The Eye of the Typer: A Benchmark and Analysis of Gaze Behavior during Typing

Alexandra Papoutsaki
Computer Science Department
Pomona College
Claremont, CA
apaa2017@pomona.edu

Aaron Gokaslan
Computer Science Department
Brown University
Providence, RI
agokasla@cs.brown.edu

James Tompkin
Computer Science Department
Brown University
Providence, RI
james_tompkin@brown.edu

Yuze He
Computer Science Department
Brown University
Providence, RI
yhe10@cs.brown.edu

Jeff Huang
Computer Science Department
Brown University
Providence, RI
jeff_huang@brown.edu

ABSTRACT

We examine the relationship between eye gaze and typing, focusing on the differences between touch and non-touch typists. To enable typing-based research, we created a 51-participant benchmark dataset for user input across multiple tasks, including user input data, screen recordings, webcam video of the participant's face, and eye tracking positions. There are patterns of eye movements that differ between the two types of typists, representing glances at the keyboard, which can be used to identify touch-.typed strokes with 92% accuracy. Then, we relate eye gaze with cursor activity, aligning both pointing and typing to eye gaze. One demonstrative application of the work is in extending WebGazer, a real-time web-browser-based webcam eye tracker. We show that incorporating typing behavior as a secondary signal improves eye tracking accuracy by 16% for touch typists, and 8% for non-touch typists.

CCS CONCEPTS

• **Human-centered computing** → **Keyboards**; **Pointing devices**; *Empirical studies in HCI*;

KEYWORDS

Typing, eye tracking; webcams; user behavior

ACM Reference Format:
Alexandra Papoutsaki, Aaron Gokaslan, James Tompkin, Yuze He, and Jeff Huang. 2018. The Eye of the Typer: A Benchmark and Analysis of Gaze Behavior during Typing. In *ETRA '18: 2018 Symposium on Eye Tracking Research and Applications, June 14–17, 2018, Warsaw, Poland.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3204493.3204552

ETRA '18, June 14–17, 2018, Warsaw, Poland

ACM ISBN 978-1-4503-5706-7/18/06…$15.00
https://doi.org/10.1145/3204493.3204552

1 INTRODUCTION

The relationship between typing and sight has long been studied. Psychologists investigated the *eye-hand span*—how the eye leads the hand when copying text on a typewriter [Butsch 1932]—and hypothesized a "supply line" of information held in a buffer until it can be typed [Logan 1983]. These past studies have shown how our minds process information cognitively when typing. But less is known about *where* a person is looking during typing—a measure that is difficult to capture without a high-precision eye tracker.

To shed light on this topic, we start by preparing a dataset to enable this investigation. We capture behavior from participants in a lab study across multiple interaction tasks, including mouse and keyboard input data, screen recordings, webcam video of the participant's face, and eye tracking positions. This dataset serves as the foundation for the analyses in this paper, and allows other researchers to replicate and compare against our work.

From the dataset, we compute the distance between eye gaze and the caret location of 51 participants as they type. We assess both the temporal and the spatial relationships between gaze and key press as fundamental measures. There is a condition where a person's behavioral patterns are quite different—that between touch typists and typists who look at the keyboard to see the key being pressed, e.g., those with a "hunt and peck" typing strategy.

We investigate the patterns of eye movements that differ between the touch and non-touch typists. As expected, touch-typists are looking at the text when pressing a key, but that instant is not the most likely time they are looking at the typed character, which comes a moment later. We explore how touch typists stay focused on the line on the screen that the text is written on zeroing in on the characters typed just after a key press, while non-touch typists look straight down just before the key is pressed. Such patterns are presented as aggregated means along with individual examples.

Then, we develop a classifier using a supervised learning algorithm to automatically discern touch typists from non-touch typists. The classification works without any special equipment or software, by using *WebGazer*, an open-source webcam-based eye tracker [Papoutsaki et al. 2016]. This not only enables applications that are targeted towards touch-typists, such as the "Flat-Glass" text input

method [Findlater et al. 2011], but also allows the *inverse application*: improving eye tracking using typing behavior. We show that it is possible to use typing to help determine where a person is looking on the screen. Webgazer currently uses mouse pointing and clicking for this purpose, but adding typing as a cue for touch typists leads to an eye tracking accuracy improvement of 16% for touch-typists, and 8% for non-touch typists.

Our contributions are: 1) describing the temporal and spatial relationship between gaze and caret during typing, 2) the automatic identification of touch typists based on gaze behavior, 3) incorporating the gaze and typing relationship into an eye tracker to improve its accuracy, and 4) the preparation and release of a 51-participant dataset for studying the gaze-typing relationship.

2 RELATED WORK

Eye tracking provides insights into visual attention and human behavior. For example, eye tracking lab studies on Web browsing are often used to investigate visual attention and its correlation with user interactions [Atterer et al. 2006]. Most research on eye tracking and user interactions has focused on cursor movements rather than typing. We describe literature that shows a strong alignment between eye and hand coordination and the need for more naturalistic datasets to better understand the different processes that take place when typing.

2.1 Gaze and User Interactions

Past research has repeatedly found a correlation between gaze and cursor, with the mouse having been characterized as the "poor man's eye tracker" [Cooke 2006]. Chen et al. [2001] investigated this relationship in Web navigation and showed that the dwell time and movement of the cursor is strongly linked to how likely it is that a user will look at that region. In Web search, Rodden et al. [2008] and Guo and Agichtein [2010] found that the distance between cursor and gaze positions was larger along the x-axis. Smith et al. [2000] and Liebling and Dumais [2014] examined the temporal relationship between hand and gaze relationship and showed that the eyes lead the cursor most of the time. Weill-Tessier et al. [2016] were the first to investigate this alignment in the context of tablets, finding that, like on desktop computers, users fixate on the location of a tap before it happens. Although our focus is on typing, analyzing our dataset allowed us to uncover similar patterns on the relationship of eye gaze, cursor movement, and clicks.

2.2 Gaze and Typing

The relationship between gaze and typing has captured the attention of researchers for almost a century, but it has largely focused on copy-typing. Copy-typing is an artificial process of copying by retyping which differs from the common everyday process of producing original text, e.g., when writing an email or a report. In one of the first publications on copy-typing with a typewriter, Butsch [1932] investigated the "eye-hand span", the number of characters the eye is ahead of the hand, and the time interval that it takes to type a character after seeing it. Inhoff et al. further explored copy-typing and, similar to Butsch, found that the eye is 5–7 characters ahead of the hand [Inhoff and Gordon 1997; Inhoff and Wang 1992]. In their findings, they note that this time interval is not consistently one second and independent of the typing speed. Johansson et al. [2010] studied typing as a creative writing activity and, using insights from a head-mounted eye tracker, divided participants into two groups: "monitor gazers" and "keyboard gazers", who can be closely linked to touch and non-touch typists. Focusing on the productivity of the different types of gazers, they found that monitor gazers are faster and more productive typists. Wengelin et al. [2009] discovered that some writers fixate on text produced prior to the location of the cursor, perhaps to process or edit it.

Feit et al. [2016] observed behavioral differences across touch typists and non-touch typists, such as in gaze location and finger placement. Rabbitt [1978] observed that even proficient touch typists tend to look at the screen for error correction. Our approach uses these distinctions to identify touch typists, plus allows us to augment gaze estimators when the user behavior allows it.

This repeatedly-observed coordination of eye and hand has been harnessed to *infer* the gaze. For example, PACE [Huang et al. 2016] is an offline eye tracker that combined mouse and keyboard interactions to predict the gaze with an accuracy of 2.56° in visual angle, after being trained on more than 1000 interactions. Similarly, WebGazer [Papoutsaki et al. 2016] uses cursor movements and clicks to infer the gaze in real time, achieving an accuracy of 4.17°. In this paper, we extend WebGazer and demonstrate how typing can improve its accuracy, especially when recognizing the differences across touch and non-touch typists.

2.3 Remotely Gathering Eyetracking Datasets

The webcam eye tracking community has focused on gathering large datasets for offline training with machine learning. Lebreton et al. [2015] used Amazon Mechanical Turk to crowdsource a webcam eye tracking calibration dataset consisting of more than 200 participants. The experiment sent telemetry to a remote webserver to perform the eyetracking computation. TurkerGaze [Xu et al. 2015] similarly released a game on Amazon Mechanical Turk to gather data that were used to predict image saliency. Krafka et al. [2016] developed an iOS app to crowdsource GazeCapture, a dataset of 2.5M frames from 1450 participants. They used it to perform eye tracking on iPhones and iPads. Unlike our dataset, none of these works includes naturalistic tasks, and typing is absent. For example, GazeCapture consists of frames collected only when participants looked at a stimulus on the screen.

3 DATASET

We created an eye tracking dataset to enable replication of our study and to enable new research. The dataset is publicly available at `https://webgazer.cs.brown.edu/data`. Our focus in this paper is on the typing behavior exhibited in the dataset, and its applications to webcam eye tracking, but researchers with other interests may find this dataset useful as it contains data from a diverse set of tasks. For the eye tracking community, this dataset provides a curated benchmark that includes videos of 51 users, interaction logs, and gaze predictions by a commercial eye tracker.

3.1 Experiment Design

Over the span of three weeks, we recruited participants to complete a series of browser-based tasks: two calibration, one pointing

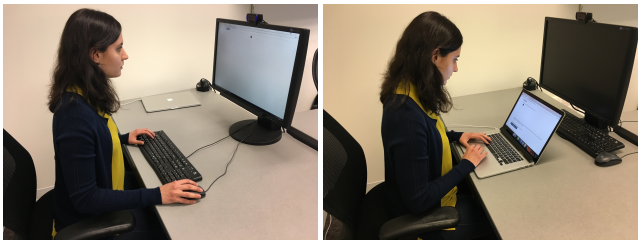

Figure 1: The two options for the experimental setting: a PC with an external webcam (left) or a MacBook Pro (right). Tobii Pro X3-120 is attached at the bottom of both screens.

and clicking, four search, and four creative writing tasks. Throughout the experiment, we recorded participant faces, screens, logged each of their mouse and keyboard interactions, and collected demographic information. For ground truth information of where a person was looking, we used the Tobii Pro X3-120 to record the participants' point of gaze on the screen throughout the experiment. This is a high-end remote eye tracker with a reported gaze sampling frequency of 120 Hz, accuracy of 0.4°, and precision of 0.24°. Contrary to most eye tracking studies, participants were free to move their heads and change their posture, prompting more naturalistic user behavior.

Following procedures reviewed by our Institutional Human Subjects Review Board, participants agreed to have the video, audio, and logs of the study recorded and released for research purposes in a publicly-available dataset. Each participant was asked if they were familiar with touch typing, an ability that was later visually confirmed by the experimenter. Upon consent, participants were randomly assigned to a lighting condition: natural light from two directly-facing windows, or typical artificial office light with the blinds of the windows closed. In the case of natural light, the experimenter noted down if the day was sunny or cloudy; the study always took place during daylight. A white projector screen was used as a uniform background.

Participants could perform the study on a desktop PC or a MacBook Pro laptop with either an external mouse or the built-in touchpad as shown in Figure 1. The laptop included a webcam, while an external Logitech C920 HD Pro webcam was attached to the desktop PC monitor. The Tobii Pro X3-120 eye tracker was mounted at the bottom of the screen for both settings. The desktop PC ran Windows 10 and had a 24-inch Samsung SyncMaster 2443 monitor with a resolution of 1920 × 1200 pixels (94 PPI). The MacBook Pro (Retina, 15-inch, Late 2013) ran macOS Sierra 10.12.5 at a resolution of 1440 × 900 pixels (111 PPI). For both settings, we used the Google Chrome browser (v. 56.0.2924) in a maximized window.

The study started with the calibration of the Tobii Pro X3-120; a stimulus appeared in five fixed locations on the screen. The experimenter would judge if the calibration was successful based on visual cues provided by the Tobii interface and would ensure that the face of the participant was within the webcam's field of view, that all interactions were logged, and that the screen was recorded.

Each participant completed the same sequence of tasks. After the completion of a task, its corresponding webcam video feed was automatically downloaded through Chrome. The first task, which

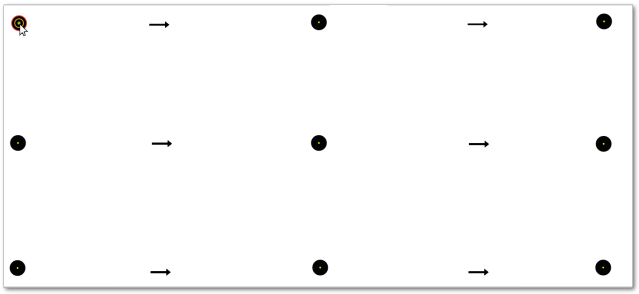

Figure 2: Composite image of the 9 stimuli locations in the Dot Test. The stimuli appear one at a time, in western order, with the next appearing after the user clicks its center.

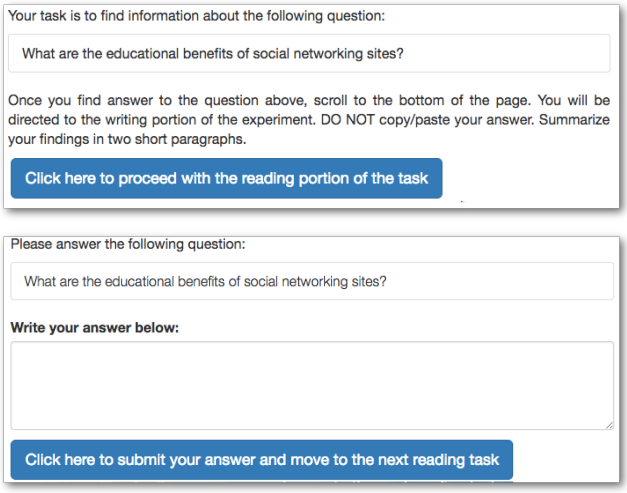

Figure 3: The instructions for the reading (top) and writing (bottom) portion of the search task that corresponds to the "educational advantages of social networking tasks" query.

we refer to as the "Dot Test", is a simple target selection task that forces the user to look closely at the target to successfully aim at it. It began with a black circle and a concentric smaller yellow circle appearing at the top left corner of the screen. The goal was to click inside the yellow circle. Because of its small size, we provided assisting visual cues. If they clicked at it, the whole black circle would move to one of 9 locations within a 3×3 grid, from the top left corner all the way to the bottom right corner of the screen, as seen in Figure 2. The Dot Test was followed by a standard Fitts' Law study using the multidirectional tapping task suggested by the ISO9241-9 standard [Soukoreff and MacKenzie 2004].

The next batch of tasks aimed to replicate a realistic scenario of reading, searching for information, and typing the solution. Participants were given four questions and a query with its corresponding search engine result page (SERP). Table 1 shows the four questions and their queries in the order they were given to all participants. The questions and queries were found in the TREC 2014 Web Track organized by NIST [of Standards and Technology 2017]. For each query, we downloaded the first Google SERP and confirmed that at

Table 1: The four questions selected from the TREC 2014 Web Track and the corresponding queries given to participants.

Task Description	Query
How is running beneficial to the health of the human body?	benefits of running
What are the educational benefits of social networking sites?	educational advantages of social networking sites
What are the best places to find morel mushrooms growing?	where to find morel mushrooms
What treatments are available for a tooth abscess?	tooth abscess

least one of its links contained the answer. Participants could visit multiple links, but were not allowed to alter the query or go beyond the first page of results. After being satisfied with their search, they would scroll to the bottom of the search result page, where a button would take them to the writing portion of the task. There, they would type the answer they synthesized. We prohibited the action of copy-and-pasting text to reveal the true interactions that take place during the creative production and typing of text. Figure 3 shows the instructions for the reading and writing tasks that correspond to the "advantages of social networking sites" query.

The final task, "Final Dot Test" was similar to the Dot Test and we use it as a measure of the accuracy of eye tracking systems. Instead of clicking on the black circle, participants watched it move on its own within a 3×3 grid, remaining for 3 seconds in each location. Participants were explicitly instructed to look at the circle as it moved. On average, this task took place 20 minutes after the calibration, therefore it can be used to estimate drift. Participants were given a brief demographics survey and compensated $20 (USD).

3.2 Participants

We recruited 64 participants (32 female, 32 male) through campus-wide mailing lists. The study lasted 21 minutes on average. Of those 64 participants, 13 were excluded from the curated dataset and its analysis due to technical difficulties during the experiment: issues with the Tobii Pro X3-120 or the screen recording, or interruptions throughout the study by the participant. This resulted in 51 participants whose data we use in this paper, unless otherwise specified. Their ages ranged from 21 to 58 years (M = 27.04, SD = 5.64). Out of the 64 participants, 26 had normal vision, 19 wore eye glasses, and 6 wore contact lenses. Across all participants, there were 4,801 clicks, 109,640 mouse movements, 71,412 key presses, and 4,501,959 gaze predictions made by the eye tracker.

At the end of the study, we surveyed participants for their gender, age, dominant hand, eye color, if they have normal vision, wear eye glasses or contacts, and to self-report their race, and skin color. The provided race categories were American Indian or Alaska Native, Asian, Black or African American, White, or Other. For the skin color, we matched a color bar obtained from Ho and Robinson [2015] to the color of the inside part of their upper arm. Finally, the experimenter made observations about any type of facial hair (none, little, beard) and classified the participants into touch typists or non-touch typists based on the frequency that the participant glanced at the keyboard.

3.3 Dataset Limitations

Reliably capturing and playing webcam video frames with precise timestamps is arguably impossible with current web standards and browsers. As such, there is a variable gap between when a frame is captured and when it is played back. This is approximately within a video frame of time (≈1/30th second). Contrast this to a Tobii X3-120 timestamp, which is approximately within 1/120th second. This typically means that saccades have incorrect instantaneous webcam-based gaze, but that fixations are correct (within error).

4 USER INTERACTIONS VERSUS EYE GAZE

Our dataset includes both specific-target selection tasks, and naturalistic tasks such as web search, reading, and typing. With this, we will confirm literature findings on mouse cursor and gaze alignment, and gain insights into the relationship between typing and gaze attention. For this, we assume that the gaze predictions obtained from the Tobii Pro X3-120 correspond to the true gaze locations.

4.1 Mouse Clicks versus Eye Gaze

Mouse click location and gaze point have been shown to approximately agree spatially, e.g., Huang et al. [2012] found that the median Euclidean distance between gaze and clicks is 74 pixels. Figure 4 reports the average distance between gaze and clicks for all tasks in our study. The mean Euclidean distance is 137 pixels, which is nearly double that found by Huang et al., and may be due to the near doubling of screen pixel density since then.

We also identify the temporal lag between gaze and mouse when a click occurs. We average across all tasks and participants the distances between the mouse location and Tobii Pro X3-120 gaze locations for 3 seconds before and after every click. This is smallest 480 ms before the click, at 110 pixels, with corresponding mean center offset $\Delta x = -20$ and $\Delta y = -2$ (Figure 4). These numbers are small, showing that the user looks at the target before the click. However, the half-second time lag indicates that, at click time, the gaze has already started to drift away.

4.2 Mouse Cursor Movement versus Eye Gaze

Past research has shown that equating cursor location to gaze location is usually imprecise, with the average distance between gaze and active cursor movements being about twice as far as during a click [Huang et al. 2012]. In our data, the mean Euclidean distance of a cursor movement and the Tobii Pro X3-120 predictions is 206 pixels (Figure 4). It is reasonable that the distance is higher, as our analysis also includes 'non-action' cursor movements. The distance between the cursor location and the corresponding Tobii Pro X3-120 prediction is smallest 100 ms before the cursor moves. Unlike mouse clicks, the temporal shift is small. This is perhaps due to the magnitude of cursor events that happen continuously and before an action has been completed. On average, the user looks

Interaction	Time (ms)	Mean dist. (px)	Offset x, y (px)
Mouse click	0	137	-25, -7
Closest dist.	-480	110	-20, -2
Mouse cursor move	0	206	-67, -17
Closest dist.	-100	-100	-66, -17
Typing—all	0	192	-17, 149
Closest dist.	210	178	-13, 142
Touch typists	0	160	-9, 119
Closest dist.	210	151	-6, 116
Non-touch typists	0	352	-55, 299
Closest dist.	540	294	-39, 239

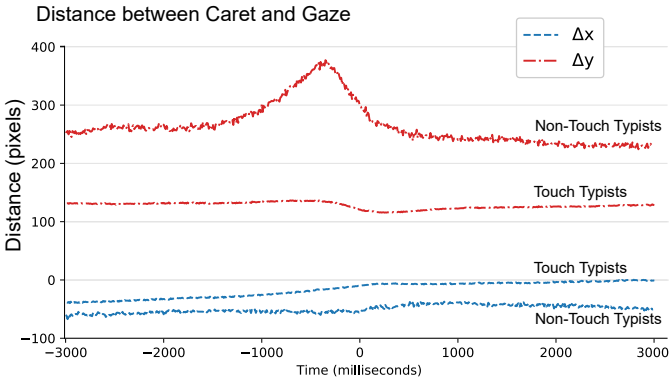

Figure 4: Comparing mean Euclidean distances between interaction locations and Tobii Pro X3-120 gaze predictions, both at the moment the interaction occurred (time offset = 0) and at the time when the interaction location and gaze prediction were closest within a 6-second window ('*Closest dist.*' in table; axes to right), with negative numbers before the interaction. The distance between the mean gaze prediction and the location of the interaction is also reported as the offset (bias) between the positions in x and y. A negative x places gaze predictions to the left of the interaction, with negative y correspondingly above.

above and left of the cursor when the distance between a cursor movement and the gaze is minimized ($\Delta x = -66$, $\Delta y = -17$).

4.3 Typing Caret versus Eye Gaze

The relationship between typing and gaze activity is less researched than between mouse clicks/movements and gaze, i.e., for copy-typing only [Inhoff and Gordon 1997], and so we begin by investigating the alignment of key presses and gaze in free-typing. In our analysis, we report numbers for all participants, and for touch typists and non-touch typists based on the labels assigned during the study by the experimenter.

During typing, the caret is typically a blinking cursor at the position where text is being inserted. On average, the distance between the caret during a key press and its corresponding gaze prediction is 192 pixels (Figure 4). There is a substantial difference between touch typists (160 pixels) and non-touch typists (352 pixels). This difference remains when examining the closest distance within a 6-second window: 210 ms after a key was pressed, the average gaze is 178 pixels away from the caret. For touch typists, on average, the Euclidean distance between a key press and the Tobii Pro X3-120 prediction is smallest 210 ms after the key press, at 151 pixels away. At that moment, the corresponding average values for the x and y axes are -6 and 116 pixels, respectively. On the other hand, for non-touch typists, the distance between a key press and the eye gaze is minimized 540 ms after the key press, at a distance of 294 pixels. The corresponding Δx and Δy for the same moment are -39 and 239 pixels, respectively. The difference between touch typists and non-touch typists can be explained: non-touch typists have to look at the keyboard far more often that touch typists, therefore the Δy is substantially greater as they look down.

As the eyes move quickly when typing, we examine the gaze-caret distance at times surrounding the key press. Contrary to clicks and cursor movements, the distance between key presses and gaze is shortest after the event has occurred. Even touch typists might look toward the character they just inserted after a short delay. At that time, on average, touch typists look to the left of the caret, which agrees with Johansson et al.'s finding [2010]. Since our study was conducted in English, where text is inserted from left to right, we expect that users examine the text they have just written as they type new characters, e.g., to confirm correct spelling. Regardless of their ability to touch type, participants on average looked below the inserted character. The distance on the y-axis is greater for non-touch typists. Note that the eye tracker can only identify the area that the fovea of the eye is focusing on. In practice, the user can still recognize characters and words within a certain radius from the foveal point of focus.

For touch typists, Figure 4, right, shows the 'visual check' for the typed keystroke appearing on the screen by the valley in the distance which occurs 210 ms after a key press. For non-touch typists, Figure 4 shows gaze looking down from the caret towards the keyboard about 200 ms before the key press.

Among touch typists, there is little variation along the y-axis; this is not the case for non-touch typists, who look below the caret while typing. Figure 5 illustrates this by comparing one touch and one non-touch typist as they perform the same writing task. The touch typist reliably looks close to the location of the caret while they type, but the non-touch typist alternates their gaze between the caret location on the screen and the keyboard. Note that these are example typists and other typists had a range of glance patterns and timings, making it difficult to detect them solely by thresholds.

5 IDENTIFYING TOUCH-TYPISTS

With this new understanding of behavioral differences across touch and non-touch typists, we examine whether we can automatically classify users in our dataset into these two categories. Doing so allows applications to know which of their users are touch typists without having to ask them explicitly. The approach is to classify each keystroke individually as touch/non-touch typing, then to average these classifications over time to classify the individual.

For each keystroke, one simple approach is to compute the distance between the text caret and the gaze location in screen space, measured durin a key press. If a user is a touch typist, then the

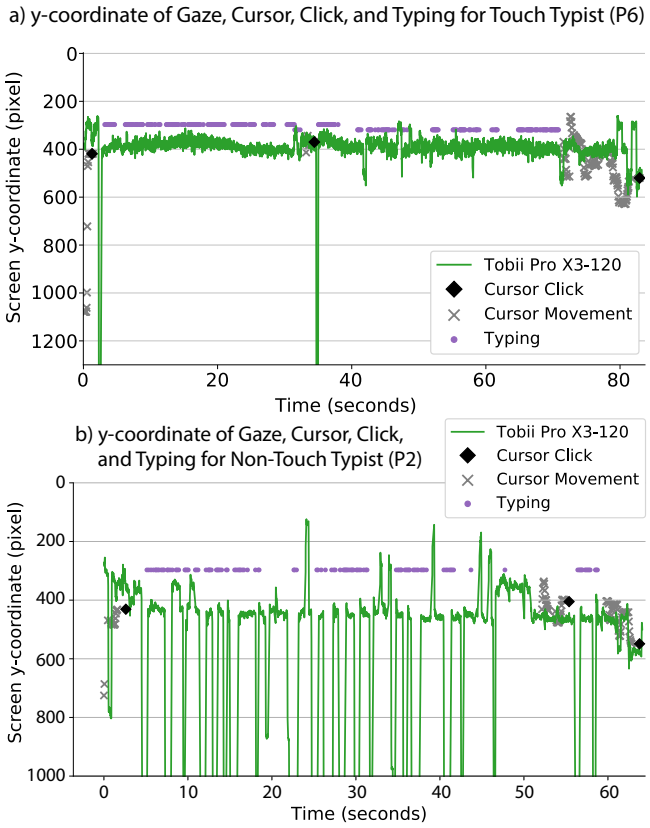

Figure 5: Gaze activity on the y-axis for a) P6, a touch typist and b) P2, a non-touch typist writing the answer to "How is running beneficial to the health of the human body?" Touch typists rarely look at the keyboard so their gaze maps closely the caret as they type, while non-touch typists look back and forth between the keyboard and text.

distance will be small as they are looking at the screen; if they are not, the distance will be large as they are looking at the keyboard.

Given all 56,000 keystrokes in our curated database and the touch/non-touch typist identification, we optimize a distance classification threshold: 540 pixels on the laptop, and 724 pixels on the desktop. These distances approximate half the screen. For the Tobii Pro X3-120, this threshold correctly classified touch typists from non-touch typists with 74.5% accuracy. More practically, for gaze predictions from the WebGazer open source online webcam eye tracker, this classifier is 62.5% accurate. Thus, this simple heuristic is insufficient, so more complex behaviors or gaze errors are present.

We improve upon this baseline by including gaze predictions from one second before and one second after each key press event. We concatenated these predictions from the over the keystrokes in our dataset and trained a classifier using auto-sklearn [Feurer et al. 2015]. Auto-sklearn is an automated machine learning toolkit which selects the best classification model for a given problem. Models were evaluated using a randomly-selected test set from our database, with five-fold cross validation. A random forest classifier was most successful. With gaze data from the Tobii eyetracker, this model achieved 92% accuracy, while with gaze predictions from Webgazer, this model achieved 91% accuracy. For comparison, this accuracy approximately matches that of our participant self-reported touch/non-touch typing values, in which 46/51 participants correctly self-reported their touch typing ability (verified visually by the experimenter).

6 TYPING FOR GAZE ESTIMATION

So far, we have examined the alignment between clicks, cursor movements, key presses and gaze, and demonstrated that their combination can automatically identify touch typists. Next, we use this knowledge to improve an application: WebGazer [Papoutsaki et al. 2016], an open source JavaScript-based webcam eye tracker.

WebGazer makes assumptions about where the user is looking by using user interactions as a proxy for gaze. Currently, WebGazer uses a facial detection library to identify the face and eyes of the user in real time and continuously trains a model that maps the gaze to the screen by matching the eye appearance to known interaction locations. Its basic regression model, a ridge regression, considers the location of clicks as permanent training points based on the assumption that the gaze is strongly aligned with the location of the cursor at the moment of an intentional action such as of a click. When it comes to cursor movements, WebGazer adds them only temporarily (for less than a second) to its regression model; cursor movements often do not follow the gaze activity, e.g., when pushing the cursor aside while reading text and given their magnitude compared to clicks, they could wrongly steer the gaze estimations. The analysis of our dataset further supports those assumptions, with the gaze and cursor distance being minimized right before a click and the cursor following the eye more loosely. We call this baseline model currently in WebGazer the "Cursor Model" since it primarily uses cursor features as a proxy for eye gaze.

We altered WebGazer so that it can accept the offline webcam video feed that we collected for every task page in the study dataset. This way the results could be deterministic and computed efficiently compared to WebGazer's usual live webcam mode. We also simulated the collected user interaction logs and synchronized them with the corresponding video frames. This allows us to replicate the entire user study as if it happened in real time, with WebGazer predicting the point of gaze given the recorded click and cursor interactions and the corresponding appearance of the detected eyes in the offline videos.

After applying WebGazer on the curated dataset, we discovered that its facial feature detection library clmtrackr [Mathias 2014] failed to properly apply the facial contour on the videos to some participants. Out of the 51 participants, 22 had faces that could not be consistently detected by the detection library. As the problem of facial detection is outside of the scope of this paper, we refrained from investigating the reasons that clmtrackr failed in these instances and leave it as future work. Nevertheless, we observed that it performed poorly under uneven lighting conditions and on participants with darker skin color. Even across the 29 participants where facial feature detection was successful, the facial model that clmtrackr fits often failed to align correctly for a few seconds, especially when the participant moved or their face partially came out

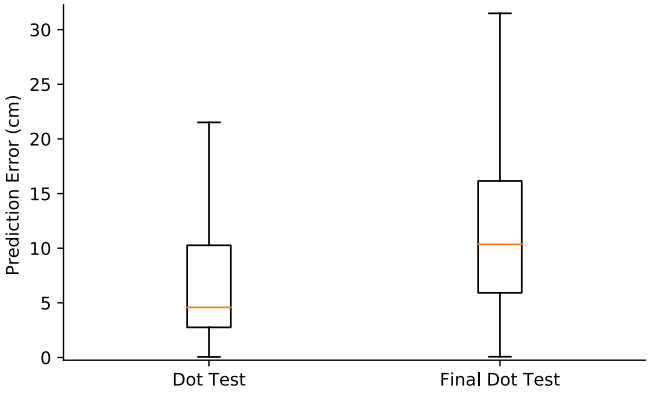

Figure 6: The WebGazer baseline Cursor Model is applied during the Dot Test and Final Dot Test tasks after being trained explicitly by at least 9 clicks. The Final Dot Test appears about twenty minutes after the Dot Test. The boxplots illustrate the prediction error of the Cursor Model (using the Tobii Pro X3-120 gaze estimations as ground truth).

of the webcam field of view. The size of the dataset is not varied enough to apply meaningful between-subject comparisons.

As a first step, we applied the Cursor Model on the two pages hosting the Dot Test and the Final Dot Test for each of the 29 participants. Since the task on the Dot Test page is to successfully click at the center of a circle appearing in 9 locations, each participant will click at least 9 times. We use these clicks to train WebGazer. Following this step, we evaluate its prediction error during the Final Dot Test, where participants just observe the stimulus moving on its own around the screen. It is worth noting, that working with offline videos allows us to train and test WebGazer using the dataset videos in any order. For example, in practice, the Final Dot Test would have happened approximately twenty minutes after the Dot Test, but we still use it as an evaluation step, since it allows us to focus on the basic functionality of WebGazer. Nevertheless, during this timespan the participants have moved in their seat, changed their posture, the lighting is not the same, etc. These factors can affect the reported prediction error, both by making the Tobii Pro X3-120 eye tracker less reliable as the ground truth, and leaving less informative parameters in the WebGazer model.

Figure 6 illustrates in boxplots the distribution of the prediction error during the Dot Test and Final Dot Test for the baseline Cursor Model regression of WebGazer. The prediction error is calculated as the Euclidean distance between the prediction made by WebGazer and the corresponding prediction from Tobii Pro X3-120. Since their sampling rates differ, we group all predictions into 10 millisecond bins. The error is translated from pixels to physical distance (centimeters) according to the pixel density of the PC monitor and laptop screens. The average prediction error is 8.24 cm during the Dot Test and 12.18 cm during the Final Dot Test. As expected, the error during the Final Dot Test is higher.

Figure 7 shows the gaze activity of P46 during the Dot Test across the x and y axes. Similarly, Figure 8 shows the gaze activity of the same participant during the Final Dot Test. We observe that the Cursor Model traces the eye gaze closely. These figures show that

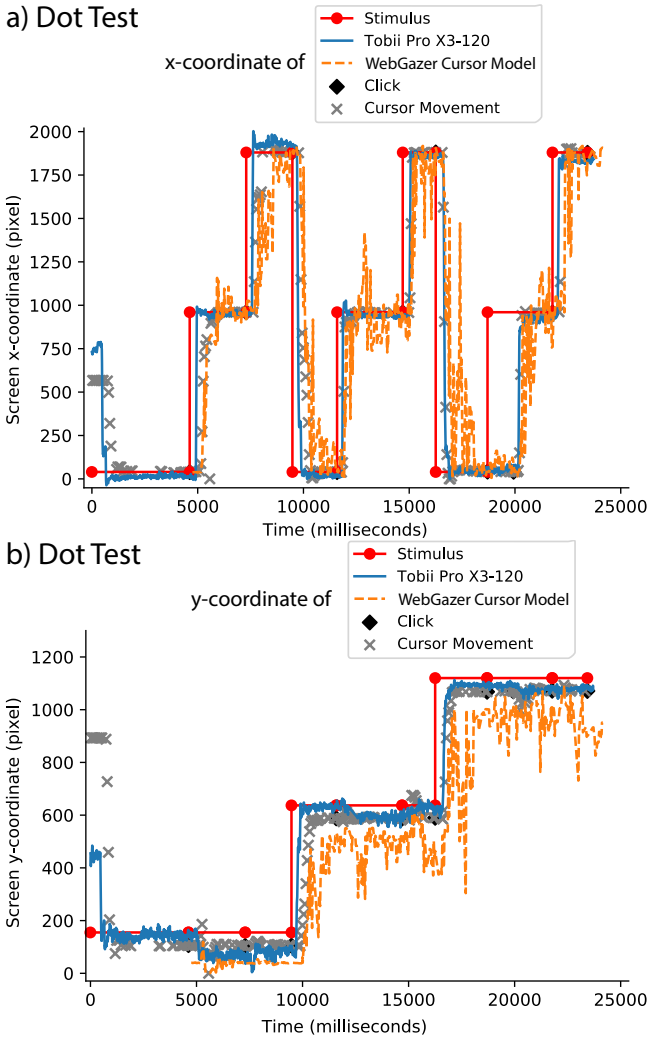

Figure 7: Gaze activity in a) the x and b) the y-axis, as predicted by Tobii (solid blue) and the baseline Cursor Model of WebGazer (dashed orange) during the Dot Test for participant P46. The 9 locations of the stimulus are shown in red.

there is alignment between user interactions and eye gaze during the two Dot Test tasks.

As a next step, we explored typing as a new cue to represent gaze. WebGazer's baseline Cursor Model uses mouse cursor interaction to map eye appearance to screen locations. We attempted to use key presses as equivalent interactions to cursor movements by including the caret location in the regression model during those instances—this is the Cursor+Typing Model. A key press contributes as training when the user is actively typing and for half a second afterwards. This approach prevents over-training the regressor as would happen if all key presses were added permanently, as they all derive from very similar screen locations. We used different tasks from our dataset to evaluate the effectiveness of this approach. In addition to the Dot Test and Final Dot Test, we used the writing portion of the "How is running beneficial to the health of the human

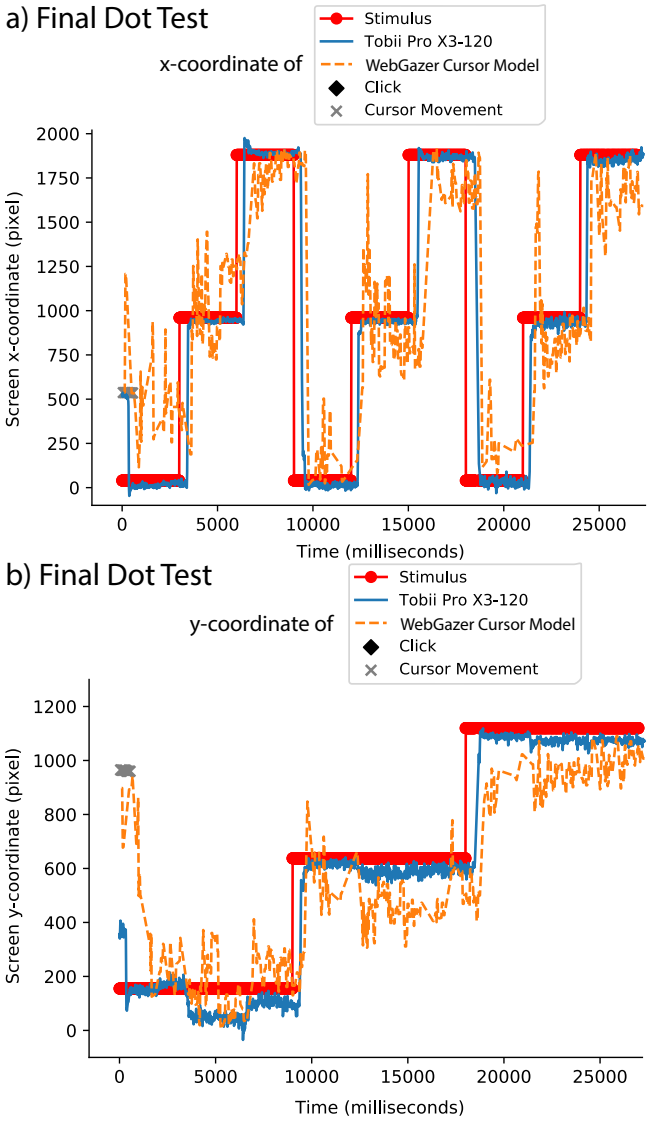

Figure 8: User interactions along a) the x and b) the y-axis, as predicted by Tobii (solid blue) and the baseline Cursor Model of WebGazer (dashed orange) during the Final Dot Test for participant P46. The stimulus (red) appears for 3 seconds in each of the 9 locations within a 3×3 grid. WebGazer's baseline Cursor Model was only trained during the Dot Test.

body?" question, with the writing task placed between the two dot tests as a task comprising typing that trained the model.

Figure 9 breaks down the differences between the Cursor Model and the Cursor+Typing Model when taking into account one's ability to touch type. On average, the Euclidean distance (error) between WebGazer's and Tobii Pro X3-120's predictions for touch typists dropped from 7.77 cm to 6.55 cm (16%). An independent-samples t-test revealed a significant difference in the error for the Cursor (M = 7.77, SD = 7.12) and the Cursor Typing Model (M = 6.55, SD = 7.10); $t(136158)$=31.65, p<0.01, d=0.17. Similarly, incorporating

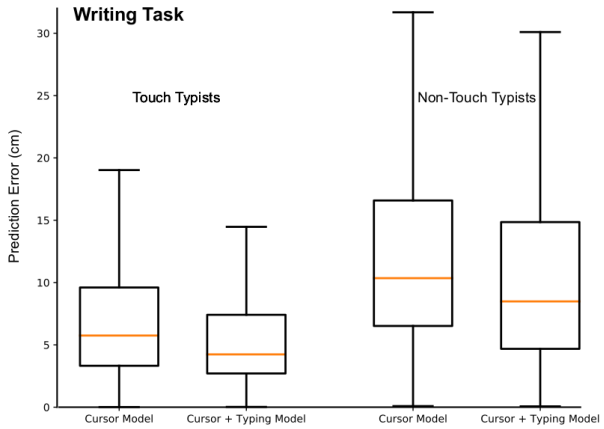

Figure 9: Euclidean distance between WebGazer and Tobii Pro X3-120 predictions during the writing tasks, split across a) touch typists and b) non-touch typists. Incorporating typing improves the gaze estimation for both types of users.

key presses decreased the error across non-touch typists from 12.78 cm to 11.76 cm (8%). A t-test revealed significant difference for the Cursor (M = 12.78, SD = 9.11) and the Cursor Typing Model (M = 11.76, SD = 10.84); $t(80812)$=14.45, p<0.01, d=0.10. Overall, knowing about typing behavior and its relation to eye gaze is a feature that can then be used to improve eye tracking. The WebGazer eye tracker can improve its prediction of where the user is looking, which can in turn lead to more impactful eye tracking applications.

7 CONCLUSION

Typing is a required task for most computer use, and it is important to better understand the different processes that occur as users create text. This work adds to the understanding of human attention and behavior by analyzing the relationship of typing and gaze. To that end, we provide and analyze a benchmark dataset with a variety of tasks, including target selection, calibration, search, and writing.

The analysis of the data confirmed prior knowledge about the spatial alignment of gaze with cursor movement and clicks. We also show the relationship of the caret's location during key presses and that of the gaze, and focus on differences across touch typists and non-touch typists. These differences are substantial, such as the habit of checking written text for touch typists and the habit of glancing down before a key press for non-touch typists. The behavioral patterns inform a method to automatically distinguish between the types of users. We use these findings to incorporate typing as a user interaction in WebGazer, a browser-based eye tracker, by altering its underlying model.

ACKNOWLEDGMENTS

We thank our study participants for allowing us to release their data. This research is supported by NSF grants IIS-1464061, IIS-1552663, and the Brown University Salomon Award.

REFERENCES

Richard Atterer, Monika Wnuk, and Albrecht Schmidt. 2006. Knowing the user's every move: user activity tracking for website usability evaluation and implicit interaction. In *Proceedings of the 15th international conference on World Wide Web (WWW '06)*. ACM, 203–212.

Russell LC Butsch. 1932. Eye movements and the eye-hand span in typewriting. *Journal of Educational Psychology* 23, 2 (1932), 104.

Mon Chu Chen, John R. Anderson, and Myeong Ho Sohn. 2001. What Can a Mouse Cursor Tell Us More?: Correlation of Eye/Mouse Movements on Web Browsing. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems (CHI EA '01)*. ACM, New York, NY, USA, 281–282. https://doi.org/10.1145/634067.634234

Lynne Cooke. 2006. Is the Mouse a "Poor Man's Eye Tracker"?. In *Annual Conference-Society for Technical Communication*, Vol. 53. 252.

Anna Maria Feit, Daryl Weir, and Antti Oulasvirta. 2016. How We Type: Movement Strategies and Performance in Everyday Typing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 4262–4273. https://doi.org/10.1145/2858036.2858233

Matthias Feurer, Aaron Klein, Katharina Eggensperger, Jost Springenberg, Manuel Blum, and Frank Hutter. 2015. Efficient and Robust Automated Machine Learning. In *Advances in Neural Information Processing Systems 28*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett (Eds.). Curran Associates, Inc., 2962–2970. http://papers.nips.cc/paper/5872-efficient-and-robust-automated-machine-learning.pdf

Leah Findlater, Jacob O Wobbrock, and Daniel Wigdor. 2011. Typing on flat glass: examining ten-finger expert typing patterns on touch surfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2453–2462.

Qi Guo and Eugene Agichtein. 2010. Towards Predicting Web Searcher Gaze Position from Mouse Movements. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems (CHI EA '10)*. ACM, New York, NY, USA, 3601–3606. https://doi.org/10.1145/1753846.1754025

Byron K Ho and June K Robinson. 2015. Color Bar Tool for Skin Type Self-Identification: a cross sectional study. *Journal of the American Academy of Dermatology* 73, 2 (2015), 312.

Jeff Huang, Ryen White, and Georg Buscher. 2012. User See, User Point: Gaze and Cursor Alignment in Web Search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1341–1350. https://doi.org/10.1145/2207676.2208591

Michael Xuelin Huang, Tiffany C.K. Kwok, Grace Ngai, Stephen C.F. Chan, and Hong Va Leong. 2016. Building a Personalized, Auto-Calibrating Eye Tracker from User Interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5169–5179. https://doi.org/10.1145/2858036.2858404

Albrecht W Inhoff and Andrew M Gordon. 1997. Eye movements and eye-hand coordination during typing. *Current Directions in Psychological Science* 6, 6 (1997), 153–157.

Albrecht W Inhoff and Jian Wang. 1992. Encoding of text, manual movement planning, and eye-hand coordination during copytyping. *Journal of Experimental Psychology: Human Perception and Performance* 18, 2 (1992), 437.

Roger Johansson, Åsa Wengelin, Victoria Johansson, and Kenneth Holmqvist. 2010. Looking at the keyboard or the monitor: relationship with text production processes. *Reading and Writing* 23, 7 (2010), 835–851. https://doi.org/10.1007/s11145-009-9189-3

Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. 2016. Eye Tracking for Everyone. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '16)*. IEEE Computer Society, Washington, DC, USA, 2176–2184. https://doi.org/10.1109/CVPR.2016.239

Pierre Lebreton, Isabelle Hupont, Toni Mäki, Evangelos Skodras, and Matthias Hirth. 2015. Bridging the gap between eye tracking and crowdsourcing. , 9394W pages. https://doi.org/10.1117/12.2076745

Daniel J. Liebling and Susan T. Dumais. 2014. Gaze and Mouse Coordination in Everyday Work. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp '14 Adjunct)*. ACM, New York, NY, USA, 1141–1150. https://doi.org/10.1145/2638728.2641692

Gordon D Logan. 1983. Time, information, and the various spans in typewriting. *Cognitive aspects of skilled typewriting* (1983), 197–224.

Audun Mathias. 2014. clmtrackr: Javascript library for precise tracking of facial features via Constrained Local Models. https://github.com/auduno/clmtrackr. [Online; accessed 2016-09-08].

National Institure of Standards and Technology. 2017. TREC 2014 Web Track. http://trec.nist.gov/data/web2014.html. [Online; accessed 2017-06-19].

Alexandra Papoutsaki, Patsorn Sangkloy, James Laskey, Nediyana Daskalova, Jeff Huang, and James Hays. 2016. WebGazer: Scalable Webcam Eye Tracking Using User Interactions. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, 2016, New York, NY, USA, 9-15 July 2016 (IJCAI '16)*. 3839–3845.

Patrick Rabbitt. 1978. Detection of errors by skilled typists. *Ergonomics* 21, 11 (1978), 945–958.

Kerry Rodden, Xin Fu, Anne Aula, and Ian Spiro. 2008. Eye-mouse Coordination Patterns on Web Search Results Pages. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems (CHI EA '08)*. ACM, New York, NY, USA, 2997–3002. https://doi.org/10.1145/1358628.1358797

Barton A. Smith, Janet Ho, Wendy Ark, and Shumin Zhai. 2000. Hand Eye Coordination Patterns in Target Selection. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications (ETRA '00)*. ACM, New York, NY, USA, 117–122. https://doi.org/10.1145/355017.355041

R. William Soukoreff and I. Scott MacKenzie. 2004. Towards a Standard for Pointing Device Evaluation, Perspectives on 27 Years of Fitts' Law Research in HCI. *International Journal of Human-Computer Studies* 61, 6 (Dec. 2004), 751–789. https://doi.org/10.1016/j.ijhcs.2004.09.001

Pierre Weill-Tessier, Jayson Turner, and Hans Gellersen. 2016. How Do You Look at What You Touch?: A Study of Touch Interaction and Gaze Correlation on Tablets. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '16)*. ACM, New York, NY, USA, 329–330. https://doi.org/10.1145/2857491.2888592

Åsa Wengelin, Mark Torrance, Kenneth Holmqvist, Sol Simpson, David Galbraith, Victoria Johansson, and Roger Johansson. 2009. Combined eyetracking and keystroke-logging methods for studying cognitive processes in text production. *Behavior Research Methods* 41, 2 (2009), 337–351. https://doi.org/10.3758/BRM.41.2.337

Pingmei Xu, Krista A Ehinger, Yinda Zhang, Adam Finkelstein, Sanjeev R Kulkarni, and Jianxiong Xiao. 2015. TurkerGaze: Crowdsourcing Saliency with Webcam based Eye Tracking. *arXiv preprint arXiv:1504.06755* (2015).