming speed, 6% of the feedback was not congruent (5 out of 84). It is clear that the subjects did not change their mind because of speed, but just as a result of punctual misunderstanding or lack of concentration. Figure 13 shows that congruence is not related with speed neither with topic.

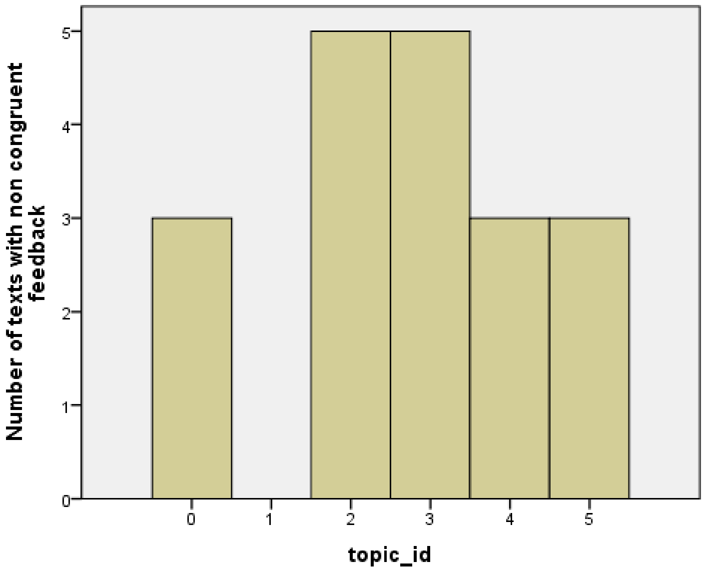
6.1 Fixation-derived features and perceived relevance

We were interested in seeing whether the few features extracted from fixations and saccades could be used to infer interest in our setup. We were also interested to see how the speed influenced their behavior. As already pointed out earlier in this thesis, the features analyzed were: number of fixations, average time fixation, regression ratio (number of regressions/number of saccades) and average forward saccades length.

In order to have a preliminary overall idea we first ran a linear mixed model procedure in SPSS, including the four gaze derived features as dependent variables and the binary response from the participant as independent variable. The exact SPSS syntax as well as some of the relevant output tables can be found in the appendices. We ran the same procedure splitting the output by instructed speeds in order to see whether there was some difference when the texts were read at different speeds. Looking at the overall behavior any of these features seemed to be strictly related to relevance feedback. When splitting by instructed speed though, there was an indication that when reading at skimming speed, the number of fixations was indicative of the relevance. We then averaged the number of fixations in the read texts assessed as relevant and the ones assessed as non-relevant, for each of the subjects. Then, we ran Wilcoxon signed-rank test on the paired-samples (relevant vs. non relevant). It turned out that when reading at skimming speed, the total number of fixations when reading texts assessed as relevant (Mdn = 59.71) was significantly higher than when reading texts assessed as non relevant (Mdn = 52.5), $z = -2.197$, $p < 0.05$, $r = -0.83$.



(a) Number of non congruent feedback given for articles read at Normal(0), Fast(1) and Skimming(2) speed



(b) Number of non-congruent feedback given for articles belonging to different topics (0 to 5)

Figure 13: Frequencies of non-congruent feedback according to reading speed and topic

To deepen in the understanding of the behavior of the number of fixations, we analyzed if reading time was also related with the relevance when skimming. A first approach by plotting the 95% confidence interval of reading time at the three different instructed speeds was carried out (see figures 14 to 16).

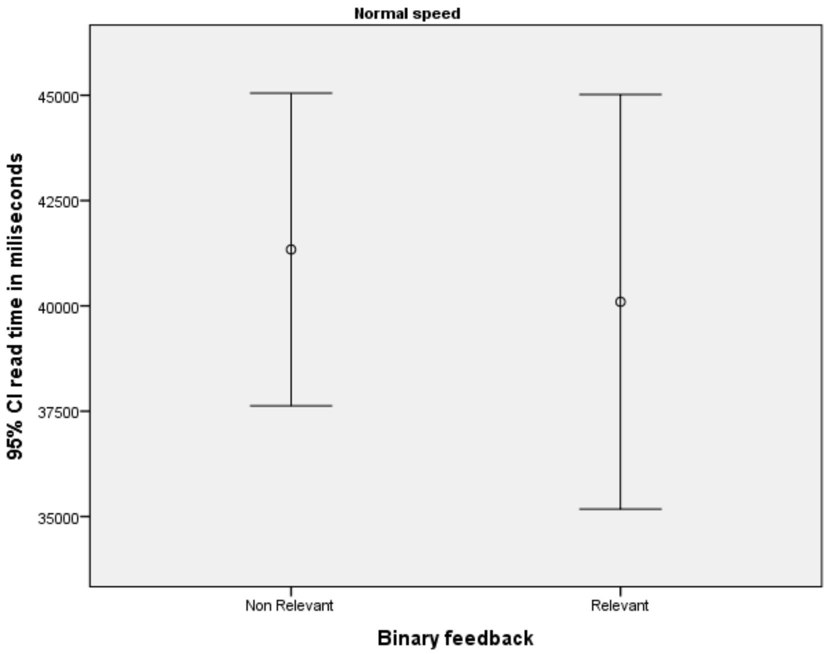


Figure 14: Time spent reading at normal speed relevant and non relevant documents.

Looking at the graphs we observe a different behavior when skimming, compared with the other two instructed speeds. We ran a similar Wilcoxon signed-rank test for the total reading time while skimming, finding out that the amount of time spent reading texts assessed as relevant (Mdn = 2238.5 ms) was significantly higher that the amount of time spent while reading texts assessed as non relevant (Mdn = 17657.8 ms), $z = -2.028$, $p < 0.05$, $r = -0.54$. On the other side, no statistical significance was found when running Wilcoxon signed-rank test on the time spent to decide whereas the text was relevant (Mdn = 6166 ms) and non relevant (Mdn = 8011.4 ms), $z = 0.169$, $r = 0.04$, while skimming. We can then assume that the increase in number of fixations was strictly related to the increase in the reading time after having assessed the binary feedback. Of course, the more time a subject spends reading a text, the more fixations on it.

We ran Wilcoxon signed-rank test on regression ratio, average time of fixations and average forward saccade length as well, both in the overall data set and splitting

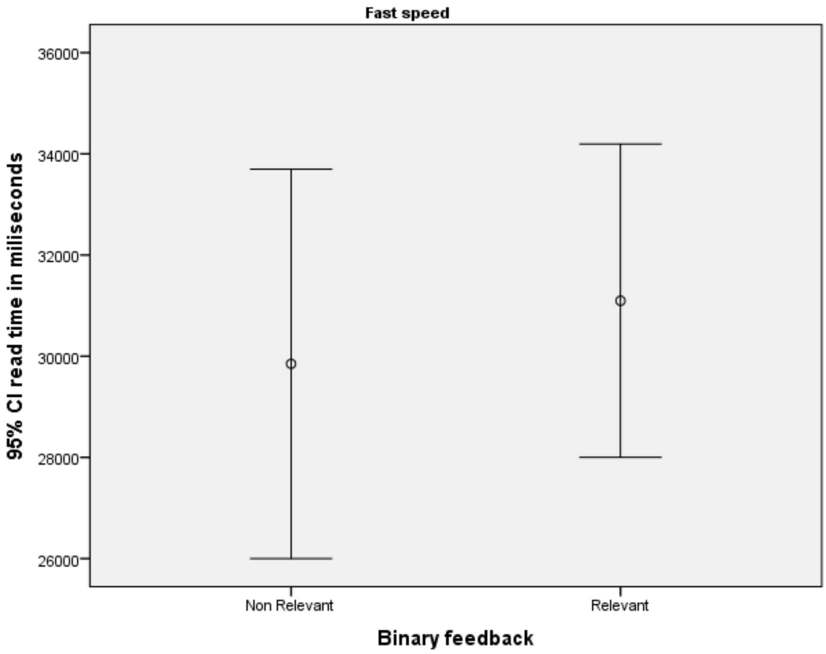


Figure 15: Time spent reading at fast speed relevant and non relevant documents.

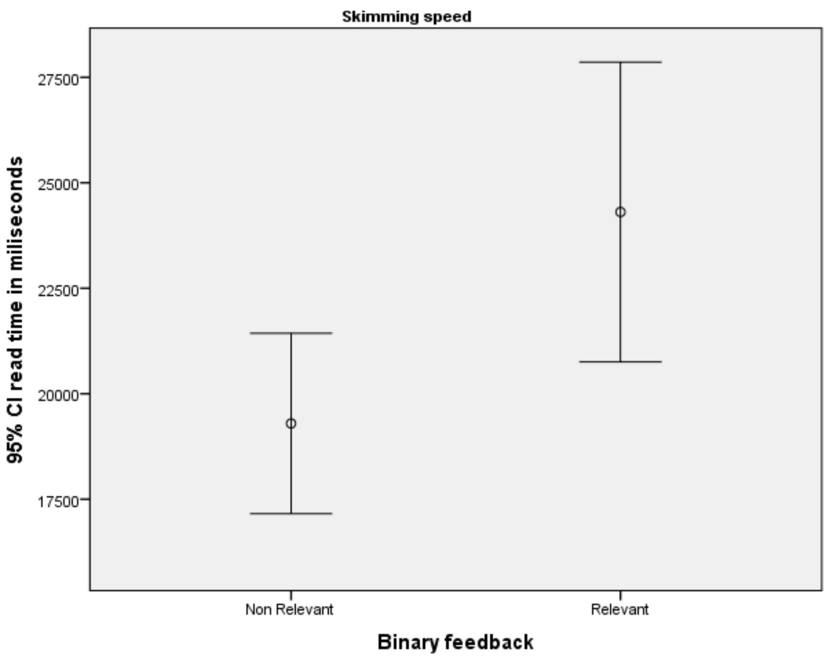


Figure 16: Time spent reading at skimming speed relevant and non-relevant documents.

the analysis by considering the addressed speed. As the mixed linear model had let to the assumption, the test did not show, in any case, significant difference between texts perceived as relevant and non-relevant. The specific test's output tables can be found in the appendices.

6.2 Fixation-derived features and reading speed

To study whether the eye-derived features varied significantly at different reading speeds, we ran a linear mixed models procedure in SPSS in order to have a quick understanding of the data. The procedure carried out was similar to the previous one but, in this case, we set the independent variable to be the instructed speed. The specific SPSS syntax of the procedure and relevant output tables can be found in the appendices.

The number of fixations appeared to be, of course, intrinsically related with the speed. This fact confirmed the previous assumption that increasing in reading time derived in an increasing number of fixations (see Figure 17). We ran a Friedman's ANOVA test on the number of fixations data and saw that, indeed, the number of fixations changed significantly when reading at different speeds ($\chi^2(2)$ = 12,286, p < 0.05). We ran Wilcoxon signed-rank test on the three pairs of number of fixations, applying the *Bonferroni correction*, i.e. we considered as statistically significant p values under 0.05 / number of tests, so 0.05/3 = 0.0167. The *post hoc* tests showed that the number of fixations when reading at normal speed (Mdn = 107.78) was significantly higher than when reading at fast speed (Mdn = 70.27), z = -2,366, p < 0.0167, r = -0,63. Number of fixations when reading at normal speed was also significantly higher than when reading at skimming speed (Mdn = 59.9), z = -2.366, p < 0.0167, r =-0.63. Nevertheless, no significance was found when comparing texts read at fast and skimming speed, z = -2.028, r = -0.54 p < 0.05.

When running Friedman's ANOVA test for average time of fixations a significant difference was reported for the different reading speeds ($\chi^2(2)$ = 10.571, p < 0.05). Wilcoxon test were used to follow up this finding. As previously, a *Bonferroni correction* was applied and so all effects were reported at a 0.0167 level of significance. Similarly as with the number of fixations, average fixations time appeared to be significantly higher when reading at normal speed (Mdn = 171.42 ms) than when reading at fast speed (Mdn = 154.24 ms), z = -2.366, p < 0.0167, r = -0.63. Average fixation time when reading at normal speed was also significantly higher than when reading at skimming speed (Mdn = 159.31 ms), z = -2.366, p < 0.0167, r =-0.63.

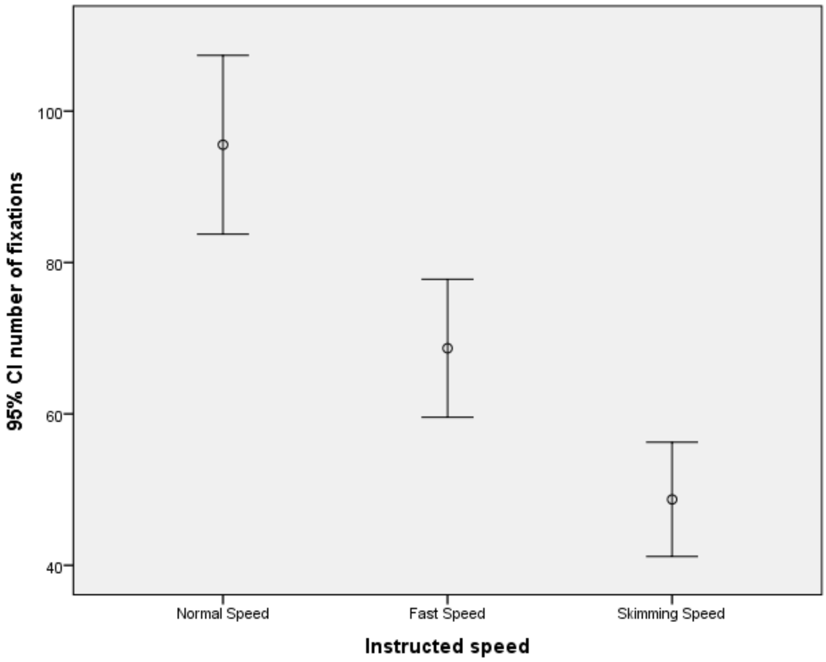


Figure 17: 95% Confidence Interval of number of fixations, by instructed speed

Nevertheless, average fixation time did not significantly changed while reading at fast speed or skimming speed, z = -1.014, r = -0.27. Figure 18 illustrates this behavior.

Friedman's ANOVA test did not report significant difference for regression ratio when reading at difference speeds ($\chi^2(2) = 0.857$, $p > 0.05$) even though the preliminary mixed model procedure indicated some possible relationship. Therefore there was no need to compute *post hoc* procedures. Average forward saccades length was expected to report statistical significant difference when comparing its values from the different addressed reading speeds. Friedman's ANOVA did report such a behavior ($\chi^2(2) = 8.857$, $p < 0.05$). When running Wilcoxon test we did find statistical significance between normal (Mdn = 208.7 pixels) and fast speed (Mdn = 231.32 pixels), z = -2.366, $p < 0.0167$, r = -0.63. Wilcoxon test did not report statistical significance when comparing fast (Mdn = 231.32 pixels) to skimming (Mdn = 277.75 pixels), z = -1.352, $p > 0.0167$, r = -0.36 or normal (Mdn = 208.7 pixels) to skimming (Mdn = 277.75 pixels) speeds, z = -2.028, $p > 0.0167$, r = -0.54. Nonetheless, figure 19 shows the tendency of forward saccade length to increase along with speed.

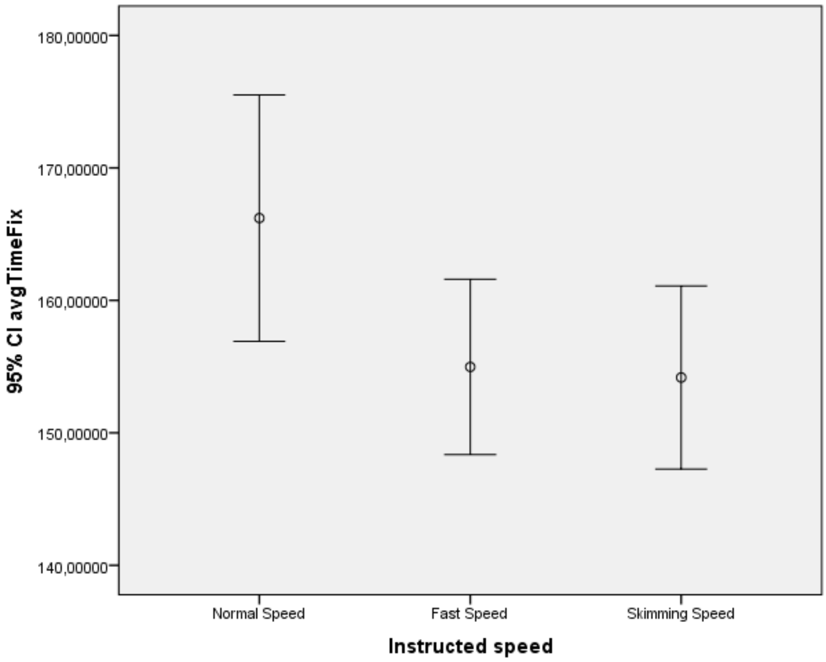


Figure 18: 95% Confidence Interval of average fixation duration, by instructed speed

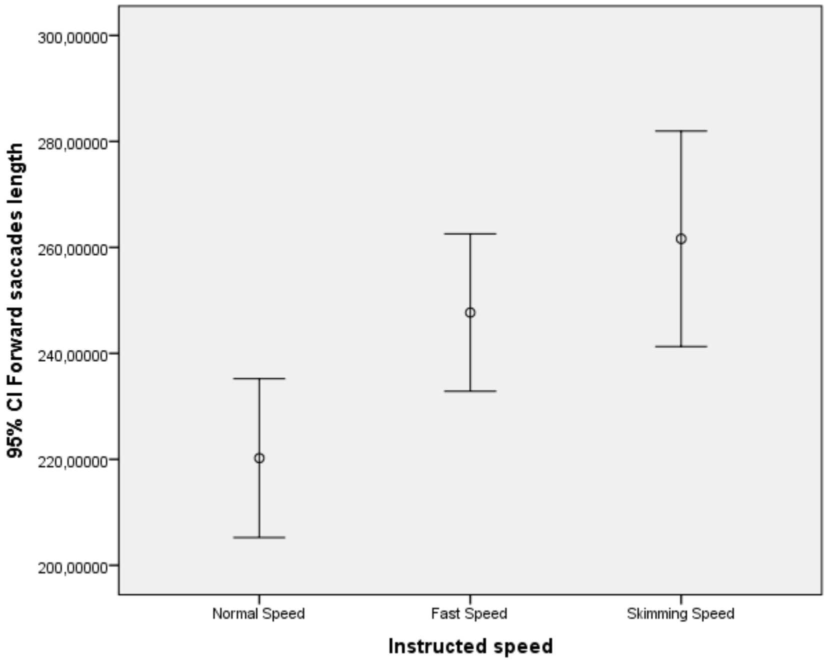


Figure 19: 95% Confidence Interval of average forward saccade length, by instructed speed

6.3 Discussion

When the user was explicitly asked to read at a given speed, his reading behavior was directly modified by the conditions of the experimental setup, therefore, part of his natural behavioral reading patterns changed. We believe that this fact mainly explains the lack of findings when trying to replicate earlier results on interest inference through features like regression ration, forward average saccades and average time of fixations. Additionally, the fact of studying eye movements in the segment scope, i.e. when reading few amount of text, was also determinant in the failure of earlier results replication. We believe that longer amount of text is required for these features to present significant difference when reading documents perceived to be relevant and non relevant.

The number of fixations showed, when skimming, statistical significance between texts perceived as relevant and non-relevant. Nevertheless, as already pointed out earlier, we believe this statistical indication to be a side effect as well of the specification of our experiment. We have already attributed the significant difference of number of fixations found when skimming to the time spent reading after assessing the binary feedback. As a consequence of the extremely high time pressure, it is likely that when a reader decided that a given text was non relevant, he would skim the remaining part of the text really fast as the decision of grading 0 in the scale of 0 to 9 would already have been taken. On the other hand, when a reader assessed the document as to be relevant, even having extremely high time pressure, he would have still tried to understand a bit of the remaining text in order to be sure to give a more precise feedback at the end. It looks like non-relevant documents were completely off topic but relevant documents were positively related to the topic in different measures.

We have stressed the need to infer interest without being aware of word positions, as well as the importance of analyzing the behavior of eye-derived features when addressing texts at different reading speeds. These two constraints have highly influenced the designing of the experiment, letting out of the analysis a wide range of eye-derived features. For instance, the whole range of word-level metrics, which proved to be of great value when inferring user interest [SPK04] [AHK+09], could not be analyzed. Thorough reading ratio (explained in section 4.4) was not suitable for this setup either, as it tries to measure the amount of text that is skimmed or read. Therefore, there was no point in analyzing this feature when controlling reading speed.

We based the second part of our fixation-derived features analysis in the understanding of their behavior when reading at different speeds, independently of their ability to predict relevance. Our aim was to observe whether the change in speed of reading would cause any effect on their behavior. Number of fixations showed a clear relationship with speed of reading. Statistical significant difference was reported when reading at normal and fast speed as well as when comparing reading at normal and skimming speed. Because of possible noise due to the difficulty found by few participants to read at two different fast speed conditions, we did not report statistically significant difference between skimming and fast speed. Taking a look at figure 17, a clear tendency is observed though.

Similar results were reported when analyzing the average fixation time, as the measure was only reported to be significantly higher when comparing texts read at normal speed with texts read at fast and skimming speed. In our understanding, in this case, the lack of significant difference when addressing texts at fast speed and skimming speed was intrinsically related to the physiology of the eye. Figure 18 shows a clear decrease in average fixation time when switching from normal to fast speed. We noticed though, that once reached a minimal fixation duration time, fixation duration does not decrease according to increase in reading speed. Our fixation recognition algorithm consider as a fixation consecutive gaze points nearly located for a minimum time of 83 ms (see section 3.4). In both fast and skimming speeds the means (155.6 milliseconds and 153.66 milliseconds accordingly) and medians (154.24 milliseconds and 159.31 milliseconds accordingly) were far away from the minimum fixation threshold of our algorithm. Therefore, we are able to claim that the lack of changes in these two speeds was not due to our fixation detection algorithm but due to the eye's physiology. We corroborate then the results of Masson [Mas83], which discussed the existence of such a minimal fixation duration time.

When analyzing the influence of reading speed in forward saccades length, the not congruent reported results required closer inspection. The analysis showed statistically significance when comparing the behavior of forward saccades length at normal and fast speeds but did not found any significance when comparing the measure at normal and skimming speeds. Additionally, when looking at figure 19 a clear difference on the means of both groups seemed to exist. We identified the source of this non-congruence as being the few amount of participants at our disposal for the study. Only data of seven subjects was taken into account. Looking at their behavior separately, it turned out that one of the users, for some reason, made shorter forward saccades when reading at skimming speed than in any other speed. When applying statistical methods to compare group means in such a limited amount of subjects, the fact of only having one participant behaving in a non conventional way can generate this kind of inconsistencies in the statistical tests results.

As when analyzing the behavior of regression ration for predicting perceived relevance, no significant change was reported when addressing texts at different speeds. As already exposed, we are positive that this measure is not suitable to be analyzed when controlling reading speed as we believe that is highly dependent on it. We also think that longer portions of read texts are required in order to analyze properly and consistently the measure.

7 Analysis of psychophysiological data

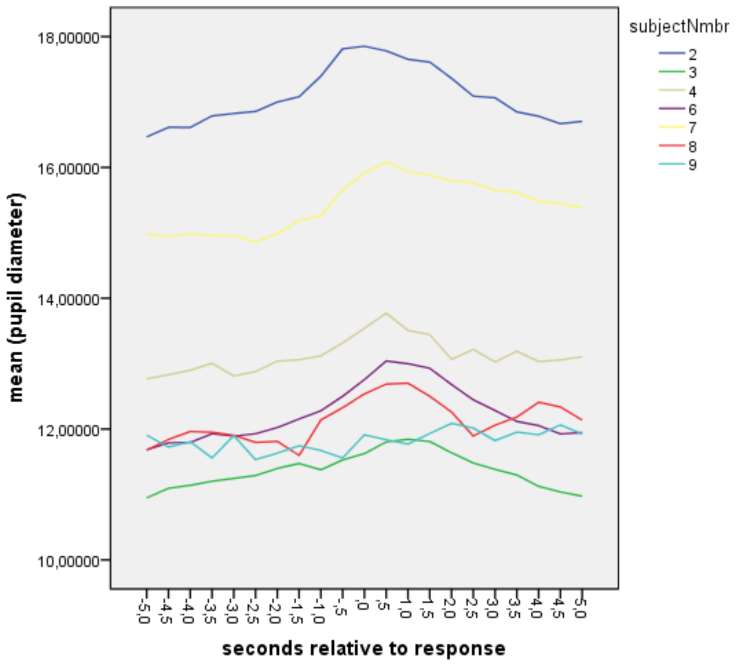
The experimental instructions were so that the participants were asked to assess the relevancy of the text as soon as they had an indication of it. When analyzing psychophysiology, we made the assumption that a specific cognitive process was happening around the response moment and that this effect could be observed in the psychophysiological signals studied. The hypothesis was that the observed effect in physiology would be different when assessing documents as relevant and non-relevant, being susceptible to change when reading texts at different speeds.

7.1 Pupil Size

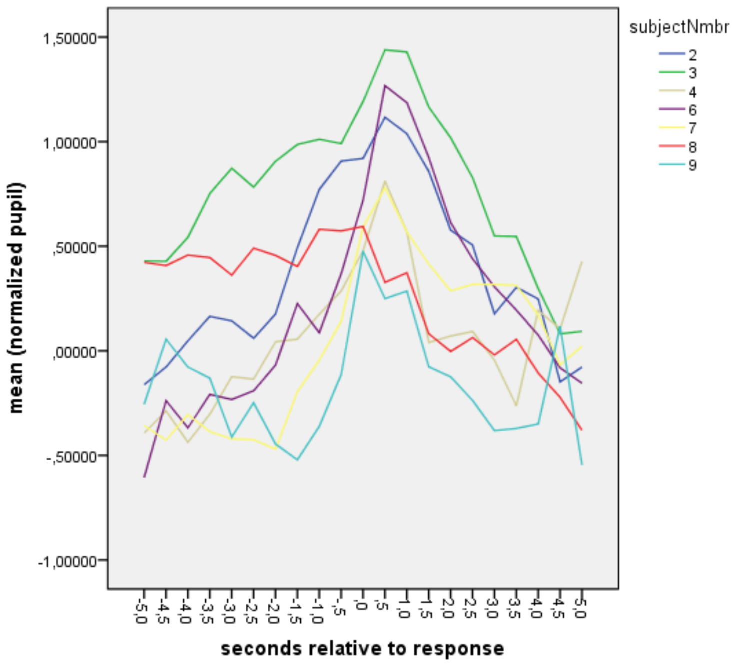
One of the main interests while analyzing the data collected during the experiment was to see whether changes in pupil size could indicate perceived relevance of documents, and if these differences in the pupil diameter would behave in different ways when reading at different speeds.

The device provided us with the information of each pupil separately. For each subject we used the data of the pupil that had less missing values. Even if calibration was restarted at the beginning of each session to preserve congruence within a subject's pupils, we used only one of the two pupils for each participant. The left pupil data was used for five out of seven subjects.

For each abstract we took a time window of 10 seconds and averaged the values of the pupil each 500 milliseconds. This is, we took five seconds before and five seconds after the subject assessed the binary feedback. We normalized the pupil data in each text by subtracting the pupil size mean of the whole document, in order to get rid



(a) Pupil diameter without normalization



(b) Normalized pupil values

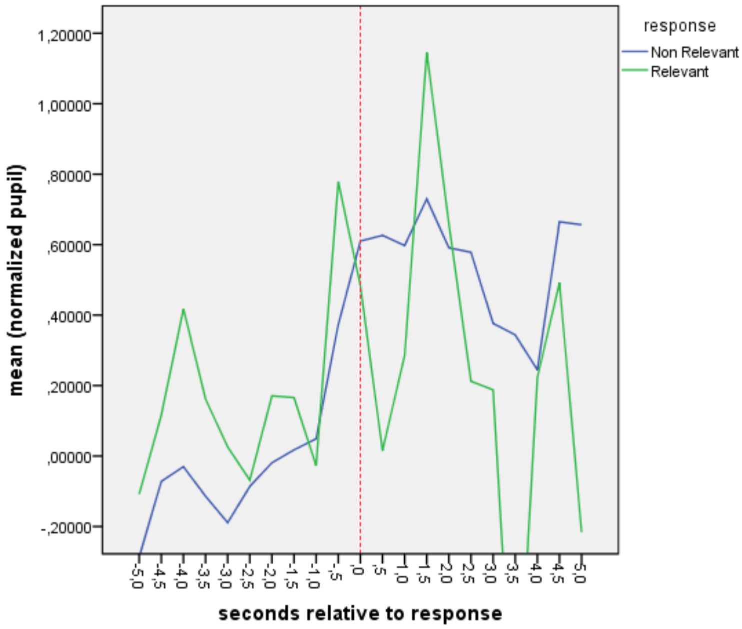
Figure 20: Pupil behavior for each of the seven subjects, in a time window of 10 seconds. Binary feedback was assessed at time 0

of subject particularities and noise (see figure 20). In the same way as with the rest of analysis, only the data of texts where the feedback was congruent and the addressed certainty was higher than 6 were taken into account. Figure 21 shows the pupil behavior when the answer was not congruent and when the subject was not sure about the answer, respectively.

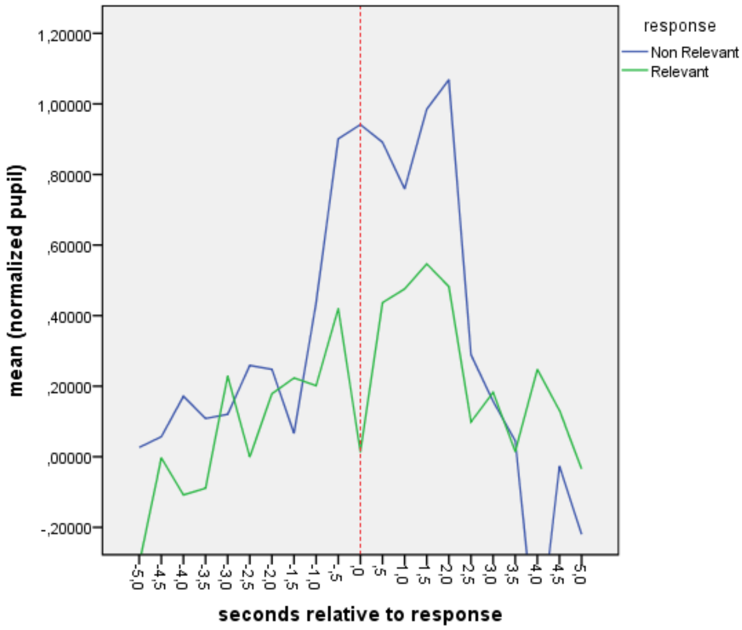
When looking at the cases where the user assessed a highly confident and congruent feedback, the behavior of the pupil size was dramatically different (see figure 22). We observed a clear spike when the user decided the relevancy of the text with the maximal dilation located around 1 to 1.5 seconds later. This was exactly what we expected to find since the maximal pupil dilation when an event attracting attention occurs has been reported to be between the moment of the event and 1.3 seconds later [JC93]. Looking at the figures 23 to 25 we can observe that this kind of behavior is repeated in the three different speeds.

We formulated the hypothesis that this event-related spike would be different when assessing relevant and non-relevant documents, and that its behavior would be affected by the addressed reading speed. The graphs let to the intuition of an existing relationship between the spike and the response type. In order to seek for statistical significance, similarly as done when analyzing fixation-derived measures, we carried out the three following steps:

- For each abstract we took the mean value of the normalized pupil size in the time window of 0 to 1.3 seconds after the response time.
- For each speed and each condition (the user answered relevant or non relevant) we averaged the values of each subject. This was also done for the overall texts read by the user, without taking into account the speed. This resulted, for each subject, in 4 pair of values:
  - Normal speed and assessed relevant — Normal speed and assessed non relevant
  - Fast speed and assessed relevant — Fast speed and assessed non relevant
  - Skimming speed and assessed relevant — Skimming speed and assessed non relevant
  - All texts assessed relevant — All texts assessed non relevant
- We performed Wilcoxon signed-rank text on the resulting paired samples.



(a) Behavior of pupil for articles rated incongruently



(b) Behavior of pupil when the assessed certainty is under seven

Figure 21: Behavior of pupil in the two excluded scenarios. The red line indicates the rating moment

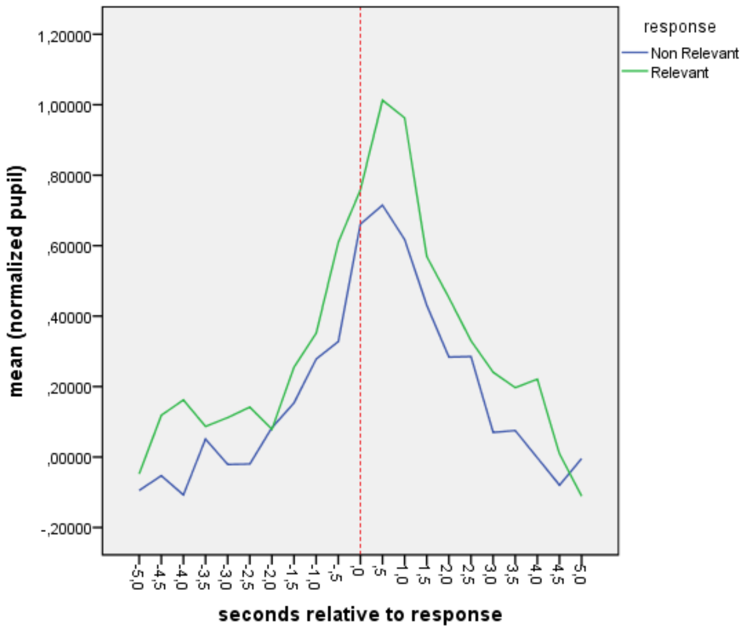


Figure 22: Behavior of pupil for articles rated congruently and with assessed certainty over six

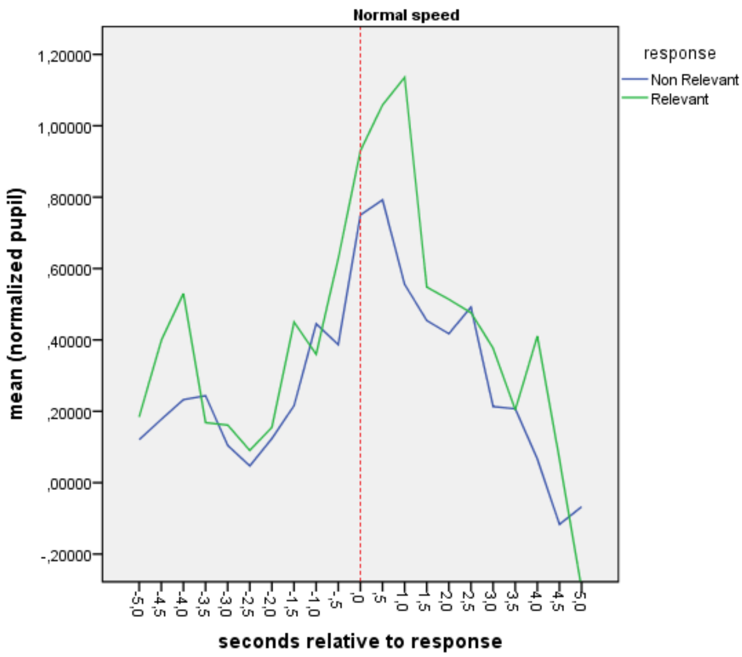


Figure 23: Behavior of pupil for articles read at Normal speed, rated congruently and with assessed certainty over six

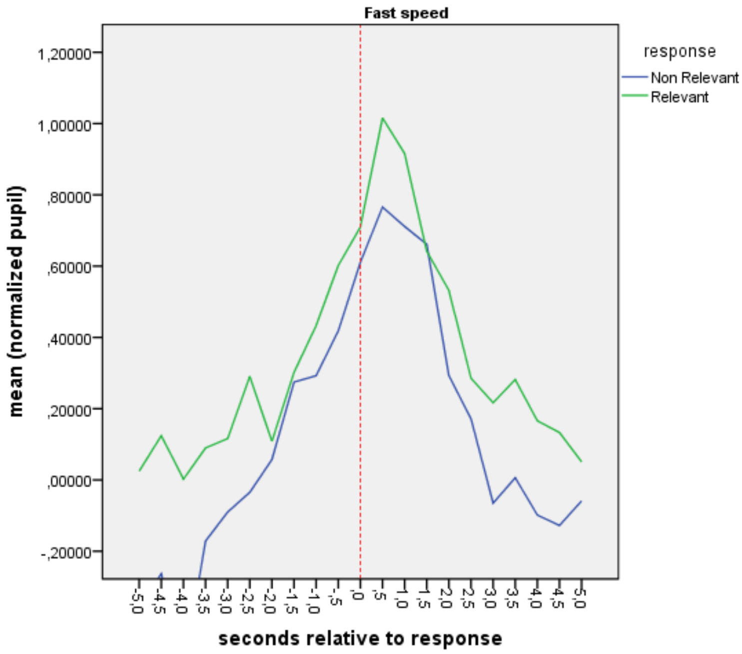


Figure 24: Behavior of pupil for articles read at Fast speed, rated congruently and with assessed certainty over six

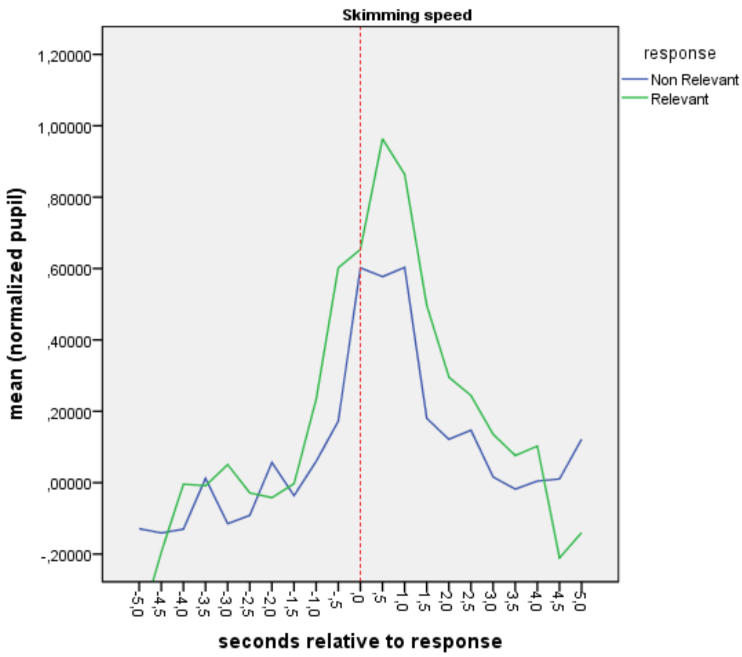


Figure 25: Behavior of pupil for articles read at Skimming speed, rated congruently and with assessed certainty over six

In overall, normalized 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, normalized 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 Skimming speed –relevant (Mdn = 66), non relevant (Mdn = 0.59), z = -0.676, r = -0.18– no statistical significance was found.

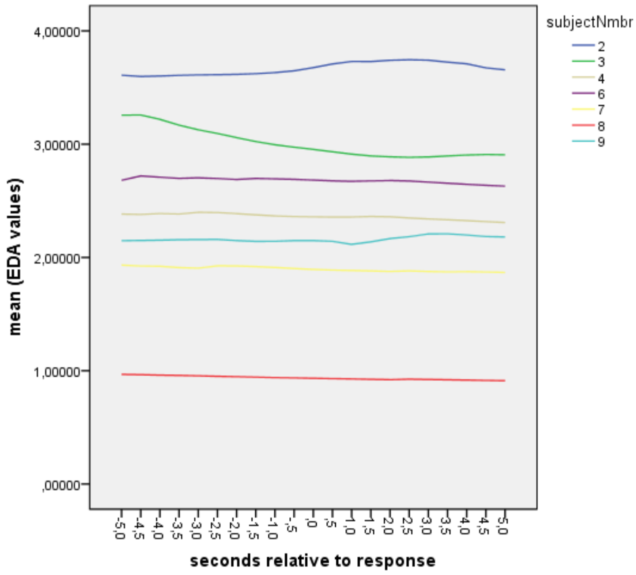
7.2 Electrodermal activity

Before running the analysis, in the same way as with the rest of analyzed features, we discarded from the dataset non-congruently assessed texts as well as texts rated with a level of certainty below seven. We then proceeded to the same normalization procedure carried out when analyzing pupil size. That is, we subtracted the mean of the whole document for every EDA value within a read text.

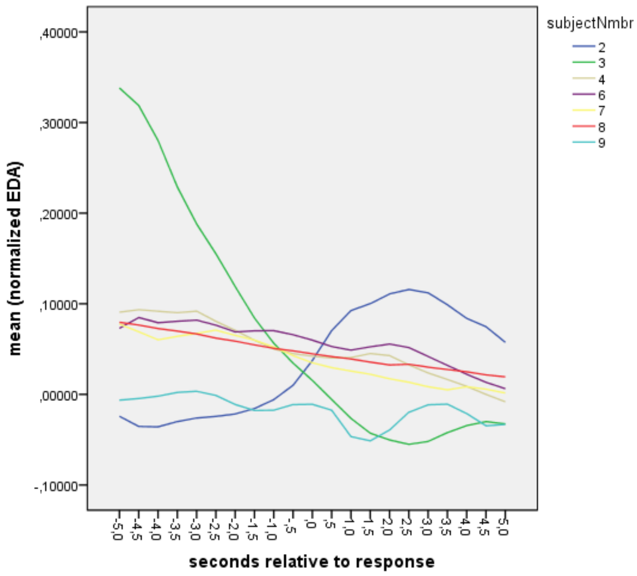
As already mentioned in section 4.3 electrodermal activity is extremely variable between subjects (see figure 26). Plotting the mean of all subjects at different speeds would then not make much sense. To have an overall idea of how this signal behaved, we plotted each subject in different plots, separating the documents assessed as relevant and non-relevant, and splitting the output by speeds. Figure 27 show the behavior of two different subjects selected randomly when reading at different speeds.

The observed EDA behavior was similar within subjects but heterogeneous between subjects (see figure 27). Even though, we were able to observe that something was happening in the approximate interval of half a second prior to the response to 2.5 seconds after the response. The selected time window to compute the mean of normalized EDA values for a each of the articles was then of -0.5 to 2.5 seconds relative to the response time. We ran a similar analysis than with pupil size.

Wilcoxon signed-rank test was used to analyze the resulting means. When looking at the overall behavior, without taking into account the addressed speed, the mean of normalized EDA values appeared to be significantly higher when reading texts assessed as relevant (Mdn = 0.041) than when reading texts assessed as non relevant (Mdn = 0.032), z = -2.028, $p < 0.05$, r = -0.54. When looking a the

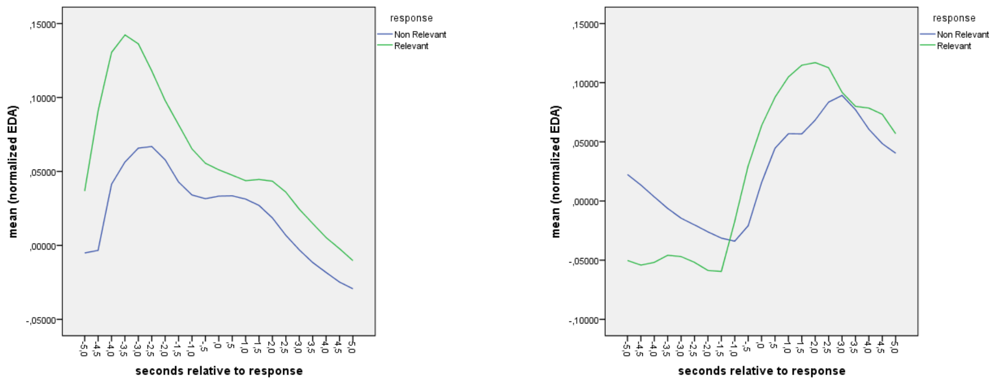


(a) EDA without normalization

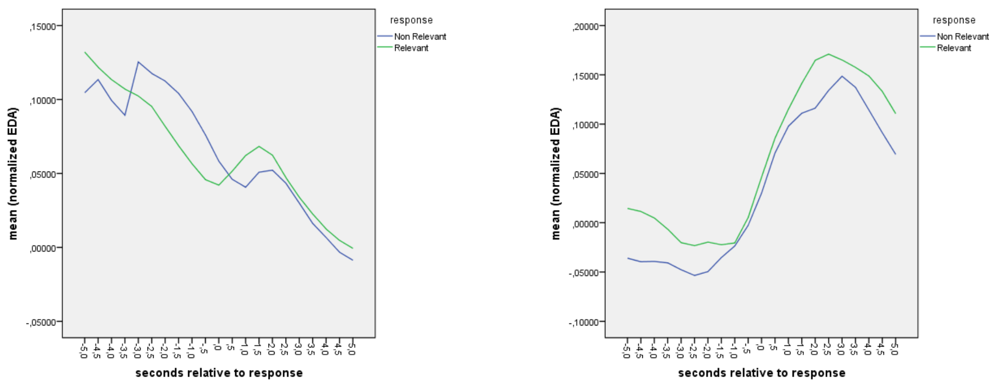


(b) Normalized EDA values

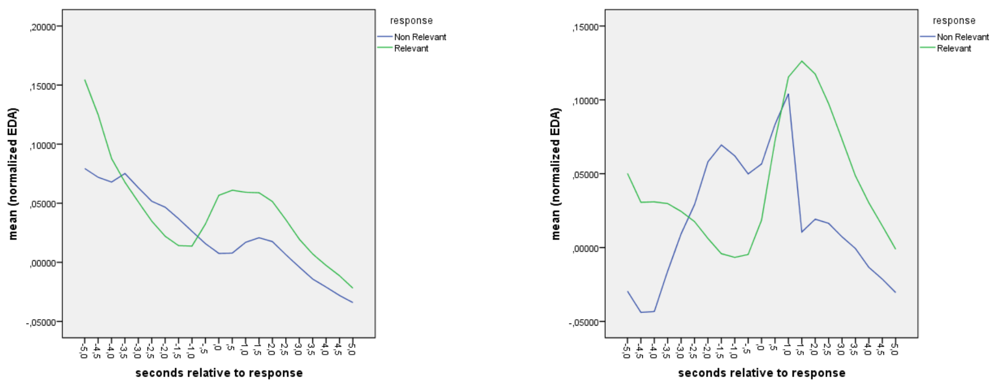
Figure 26: EDA values for each of the seven subjects, in a time window of 10 seconds. Binary feedback was assessed at time 0



(a) Normal speed



(b) Fast speed



(c) Skimming speed

Figure 27: EDA behavior when assessing relevance of texts read at different speeds. Data extracted from two randomly selected participants (left and right columns)

data split by addressed speed, similarly as with the pupil size, texts read at normal speed and assessed as relevant (Mdn = 0.09) shown a mean significantly higher than when assessing non relevant texts (Mdn = 0.044), z = -2.366, $p < 0.05$, r = -0.63. Nonetheless, when reading at fast speed, no significance was found when comparing normalized EDA values in texts assessed as relevant (Mdn = 0.037) and non relevant (Mdn = 0.045), z = -0.338, r = -0.09. When reading at skimming speed, Wilcoxon signed-rank test did not report statistical significance when comparing normalized EDA values in texts assessed as relevant (Mdn = 0.047) and non relevant (Mdn = 0.013), z = -1.014, r = 0.27.

7.3 Discussion

The results when analyzing pupil size showed a clear relationship between the pupil dilation when assessing documents as relevant and when doing it as non relevant. Additionally, 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, the results indicated that the response-related spike was significantly bigger 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 speed of reading, the data showed statistical significance only when the user was reading at normal speed. That is, when the participant was reading without time pressure, taking the amount of time necessary to understand the entirety of the text read. When the subject was given the instruction to read at faster rates than the comfortable normal reading speed, the response-related spike did not show statistically significant different behavior when perceiving documents as relevant or non-relevant.

We would like to stress the fact that we specifically analyzed the response-related spike in order to seek for relevant differences. We based our analysis in the assumption of cognitive effect happening when the user made the decision of relevance (we can think about it as the *ahá! moment*). It would be interesting to base further analysis in the extracting of such cognition-related spike out of the whole amount of raw gaze data. In the goal of developing highly adaptive information systems, one will not be given the response moment, automatic detection of such being crucial for the application of our findings.

Assuming that such automatic detection of response was possible, our results indicate that a hypothetical adaptive information retrieval system would also need to be aware of the speed of reading when basing the predictions of user perception of relevance on pupil dilation. In other words, the results indicate that a pupil size-based adaptive system would loose accuracy if not being aware of reading speed at every moment.

Electrodermal activity results indicated that both in overall and normal speed, arousal when facing a stimulus of the kind “relevant text” was significantly higher than arousal when addressing “non relevant texts”. The results were similar than those reported for pupil size, but in the case of EDA we would like to be extremely cautious when taking them for granted. When looking at the graphs for all the subjects (see figure 27), EDA did not showed such a clear response-related effect than pupil size. Thus, the analyzed data was probably not only caused by cognitive response-related effects, but also influenced by other environmental factors. Additionally, the highly variance of this measure between subjects, let to the possible non-optimality of this its recording. In our understanding, EDA would demand a bigger sample of the population and even a more controlled recording method for the results to be trustful.

8 Conclusions

This thesis has explored the relationship between fixation-based measures of predicted relevance, physiology and reading speed. We have first reviewed the background regarding perception of relevance inference and reading behavioral patterns in order to expose coherently the motivations and research questions of this thesis. An experiment has been designed to address the mentioned questions and both the analysis and results have been reported and thoroughly discussed.

Digital era is bringing new reading paradigms. The way we seek for information has changed radically, due to the huge and highly structured amount of information available. Nowadays, from a simple personal computer one can access digital libraries from all around the world, in few minutes. The reader then faces the need of selecting relevant information from a ridiculously big amount of documents. In order to do so, a small portion of documents viewed during the search task is fully read, the majority being just scanned so to get a quick overview of its content. It is our belief that adaptive information retrieval systems need to be aware of different reading speeds in order to make more accurate eye-gaze based predictions.

Eye movements as well as fixation-derived features have well been studied in information retrieval as well as other computer science related fields. Interesting findings have been discussed in the thesis regarding whether the eyes can be used as indicators of user interest. We have first reviewed the literature in the domain of inferring perceived relevance in documents from user eyes. Then we have selected fixation-based features suitable to be studied in our experiment. We were mostly interested in those features who did not require the knowledge of word positions and semantic meaning as we understand that it is extremely difficult for a highly dynamic and adaptive information retrieval system to be aware of such low-level parameters.

We were also highly interested in looking at the psychophysiology of the eyes. Pupil size has been proved to react to different cognitive processes, but when in the goal of inferring interest, it has been explored in a smaller measure than other eye-derived metrics. We accompanied the analysis of pupil size with the one of electrodermal activity. It is difficult to predict relevance from EDA by itself, but we are positive that the understanding of its behavior while addressing texts at different speeds will help to have a better insight into the measure.

We have mainly based our analysis on non-parametric statistical tests to compare group means. We have observed differences when reading texts at different speeds in features such as number of fixations, fixation duration and length of forward saccades. Studied psychophysiology also has proven to react differently to stimuli when changing the reading speed. When reading at comfortable speed, pupil size as well as electrodermal activity has been found to behave in different ways depending in the assessed relevance but, when increasing the reading speed, no difference has been reported. Therefore, we consider the overall analysis outcomes as to be positive regarding the usefulness of being aware of user's reading speed when aiming to the inferring of perceived relevance in documents. We consequently believe that it is worthy to take into account reading speed when designing intelligent adaptive systems.

Nevertheless, we think that the reduced amount of subjects as well as the specifications and design of our experiment have limited the number and the robustness of our findings. In the same way as we think that it is important to study the effects, hence control reading speed, we have also found our controlled factors to have a strong impact in the behavior of the readers. We encourage the carrying out of future experiments in the same field while trying to minimize the side effects of controlling reading speed. Word-level fixation-derived metrics are useful in specific contexts and we would also like to encourage the carrying out of word-level studies under similar conditions, i.e. taking into account the speed of reading. A much broader range of features and their relationship with reading speed could then be analyzed.

The outcomes of this thesis show, as already pointed by Oliveira and Russell [OAR09], that pupil size might be an important carrier of cognitive information when assessing articles regarding their relevancy to a given topic. We believe that further specific studies addressing pupil size for inferring perceived relevance need to be carried out, controlling as much as possible all the factors influencing that measure in order to get precise and trustful results. We also encourage the combination of pupil size analysis with other eye-gaze derived implicit signals as well as psychophysiological signals such as EDA. We also think that it would be useful to take into account reading speed when addressing further research on perceived relevance inference through *traditional* implicit indicators as well as other psychophysiological signals such as EMG or ECG.

Last, we would like to stress the fact that plenty of algorithms capable of tracking reading behavior, namely reading speed, have already been designed and tested [CM01] [BDvE08]. Therefore we understand that a system can easily apply such methods in order to be aware of the speed of reading, deriving in a better understanding of user behavior, thus, in more accurate predictions.

References

ABD06 Agichtein, E., Brill, E. and Dumais, S., Improving web search ranking by incorporating user behavior information. *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval.* ACM, 2006, pages 19–26.

ADS05 Ashmore, M., Duchowski, A. T. and Shoemaker, G., Efficient eye pointing with a fisheye lens. *Proceedings of Graphics interface 2005.* Canadian Human-Computer Communications Society, 2005, pages 203–210.

AHK+09 Ajanki, A., Hardoon, D. R., Kaski, S., Puolamäki, K. and Shawe-Taylor, J., Can eyes reveal interest? implicit queries from gaze patterns. *User Modeling and User-Adapted Interaction*, 19,4(2009), pages 307–339.

BBH+09 Bowyer, K. W., Baker, S. E., Hentz, A., Hollingsworth, K., Peters, T. and Flynn, P. J., Factors that degrade the match distribution in iris biometrics. *Identity in the Information Society*, 2,3(2009), pages 327–343.

BDBE12 Buscher, G., Dengel, A., Biedert, R. and Elst, L. V., Attentive documents: Eye tracking as implicit feedback for information retrieval and beyond. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 1,2(2012), page 9.

BDvE08 Buscher, G., Dengel, A. and van Elst, L., Eye movements as implicit relevance feedback. *CHI'08 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2008, pages 2991–2996.

BLW00 Beatty, J. and Lucero-Wagoner, B., The pupillary system. *Handbook of psychophysiology*, New York, NY, US, 2000, Cambridge University Press, pages 142–162.

BMA+01 Bersak, D., McDarby, G., Augenblick, N., McDarby, P., McDonnell, D., McDonald, B. and Karkun, R., Intelligent biofeedback using an immersive competitive environment. Paper at the Designing Ubiquitous Computing Games Workshop at UbiComp, 2001.

BPB+06 Brooks, P., Phang, K. Y., Bradley, R., Oard, D., White, R. and Guimbretire, F., Measuring the utility of gaze detection for task modeling: A preliminary study. *Workshop on Intelligent Interfaces for Intelligent Analysis*, 2006.

CC73 Cornsweet, T. N. and Crane, H. D., Accurate two-dimensional eye tracker using first and fourth purkinje images. *JOSA*, 63,8(1973), pages 921–928.

CLWB01 Claypool, M., Le, P., Wased, M. and Brown, D., Implicit interest indicators. *Proceedings of the 6th international conference on Intelligent user interfaces*. ACM, 2001, pages 33–40.

CM01 Campbell, C. S. and Maglio, P. P., A robust algorithm for reading detection. *Proceedings of the 2001 workshop on Perceptive user interfaces*. ACM, 2001, pages 1–7.

DH00 Dyson, M. and 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(2000), pages 210–223.

DH01 Dyson, M. C. and Haselgrove, M., The influence of reading speed and line length on the effectiveness of reading from screen. *International Journal of Human-Computer Studies*, 54,4(2001), pages 585–612.

Dil92 Dillon, A., Reading from paper versus screens: A critical review of the empirical literature. *Ergonomics*, 35,10(1992), pages 1297–1326.

DK98 Dyson, M. C. and Kipping, G. J., The effects of line length and method of movement on patterns of reading from screen. *Visible Language*, 32,2(1998), pages 150–181.

DMR88 Duffy, S. A., Morris, R. K. and Rayner, K., Lexical ambiguity and fixation times in reading. *Journal of memory and language*, 27,4(1988), pages 429–446.

DP09 Duggan, G. B. and Payne, S. J., Text skimming: The process and effectiveness of foraging through text under time pressure. *Journal of Experimental Psychology: Applied*, 15,3(2009), page 228.

DSF07 Dawson, M. E., Schell, A. M. and Filion, D. L., The electrodermal system. *Handbook of psychophysiology*, 2007, page 159.

Duc02 Duchowski, A. T., A breadth-first survey of eye-tracking applications. *Behavior Research Methods*, 34,4(2002), pages 455–470.

Duc07 Duchowski, A., *Eye tracking methodology: Theory and practice*, volume 373. Springer, 2007.

FKM+05 Fox, S., Karnawat, K., Mydland, M., Dumais, S. and White, T., Evaluating implicit measures to improve web search. *ACM Transactions on Information Systems (TOIS)*, 23,2(2005), pages 147–168.

GAF+87 Gould, J. D., Alfaro, L., Finn, R., Haupt, B. and Minuto, A., Why reading was slower from crt displays than from paper. *ACM SIGCHI Bulletin*, 18,4(1987), pages 7–11.

GJ84 Gale, A. and Johnson, F., Operational problems in analysing eye movements. *European Conference on Eye Movements: Theoretical and applied aspects of eye movement research*, volume 22. Access Online via Elsevier, 1984, page 21.

GK99 Goldberg, J. H. and Kotval, X. P., Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*, 24,6(1999), pages 631–645.

GPS99 Golovchinsky, G., Price, M. N. and Schilit, B. N., From reading to retrieval: freeform ink annotations as queries. *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval.* ACM, 1999, pages 19–25.

GW03 Goldberg, J. H. and Wichansky, A. M., Eye tracking in usability evaluation: A practitioner's guide. *The mind's eye: cognitive and applied aspects of eye movement research*, 2003, pages 573–605.

HH02 Hornof, A. J. and Halverson, T., Cleaning up systematic error in eye-tracking data by using required fixation locations. *Behavior Research Methods, Instruments, & Computers*, 34,4(2002), pages 592–604.

HJM+89 Hutchinson, T. E., Jr, K. P. W., Martin, W. N., Reichert, K. C. and Frey, L. A., Human-computer interaction using eye-gaze input. *Systems, Man and Cybernetics, IEEE Transactions on*, 19,6(1989), pages 1527–1534.

HKTR04 Herlocker, J. L., Konstan, J. A., Terveen, L. G. and Riedl, J. T., Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22,1(2004), pages 5–53.

HP64 Hess, E. H. and Polt, J. M., Pupil size in relation to mental activity during simple problem-solving. *Science*, 143,3611(1964), pages 1190–1192.

HPI12 Hajimirza, S. N., Proulx, M. J. and Izquierdo, E., Reading users' minds from their eyes: A method for implicit image annotation. *Multimedia, IEEE Transactions on*, 14,3(2012), pages 805–815.

HTIH96 Horton, W., Taylor, L., Ignacio, A. and Hoft, N. L., The web page design cookbook: all the ingredients you need to create 5-star web pages. *New York: Wiley,| c1996*, 1.

Jac90 Jacob, R. J., What you look at is what you get: eye movement-based interaction techniques. *Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people*. ACM, 1990, pages 11–18.

Jac93 Jacob, R. J., Eye movement-based human-computer interaction techniques: Toward non-command interfaces. *Advances in human-computer interaction*, 4, pages 151–190.

JC93 Just, M. A. and Carpenter, P. A., The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expé rimentale*, 47,2(1993), page 310.

KB01 Kelly, D. and Belkin, N. J., Reading time, scrolling and interaction: exploring implicit sources of user preferences for relevance feedback. *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2001, pages 408–409.

Kel05 Kelly, D., Implicit feedback: Using behavior to infer relevance. *New directions in cognitive information retrieval*, 2005, pages 169–186.

KLT+10 Kuikkaniemi, K., Laitinen, T., Turpeinen, M., Saari, T., Kosunen, I. and Ravaja, N., The influence of implicit and explicit biofeedback in first-person shooter games. *Proceedings of the 28th international conference on Human factors in computing systems*. ACM, 2010, pages 859–868.

KMM+97 Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R. and Riedl, J., Grouplens: applying collaborative filtering to usenet news. *Communications of the ACM*, 40,3(1997), pages 77–87.

KOR00 Kim, J., Oard, D. W. and Romanik, K., User modeling for information access based on implicit feedback. *Technical Report: HCIL-TR-2000-11/UMIACS-TR-2000-29/CS-TR-4136*, University of Maryland, 2000, College Park.

KT03 Kelly, D. and Teevan, J., Implicit feedback for inferring user preference: a bibliography. *ACM SIGIR Forum*, volume 37. ACM, 2003, pages 18–28.

LBB11 Loboda, T. D., Brusilovsky, P. and Brunstein, J., Inferring word relevance from eye-movements of readers. *Proceedings of the 16th international conference on Intelligent user interfaces*. ACM, 2011, pages 175–184.

LF00 Liversedge, S. P. and Findlay, J. M., Saccadic eye movements and cognition. *Trends in cognitive sciences*, 4,1(2000), pages 6–14.

Liu05 Liu, Z., Reading behavior in the digital environment: Changes in reading behavior over the past ten years. *Journal of Documentation*, 61,6(2005), pages 700–712.

LN02 Lisetti, C. L. and Nasoz, F., Maui: a multimodal affective user interface. *Proceedings of the tenth ACM international conference on Multimedia*. ACM, 2002, pages 161–170.

LS01 Li, M. and Selker, T., Eye pattern analysis in intelligent virtual agents. *Intelligent Virtual Agents*. Springer, 2001, pages 23–35.

Mas82 Masson, M. E., Cognitive processes in skimming stories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8,5(1982), page 400.

Mas83 Masson, M. E., Conceptual processing of text during skimming and rapid sequential reading. *Memory & cognition*, 11,3(1983), pages 262–274.

Mir13 http://www.mirametrix.com, 2013.

MJL07 Moe, K. K., Jensen, J. M. and Larsen, B., A qualitative look at eye-tracking for implicit relevance feedback. *Proceedings of the Workshop on Context-Based Information Retrieval*, 2007, pages 36–47.

MKAF00 Morimoto, C. H., Koons, D., Amir, A. and Flickner, M., Pupil detection and tracking using multiple light sources. *Image and Vision Computing*, 18,4(2000), pages 331–335.

MLTB82 Muter, P., Latrémouille, S. A., Treurniet, W. C. and Beam, P., Extended reading of continuous text on television screens. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 24,5(1982), pages 501–508.

MM91 Muter, P. and Maurutto, P., Reading and skimming from computer screens and books: the paperless office revisited? *Behaviour & Information Technology*, 10,4(1991), pages 257–266.

MR02 Majaranta, P. and Räihä, K.-J., Twenty years of eye typing: systems and design issues. *Proceedings of the 2002 symposium on Eye tracking research & applications*. ACM, 2002, pages 15–22.

MS94 Morita, M. and Shinoda, Y., Information filtering based on user behavior analysis and best match text retrieval. *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval*. Springer-Verlag New York, Inc., 1994, pages 272–281.

Nel01 Nelson, D., The uptake of electronic journals by academics in the uk, their attitudes towards them and their potential impact on scholarly communication. *Information Services and Use*, 21,3(2001), pages 205–214.

NHJW06 Nicholas, D., Huntington, P., Jamali, H. R. and Watkinson, A., The information seeking behaviour of the users of digital scholarly journals. *Information Processing & Management*, 42,5(2006), pages 1345–1365.

Nic98 Nichols, D. M., Implicit rating and filtering. *In Proceedings of the Fifth DELOS Workshop on Filtering and Collaborative Filtering*, 1998, pages 31–36.

OAR09 Oliveira, F. T., Aula, A. and Russell, D. M., Discriminating the relevance of web search results with measures of pupil size. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2009, pages 2209–2212.

OK01 Oard, D. W. and Kim, J., Modeling information content using observable behavior. *Proceedings of the 64th Annual Meeting of the American Society for Information Science and Technology*, 2001.

OM96 Oard, D. W. and Marchionini, G., A conceptual framework for text filtering process. *Technical Report EE-TR-96-25 CAR-TR-830 CLIS-TR-9602 CS-TR-3643*. University of Maryland, May 1996.

PB05 Poole, A. and Ball, L. J., Eye tracking in HCI and usability research. *Encyclopedia of Human-Computer Interaction*, Pennsylvania, 2005, C. Ghaoui (ed.), Idea Group.

PSS+05 Puolamäki, K., Salojärvi, J., Savia, E., Simola, J. and Kaski, S., Combining eye movements and collaborative filtering for proactive information retrieval. *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2005, pages 146–153.

Ray98 Rayner, K., Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124,3(1998), page 372.

RM76 Rayner, K. and McConkie, G. W., What guides a reader's eye movements? *Vision research*, 16,8(1976), pages 829–837.

RNW+08 Rowlands, I., Nicholas, D., Williams, P., Huntington, P., Fieldhouse, M., Gunter, B., Withey, R., Jamali, H. R., Dobrowolski, T. and Tenopir, C., The google generation: the information behaviour of the researcher of the future. *Aslib Proceedings*, volume 60. Emerald Group Publishing Limited, 2008, pages 290–310.

Rob68 Robinson, D. A., The oculomotor control system: A review. *Proceedings of the IEEE*, 56,6(1968), pages 1032–1049.

RP81 Rayner, K. and Pollatsek, A., Eye movement control during reading: Evidence for direct control. *The Quarterly journal of experimental psychology*, 33,4(1981), pages 351–373.

RP89 Rayner, K. and Pollatsek, A., *The psychology of reading.* Routledge, 1989.

RR50 Ratliff, F. and Riggs, L. A., Involuntary motions of the eye during monocular fixation. *Journal of experimental psychology*, 40,6(1950), pages 687–701.

RS13 Richardson, D. C. and Spivey, M. J. *Eye Tracking: Characteristics and Methods*, pages 568–572. Taylor & Francis, February 2013.

SBL+09 Salimpoor, V. N., Benovoy, M., Longo, G., Cooperstock, J. R. and Zatorre, R. J., The rewarding aspects of music listening are related to degree of emotional arousal. *PloS one*, 4,10(2009), page e7487.

SBSL02 Selker, T., Burleson, W., Scott, J. and Li, M., Eye-bed. *Workshop on Multimodal Resources and Evaluation in conjunction with the Third International Conference on Language Resources and Evaluation (LREC)*, June 2002.

SG00 Salvucci, D. D. and Goldberg, J. H., Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the 2000 symposium on Eye tracking research & applications*. ACM, 2000, pages 71–78.

Sil08 Silvia, P. J., Interest—the curious emotion. *Current Directions in Psychological Science*, 17,1(2008), pages 57–60.

SKSK03 Salojärvi, J., Kojo, I., Simola, J. and Kaski, S., Can relevance be inferred from eye movements in information retrieval. *Proceedings of WSOM*, volume 3, 2003, pages 261–266.

SM95 Shardanand, U. and Maes, P., Social information filtering: algorithms for automating "word of mouth". *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press/Addison-Wesley Publishing Co., 1995, pages 210–217.

Smi03 Smith, E. T., Changes in faculty reading behaviors: the impact of electronic journals on the university of georgia. *The Journal of Academic Librarianship*, 29,3(2003), pages 162–168.

SPK04 Salojärvi, J., Puolamäki, K. and Kaski, S., Relevance feedback from eye movements for proactive information retrieval. *Workshop on Processing Sensory Information for Proactive Systems*, 2004.

SZM13 Schomisch, S., Zens, M. and Mayr, P., Are e-readers suitable tools for scholarly work? *Online Information Review*, 37,3(2013), pages 4–4.

Tho65 Thomas, J., Use of a piezo-accelerometer in studying eye dynamics. *JOSA*, 55,5(1965), pages 534–537.

TJ00 Tanriverdi, V. and Jacob, R. J., Interacting with eye movements in virtual environments. *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2000, pages 265–272.

WKMB99 Walczyk, J. J., Kelly, K. E., Meche, S. D. and Braud, H., Time limitations enhance reading comprehension. *Contemporary educational psychology*, 24,2(1999), pages 156–165.

You63 Young, L. R., Measuring eye movements. *The American journal of medical electronics*, volume 2, 1963, pages 300–307.

ZMI99 Zhai, S., Morimoto, C. and Ihde, S., Manual and gaze input cascaded (magic) pointing. *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*. ACM, 1999, pages 246–253.

ZRZ08 Zhang, X., Ren, X. and Zha, H., Improving eye cursor's stability for eye pointing tasks. *CHI-CONFERENCE-*, volume 1. Association for computing machinery INC, 2008, page 525.

ZZ06 Zigoris, P. and Zhang, Y., Bayesian adaptive user profiling with explicit & implicit feedback. *Proceedings of the 15th ACM international conference on Information and knowledge management*. ACM, 2006, pages 397–404.

Appendix 1. Fixation-recognition algorithm: Code

This appendix includes the code in *python* of the algorithm described in section 3.4 used to extract fixations from the raw eye gaze data provided by the mirametrix device. The function *find_ nth* is used in order to extract the appropriate fields from the raw gaze file. The function *fitInBox* return true if all the gaze points identified by the vectors x and y fit in a *size* x *size* square pixel box. The more important function is the one named *findNextFixation*. This function is charged to find the next fixation in the raw gaze data field. If a new fixation is found, it is written in the output file and the function returns 1. If there are no further fixations, the function returns 0. The main initialize the parameters and run the function *findNextFixation* until no more fixations are found in the raw data file.

```
import sys
from decimal import Decimal
BPOGX_POS = 15
BPOGY_POS = 16
BPOGV_POS = 17
TIME_POS = 2

SCREEN_WIDTH = 1280
SCREEN_HEIGHT = 1024

FIX_BORDER = 30
FIX_MARGIN = 20

def find_nth(line,n):
    start = line.find(',')
    if n == 0:
        return line[0:start]
    while n > 1:
        start = line.find(',', start+1)
        n -= 1
    end = line.find(',', start+1)
    return line[start+1:end]

def fitInBox(x,y,size):
```

```
    return (min(x) >= max(x) -size and min(y) >= max(y) -size)


def findNextFixation(inFile,f,x,y,t):
    while (not fitInBox(x,y,FIX_BORDER)):
        x.pop(0)
        y.pop(0)
        t.pop(0)
        line = inFile.readline()
        if not line:
            return 0
        while (find_nth(line,BPOGV_POS) == '0'):
            line = inFile.readline()
            if not line:
                return 0
        x.append(Decimal(find_nth(line,BPOGX_POS))*SCREEN_WIDTH)
        y.append(Decimal(find_nth(line,BPOGY_POS))*SCREEN_HEIGHT)
        t.append(int(find_nth(line,TIME_POS)))

    numOutliers = 0
    auX = []
    auY = []
    auT = []
    while (numOutliers < 5):
        line = inFile.readline()
        if not line:
            f.write(str(t[0])+','+str(t[len(t)-1] - t[0])+','+
        str(round((sum(x)/len(x))/SCREEN_WIDTH,5)) + ','+
        str(round((sum(y)/len(y))/SCREEN_HEIGHT,5))+'\n')
            return 0
        while (find_nth(line,BPOGV_POS) == '0'):
            line = inFile.readline()
            if not line:
                f.write(str(t[0])+','+str(t[len(t)-1] - t[0])+','+
        str(round((sum(x)/len(x))/SCREEN_WIDTH,5)) + ','+
        str(round((sum(y)/len(y))/SCREEN_HEIGHT,5))+'\n')
```

```
                return 0
            x.append(Decimal(find_nth(line,BPOGX_POS))*SCREEN_WIDTH)
            y.append(Decimal(find_nth(line,BPOGY_POS))*SCREEN_HEIGHT)
            t.append(int(find_nth(line,TIME_POS)))
            if (not fitInBox(x,y,FIX_BORDER+FIX_MARGIN)):
                numOutliers = numOutliers + 1
                auX.append(x.pop())
                auY.append(y.pop())
                auT.append(t.pop())
            else:
                numOutliers = 0
                auX = []
                auY = []
                auT = []

        f.write(str(t[0])+','+str(t[len(t)-1] - t[0])+','+ str(round((sum(x
        x[:] = auX
        y[:] = auY
        t[:] = auT

        return 1



for filename in sys.argv[1:len(sys.argv)]:
    outName = 'fix-'+filename
    f = open(outName,'w')
    f.write('FIX_INIT_TIME,FIX_DURATION,FIX_X,FIX_Y\n')

    inFile = open(filename, 'r')
    line = inFile.readline()   #we jump the headers
    count = 0
    fixInit = -1
    fixDuration = -1
    x = []
    y = []
    t = []
```

```
#we find the initial points of the fixation (we need at least 5)
done = 0
while (len(x) < 5):
    line = inFile.readline()
    if not line:
        #inFile.close()
        print('EEEROR1 '+ outName)
        done = 1
        break
    if (find_nth(line,BPOGV_POS) == '0'):
        continue
    x.append(Decimal(find_nth(line,BPOGX_POS))*SCREEN_WIDTH)
    y.append(Decimal(find_nth(line,BPOGY_POS))*SCREEN_HEIGHT)
    t.append(int(find_nth(line,TIME_POS)))

#we write all the fixations
if (done == 0):
    while findNextFixation(inFile,f,x,y,t):
        continue

inFile.close()
f.close()
```

Appendix 2. Sample of values of interest extracted from the ePrime logs

```
SubjectNmbr,SessionNmber,numWords,relevant,speed,topic,
target-onsetTime,response-time,response,target-closeTime,feedbackTime,
readingTime,grading,sure
3,1,142,1,2,1,21304070162,21332406963,1,21358073314,9062,17270,6,6
3,1,205,0,2,1,21442286335,21465551156,0,21497999952,7440,17817,0,9
3,1,189,0,1,1,21582847753,21602622850,0,21687308047,6324,33406,0,9
3,1,189,0,0,1,21724140886,21762509078,0,21889224175,12270,52793,0,8
3,1,151,1,0,1,21944987823,21965053731,1,22120415201,6417,56101,9,9
3,1,164,1,1,1,22167498320,22191038316,1,22273497095,7528,33898,8,8
3,1,190,0,2,0,22327547151,22353454280,0,22395609259,8285,21766,2,5
3,1,198,1,0,0,22449249680,22486173202,1,22576812192,11808,40794,7,8
3,1,157,0,0,0,22613723205,22635902959,0,22752186411,7093,44280,0,9
3,1,214,1,1,0,22784972922,22799194482,1,22921953934,4548,43806,8,9
3,1,162,0,2,0,22959190154,22983974693,0,23019950736,7926,19431,0,6
3,1,142,1,1,0,23074432318,23096105499,1,23170305892,6931,30660,5,8
3,1,176,0,0,2,23228489832,23280532360,0,23379804975,16643,48390,1,8
3,1,169,0,2,2,23415315096,23436369134,0,23480450340,6733,20830,1,7
3,1,171,0,2,2,23522949288,23555829609,0,23602462441,10515,25428,0,7
3,1,144,1,0,2,23649235987,23670133675,1,23780788540,6683,42070,7,8
3,1,141,1,1,2,23819391257,23847384089,1,23929967948,8952,35362,7,8
3,1,204,1,1,2,23961203471,23982329429,1,24112696852,6756,48447,7,7
```

This is the file extracted from the first session of subject number three. This file contains all the relevant ePrime extracted information needed to parse the eye tracking raw data, as well as for the understanding of the user's general behavior. The field names should be self-explanatory. The keyword *target* refers to the stimuli, the shown abstract. The fields *target-onsetTime*, *response-time* and *target-closeTime* are expressed in system clock ticks.

Appendix 3. Syntax and relevant output tables from Mixed Models procedures in SPSS

This appendix contains the syntax as well as the output tables corresponding to the fixed effects and the correlation parameters estimates, for the mixed models procedure ran when exploring the data. All the procedures were ran using number of fixation, average time of fixation, forward saccade length and regression ratio as covariates. Different subjects were indicated with the subject identifier.

The first two procedures were run with the binary response as fixed effect. The first one was ran considering all the data, and the second one was ran splitting the output by instructed speed (variable speedClass, 0 -> Normal speed, 1 -> Fast speed, 2 -> Skimming speed).

The third procedure was run in order to explore the behavior at different reading speeds, so the instructed speed was set as the fixed effect.

```
MIXED response WITH numFix avgTimeFix regressionRatio fwdAvgSaccade
 /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.000000000001) HCONVERGE(0,
 /FIXED=numFix avgTimeFix regressionRatio fwdAvgSaccade | SSTYPE(3)
 /METHOD=REML
 /PRINT=CORB  SOLUTION
 /RANDOM=INTERCEPT | SUBJECT(subjectNmbr) COVTYPE(VC).
```

Fixed Effects

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	206,000	2,268	,134
numFix	1	206,000	,780	,378
avgTimeFix	1	206,000	,492	,484
regressionRatio	1	206,000	,467	,495
fwdAvgSaccade	1	206,000	,068	,794

a. Dependent Variable: response.

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% ...
						Lower Bound
Intercept	,408129	,271022	206,000	1,506	,134	-,126204
numFix	,000903	,001022	206,000	,883	,378	-,001113
avgTimeFix	-,000859	,001224	206,000	-,702	,484	-,003272
regressionRatio	,318934	,466501	206,000	,684	,495	-,600795
fwdAvgSaccade	,000165	,000632	206,000	,261	,794	-,001080

Estimates of Fixed Effects[a]

Parameter	95% ...
	Upper Bound
Intercept	,942461
numFix	,002918
avgTimeFix	,001555
regressionRatio	1,238664
fwdAvgSaccade	,001410

a. Dependent Variable: response.

Correlation Matrix for Estimates of Fixed Effects[a]

Parameter	Intercept	numFix	avgTimeFix	regressionRatio	fwdAvgSaccade
Intercept	1	-,271	-,651	-,135	-,679
numFix	-,271	1	-,334	-,105	,504
avgTimeFix	-,651	-,334	1	-,147	,161
regressionRatio	-,135	-,105	-,147	1	-,313
fwdAvgSaccade	-,679	,504	,161	-,313	1

a. Dependent Variable: response.

```
 SORT CASES  BY speedClass.
SPLIT FILE SEPARATE BY speedClass.
MIXED response WITH numFix avgTimeFix regressionRatio fwdAvgSaccade
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.000000000001) HCONVERGE(0,
  /FIXED=numFix avgTimeFix regressionRatio fwdAvgSaccade | SSTYPE(3)
  /METHOD=REML
  /PRINT=CORB  SOLUTION
  /RANDOM=INTERCEPT | SUBJECT(subjectNmbr) COVTYPE(VC).
```

Fixed Effects

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	68,000	,016	,900
numFix	1	68,000	2,271	,136
avgTimeFix	1	68,000	,629	,430
regressionRatio	1	68,000	,334	,565
fwdAvgSaccade	1	68,000	2,718	,104

a. speedClass = 0

b. Dependent Variable: response.

Estimates of Fixed Effects[a,b]

Parameter	Estimate	Std. Error	df	t	Sig.	95% ...
						Lower Bound
Intercept	-,061711	,487246	68,000	-,127	,900	-1,033995
numFix	,002639	,001751	68,000	1,507	,136	-,000855
avgTimeFix	-,001189	,001499	68,000	-,793	,430	-,004181
regressionRatio	-,559097	,966860	68,000	-,578	,565	-2,488437
fwdAvgSaccade	,002487	,001508	68,000	1,649	,104	-,000523

Estimates of Fixed Effects[a,b]

Parameter	95% ...
	Upper Bound
Intercept	,910573
numFix	,006132
avgTimeFix	,001803
regressionRatio	1,370243
fwdAvgSaccade	,005496

a. speedClass = 0

b. Dependent Variable: response.

Correlation Matrix for Estimates of Fixed Effects[a,b]

Parameter	Intercept	numFix	avgTimeFix	regressionRatio	fwdAvgSaccade
Intercept	1	-,612	-,535	-,064	-,689
numFix	-,612	1	-,106	-,360	,741
avgTimeFix	-,535	-,106	1	,070	,035
regressionRatio	-,064	-,360	,070	1	-,522
fwdAvgSaccade	-,689	,741	,035	-,522	1

a. speedClass = 0

b. Dependent Variable: response.

Fixed Effects

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	68,000	1,023	,315
numFix	1	68	,007	,932
avgTimeFix	1	68,000	,184	,669
regressionRatio	1	68	,028	,868
fwdAvgSaccade	1	68	,015	,904

a. speedClass = 1

b. Dependent Variable: response.

Estimates of Fixed Effects[a,b]

Parameter	Estimate	Std. Error	df	t	Sig.	95% ... Lower Bound	95% ... Upper Bound
Intercept	,681074	,673348	68,000	1,011	,315	-,662571	2,024718
numFix	,000218	,002543	68	,086	,932	-,004856	,005291
avgTimeFix	-,001541	,003592	68,000	-,429	,669	-,008709	,005628
regressionRatio	,144669	,867727	68	,167	,868	-1,586854	1,876193
fwdAvgSaccade	,000164	,001353	68	,121	,904	-,002536	,002863

a. speedClass = 1

b. Dependent Variable: response.

Correlation Matrix for Estimates of Fixed Effects[a,b]

Parameter	Intercept	numFix	avgTimeFix	regressionRatio	fwdAvgSaccade
Intercept	1	,140	-,793	-,023	-,735
numFix	,140	1	-,613	-,073	,263
avgTimeFix	-,793	-,613	1	-,136	,339
regressionRatio	-,023	-,073	-,136	1	-,343
fwdAvgSaccade	-,735	,263	,339	-,343	1

a. speedClass = 1

b. Dependent Variable: response.

Fixed Effects

Type III Tests of Fixed Effects[a,b]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	19,888	1,058	,316
numFix	1	58,787	4,828	,032
avgTimeFix	1	35,244	,562	,458
regressionRatio	1	52,879	,427	,516
fwdAvgSaccade	1	51,572	,380	,540

a. speedClass = 2

b. Dependent Variable: response.

Estimates of Fixed Effects[a,b]

Parameter	Estimate	Std. Error	df	t	Sig.	95% ... Lower Bound	95% ... Upper Bound
Intercept	,531716	,517021	19,888	1,028	,316	-,547161	1,610592
numFix	,006837	,003112	58,787	2,197	,032	,000610	,013064
avgTimeFix	-,002868	,003825	35,244	-,750	,458	-,010631	,004895
regressionRatio	,524965	,803360	52,879	,653	,516	-1,086457	2,136388
fwdAvgSaccade	-,000507	,000822	51,572	-,617	,540	-,002156	,001143

a. speedClass = 2

b. Dependent Variable: response.

Correlation Matrix for Estimates of Fixed Effects[a,b]

Parameter	Intercept	numFix	avgTimeFix	regressionRatio	fwdAvgSaccade
Intercept	1	,347	-,812	,195	-,584
numFix	,347	1	-,690	,233	,097
avgTimeFix	-,812	-,690	1	-,502	,218
regressionRatio	,195	,233	-,502	1	-,265
fwdAvgSaccade	-,584	,097	,218	-,265	1

a. speedClass = 2

b. Dependent Variable: response.

```
 SPLIT FILE OFF.
MIXED speedClass WITH numFix avgTimeFix regressionRatio fwdAvgSaccade
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.000000000001) HCONVERGE(0,
  /FIXED=numFix avgTimeFix regressionRatio fwdAvgSaccade | SSTYPE(3)
  /METHOD=REML
  /PRINT=CORB  SOLUTION
  /RANDOM=INTERCEPT | SUBJECT(subjectNmbr) COVTYPE(VC).
```

Fixed Effects

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	90,852	20,041	,000
numFix	1	205,987	61,101	,000
avgTimeFix	1	205,945	3,230	,074
regressionRatio	1	204,186	4,486	,035
fwdAvgSaccade	1	202,404	,012	,911

a. Dependent Variable: speedClass.

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% ...
						Lower Bound
Intercept	2,026040	,452574	90,852	4,477	,000	1,127038
numFix	-,012508	,001600	205,987	-7,817	,000	-,015663
avgTimeFix	-,003463	,001927	205,945	-1,797	,074	-,007262
regressionRatio	1,445322	,682364	204,186	2,118	,035	,099938
fwdAvgSaccade	,000100	,000896	202,404	,112	,911	-,001666

Estimates of Fixed Effects[a]

Parameter	95% ...
	Upper Bound
Intercept	2,925042
numFix	-,009353
avgTimeFix	,000336
regressionRatio	2,790705
fwdAvgSaccade	,001867

a. Dependent Variable: speedClass.

Correlation Matrix for Estimates of Fixed Effects[a]

Parameter	Intercept	numFix	avgTimeFix	regressionRatio	fwdAvgSaccade
Intercept	1	-,317	-,628	-,097	-,566
numFix	-,317	1	-,144	-,178	,472
avgTimeFix	-,628	-,144	1	-,167	,112
regressionRatio	-,097	-,178	-,167	1	-,282
fwdAvgSaccade	-,566	,472	,112	-,282	1

a. Dependent Variable: speedClass.

Appendix 4. Syntax and output tables from Wilcoxon's signed-rank tests in SPSS

Below are attached the output of Wilcoxon's signed-rank test for average fixation duration, regression ratio and forward saccades length, when comparing the texts assessed as relevant with the ones assessed as non relevant. The quartiles tables are included as well. The first set of tables refer to the test ran without taking into account the addressed reading speed. The following set of tables refers to the dataset split by addressed speed of reading. In section 6.1 we just mentioned that no statistical significance was found for any of these measures in any of these situations, without reporting the exact results for these tests (contrarily as with the rest of reported tests). This is the reason of including these output tables here.

```
NPAR TESTS
  /WILCOXON=avgTimeFix_NonRelevant regressionRatio_NonRelevant fwdAvgSaccade_NonRelevant WIT
  /STATISTICS QUARTILES
  /MISSING ANALYSIS
  /METHOD=EXACT TIMER(5).
```

NPar Tests

```
[$DataSet] C:\Users\Oswald\Desktop\SPSS\FeaturesTTestGlobal.sav
```

Descriptive Statistics

		Percentiles		
	N	25th	50th (Median)	75th
avgTimeFix_NonRelevant	7	136,5418	166,4970	183,1409
regressionRatio_NonRelevant	7	,2110	,2507	,3099
fwdAvgSaccade_NonRelevant	7	232,6308	237,1605	279,7337
avgTimeFix_Relevant	7	138,7540	157,7240	183,3297
regressionRatio_Relevant	7	,2385	,2515	,2970
fwdAvgSaccade_Relevant	7	198,7546	225,9714	288,3988

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
avgTimeFix_Relevant - avgTimeFix_NonRelevant	Negative Ranks	3[a]	5,00	15,00
	Positive Ranks	4[b]	3,25	13,00
	Ties	0[c]		
	Total	7		
regressionRatio_Relevant - regressionRatio_NonRelevant	Negative Ranks	3[d]	3,33	10,00
	Positive Ranks	4[e]	4,50	18,00
	Ties	0[f]		
	Total	7		
fwdAvgSaccade_Relevant - fwdAvgSaccade_NonRelevant	Negative Ranks	4[g]	3,75	15,00
	Positive Ranks	3[h]	4,33	13,00
	Ties	0[i]		
	Total	7		

a. avgTimeFix_Relevant < avgTimeFix_NonRelevant

b. avgTimeFix_Relevant > avgTimeFix_NonRelevant

c. avgTimeFix_Relevant = avgTimeFix_NonRelevant

d. regressionRatio_Relevant < regressionRatio_NonRelevant

e. regressionRatio_Relevant > regressionRatio_NonRelevant

f. regressionRatio_Relevant = regressionRatio_NonRelevant

g. fwdAvgSaccade_Relevant < fwdAvgSaccade_NonRelevant

h. fwdAvgSaccade_Relevant > fwdAvgSaccade_NonRelevant

i. fwdAvgSaccade_Relevant = fwdAvgSaccade_NonRelevant

Test Statistics[a]

	avgTimeFix_Relevant - avgTimeFix_NonRelevant	regressionRatio_Relevant - regressionRatio_NonRelevant	fwdAvgSaccade_Relevant - fwdAvgSaccade_NonRelevant
Z	-,169[b]	-,676[c]	-,169[b]
Asymp. Sig. (2-tailed)	,866	,499	,866
Exact Sig. (2-tailed)	,938	,578	,938
Exact Sig. (1-tailed)	,469	,289	,469
Point Probability	,063	,055	,063

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

c. Based on negative ranks.

```
NPAR TESTS
  /WILCOXON=avgTimeFix_NonRelevant_Normal avgTimeFix_NonRelevant_Fast avgTimeFix_NonRelevant
fwdAvgSaccade_Relevant_Fast fwdAvgSaccade_Relevant_Skim (PAIRED)
  /STATISTICS QUARTILES
  /MISSING ANALYSIS
  /METHOD=EXACT TIMER(5).
```

NPar Tests

[$DataSet] C:\Users\Oswald\Desktop\SPSS\FeaturesTTestSpeeds.sav

Descriptive Statistics

	N	Percentiles		
		25th	50th (Median)	75th
avgTimeFix_NonRelevant_Normal	7	146,5938	167,2534	193,1214
avgTimeFix_NonRelevant_Fast	7	133,7003	150,8658	181,5441
avgTimeFix_NonRelevant_Skim	7	127,6763	159,7594	172,0029
regressionRatio_NonRelevant_Normal	7	,2245	,2440	,2869
regressionRatio_NonRelevant_Fast	7	,1870	,2429	,3139
regressionRatio_NonRelevant_Skim	7	,2107	,2707	,2949
fwdAvgSaccade_NonRelevant_Normal	7	182,2286	218,6420	226,8624
fwdAvgSaccade_NonRelevant_Fast	7	207,3456	235,1222	313,6951
fwdAvgSaccade_NonRelevant_Skim	7	251,6463	279,4897	318,3987
avgTimeFix_Relevant_Normal	7	139,9585	155,5237	187,7391
avgTimeFix_Relevant_Fast	7	138,4880	157,0468	178,0733
avgTimeFix_Relevant_Skim	7	136,2044	157,6108	192,8217
regressionRatio_Relevant_Normal	7	,2359	,2592	,2683
regressionRatio_Relevant_Fast	7	,2205	,2414	,3002
regressionRatio_Relevant_Skim	7	,2436	,2572	,3362
fwdAvgSaccade_Relevant_Normal	7	178,6427	205,3599	257,5503

Descriptive Statistics

	N	Percentiles		
		25th	50th (Median)	75th
fwdAvgSaccade_Relevant_Fast	7	196,5073	229,7423	308,5451
fwdAvgSaccade_Relevant_Skim	7	218,1933	259,3181	323,3289

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
avgTimeFix_Relevant_Normal - avgTimeFix_NonRelevant_Normal	Negative Ranks	2[a]	6,00	12,00
	Positive Ranks	5[b]	3,20	16,00
	Ties	0[c]		
	Total	7		
avgTimeFix_Relevant_Fast - avgTimeFix_NonRelevant_Fast	Negative Ranks	3[d]	5,00	15,00
	Positive Ranks	4[e]	3,25	13,00
	Ties	0[f]		
	Total	7		
avgTimeFix_Relevant_Skim - avgTimeFix_NonRelevant_Skim	Negative Ranks	3[g]	3,33	10,00
	Positive Ranks	4[h]	4,50	18,00
	Ties	0[i]		
	Total	7		
regressionRatio_Relevant_Normal - regressionRatio_NonRelevant_Normal	Negative Ranks	4[j]	3,75	15,00
	Positive Ranks	3[k]	4,33	13,00
	Ties	0[l]		
	Total	7		
regressionRatio_Relevant_Fast - regressionRatio_NonRelevant_Fast	Negative Ranks	3[m]	4,33	13,00
	Positive Ranks	4[n]	3,75	15,00
	Ties	0[o]		
	Total	7		
regressionRatio_Relevant_Skim - regressionRatio_NonRelevant_Skim	Negative Ranks	3[p]	4,33	13,00
	Positive Ranks	4[q]	3,75	15,00
	Ties	0[r]		
	Total	7		
fwdAvgSaccade_Relevant_Normal - fwdAvgSaccade_NonRelevant_Normal	Negative Ranks	3[s]	4,00	12,00
	Positive Ranks	4[t]	4,00	16,00
	Ties	0[u]		
	Total	7		
fwdAvgSaccade_Relevant_Fast - fwdAvgSaccade_NonRelevant_Fast	Negative Ranks	4[v]	3,50	14,00
	Positive Ranks	3[w]	4,67	14,00
	Ties	0[x]		
	Total	7		
fwdAvgSaccade_Relevant_Skim - fwdAvgSaccade_NonRele	Negative Ranks	4[y]	4,50	18,00
	Positive Ranks	3[z]	3,33	10,00

Ranks

	N	Mean Rank	Sum of Ranks
Ties	0[aa]		
Total	7		

a. avgTimeFix_Relevant_Normal < avgTimeFix_NonRelevant_Normal

b. avgTimeFix_Relevant_Normal > avgTimeFix_NonRelevant_Normal

c. avgTimeFix_Relevant_Normal = avgTimeFix_NonRelevant_Normal

d. avgTimeFix_Relevant_Fast < avgTimeFix_NonRelevant_Fast

e. avgTimeFix_Relevant_Fast > avgTimeFix_NonRelevant_Fast

f. avgTimeFix_Relevant_Fast = avgTimeFix_NonRelevant_Fast

g. avgTimeFix_Relevant_Skim < avgTimeFix_NonRelevant_Skim

h. avgTimeFix_Relevant_Skim > avgTimeFix_NonRelevant_Skim

i. avgTimeFix_Relevant_Skim = avgTimeFix_NonRelevant_Skim

j. regressionRatio_Relevant_Normal < regressionRatio_NonRelevant_Normal

k. regressionRatio_Relevant_Normal > regressionRatio_NonRelevant_Normal

l. regressionRatio_Relevant_Normal = regressionRatio_NonRelevant_Normal

m. regressionRatio_Relevant_Fast < regressionRatio_NonRelevant_Fast

n. regressionRatio_Relevant_Fast > regressionRatio_NonRelevant_Fast

o. regressionRatio_Relevant_Fast = regressionRatio_NonRelevant_Fast

p. regressionRatio_Relevant_Skim < regressionRatio_NonRelevant_Skim

q. regressionRatio_Relevant_Skim > regressionRatio_NonRelevant_Skim

r. regressionRatio_Relevant_Skim = regressionRatio_NonRelevant_Skim

s. fwdAvgSaccade_Relevant_Normal < fwdAvgSaccade_NonRelevant_Normal

t. fwdAvgSaccade_Relevant_Normal > fwdAvgSaccade_NonRelevant_Normal

u. fwdAvgSaccade_Relevant_Normal = fwdAvgSaccade_NonRelevant_Normal

v. fwdAvgSaccade_Relevant_Fast < fwdAvgSaccade_NonRelevant_Fast

w. fwdAvgSaccade_Relevant_Fast > fwdAvgSaccade_NonRelevant_Fast

x. fwdAvgSaccade_Relevant_Fast = fwdAvgSaccade_NonRelevant_Fast

y. fwdAvgSaccade_Relevant_Skim < fwdAvgSaccade_NonRelevant_Skim

z. fwdAvgSaccade_Relevant_Skim > fwdAvgSaccade_NonRelevant_Skim

aa. fwdAvgSaccade_Relevant_Skim = fwdAvgSaccade_NonRelevant_Skim

Test Statistics[a]

	avgTimeFix_Relevant_Normal - avgTimeFix_NonRelevant_Normal	avgTimeFix_Relevant_Fast - avgTimeFix_NonRelevant_Fast	avgTimeFix_Relevant_Skim - avgTimeFix_NonRelevant_Skim	regressionRatio_Relevant_Normal - regressionRatio_NonRelevant_Normal
Z	-,338[b]	-,169[c]	-,676[b]	-,169[c]
Asymp. Sig. (2-tailed)	,735	,866	,499	,866
Exact Sig. (2-tailed)	,813	,938	,578	,938
Exact Sig. (1-tailed)	,406	,469	,289	,469
Point Probability	,063	,063	,055	,063

Test Statistics[a]

	regressionRatio_Relevant_Fast - regressionRatio_NonRelevant_Fast	regressionRatio_Relevant_Skim - regressionRatio_NonRelevant_Skim	fwdAvgSaccade_Relevant_Normal - fwdAvgSaccade_NonRelevant_Normal	fwdAvgSaccade_Relevant_Fast - fwdAvgSaccade_NonRelevant_Fast
Z	-,169[b]	-,169[b]	-,338[b]	,000[d]
Asymp. Sig. (2-tailed)	,866	,866	,735	1,000
Exact Sig. (2-tailed)	,938	,938	,813	1,000
Exact Sig. (1-tailed)	,469	,469	,406	,531
Point Probability	,063	,063	,063	,063

Test Statistics[a]

	fwdAvgSaccade_Relevant_Skim - fwdAvgSaccade_NonRelevant_Skim
Z	-,676[c]
Asymp. Sig. (2-tailed)	,499
Exact Sig. (2-tailed)	,578
Exact Sig. (1-tailed)	,289
Point Probability	,055

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

c. Based on positive ranks.

d. The sum of negative ranks equals the sum of positive ranks.