

Can we detect English proficiency through reading behavior? A preliminary study

Inês Garganta Silva, Carla Teixeira Lopes
INESC TEC and
DEI, Faculty of Engineering, University of Porto
Porto, Portugal
{ei10162,ctl}@fe.up.pt

Maria Ellison
Faculty of Arts, University of Porto
Porto, Portugal
mellison@letras.up.pt

Abstract — If it were possible to automatically detect proficiency in languages using data from eye movements, new levels of customizing computer applications could possibly be achieved. An example in case is web searches where suggestions and results could be adjusted to the user's knowledge of the language. The objective of this study is to compare the reading habits of users with high and low English language proficiency, having in mind the possible automatic detection of the English proficiency level through reading. For this purpose, a study was conducted with two types of user, those with a high level of proficiency (Proficient Users), and those with low proficiency (Basic Users) in the English language. An eye-tracker was used to collect users' eye movements while reading a text in English. Results show that users with high proficiency engage in more careful reading. In contrast, low English proficiency users take more time to read, revisit sentences and paragraphs more often, have more and longer fixations and also a higher number of saccades. As expected, these users have more difficulties in understanding the text.

Keywords – human-computer interaction; eye-tracker; English proficiency; user study.

I. INTRODUCTION

Screen reading has grown considerably in recent years. As an essential daily activity in most people lives, this action has become an object of study. This is visible in a number of investigations that try to comprehend the way we read. This research has numerous goals, from trying to understand types of reader [1], to what font type facilitates or hinders the reading process [2], see [3] for more details.

Eye tracking is the process of measuring either the point of gaze (where one is looking) or the motion of an eye in relation to the position of the head. An eye-tracker is a device for measuring eye positions and eye movements. Consequently, studies may be conducted which involve the collection and analysis of eye movements.

When reading, looking at a picture or landscape, or searching for an object, we continually make eye movements known as 'saccades' [4], which can reach velocities as high as 500° per second. Between saccades, our eyes remain relatively stationary, a state known as fixation which may last about 200-300 ms [4]. Saccades carry the gaze forward in a text. Fixations can be on a word or word group and indicate that the brain is processing the visual information [2]. Longer fixations may indicate a greater effort to understand the text. About 10 to 15 per cent of the saccades are regressions, which are right-to-left

movements along the line, or movements back to previously read lines [4]. According to Reyner "Many regressions tend to be only a few letters long and could be due making too long of a saccade, in which case a short saccade to the left may be necessary for reading to proceed efficiently. Short within-word regressive saccades may also be due to problems that the reader has with processing the currently fixated word. Longer regressions (more than 10 letter spaces back along the line or to another line) occur because the reader did not understand the text." [4].



Figure 1. Key eye-tracking measures (Taken from [8])

This study aims to compare the reading habits of users of low and high proficiency in English language using eye-tracking reading statistics. Our ultimate goal is to determine if it is possible to automatically detect the proficiency of users while reading a text on a computer screen. If this is possible, new levels in customizing computer applications in search engines on the web where suggestions and results could be adjusted to the user's knowledge of the language may be achieved.

We begin this article with a literature review about the use of eye-tracking devices to study reading behavior in different contexts. This is followed by a description of the current study, the main findings, and finally, our conclusions.

II. EYE-TRACKING TO STUDY READING BEHAVIOR

Several studies have been conducted into the reading methods or patterns of native and non-native readers without the use of the eye-tracker device ([9], [10], [11], [12]). These reveal differences in comprehension scores and reading strategies used.

The eye-tracker measures eye positions and movements which allows for the study of the reading pattern of users.

A variety of studies have been carried out using the eye-tracker device which demonstrate that the eye movements are

related to the difficulty of the text passage and inconsistencies in the text [5]; another presents an analysis on how the inclusion of images, which may or not be related to the text, affect online reading [7]; other article presents a study on how font size and type can affect online reading [2]; and further studies analyse the cognitive processes involved in reading tests [8] [13].

The studies mentioned above have different objectives from ours. However, there are others with similar objectives. One of them presents a framework which may be used to detect the level of understanding and proficiency of a reader in a given language using the reading behavior analysis obtained by eye movements. Thus, the hypothesis was that: "A combination of eye-movement features and document characteristics are predictive of language skill and level of understanding." [1]. The authors tested factors such as the number of fixations, time of reading, average time of fixations, average scanpath length and number of regressions. The study consisted of two parts: (1) Analysis based on the proficiency level that the participants claim to have in English; (2) Analysis based on the results of tests to determine the exact English proficiency of the tested participants.

One of the main difficulties encountered in this study was the level that the participants said they had in English, since it did not correspond with the results obtained in the tests. In this way, the authors fail to reach conclusive results.

No significant result emerged for the exact quantification of different English proficiency levels. The authors reached only one significant result in the discrimination between readers with low and high level of understanding. Readers with lower level of understanding had larger average fixation durations and large variance in this duration. They also had smaller saccade median length and fixation mean acceleration.

Another article compares the online reading pattern and comprehension by readers of type L1 (English as their mother language) and readers of type L2 (English as a second language). Three research questions specifically addressed online reading efficiency and effectiveness: "Rapid" – How fast can L1 and L2 readers read?; "Purposeful" – What do L1 and L2 readers read and how much do they read? "Comprehending" – How well do L1 and L2 readers perform on reading comprehension tests?[6]. The results showed that not only are there no differences in understanding the texts between the two groups of readers, but there is also a similarity between the distribution of attention by stimuli and by the content read. The only difference that is important to highlight is the average reading time of the L2 group which is 62% higher than the L1 group [6].

In the current study we use measures that are not tested in these articles and have an *a priori* knowledge of the English proficiency level of the participants which allows us to recruit only those with low and high proficiency levels.

III. USER STUDY

To achieve our goals, we conducted a user study with the Red 120hz eye-tracker from the SensoMotoric Instruments (SMI).

In order to select users for both groups, we asked about 83 students to answer an English proficiency test used and validated by the Faculty of Arts of the University of Porto. This test places users in the Common European Framework (CEF) levels. The set of students was composed by 43 students were enrolled in the 5th year of the Masters programme in Informatics Engineering and 40 students were enrolled in the 3rd year of the under-graduate degree in Information Science.

Based on the results of this test, we selected the basic and proficient users according to the CEF reference. Of the group with low proficiency, 13 Basic Users (grades <10) were recruited and a further 12 Proficient Users (grades >16). All students were Portuguese and all but one (aged 52) are aged between 20 and 25 years and have studied English as a foreign language. Of these 16 have female gender and 9 male gender. The age distribution is showed in Fig. 2.

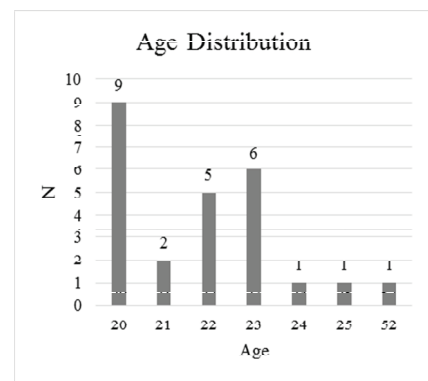


Figure 2. Age Distribution

The experiment was divided into four stages and had no time limit. The goals of the study and how the experiment would be performed were explained to all students. In order to illustrate the procedure of the experiment, we asked users to read a very small text (Fig. 3) on the computer screen to show them what the experience would be like. This allowed us to clarify any doubts users might have.

The third stage was the actual experiment which consisted of asking the user to read a selected text of intermediate difficulty on a computer screen. This text, as well as the first mentioned, were selected from a pedagogical book for English teaching, the "Cambridge First Certificate in English 3 for Updated Exam Student's Book with answers - Examination Papers from University of Cambridge ESOL Examinations", with ISBN: 9780521739306. The Fig. 3 text was withdrawn from page 2 and the Fig. 4 text, entitled "Gary and Me", from page 52.

The second text was formatted with Times New Roman and a font size of 32 pt. It was divided into 3 pages and each page was the size of the screen. The final page of both the experiment test and the real experiment contained only the expression "The End" marking the end of the text. This was used to allow the intervening return to re-read the text if they thought necessary. This way, the experiment was not ended prematurely. After the participant read and reread what they wanted, they would end the experiment by answering a multiple choice questionnaire. This questionnaire was

withdrawn from the same book as the text, the book also provided the questionnaire answers, which were used to correct the responses given by the participants. The questionnaire allow us to assess how well they understood the text.

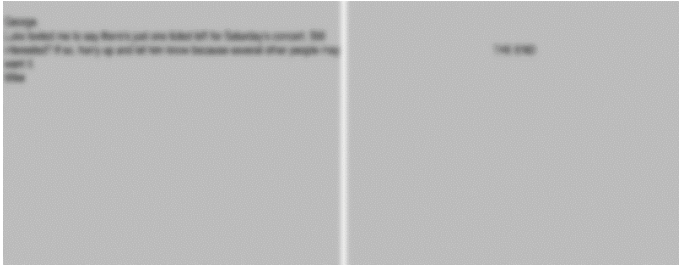


Figure 3. Experiment-test text (intentionally blurred for copyright issues).



Figure 4. Real experiment text (intentionally blurred for copyright issues).

IV. RESULTS AND ANALYSIS

For the analysis, were considered twenty-nine reading metrics. They were divided in full text metrics and paragraph/phrase metrics, allowing to study the reading patterns in the all text and also by paragraphs and phrases.

The metrics studied can be seen in Tables I and II.

TABLE I. PARAGRAPH AND PHRASE METRICS

Category	Description
<i>Revisits</i>	The revisits are the number of revisits to the paragraphs/phrases.
<i>First Fixation Duration</i>	The first fixation duration is the duration of the first look at a paragraph or phrase measure in ms.
<i>Average Fixation Duration</i>	The average fixation duration is the average fixation time in ms.
<i>Dwell Time</i>	The dwell time is the sum of durations from all fixations and saccades that hit the paragraphs/phrases measures in ms.
<i>Glance Duration</i>	The glance duration is the sum of the dwell time plus the duration of the entering saccade measure in ms.

TABLE III. FULL TEXT METRICS

Category	Metric	Description
<i>End Time</i>		The end time is the total time of the experiment in ms.
<i>Fixation</i>	<i>Count</i>	The fixation count counts the total number of fixations.
	<i>Frequency</i>	The fixation frequency measures the number of fixations per second.
	<i>Duration</i>	The fixation duration measures how long a fixation lasts in ms. Analysed statistics: sum, average, maximum and minimum.
	<i>Dispersion</i>	The fixation dispersion measures the length of the distribution path in px. Analysed statistics: sum, average, maximum and minimum.
<i>Saccade</i>	<i>Count</i>	The saccade count counts the total number of saccades.
	<i>Frequency</i>	The saccades frequency measures the number of saccades per second.
	<i>Duration</i>	The saccade duration measures how long a saccade lasts in ms. Analysed statistics: sum, average, maximum and minimum.
	<i>Amplitude</i>	The saccade amplitude measures the amplitude of the saccades in degrees. Analysed statistics: sum, average, maximum and minimum.
	<i>Velocity</i>	The saccade velocity measures velocity of the saccades in degrees per second. Analysed statistics: sum, average and maximum.
	<i>Latency</i>	The saccade latency measure the average latency of the saccades in ms.
<i>Scanpath</i>		The scanpath is the total length of the eye path in px.

To collect and analyse the data of the experiment we used the SMI Experiment Center 3.5 and the SMI BeGaze 3.5. The system first calibrates the eye-tracker to the subject using 5 points.

To determine if the differences between the two groups found were significant or not, we first checked if we could apply a parametric test. We applied the Shapiro-Wilk test to see if the groups followed a normal distribution, and the Levene test to analyze the homogeneity of variances. If the data met the assumptions of a parametric test, we applied the two independent sample t-test. If not, we applied the Mann-Whitney test. Any result with a p-value under 0.05 was considered significant.

Differences between the readings of the two groups were found. Only significant results are presented in the tables; the graphs show both significant (in green) and non-significant results.

Fig. 5 is a Heat Map, which displays a range of colors from grey through to green and to red to represent the aggregate amount of time a participant spends focusing on a particular area of the input [8]. Red represents an area with a lot of frequency (hot) and grey means a low or even zero frequency (cold). As can be seen in the figure, the Proficient Users group has a more homogeneous and embracing reading of the text (wider green, yellow and red zones) than the Basic Users group, where we can find cold zones in the text area (grey and blue zones).

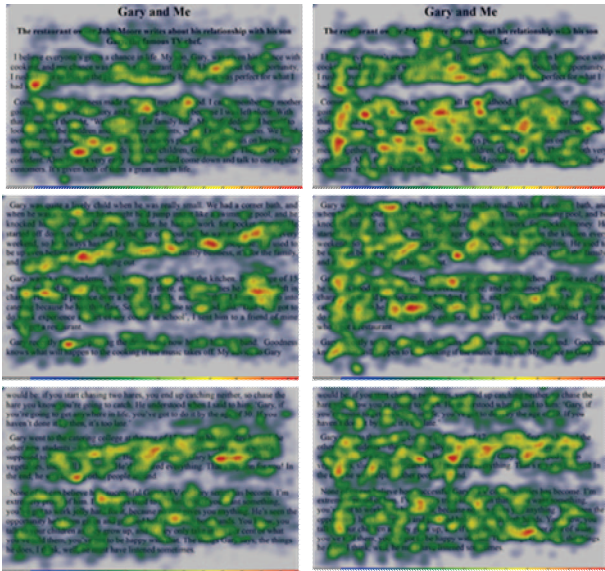


Figure 5. Heat Map of the stimulus. The left images are from the group of Basic Users and the ones on the right are from the Proficient Users' group.

In Table III we can see that the Basic Users group has a total reading time of about 3 times higher than the Proficient Users. It also shows that the Basic Users group had more fixations and spent twice as much time on these. This indicates that the former had more difficulties in processing information than the latter group. Regarding the fixation duration maximum and the fixation dispersion total metrics, we found that the Basic Users group needed more time and had more difficulty in processing information than the Proficient Users group. The results of the number of saccades and their total velocity indicates that the Basic Users group wandered over the text and reread more parts (more regressions).

TABLE III. SIGNIFICANT DIFFERENCES BETWEEN GROUPS

Measures	Basic Users	Proficient Users	P-value
<i>End Time [ms]</i> **	391533	123909	6,15E-04
<i>Fixation Count</i> **	16591	10134	7,88E-03
<i>Fixation Frequency [count/s]</i> **	2,9	3,6	4,80E-03
<i>Fixation Duration Total [ms]</i> **	4489068,4	2212148,7	2,25E-03
<i>Fixation Duration Average [ms]</i> *	270,6	218,3	4,34E-02
<i>Fixation Duration Maximum [ms]</i> **	3683,2	1100	6,61E-04
<i>Fixation Dispersion Total [px]</i> *	1208710,7	685888	4,78E-03
<i>Fixation Dispersion Average [px]</i> *	72,9	67,7	4,88E-02
<i>Saccade Count</i> **	16962	10260	7,54E-03
<i>Saccade Frequency [count/s]</i> *	3	3,7	1,99E-02
<i>Saccade Duration Total [ms]</i> *	738658,4	429603,9	1,21E-02
<i>Saccade Amplitude Total [A°]</i> *	116760,2	62696,5	1,46E-02
<i>Saccade Velocity Total [A°/s]</i> **	1667480,6	1017755,4	9,39E-03
<i>Saccade Latency Average [ms]</i> *	290,2	224,7	1,14E-02

(*p-value <0.05; **p-value <0.01)

In this study part the measures were analysed by paragraph (Table IV) or phrase (Table V) instead of considering the complete text.

It can be observed that in both cases all variables are higher in the Basic Users group. This indicates that the Basic Users group took longer on each of the text components and also had a higher number of revisits in these. These results corroborate the results shown in Table III.

TABLE IV. RESULTS WITH STATISTICAL SIGNIFICANCE [PARAGRAPHS]

Measures	Basic Users	Proficient Users	P-value
<i>Glance Duration [ms]</i> **	23960,11	10971,52	1,27E-03
<i>Revisits</i> **	31,59	16,48	2,15E-05
<i>Average Fixation [ms]</i> **	3000,56	1918,75	4,00E-03
<i>Dwell Time [ms]</i> **	22432,67	10150,96	2,01E-03

(*p-value <0.05; **p-value <0.01)

TABLE V. RESULTS WITH STATISTICAL SIGNIFICANCE [PHRASES]

Measures	Basic Users	Proficient Users	P-value
<i>Glance Duration [ms]</i> **	23207,24	10692,92	3,25E-15
<i>First Fixation Duration [ms]</i> *	3015,42	1960,48	3,53E-02
<i>Revisits</i> **	33,03	17,66	2,2E-16
<i>Average Fixation [ms]</i> **	3012,66	1951,23	1,33E-09
<i>Dwell Time [ms]</i> **	21600,51	9832,27	6,17E-14

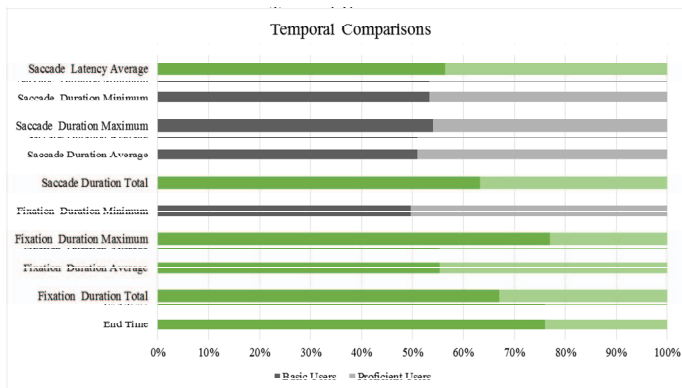
(*p-value <0.05; **p-value <0.01)

The results of the questionnaire (Table VI) have statistical significance. This shows that though the Basic Users group, took longer to read and reread the text, they did not understand well.

TABLE VI. AVERAGE, MAXIMUM AND MINIMUM GRADES OF EACH GROUP

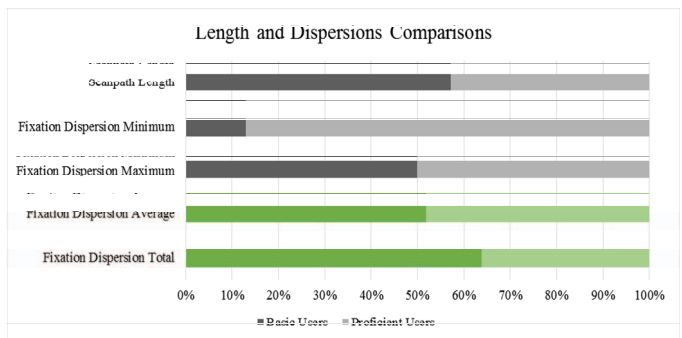
Measures	Basic Users	Proficient Users	P-value
Grades Average	7,14	15,56	9.023E-06
Grades Maximum	10	20	NA
Grades Minimum	0	13,33	NA

Graph 1 shows that the Basic Users group always had a larger length of time in the studied factors than the Proficient Users group (except for the fixation duration minimum).



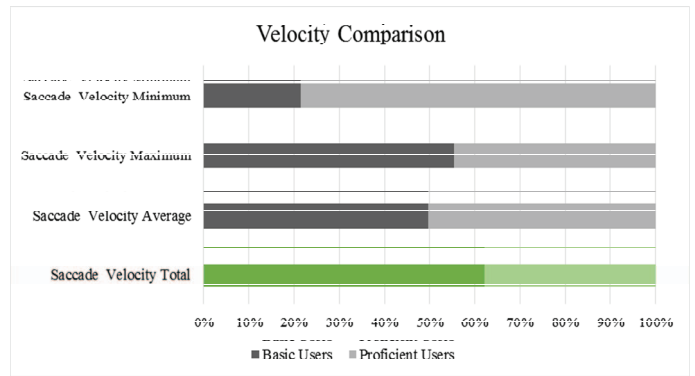
Graphic 1. Temporal Comparisons

Graph 2 compares the scan path length between the two groups. Here we see that although this result is not statistically significant, it is considerably higher for the Basic Users group indicating that they reread the text more times.



Graph 2. Length and Dispersion Comparisons

Graph 3 illustrates the data collected from the saccades and reveals that although the average velocity of the saccades is similar in both groups, the total velocity is significantly different, which is superior for the Basic Users group, as already stated above.



Graph 3. Velocity Comparisons

V. CONCLUSIONS

In this study we compared the reading habits in English for users with low and high proficiency in this language. The heat map shows that the Proficient Users demonstrates more careful reading, since the heat pattern is better distributed through the text. Regarding total time reading, the Basic Users group has a value about 3 times that of the Proficient Users group. The number and average duration of the fixations are, once again, bigger for the Basic Users group, showing a major difficulty in text processing. Regarding the number of saccades, this is highest for the Basic Users group, which indicates, possibly, a bigger number of regressions in reading. The time pauses and revisits to each paragraph and sentences in the text, bigger by the Basic Users group corroborate the previous results. Finally, the result from the questionnaire (which served to test the level of understanding of the text read), revealed a large discrepancy in the difficulty of text comprehension by the Basic Users group compared to the Proficient Users group.

From this study it is concluded that there are significant differences in the reading patterns among users with high proficiency and low proficiency in English. This is preliminary work which contributes to the automatic detection of language proficiency through screen reading. In the future, we would like to conduct this study with a larger sample and analyze how well the features associated with the significant differences can be used to predict proficiency in English.

ACKNOWLEDGMENT

This work was partially supported by the Portuguese government, through the National Foundation for Science and Technology – FCT (Fundação para a Ciência e Tecnologia), the European Union (COMPETE, QREN and FEDER) through the project REC I/EEI-SII/0360/2012 entitled “MASSIVE - Multimodal Acknowledgeable multisensorial Immersive Virtual Environments” and by the Master in Informatics and Computing Engineering of the Faculty of Engineering of the University of Porto.

REFERENCES

- [1] P. Martínez-Gómez and A. Aizawa, “Recognition of understanding level and language skill using measurements of reading behavior,” Proceedings of the 19th international conference on Intelligent User Interfaces - IUI '14, pp. 95–104, 2014.

- [2] D. Beymer, D. Russell, and P. Orton, "An Eye Tracking Study of How Font Size and Type Influence Online Reading," *Br. Comput. Soc.*, pp. 15–18, 2008.
- [3] S. Schroeder, J. Hyönä, and S. P. Liversedge, "Developmental eye-tracking research in reading: Introduction to the special issue," *J. Cogn. Psychol.*, vol. 27, no. 5, pp. 500–510, 2015.
- [4] K. Rayner, "Eye movements in reading and information processing: 20 years of research.," *Psychol. Bull.*, vol. 124, no. 3, pp. 372–422, 1998.
- [5] K. Rayner, K. H. Chace, T. J. Slattery, and J. Ashby, "Eye Movements as Reflections of Comprehension Processes in Reading," *Sci. Stud. Read.*, vol. 10, no. November, pp. 241–255, 2006.
- [6] H. Kang, "Understanding online reading through the eyes of first and second language readers: An exploratory study," *Comput. Educ.*, vol. 73, pp. 1–8, 2014.
- [7] D. Beymer, D. M. Russell, and P. Z. Orto, "An Eye Tracking Study of How Pictures Influence Online Reading," *INTERACT 2007*, pp. 456 – 460, 2007.
- [8] T. Brunfaut and G. McCray, "Looking Into Test-takers' Cognitive Processes While Completing Reading Tasks: A mixed-method eye-tracking and stimulated recall study" 2015.
- [9] J. E. Brisbois, "Connections Between First- and Second-Language Reading" *J. Read. Behav.*, vol. 27, no. 4, pp. 565–584, 1995.
- [10] R. Sheorey and K. Mokhtari, "Differences in the metacognitive awareness of reading strategies among native and non-native readers," *Elsevier Sci. Ltd.*, vol. 29, pp. 431–449, 2001.
- [11] H. Huang, C. Chern, and C. Lin, "EFL learners' use of online reading strategies and comprehension of texts: An exploratory study," *Comput. Educ.*, vol. 52, no. 1, pp. 13–26, 2009.
- [12] C. Hill, "What can Teachers do to Improve Reading Comprehension?: An Examination of Second Language Reading Research and Implications for English Language Teaching Practices," *Arbutus Rev.*, vol. 2, no. 1, pp. 62–74, 2011.
- [13] S. Bax, "Readers' cognitive processes during IELTS reading tests: evidence from eye tracking," *ELT reserach Pap.*, no. retrieved from www.britishcouncil.org, 2013.