Systematic Drift Correction in Eye Tracking Reading Studies: Integrating Line Assignments with Implicit Recalibration

Wolf Culemann, Leana Neuber and Angela Heine

May 15, 2024

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2024) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

28th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2024)

Systematic Drift Correction in Eye Tracking Reading Studies: Integrating Line Assignments with Implicit Recalibration

Wolf Culemann[a,*], Leana Neuber[a], Angela Heine[a]

[a] *University Duisburg-Essen, Universitätsstraße 2, 45141, Essen, Germany*

Abstract

Eye tracking data is typically compromised by a systematic error that is commonly referred to as drift. In reading research, most manual and automated approaches to dealing with drift assign fixations to lines of text. However, correcting for the y dimension only means that horizontal misalignment of fixations is neglected. Available approaches for horizontal correction that involve inferring systematic error from probable fixation locations have not been used in conjunction with line assignment procedures. In this paper, we present a novel approach for extracting the systematic error across multiple reading trials. Starting with trial-by-trial line-to-word mapping for multi-line text, our approach uses a line assignment algorithm based on dynamic time warping. This initial step is followed by an extraction of systematic drift through spatial and temporal filtering to reduce artificial noise. We compare this approach with manual ground-truth line assignments and explicit validation grids. For a set of data from a reading study (N = 30), our method significantly reduced drift in both the horizontal and vertical dimensions. The agreement of a number of vertical drift correction algorithms with manual line assignment improved from 75 to 85% to over 94% by prior elimination of systematic drift with our method, even outperforming results by prior correction of drift derived from validation grids. This suggests that accounting for systematic drift over trials may lead to more accurate correction in reading studies.

Keywords: eye tracking; drift correction; line assignment; implicit recalibration; multiline reading

1. Introduction

Eye-tracking methodology has become an indispensable tool for understanding the cognitive processes involved in reading. By accurately tracking where and for how long a person looks while reading text, researchers gain insight into the processes involved in reading comprehension and engagement. Consequently, improving the accuracy and quality of eye-tracking data is crucial given that systematic error, often referred to as drift, can significantly affect the

* Corresponding author.
E-mail address: wolf.culemann@uni-due.de

1877-0509

reliability of the data. Drift is particularly problematic for reading studies, where accurate assignment of gaze data to text is essential for further analysis.

The quality of eye-tracking data is affected by factors such as individual differences in eye anatomy, environmental conditions, and the technical specifications of the eye-tracking device used. These factors result in discrepancies between measured and actual gaze directions that vary across the screen, which is particularly problematic in fields that require precise gaze tracking, such as reading studies [2].

Traditionally, drift correction methods in eye tracking-based reading research have focused primarily on correcting *vertical* drift, leaving horizontally displaced fixations uncorrected. Available methods for horizontal correction typically rely on inferring the systematic error from explicit calibration trials which require additional time during data collection and, even more problematic, which may elicit a gaze behavior different from that in reading. Obviously, a failure to properly correct for horizontal drift can lead to unsound conclusion especially in the context of reading studies. Going beyond reading research, implicit recalibration approaches that estimate drift based on probable fixation positions relative to certain visual stimuli or a visual saliency map are available [8, 11, 7, 2, 12]. Complementing the line assignment algorithms used in reading research with implicit calibration approaches seems promising to deal with known problems of the former by making them more robust across different types of datasets.

The present paper presents a novel approach that combines two perspectives on eye-tracking data correction to improve the accuracy of data from multi-page multi-line reading studies. By integrating recent advances in the use of dynamic time warping (DTW) for vertical drift correction in reading data [4] with systematic drift extraction methods, our approach can correct for vertical and horizontal displacement based on the systematic drift present over multiple consecutive trials. This method improves drift correction in that it accounts for varying vertical and horizontal drift across the screen while considering similar systematic drift patterns across multiple reading trials. As a result, it eliminates the need for additional explicit calibration grids.

1.1. Related Work

Even though originating from different research areas, all approaches to the correction of eye-tracking data are based on the calculation of offsets from known fixation positions. Aiming to improve the accuracy of determining where a user is looking at on a screen, these approaches differ mainly in the type of stimuli and tasks used as the basis for calculating drift and, as a result, in the probability that the measured offset reflects actual offset. And while approaches such as area-of-interest (AOI)-based line assignment algorithms, implicit or explicit calibration methods differ somewhat in the way spatial error corrections are performed, they are all based on similar underlying assumptions. The main difference between *explicit* and *implicit* calibration lies in the method used to obtain calibration data. In explicit calibration, the study subjects are explicitly instructed to focus on predetermined targets. While it is assumed that they will follow these instructions, the level of compliance can vary due to distractions or fatigue. In contrast, implicit calibration determines probable fixation locations based on the subject's typical interactions with a certain stimulus such as focusing on salient features [11]. Both methods aim at accurately mapping gaze points, but differ in their dependence on the cooperation of the subjects. Implicit calibration has the advantage that calibration can be seamlessly integrated into the subject's natural activities, which may increase compliance and calibration accuracy over longer sessions [8]. After determining the offsets between the eye tracker's predicted gaze positions and the probable fixation locations on the screen, an optimal transformation function for these offsets can be computed and, subsequently, applied to these and/or other data points [7, 13, 14, 2, 12, 8]. For example, Vadillo et al. [12] used a linear transformation of eye-tracking data, where the best-fitting transformation matrix is determined by an optimization routine specifically designed to align fixation coordinates with visual stimuli more accurately. Blignaut et al. [2] showed that linear regression can be successfully used for recalibration. This is not surprising, given that commercial eye trackers such as the Eyelink 1000 also use multivariate polynomial regression for calibration [10]. In reading studies, where a particularly high degree of accuracy is required for word-level analysis, the most commonly used drift correction methods are based on the assumption that lines of text are more likely to be looked at than positions between lines. Although there are some approaches to correct for horizontal drift as well (e.g. for source code reading see [3]), it is more common to correct for vertical drift only, either by manual line assignment or by automatic line assignment algorithms [4]. Carr et al. [4] recently compared a number of classical vertical drift correction algorithms for reading research and concluded that there is no single algorithm that is suitable for all cases and data. They found strong differences, especially with respect to robustness against strong drift or within or between line regressions. In addition to the established algorithms, they presented a 'warp' algorithm based on DTW, which seems to be a promising approach for the line-assignment problem. Unlike the algorithms that are based on absolute or relative position to the lines of text, 'warp' is based on the assumptions that text is read sequentially, i.e. from the beginning to the end without the reader skipping back, and that fixations aim at the center of a word. The 'warp' approach is particularly interesting because it can cope with very strong drift, and recent studies have demonstrated the high accuracy of 'warp' or DTW-based approaches [1, 9].

The assumption that words or lines of text are more likely to be fixated is analogous to the concept of probable fixation locations in implicit calibration approaches. However, these approaches come from different research fields (i.e. natural reading vs. visual search) and, thus, are based on different underlying assumptions. In reading research, in order to compute AOI-based metrics such as mean word-level fixation duration it is necessary to assign fixations to words (AOIs). AOI-based line assignment algorithms often rely heavily on the detection of line-runs, i.e. the coherent reading of a line of text [4]. They are also typically based on the well-established finding that fixations in close proximity should have similar offsets [2]. However, AOI-based line assignment algorithms do not necessarily correct for drift consistently across the screen, i.e. typically, no holistic transformation is computed for the entire screen. In addition, although relatively stable drift patterns are typically observed across trials during manual assignment (e.g., stronger drift for the upper right part of the screen), drift stability over multiple trials is not considered in classic AOI-based line assignment algorithms.

2. Our Approach

The approach presented here (cf. Figure 1) is based on recent development in AOI-based line assignment algorithms, in particular DTW-based line assignment, which has been shown to be suitable to deal even with strong drift [4]. The assignment of fixations to word centers of the 'warp' approach is also used, which enables the correction of the horizontal deviation. As demonstrated by Al-Madi et al. [1], the sensitivity of the 'warp' approach to frequent regressions within or between lines can be addressed by detecting the corresponding outliers. We consider this correction approach from the perspective of implicit calibration, i.e. a multivariate polynomial regression is fitted over the corrected fixation positions to obtain a transformation for the estimated drift across the screen for each trial. As the underlying fixation-to-AOI assignment or offset outlier detection may be somewhat inaccurate in individual cases or biased by characteristics of the spatial AOI distribution in a trial, we assume that the drift remains relatively stable over time – which is an implicit assumption in eye-tracker calibration. By calculating the average of the derived offset over multiple trials (e.g., a complete text) for the same regions of the screen, we reduce such biases. The resulting matrix of offsets across the screen is interpolated, and, in a final step, fixations can be corrected by their nearest offset on the basis this drift matrix.

2.1. Fixation to Word Mapping and Outlier Removal

To get a first estimate of the existing drift, we use word-level fixation-to-line mapping based on the 'warp' algorithm [4]. Unlike most others, this algorithm not only adjusts fixations to specific lines, but also accounts for the specific drift patterns observed across different text segments [4]. A notable feature of this approach is its inherent use of horizontal word assignment. The DTW cost matrix incorporates both x and y positions, aligning fixations to the center of words or AOIs.

However, the 'warp' algorithm has its weaknesses, particularly with respect to within-line or between-line regressions. As the standard DTW algorithm is not designed for repetitive gaze patterns, regressions can introduce significant noise. To mitigate this, we employ a heuristic approach by excluding fixations whose offsets – computed separately in the x and y directions – are greater than 1.5 standard deviations from the mean. This helps to effectively deal with outliers. Due to the nature of the DTW algorithm, multiple fixations are often assigned to a single AOI (when AOIs were skipped). For these cases, selection criteria based on both position and cost minimize errors by ensuring that only the more probable fixations are considered in the case of multiple fixations on an AOI. In particular, we only include the first fixation if it is the beginning of a line, the last fixation if it is the end of a line, and the nearest fixation for all other cases. If multiple AOIs are assigned to a single fixation (in the case of re-reading or regression) , we select the AOI with the lowest cost (closest to the original fixation position). In addition, small regressions (jumping back

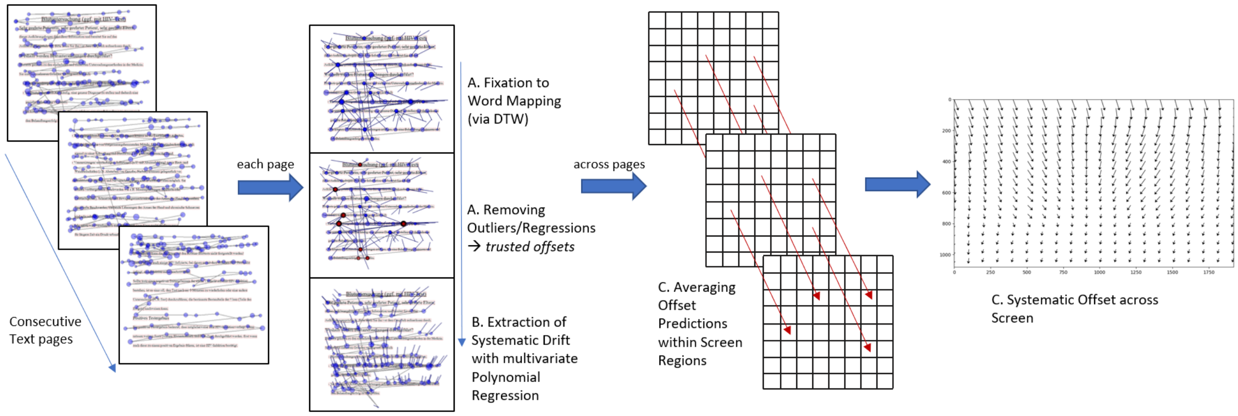

Fig. 1. Illustration of our Approach to derive systematic drift from line assignment across consecutive text pages. The main steps include Fixation to Word Mapping and Outlier Removal (A), Extraction of Systematic Drift for one Page (B) and Deriving the Mean Systematic Drift across multiple Pages (C)

to a previous word) result in crossed fixation positions, which can be easily accounted for by inverting/switching the locations of these fixations. Another challenge arises with fixations that occur above the first line, which may result in an incorrect assignment of both the first and potentially the second line fixations to the first line. To correct for this, we apply a general y-offset to these fixations when three or more occur above the first line. This adjustment ensures that the average position is correctly aligned with the intended line. If this adjustment results in an overlap below the last text line, we modify the offset by considering the lowest fixations, ie., we effectively center the fixations at the top and bottom lines.

Through the application of 'warp' in both x and y dimensions, and rigorous filtering for outliers, we ensure the integrity of the remaining fixations and their *trusted offsets*. It is worth noting that it may not be necessary to be overly careful in distinguishing between trusted and outlier offsets, as subsequent steps in systematic drift correction both across the screen and across trials help to minimize the impact of residual error.

2.2. *Extraction of Systematic Drift*

After identifying *trusted offsets* within a single reading trial, we proceed to model systematic drift. This drift is quantified using second-order multivariate polynomial regression models that describe how fixation offsets behave across the screen. We use a two-dimensional polynomial regression approach for each coordinate – horizontal (x) and vertical (y). These models are constructed to fit the *trusted offsets*, allowing us to predict and correct for drift by capturing both linear and quadratic relationships within the fixation data.

2.3. *Mean Systematic Drift across Trials*

Mean systematic drift across multiple eye tracking trials is quantified by analyzing the 2D position data predicted by the polynomial regressions fitted to each single trial. This procedure involves dividing the screen into a grid of $N \times M$ sections, and predicting the offset values for each grid cell using the polynomial models derived from each trial. The data structure resulting from this analysis is a four-dimensional matrix $D \times N \times M \times T$, where D represents the two dimensions (x and y coordinates), N is the number of cells in the x-direction, M is the number of cells in the y-direction, and T is the number of consecutive trials. N and M are chosen so as to balance sufficient overlap between fixation offsets of matching cells across trials and capture variations in drift across different parts of the screen.

To synthesize a comprehensive representation of drift across trials, this matrix is subjected to filtering. By computing the mean across the trial dimension T for each cell in the grid, this filtering effectively reduces noise and averages out trial-specific anomalies, resulting in a condensed matrix of average drift values. In a next step, gaps or regions outside the convex hull of available points (areas without direct polynomial predictions) are interpolated or filled.

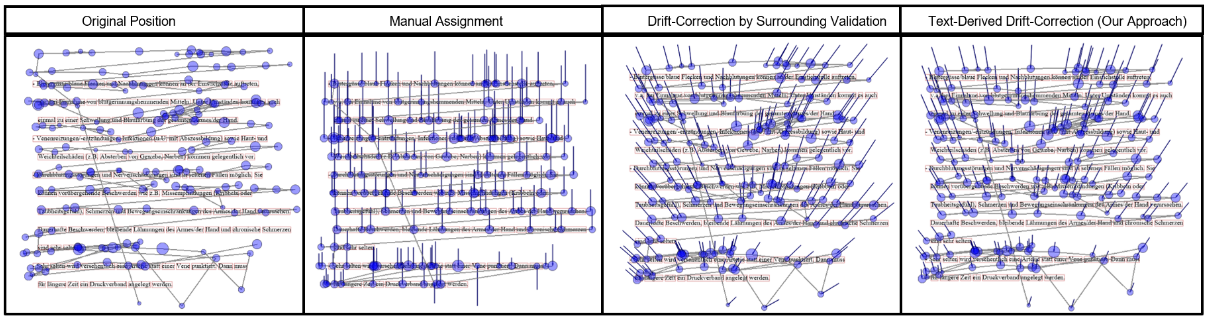

Fig. 2. Fixation Scanpaths of an exemplary trial of Original Fixation Positions, Manual Assignment (vertical only) and drift correction based on an Surrounding Validations and Text-Derived Drift (Our Approach)

Additionally, the matrix is upsampled by simple interpolation in the directions N and M to ensure a smooth and continuous drift representation over the entire screen grid. The refined matrix *TD* (Text-Derived) provides a detailed map of the drift across the screen that can be used to correct the fixation data. For each fixation, an offset correction is applied based on the closest x and y values from the *TD* matrix. This approach ensures that all fixations are adjusted according to the average drift behavior observed over multiple trials. Furthermore, it allows for further refinements, such as applying Gaussian filtering or restricting the analysis to a window of trials around a specific time window, to tailor the correction process to specific experimental conditions or desired levels of detail.

3. Evaluation Study

3.1. Participants, Task and Data Collection

We evaluated our drift correction approach using data from a reading study. The study included 30 participants (7 males and 23 females) between the ages of 18 and 30 years (mean age = 22.8 years, SD = 3 years), all with normal or corrected-to-normal vision using soft contact lenses. For this study, participants were asked to read several standard medical consent forms. Eye-tracking data were collected using an Eyelink 1000+ (SR Research) set to binocular and remote mode with a sampling rate of 1000 Hz. A chin rest was used to minimize head movement. An initial 9-point calibration procedure was repeated when necessary to maintain an average accuracy of less than 0.5 degrees during validation checks. The study was approved by the local ethics committee.

The text consisted of 16 pages altogether, but only pages containing at least 8 lines of text were selected. This resulted in 14 included pages (N = 8-13, mean = 11, SD = 1.3) for a total of 838 reading trials across subjects (excluding one page skipped by one participant). A screen resolution of 1920 x 1080 pixels was used with a refresh rate of 100 fps, and a subject-screen distance of 65 cm which resulted in approximately 42 pixels per degree of visual angle. The average text line width was 533 pixels with a line spacing of approximately 51 pixels. The text was presented in Times New Roman, 18 pt, black on a light gray background, with headings and subheadings at 40 pt and 24 pt, respectively.

Following the best practice for ensuring the quality of eye-tracking data [6], validation trials were included before and after each text segment. For this purpose, a dynamic target on a light gray background was presented. Starting from a central position to familiarize participants with the task, the target, a small white dot within a larger dark gray circle, moved –from top-left to bottom-right–in a reading-like pattern across a 9-point matrix on the screen. For each position in the validation grid the size of the target decreased to its smallest size within 300 ms, and remained on the position for 400 ms. This duration represents the time-window during which subjects are most likely to fixate the target directly and continuously. The target expanded back to its original size within 300 ms before quickly moving to the next position at a speed of 1800 pixels per second. This procedure was repeated for each grid point.

3.2. Data Preprocessing

We used the online velocity-based event detection algorithm (30 deg/s velocity and 9500 deg/sec^2 acceleration thresholds) of the SR Research Eyelink system to detect fixations. Fixations shorter than 60 ms were discarded as they are typically prone to errors due to blinking or other artifacts. Scanpath visualizations for the original fixation positions and the corrected positions are shown in 2 for an exemplary trial.

Manual Assignment (MA). To establish a reference for vertical drift correction, fixations were assigned to their most probable line manually, a common practice in reading eye-tracking studies [4]. This processing step was realized using Eyeflow Studio, a graphical software tool developed by the first author (www.eyeflow-labs.com). The tool allows interactive correction of vertical fixation positions with assignment to lines, either individually or in batches. All adjustments are visualized by arrows indicating the shift from original to new positions, thus allowing easy identification and correction of misalignments. Crucially, adjustments made to one eye's data can be mirrored to the other eye and, if necessary, checked and reciprocally adjusted there, thus ensuring consistent line assignment for both eyes.

Drift from Surrounding Validation (SV). Additional drift correction references were derived from the validation grids placed before and after the text reading trials. For the 400 ms periods when the fixation target was minimal and stabilized, fixations selected on the basis of proximity and minimal dispersion were extracted and offsets for each of the nine validation target locations computed. The resulting offsets from the two validation grid trials were averaged. Locations outside the convex hull of the validation positions were filled, and those inside were interpolated to create a 2x100x100 *Surrounding Validation* (*SV*) drift matrix containing offsets for both vertical and horizontal directions.

Text-Derived (TD) Drift. Following the approach outlined in 2, systematic drift was derived for each of the 14 text trials per subject. After an initial fixation-to-word mapping and extraction of *trusted offsets*, the screen was divided into $N \times M$ sections, with $N = M = 5$, corresponding to a visual angle of approximately to 9 degrees horizontally and 5 degrees vertically. By interpolation, we get a 2x100x100 *Text-Derived* (*TD*) drift matrix containing offsets for both vertical and horizontal directions.

AOI-based line assignment algorithms. Since the goal of preprocessing eye-tracking data for reading research is usually to obtain an exact assignment of each fixation to an AOI, while our approach does not result in such assignments, a line-assignment algorithm was applied after removing the systematic drift. For this reason, we include the evaluation of classic AOI-based line assignment algorithms before and after drift correction (both by *SV* and *TD* drift matrices). For this purpose, the algorithms and implementations described by Carr et al. [4] were used in addition to the slice algorithm [5] and the WOC approach [9]. The latter, the 'Wisdom of the Crowd' approach, uses an combination of these classical algorithms to assign fixations to their modal position. The algorithms will not be described in detail here, as we used the implementations described in detail by [4, 5, 9, 1].

3.3. Results

Using the *MA* of fixations as a reference, we evaluated the vertical offsets and their improvements in comparison to both our *TD* and the *SV* drift correction approach. Table 1 shows that both methods significantly reduced vertical drift. Our *TD* method significantly outperformed the *SV* method, improving the average vertical drift by 54.14% (from

Table 1. Mean and Maximum Offset Improvement by Drift Correction Method and Dimension

Ref.	Dimension	Correction Method	Metric	Before Correction (deg)	After Correction (deg)	Improvement (%)
MA	y	TD	Mean	0.49	0.22	54.14
MA	y	TD	Max	1.95	0.71	76.07
MA	y	SV	Mean	0.49	0.36	25.73
MA	y	SV	Max	1.95	1.24	45.86
SV	x	TD	Mean	0.37	0.26	28.54
SV	x	TD	Max	1.05	0.79	43.98
SV	y	TD	Mean	0.31	0.28	8.74
SV	y	TD	Max	1.10	0.87	81.58

Notes: MA = Manual Assignment, SV = Surrounding Validation, TD = Text Derived Drift

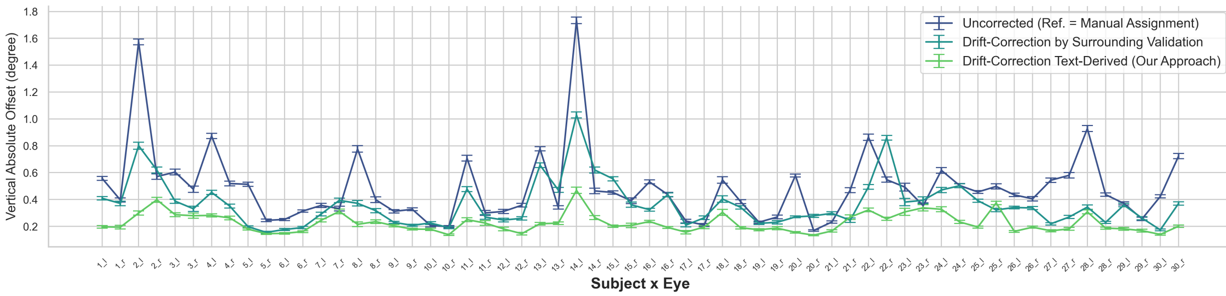

Fig. 3. Magnitude of vertical offset (mean absolute values) across subjects before and after drift correction. Reference for the uncorrected offset are the fixation positions from manual assignment. Error bars represent the 95% CI.

0.49 degrees to 0.22 degrees) and the maximum drift by 76.07% (from 1.95 degrees to 0.71 degrees). In contrast, the *SV* method improved the average offset by 25.73% and the maximum offset by 45.86%. These results are shown in Figure 3, which illustrates the reduction in vertical offset across subjects, highlighting the superior performance of the *TD* method.

For the evaluation of the horizontal offset correction, *MA* cannot serve as a ground truth reference because it only corrects vertically. Therefore, we used the horizontal offsets from the *SV* drift matrix as a ground truth reference for horizontal drift corrections. Despite the limitations inherent in not measuring drift simultaneously with the reading tasks, this approach allowed for a comparative evaluation of the horizontal offsets. According to Table 1, the *TD* method significantly reduced the horizontal offset with an improvement of 28.54% in the mean offset (from 0.37 degrees to 0.26 degrees) and 43.98% in the maximum offset (from 1.05 degrees to 0.79 degrees). Figure 4 shows a consistent reduction in the magnitude of horizontal offsets across almost all subjects, confirming the effectiveness of our *TD* method in correcting for horizontal drift as well.

Furthermore, the performance of the drift correction methods is evident from the strong positive correlations between the initial offsets and the degree of improvement. The *TD* method shows a particularly strong correlation for the vertical dimension when using the *MA* as a reference, with a correlation coefficient (r) of 0.958 and a R^2 value of 0.918, both indicating a highly effective correction with a significant p-value ($p < .001$). The *SV* method also shows a significant correlation with improvements in vertical offsets ($r = 0.786$, $R^2 = 0.618$, $p < .001$), although less so than the *TD* method. For the horizontal dimension, using *SV* as a reference, the *TD* method achieves a correlation of 0.747 and an R^2 of 0.558 ($p < .001$). These correlations strongly suggest that larger initial offsets benefit more from drift correction, which is consistent with previously reported results [2].

Figure 5 illustrates the consistency of different line assignment methods with *MA* under different conditions of drift correction. Initially, without any prior drift correction, the 'warp' method shows the highest agreement with 85.88%, demonstrating its robustness even in uncorrected scenarios. Other methods such as 'chain' and 'regress'

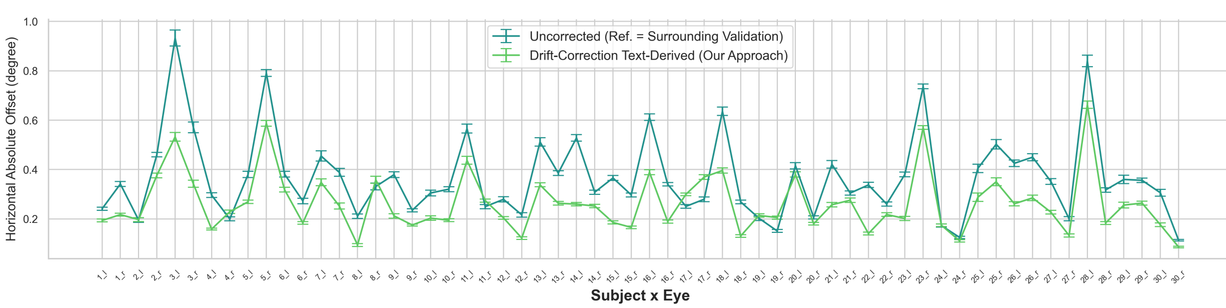

Fig. 4. Magnitude of horizontal offset (mean absolute values) across subjects before and after drift correction. Reference for the uncorrected offset are the fixation positions corrected by surrounding validation.

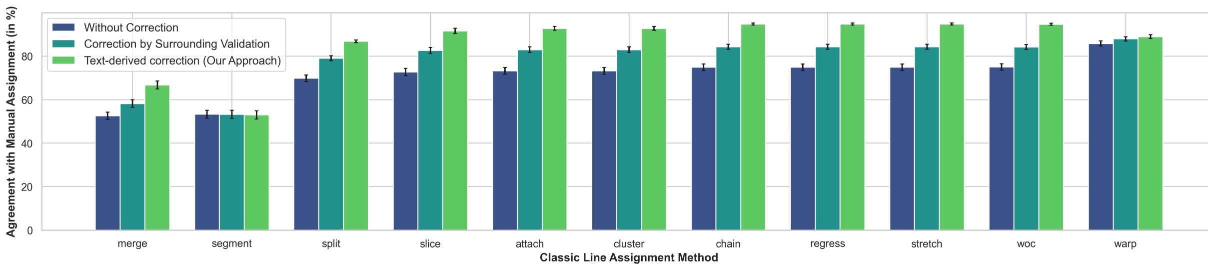

Fig. 5. Performance of classic AOI-based line assignment Algorithms (vertical drift only) measured by agreement with Manual Assignment before Drift-Correction, after correcting for drift by surrounding validation and after correction for text derived drift (our approach). Methods are sorted by uncorrected performance.

show moderate performance, with average agreements around 75%. After applying the *SV* drift correction, most methods show significant improvements. However, while 'warp' has the highest overall agreement at 87.97%, it is not the method with the greatest improvement following the *SV* correction. It is rather methods such as 'chain', 'regress', and 'stretch' that improve their performance significantly, each exceeding 84% agreement. The performance gain is even more pronounced with our *TD* drift method, where 'chain', 'regress', and 'woc' excel, each achieving close to 95% agreement. This emphasizes the effectiveness of our *TD* method in improving the accuracy of AOI-based line assignment algorithms by significantly reducing vertical drift and, thus, achieving closer alignment with the ground truth established by *MA*.

4. Discussion

We presented a novel approach to correct eye-tracking data which combines AOI-based line assignment with systematic drift extraction over multiple trials. Our approach proved to be highly effective in reducing vertical and horizontal drift, and significantly improved the performance of classical vertical line assignment methods. The superior performance of our *TD* method over the *SV* method, which is based on using explicit validation grids, suggests that systematic drift extraction based on fixation-word mapping – which is similar to implicit calibration – may be superior to explicit calibration. This may be due to explicit calibration methods not eliciting the same binocular viewing behavior as reading tasks, given that staring at dots is not comparable to normal gaze behavior in reading. Another reason could be changes in head position, which may occur more frequently when transitioning between different tasks (i.e. reading vs validation grids) even when a chin rest is used to prevent this. Future studies could investigate the influence of head movement and possible changes in visual behavior, e.g., through pseudo-reading. The observed improvements in drift correction correlated strongly with the magnitude of the initial offsets, suggesting that larger drifts benefit more from our correction approach. This has direct implications for eye-tracking reading studies, which often use data from one eye only. Our method is promising because it minimizes accuracy differences between eyes, potentially reducing bias and ensuring more reliable data interpretation across subjects.

While often neglected, the capacity to correct for horizontal drift can significantly affect the accuracy of data interpretation in most reading tasks. Even though the use of *SV* as a benchmark may not be ideal, the results of the evaluation study still indicate a positive effect of our *TD* method in correcting for horizontal drift. Future studies could integrate required and probable fixation locations across multiple trials to provide a more reliable reference for horizontal correction.

It is noteworthy that the classical line assignment algorithms achieved an accuracy of only 75-85% in our dataset, which is lower than in other studies, where in other studies accuracies of 90-95% were reported ([4, 9]). This could be due to differences in reading behavior, data quality, or text stimuli, especially since our study implemented a wide range of line widths and used data from both eyes of each subject instead of just monocular data. Undoubtedly, the improvement in performance of classic AOI-based line assignment algorithms due to the prior drift reduction using our method is remarkable, and shows that it may be beneficial to reduce systematic drift before assigning fixations to lines directly. An advantage of our approach is that it can be combined with different line assignment approaches. For example, the DTW algorithm that was used here could be replaced by new deep learning based line assignment models (see [9]), although currently such approaches do not allow for horizontal correction. Finally, the approach of extracting systematic drift from line assignments and averaging over time could also simplify and speed up the process of manually assigning fixations.

Limitations of the current evaluation study include the use of data from a high-precision eye tracker only, and the limited number of validation grids and points. Future studies could benefit from using different eye-trackers to understand how different levels of precision and systematic noise affect the effectiveness of drift correction. In addition, the method should be tested for influence related to stimulus presentation such as the number of lines of text, line width, or font size. Since our dataset may not be representative for all types of reading behavior, extending the evaluation to larger or synthetic datasets may be useful. Finally, there is room for further improvement in the data cleaning procedure to remove artifacts such as blinking, outliers in the fixation-to-word mapping, or in the type of spatial and temporal filtering of the extracted systematic drift.

In conclusion, our approach has the potential to significantly reduce drift and improve the performance of classic line assignment algorithms used in reading research. The results of the present study highlight the advantage of line assignment methods that account for systematic drift over multiple trials, rather than working on a single trial basis. In addition, the ability to correct for horizontal drift – commonly ignored by traditional approaches to – is a promising advance for reading research in that it prevents fixations from being associated with the wrong words. And, ultimately, the capacity of our method to eliminate the need for additional calibration or validation trials provides for a more ecologically valid and efficient approach to correcting eye-tracking data.

Acknowledgements

This work was funded by the German Federal Ministry for Education and Research (BMBF) under grant number 16SV8620 (project MindTheEyes).

References

[1] Al Madi, N., 2024. Advancing Dynamic-Time Warp Techniques for Correcting Eye Tracking Data in Reading Source Code. Journal of Eye Movement Research 17. URL: https://bop.unibe.ch/JEMR/article/view/10741, doi:10.16910/jemr.17.1.4.

[2] Blignaut, P., Holmqvist, K., Nyström, M., Dewhurst, R., 2014. Improving the accuracy of video-based eye tracking in real time through post-calibration regression. Current trends in eye tracking research , 77–100.

[3] Busjahn, T., 2021. Empirical Analysis of Eye Movements during Code Reading: Evaluation and Development of Methods , 226.

[4] Carr, J.W., Pescuma, V.N., Furlan, M., Ktori, M., Crepaldi, D., 2022. Algorithms for the automated correction of vertical drift in eye-tracking data. Behavior Research Methods 54, 287–310. URL: https://link.springer.com/10.3758/s13428-021-01554-0, doi:10.3758/s13428-021-01554-0.

[5] Glandorf, D., Schroeder, S., 2021. Slice: an algorithm to assign fixations in multi-line texts. Procedia Computer Science 192, 2971–2979. URL: https://linkinghub.elsevier.com/retrieve/pii/S1877050921018068, doi:10.1016/j.procs.2021.09.069.

[6] Holmqvist, K., Nyström, M., Mulvey, F., 2012. Eye tracker data quality: what it is and how to measure it, in: Proceedings of the Symposium on Eye Tracking Research and Applications, ACM, Santa Barbara California. pp. 45–52. URL: https://dl.acm.org/doi/10.1145/2168556.2168563, doi:10.1145/2168556.2168563.

[7] Hornof, A.J., Halverson, T., 2002. Cleaning up systematic error in eye-tracking data by using required fixation locations. Behavior Research Methods, Instruments, & Computers 34, 592–604. URL: http://link.springer.com/10.3758/BF03195487, doi:10.3758/BF03195487.

[8] Kasprowski, P., Harężlak, K., Skurowski, P., 2019. Implicit Calibration Using Probable Fixation Targets. Sensors 19, 216. URL: https://www.mdpi.com/1424-8220/19/1/216, doi:10.3390/s19010216.

[9] Mercier, T.M., Budka, M., Vasilev, M.R., Kirkby, J.A., Angele, B., Slattery, T.J., 2024. Dual input stream transformer for vertical drift correction in eye-tracking reading data. URL: http://arxiv.org/abs/2311.06095. arXiv:2311.06095 [cs].

[10] SR Research, 2017. SR Research EyeLink 1000 Plus User Manual. Version 1.0.12.

[11] Sugano, Y., Matsushita, Y., Sato, Y., 2010. Calibration-free gaze sensing using saliency maps, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, San Francisco, CA, USA. pp. 2667–2674. URL: http://ieeexplore.ieee.org/document/5539984/, doi:10.1109/CVPR.2010.5539984.

[12] Vadillo, M.A., Street, C.N.H., Beesley, T., Shanks, D.R., 2015. A simple algorithm for the offline recalibration of eye-tracking data through best-fitting linear transformation. Behavior Research Methods 47, 1365–1376. URL: http://link.springer.com/10.3758/s13428-014-0544-1, doi:10.3758/s13428-014-0544-1.

[13] Zhang, Y., Hornof, A.J., 2011. Mode-of-disparities error correction of eye-tracking data. Behavior Research Methods 43, 834–842. URL: http://link.springer.com/10.3758/s13428-011-0073-0, doi:10.3758/s13428-011-0073-0.

[14] Zhang, Y., Hornof, A.J., 2014. Easy post-hoc spatial recalibration of eye tracking data, in: Proceedings of the Symposium on Eye Tracking Research and Applications, ACM, Safety Harbor Florida. pp. 95–98. URL: https://dl.acm.org/doi/10.1145/2578153.2578166, doi:10.1145/2578153.2578166.