Estimation and Visualization of Webcam Eye Tracking for Text Reading

Anastasiia Grynenko[1], Olena Turuta[1], Ruslan Kasheparov[1], Olga Kalynychenko[1] and Oleksii Turuta[1]

[1] Kharkiv National University of Radio Electronics, Nauky Ave. 14, Kharkiv, 61166, Ukraine

Abstract

With increasingly popular eye-tracking research, we wondered if it was necessary to use special equipment – eye trackers – for such experiments. Webcams are inexpensive and everyone has them at home. But are they precise enough to be used in text reading experiments? We tested whether an ordinary webcam can determine where the reader's eyes are located when reading text from a screen. We found that a webcam can be used to detect a line at certain text sizes and line spacing. However, the webcam is not suitable for capturing individual words or letters. This requires more accurate eye trackers.

Keywords

Eye-tracking, gaze detection, NLP, Corpus, Artificial Intelligence Ethics, Diagnostic Accuracy

1. Introduction

Eye-tracking is the measurement of eye movements and determination of the gaze, which is a location where the person is looking. Eye movement has been of interest to scientists since the 19th century. One of the first important milestones was the discovery by Louis Émile Javal that reading is not a smooth movement over a text, but a series of short stops (fixations) and quick jumps (saccades). Since the 1970s, research on gaze tracking in reading has gained popularity, especially the Rayner study [1]. An important idea is the hypothesis put forward in 1980 by Just and Carpenter about the strong eye-mind, according to which what the gaze is fixed on is processed [2]. There are also other hypotheses, in particular the immediacy hypothesis, according to which the eyes do not move until all processing is completed. Moreover, there are opposing ideas that eye movements do not really reflect moment-by-moment cognitive processing demands during reading. In any case, scientists agree that gaze tracking can provide important data about reading.

Initially eye movement was studied to determine its role in reading and language comprehension. Later psycholinguists began to use this technology to organize the process of language learning for different social groups, to improve cross-cultural communication in business, as well as for many theoretical questions about understanding how people perceive language. Eye-tracking is not only used in reading text, but also in listening to music, typing, and visual search. There are many applications of this technology [3].

2. Relevance

The technology is widely used in education, both in the form of student eye tracking (tasks to determine student engagement, academic performance, and skills assessment) and in the form of teacher eye tracking (professional eye tracking). In addition, there is an idea to explore the methodology of performing tasks with eye tracking with the possibility of turning them into

CEUR Workshop Proceedings (CEUR-WS.org)

educational material. Eye movement has also helped in the training of some professions, such as pilots, by developing methods for modeling complex realized events, detecting errors, and assessing consequences.

Marketing, human-computer interaction, and neuroergonomics are also well-known applications of eye tracking for market research, audience research, as an input device, and for adaptation in the workplace, respectively [4]. There are ideas for using tracking data to improve NLP models, to evaluate them, and to explore the possibility of applying the models to other languages and problems [5-8]. There are also many other applications of gaze tracking systems, such as research on perception and cognitive processes, lie detection [9-10], and detection of deepfakes [11-12].

Another interesting use of the system is to diagnose a person's mental state, which is especially relevant in times of full-scale war. An example of the realization of such an idea is the Ukrainian startup Anima [13]. Indeed, many Ukrainians, especially the military, suffer from constant stress that puts a strain on their psychological state. It is important to constantly monitor and prevent possible deterioration of psychological health in order to understand the readiness of the military for combat. Gaze tracking systems can help with this, as a shifted distribution of attention to stimuli of different emotional colors may indicate certain deviations. This idea can be expanded to diagnose various other mental illnesses, as there are already publications describing ways to diagnose them by studying a person's gaze.

3. Eye-tracking features

Calibration is necessary to adapt the system algorithm before applying it to the person sitting in front of the eye-tracker.

Some systems do not need calibration, but most still require it. Calibration is possible when the point the person is looking at is known. Mobile device calibrations vary, and their instructions are in the manuals. For a screen-based calibration, you need to focus your gaze on the targets shown on the screen. The more targets in different parts of the screen are analyzed, the more accurate the system will be, but this can be time consuming and inconvenient for some experiments.

In our experiment, we want to collect data that can be useful to different professionals interested in the application of eye tracking in their field, and to evaluate the quality and accuracy of the resulting system.

Since in our experiment we will not use special high-precision equipment that may not require calibration, and the format will be screen-based, we will use points from different parts of the screen for calibration, namely 3 rows of 3 points, with one point appearing in the center of the screen and the rest in the perimeter of the screen with 4 points in the corners.

4. Datasets

Eye-tracking reading experiments often use fixations (place, duration) and saccades (length, duration, start and end point) as captured primitives. For more detailed experiments, the number of fixations per 100 words, whether they are progressive or regressive, and the frequency of regressions are observed.

There are eye-tracking datasets with 2 different types of input data: the original video with the specified characteristics of the experiment and already processed format, where with the original parameters of the eyes the resultant point on the screen is indicated [14].

Two methods are used for text experiments: rapid serial visual presentation (RSVP), in which the words are presented at a set rate in the same location, and self-paced reading, in which only several words are presented on the screen at a time and readers advance the text by pushing a button. There are variations of the self-paced method. Some experiments include the ability to look back in the text. Some display only one word, while others display several at once. Sometimes the words are in the same place, and sometimes they are presented spatially, as in a normal text. There are also studies that consider a text as a time-series process [15] or based on an attention mechanism [16].

As for textual datasets, they often consist of a single sentence and are monolingual, although there are also longer texts and multilingualism available, such as GECO [17] and MECO [18] datasets [19]. It is also important to note that most of the existing datasets are in English. The Ukrainian language is considered to be a low resource language with very few existing datasets [20]. Therefore, the text in Ukrainian will be used in our experiment.

5. Hardware

Hardware differs by specialty, human interface, scope of tracking, and technical characteristics.

Special trackers refer to any additional gaze-detection hardware, when non-special is a standard webcam. The interface can be head-stabilized, remote, and mobile (also called "head-mounted"). In terms of tracking area, devices are differentiated into those that use a computer screen as the stimulus area; those that can operate in more complex geometries, such as a multi-screen booth; and those that can operate in the real world. Technical characteristics include accuracy, sampling rate, resolution, etc.

We divided trackers by human interface, additionally adding non-special devices (webcams). We get the following groups: head-stabilized, remote, mobile and non- special.

Head-stabilized eye tracking is usually much more accurate than other types of trackers, so they are used in neurophysiological experiments, where participants' comfort is less important than system accuracy and precision, as well as in experiments with animals. However, such trackers are uncomfortable and not designed for experiments where natural responses with head movements are important.

Remote systems consist of a camera and an infrared source. Most often such cameras are mounted under the screen, as the pupil is more visible from that position due to the shape of the eye and eyelids. Remote eye tracking is useful in experiments with infants and in natural interactions. Limitations for such trackers are a fixed working area beyond which gaze tracking is difficult, the tolerance of head movements affecting the inaccuracy of results, sunlight reflected in participants' eyes, and the problem of capturing more than one participant.

Mobile gaze tracking is often in the form of goggles that include a scene recording camera, gaze capture cameras, and illuminators. Such devices are a great option for real-world research. There is the problem of sunlight, but it is solvable. In addition there is a problem of the difficulty of tracking the gaze at the periphery and the use of a relative coordinate system that differs greatly from participant to participant.

Non-special equipment does not have infrared beams or a head stabilizer, so its effectiveness is orders of magnitude lower. However, a significant advantage is the popularity of web cameras.

As a result of comparing different hardware, it was concluded that greater accuracy is given by special devices, especially with a head lock. But since webcams are much more common and do not require additional costs, we would like to investigate camera accuracy. The camera used in the experiment has a frame rate of 30 FPS and a resolution of 1280x720.

6. Software

These days, there are various programs that allow eye-tracking, both with special eye trackers and with general webcams [21]. Our focus was on those of them that can be applied to general webcams. A comparison of different eye-tracking software is shown in Table 1.

Table 1
Comparative characteristics of the software

Name	Format	Programming language	Detection objects	Accessibility
PyGaze	Toolbox for eye-tracking	Python	Coordinates on the screen, position and duration of fixations, saccades	Open source
WebGazer.js	Web application	JavaScript	Coordinates on the screen	Open source
PyGazeAnalyser	Toolbox for analysis and plotting eye-tracking data	Python	Position and duration of fixations, saccades, than fixation map, scanpath and heatmap	Open source
xLabs	Browser extension for Google Chrome	JavaScript, C++	Coordinates on the screen	No longer supported
OGAMA	Windows desktop application	C#.NET	Coordinates on the screen, position and duration of fixations, saccades	Open source, no longer supported

Although some software can detect, for example, saccades, it can only be achieved using special equipment, because the frequency of the web-cameras is not sufficient for this.

Having compared and worked with the above-mentioned programs, it was decided to use WebGazer.js for the experiment. Although this software works relatively roughly and cannot capture saccades and fixations, due to its well-developed community and available documentation it is excellent for conducting small experiments.

7. Data preparation

In preparation for an upcoming experiment, the task at hand involves the careful and meticulous preparation of data. To accomplish this, we have developed a specialized service designed to generate custom images with specific parameters. These parameters include the picture resolution, aimed at optimizing the display quality on the user's screen, as well as the desired text, font style, font size, spacing between words, spacing between sentences, distance to the screen border, and the desired position of the text on the screen.

By leveraging our innovative service, a seamless process is established for creating a high-quality PNG image that precisely adheres to the specified parameters. Additionally, a corresponding JSON file is generated, which contains metadata that comprehensively describes the various parameters of the image. This includes vital information such as the exact placement coordinates for each word within the image.

By providing this comprehensive and meticulously detailed output, our service ensures that the necessary data is readily available for the future experiment, enabling a smooth and accurate analysis of the acquired information.

The final generated image contains the text itself and additional borders. The word boundaries calculated by the program according to the text metadata are displayed in green.

Figure 1 shows the generated final image, using the data taken from the metadata.

Figure 1: The generated final image, using the data taken from the metadata
The user perceives the image without the boundaries of individual words.

8. Experiment description

The primary objective of this paper is not to conduct a comprehensive study on eye movement during text-related tasks. Instead, its main focus is to introduce a system that can potentially facilitate such experiments. The paper includes an experiment that serves two purposes: firstly, to showcase the viability of employing an eye-tracking system, and secondly, to examine the feasibility of substituting costly eye trackers with a simpler webcam solution.

The goal of the experiment is to determine the extent to which webcams can be useful using WebGazer.js to determine the position in the text, where the participant focuses. For this purpose, Ukrainian text was taken, which was projected on a 14" FullHD screen with different font sizes and line spacing. The result of the experiment was to select the optimal ratio of text parameters.

At the stage of preparation for the experiment 16 variants of text with Calibri font and font size from 24 pt to 32 pt with the step of 4 pt, and with the line spacing from 1.5 to 3 with the step of 0.5 were formed.

The experiment involved people with different eye colors without special features (e.g., glasses). Each participant had to wait for the camera to capture their face and then underwent calibration at 9 points after starting the app. During the calibration, each point had to be clicked 5 times while looking at it. After a successful calibration, the system's accuracy was calculated when looking at the center point and displayed on the screen. The system-determined point of view was displayed as a blue dot. The text parameters could be changed from the top menu bar. During the experiment, attention was paid to how accurately the place where the gaze was directed could be determined. For this purpose, hits in a word were examined.

Percentages of hits in the word was calculated using the following equation

$$P = \frac{N_H}{N_A},\qquad(1)$$

where N_H is the fraction of hits, i.e. the number of hits on the container that contains the word, and N_A is the number of the number of times the container was missed calculated for 5 seconds of gaze. The hit or miss accuracy is evaluated at regular intervals of 10 milliseconds.

To activate the accuracy calculation function, we used the right mouse click at the place where the gaze is actually directed.

9. Heat maps

Heat maps in eye tracking play a crucial role in visualizing and analyzing gaze focus data. By utilizing color intensity, heat maps provide a graphical representation that indicates the level of eye attention in specific areas of an image or screen. This enables researchers and practitioners to gain insights into the patterns of visual exploration and focus during various tasks.

Real-time display of eye movements is particularly valuable in tasks that require immediate feedback or interaction. The inclusion of real-time heat maps allows for the visualization of the current gaze location on the screen, providing instant visual feedback to both the user and the experimenter.

One of the advantages of heat maps is their ability to accumulate information over time. By highlighting areas on the screen where the gaze was directed for longer durations with stronger and more noticeable colors, heat maps effectively capture the salient regions of interest. This feature aids in reducing noise or transient eye movements, as the emphasis is placed on areas that received sustained attention.

Heat maps can be employed in various domains, including usability testing, user experience research, website optimization, and advertising analysis. They provide a valuable tool for understanding visual attention patterns, optimizing user interfaces, and enhancing the overall user experience.
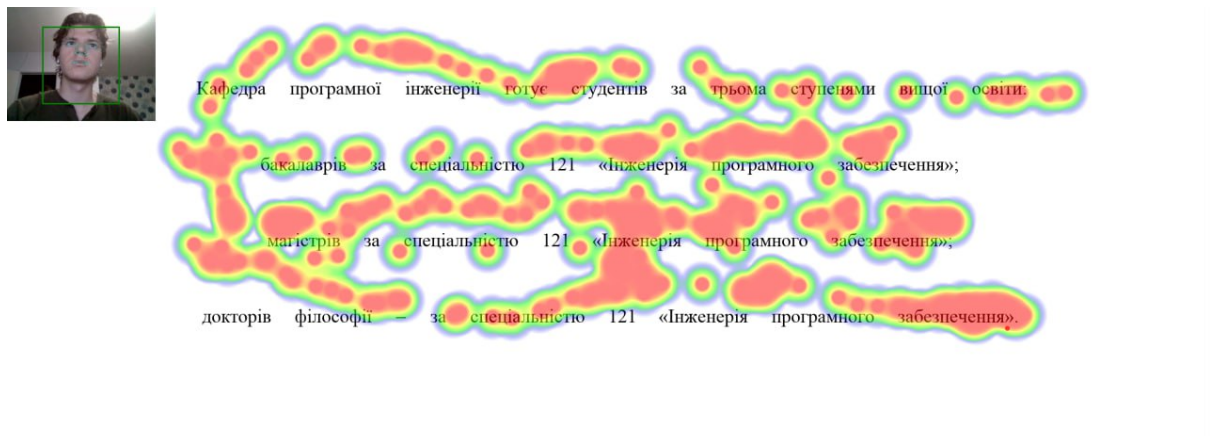
Figure 2 shows the heat map display.

Figure 2: Screenshot of the heat map

The image clearly demonstrates the informative nature of heat maps, as they vividly depict the visual information regarding the focus of attention.

10. Post-processing of received data

During the experiment, the program collects important data about the user, in particular the coordinates of his gaze, namely the x and y coordinates and the corresponding timestamps. This data is then organized and stored in a structured format in a CSV (Comma-Separated Values) file. The collected data is then transmitted to a server for further processing and analysis.

11. Results

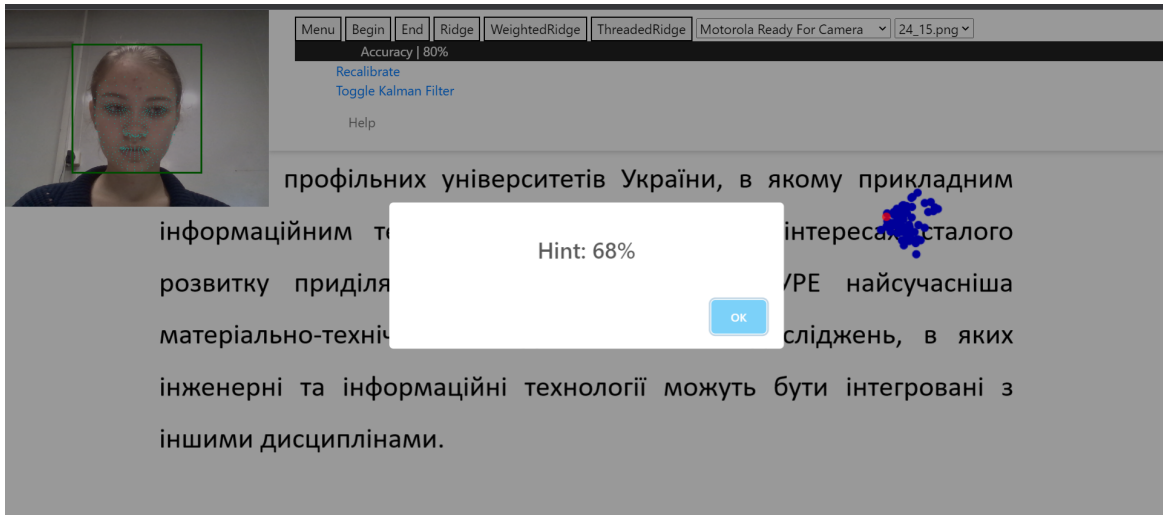Figure 3 shows the application during the accuracy assessment at a particular location.

Figure 3: Screenshot of the application during the experiment

Based on the results of the participation of several people who tested the accuracy of the three places in the text for each set of parameters, a table 2 was constructed with line spacing on the rows and font size on the columns, where their intersection shows the percentage of hits in the line. It is worth mentioning that the values were calculated by taking the arithmetic mean of accuracy across all participants.

Table 2
Experiment results

\ Font size Line spacing	24	28	32
1.5	53	60	16
2.0	64	71	31
2.5	69	64	87
3.0	41	67	84

Despite the expected linearity of the data, we got a different result. We can explain it by the inaccuracies of the program, lack of experiments and by the internal processes of the brain, which lead to uncontrolled eye movements.

According to the results of the experiment, a word can be caught with a webcam with a font size not less than 32 pt and a line spacing not less than 2.5 with an accuracy more than 84%, but to determine the letter that the participant is looking at, the webcam is not enough.

12.Conclusion

The results of the experiment showed that a webcam can be used to recognize a line that reads text. However, compared to the results that can be obtained using special equipment, webcams do not work accurately. The plan is to experiment with more accurate special trackers in order to separate the letter at which the gaze is directed with greater accuracy.

The experiment yielded a discovery regarding the feasibility of utilizing webcams for eye-tracking in the context of reading texts on general devices. This finding highlights a crucial advancement in the field, as it demonstrates the successful adaptation of commonly available webcams for capturing and analyzing eye movement data. The ability to utilize webcams for eye-tracking opens up new possibilities for researchers and practitioners, as it eliminates the need for specialized and expensive eye-tracking equipment. This not only expands the accessibility of eye-tracking technology but also presents opportunities for conducting eye-tracking studies on a larger scale and in diverse settings.

The implications of this finding extend beyond the scope of the experiment and have the potential to shape future research methodologies and applications in the field of eye-tracking.

Acknowledgements

References

[1] Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. Psychological Bulletin, 124(3), 372–422.

[2] Just, M. A., & Carpenter, P. (1980). A theory of reading: From eye fixations to comprehension. Psychological Review, 85, 109-130.

[3] Krakowczyk, D. G., Reich, D. R., Chwastek, J., Jakobi, D. N., Prasse, P., Süss, A., . . . Jäger, L. A. (2023). Pymovements: A python package for eye movement data processing. Paper presented at the Eye Tracking Research and Applications Symposium (ETRA), doi:10.1145/3588015.3590134

[4] Peysakhovich, Vsevolod & Lefrançois, Olivier & Dehais, Frédéric & Causse, Mickael. (2018). The Neuroergonomics of Aircraft Cockpits: The Four Stages of Eye-Tracking Integration to Enhance Flight Safety. Safety. 4. 8. 10.3390/safety4010008. https://www.researchgate.net/publication/323441671_The_Neuroergonomics_of_Aircraft_Coc kpits_The_Four_Stages_of_Eye-Tracking_Integration_to_Enhance_Flight_Safety

[5] Hollenstein, N., Barrett, M., & Beinborn, L. (2020). Towards Best Practices for Leveraging Human Language Processing Signals for Natural Language Processing. LINCR. https://drive.google.com/file/d/1FxZso4wgjz2PFrKsZC7Elb-L5PXYZdEJ/view

[6] Erdem, Erkut & Kuyu, Menekşe & Yagcioglu, Semih & Frank, Anette & Parcalabescu, Letitia & Plank, Barbara & Babii, Andrii & Turuta, Oleksii & Erdem, Aykut & Calixto, Iacer & Lloret, Elena & Apostol, Elena Simona & Truică, Ciprian-Octavian & Šandrih Todorović, Branislava & Martinčić-Ipšić, Sanda & Berend, Gábor & Gatt, Albert & Korvel, Gražina. (2022). Neural Natural Language Generation: A Survey on Multilinguality, Multimodality, Controllability and Learning. Journal of Artificial Intelligence Research. 73. 1131-1207. 10.1613/jair.1.12918.

[7] Hahn, M., & Keller, F. (2016). Modeling Human Reading with Neural Attention. Conference on Empirical Methods in Natural Language Processing. https://aclanthology.org/D16-1009.pdf

[8] Barkovska, O., Pyvovarova, D., Kholiev, V., Ivashchenko, H., & Rosinskyi, D. (2021). Information object storage model with accelerated text processing methods. Paper presented at the CEUR Workshop Proceedings, , 2870 286-299.

[9] Fang, X., Sun, Y., Zheng, X., Wang, X., Deng, X., & Wang, M. (2021). Assessing Deception in Questionnaire Surveys With Eye-Tracking. Frontiers in Psychology, 12. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.774961/full

[10] Ge, F., Yang, X.Q., Chen, Y.X., Huang, H.L., Shen, X.C., Li, Y., & Hu, J.M. (2020). Application of Eye Tracker in Lie Detection. Fa yi xue za zhi, 36 2, 229-232. https://pubmed.ncbi.nlm.nih.gov/32530172/

[11] Gupta, P., Chugh, K., Dhall, A., & Subramanian, R. (2020). The eyes know it: FakeET- An Eye-tracking Database to Understand Deepfake Perception. Proceedings of the 2020 International Conference on Multimodal Interaction. https://arxiv.org/pdf/2006.06961.pdf

[12] Demir, I., & Ciftci, U.A. (2021). Where Do Deep Fakes Look? Synthetic Face Detection via Gaze Tracking. ACM Symposium on Eye Tracking Research and Applications. https://arxiv.org/pdf/2101.01165.pdf

[13] Neurobiological test of mental state Anima. https://ua.anima.help/

[14] K.Krafka*, A. Khosla*, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik and A. Torralba Eye Tracking for Everyone. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016 rXiv, abs/1606.05814.

[15] A. Yerokhin, O. Turuta, A. Babii, A. Nechyporenko and I. Mahdalina, "Usage of phase space diagram to finding significant features of rhinomanometric signals," 2016 XIth International Scientific and Technical Conference Computer Sciences and Information Technologies (CSIT), 2016, pp. 70-72, doi: 10.1109/STC-CSIT.2016.7589871.

[16] Brandl, S., & Hollenstein, N. (2022). Every word counts: A multilingual analysis of individual human alignment with model attention. ArXiv, abs/2210.04963.

[17] Cop, U., Dirix, N., Drieghe, D. et al. Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. Behav Res 49, 602–615 (2017). https://doi.org/10.3758/s13428-016-0734-0.

[18] The Multilingual Eye-tracking Corpus (MECO) https://meco-read.com/category/data-news/ (Siegelman et al., 2022; Kuperman et al., 2022).

[19] D. Dashenkov, K. Smelyakov and O. Turuta, "Methods of Multilanguage Question Answering," 2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T), 2021, pp. 251-255, doi: 10.1109/PICST54195.2021.9772145.

[20] Panchenko, D., Maksymenko, D., Turuta, O., Luzan, M., Tytarenko, S., Turuta, O. (2022). Ukrainian News Corpus as Text Classification Benchmark. In: , et al. ICTERI 2021 Workshops. ICTERI 2021. Communications in Computer and Information Science, vol 1635. Springer, Cham. https://doi.org/10.1007/978-3-031-14841-5_37.

[21] Punde, P.A.; Jadhav, M.E.; Manza, R.R. A study of eye tracking technology and its applications. In Proceedings of the 2017 1st International Conference on Intelligent Systems and Information Management (ICISIM), Aurangabad, India, 5–6 October 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 86–90.