Inferring target locations from gaze data: A smartphone study

Stefanie Mueller
ZPID - Leibniz Institute for Psychology Information
Trier, Germany
stm@leibniz-psychology.org

ABSTRACT

Although smartphones are widely used in everyday life, studies of viewing behavior mainly employ desktop computers. This study examines whether closely spaced target locations on a smartphone can be decoded from gaze. Subjects wore a head-mounted eye tracker and fixated a target that successively appeared at 30 positions spaced by 10.0 x 9.0 mm. A "hand-held" (phone in subject's hand) and a "mounted" (phone on surface) condition were conducted. Linear-mixed-models were fitted to examine whether gaze differed between targets. T-tests on root-mean-squared errors were calculated to evaluate the deviation between gaze and targets. To decode target positions from gaze data we trained a classifier and assessed its performance for every subject/condition. While gaze positions differed between targets (main effect "target"), gaze deviated from the real positions. The classifier's performance for the 30 locations ranged considerably between subjects ("mounted": 30 to 93 % accuracy; "hand-held": 8 to 100 % accuracy).

CCS CONCEPTS

• **Human-centered computing** → **Smartphones**; *User studies*; *Empirical studies in HCI*; • **Applied computing** → Psychology.

KEYWORDS

fixations, mobile devices, accuracy, gaze positions

ACM Reference Format:

Stefanie Mueller. 2019. Inferring target locations from gaze data: A smartphone study. In *2019 Symposium on Eye Tracking Research and Applications (ETRA '19), June 25–28, 2019, Denver , CO, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3314111.3319847

1 INTRODUCTION

Smartphones have widely replaced desktop computers, e.g. for browsing the internet, which is mirrored in the emergence of mobile optimized websites or apps. At the same time, mobile eye trackers have evolved to lightweight, versatile devices that allow the recording of gaze behavior under natural, real-life-like conditions as opposed to constrained lab settings. Yet, only few studies exist combining these technologies to investigate human gaze in a naturalistic, every-day mobile device setting. Some studies have so far investigated gaze as input medium to control mobile devices (e.g. [Paletta et al. 2014], [Liu et al. 2015]). Studies, recording gaze to examine the observer's behavior, are however scarce. The only two studies that the author is aware of ([Tupikovskaja-Omovie et al. 2015], [Tupikovskaja-Omovie and Tyler 2018]), investigated user experience of commercial apps by analyzing gaze data in terms of dwell durations, heatmaps, and scan paths. An account on the accuracy and thus, validity, of more fine-grained positional gaze data is however lacking. The present study was conducted to close this gap by investigating gaze positions relative to given closely spaced target locations. More specifically, participants were asked to fixate target locations and it was analyzed 1) whether their measured gaze positions differed between target locations, 2) whether gaze data reflected the actual target locations, and 3) whether target locations can be identified based on gaze positions. Overall, this serves to estimate potential errors/deviations between gaze and actual target position and can be used to correct for them.

Tracking gaze in human-smartphone-interactions poses a variety of challenges. Besides the smaller screen, the downwards viewing angle causes the eyelid to close, potentially covering parts of the pupil. Moreover, both the smartphone and the head are more prone to movement than in a traditional lab setting in which a chin rest and a fixed monitor are used, thus potentially affecting the mapping between pupil position and world which is established during calibration.

To account for these facts, two conditions were employed. In the "mounted" condition, the smartphone was clamped in front of the laptop used for calibrating the eye tracker. Thus, the smartphone was fixed during the experiment and remained in the plane of calibration. In the "hand-held" condition, participants held the mobile device in their left hand which allowed adopting a natural hand position and free movement of head and hand.

2 METHODS

2.1 Participants

Five volunteers (1 female) participated in the experiment, aged 27 - 39 years (mean, M = 32 yr; standard deviation, SD = 4.85 yr). Written informed consent was provided before participation. One subject had to be excluded from the final sample because tracking was severely disrupted by the eye lashes covering the pupil when the subject gazed downwards. The final sample (1 female) consisted of four participants aged between 28 and 39 years (M = 33.25 yr, SD = 4.57 yr).

2.2 Materials and Apparatus

Participants were seated at a table. Their gaze data was recorded with an eye tracker by pupil labs ([Kassner et al. 2014]) consisting

ETRA '19, June 25–28, 2019, Denver , CO, USA

ACM ISBN 978-1-4503-6709-7/19/06…$15.00
https://doi.org/10.1145/3314111.3319847

of a headset equipped with three cameras: Two eye cameras recording gaze data at a sampling rate of 120 Hz, and a world camera recording a video from the observer's point of view (field of view = 100 deg). The calibration of the eye tracker was performed on a laptop with the Pupil Capture App (v1.9-7) using the 5-points screen marker calibration (3D eye model). The laptop's screen (14 inch) was tilted at an angle of 225 deg relative to the table surface and was located at a distance of 26.5 cm relative to the edge of the table, rendering an eye-to-screen distance of approximately 43 cm. The screen orientation was intended to be similar to the viewing plane of a smartphone screen. The calibration was run before each condition and repeated until results were satisfying (< 2.0 deg accuracy; 1.5 - 2.5 deg is denoted as normal range in the documentation of the software).

The experiment was performed on a smartphone (Motorola Moto z3 play; display diagonal = 6 inch, display size = 136 x 68 mm, display resolution = 2160 x 1080 pixel, 402 PPI, Android 8.1.0). A frame of white cardboard surrounded the display of the smartphone. On the four edges of the frame, fiducial markers were placed, thus defining a surface (148 x 148 mm) for later analyses. Figure 1 illustrates the layout of surface, smartphone, and target locations. The experimental procedure was programmed with the software Expyriment ([Krause and Lindemann 2014]) utilizing the Expyriment Android Runtime application (v0.1.0). After recording, gaze data was analyzed offline with the Pupil Player App (v1.9-7) and exported for further analyses that were conducted with the software R ([R Core Team 2018], packages used: [Wickham 2018], [Williams et al. 2018], [Dimitriadou et al. 2008], [Bates et al. 2014]).

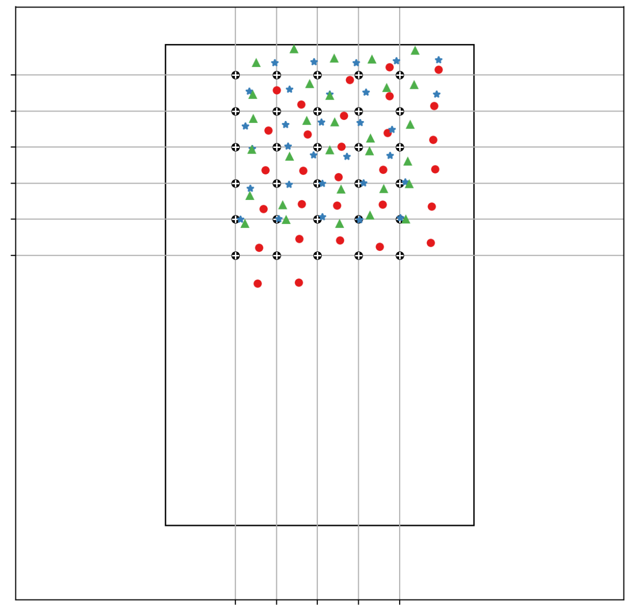

Figure 1: Target configuration (black circles) relative to the smartphone (inner rectangle) and to the defined surface (outer rectangle). Gaze locations of participant S2 in the three blocks (different symbols) of the "hand-held" condition are superimposed. The point of origin was in the lower left corner of the surface defined by the fiducial markers (outer rectangle).

2.3 Procedure

Participants received an oral instruction and watched a couple of practice trials on the mobile phone until they felt familiar with the task. The task was to fixate the center of the target as long as the target was visible. After practice, participants put on the eye tracker. The experimenter adjusted the cameras until the pupil was detected reliably for different eye positions. Calibration was performed and repeated until an accuracy of at least 1.9 deg was achieved. In the "mounted" condition, the experimenter placed the smartphone in the foamed plastics frame and mounted it on the screen of the laptop. The screen of the laptop was fully covered. In the "hand-held" condition, the laptop was moved to the side and turned away from the participant. The mobile phone was given to the participants who were asked to hold it in their left hand without any further instructions about its positioning.

The experimenter started the recording of the eye tracker and the experiment. In every trial, the target was shown for 2000 ms followed by 200 ms of a blank screen, before the next trial began. Every twelve trials, a break of 5000 ms was included as signaled by the screen changing to a black background and showing a countdown from five to one seconds. In total, 180 trials were performed resembling three blocks, each consisting of 30 trials, in two conditions. Blocks differed by the order in which target positions were shown. In block 1 and 2, the presentation of target locations followed a pattern: The first trial showed the target at the upper left corner and then it jumped to the adjacent position in the next trial, either in a horizontal or vertical direction (counterbalanced across participants). After the row/column was completed, the target jumped to the closest position of the next row/column. In block 3, the order of target locations was randomized. Both conditions were performed successively on the same day. Overall, the experiment took about 35 minutes to complete; 4 min per condition and approximately 25 min for setting up the eye tracker.

2.4 Analysis

Gaze data exported within the defined surface was analyzed in R. The lower left corner of the surface constituted as the point of origin (0,0) for gaze and target locations in mm. The trial timing controlled by the smartphone and the gaze data recorded by the laptop were synchronized offline based on the video of the world camera. Specifically, the first time frame in the video that showed the smartphone with a white screen (screen changed from black to white when the experiment started) defined the onset of the experiment (time = 0) for the gaze data. The mean gaze position of every trial was computed by averaging across gaze positions occurring in a time window of 500 to 2000 ms relative to trial onset. Positions exceeding the trial mean by +/- 2 standard deviations were excluded and the trial mean was recalculated.

In order to test whether mean gaze positions varied as a function of condition and target location, linear mixed effect models were fitted to horizontal and vertical gaze positions. Condition

Table 1: Key statistics as summarized by confusion matrices.

		Accuracy Score	95 % Confidence Interval	No information rate (NIR)	P-Value (Acc > NIR)
"mounted"	S1	0.30	0.15, 0.49	0.17	.051
	S2	0.87	0.69, 0.96	0.07	< .001
	S3	0.43	0.25, 0.63	0.10	< .001
	S4	0.93	0.78, 0.99	0.07	< .001
"hand-held"	S1	0.23	0.10, 0.42	0.13	.095
	S2	0.63	0.44, 0.80	0.10	< .001
	S3	0.08	0.01, 0.26	0.28	.997
	S4	1.00	0.88, 1.00	0.03	< .001

(mounted/hand-held) and target locations (5 horizontal; 6 vertical) composed the fixed effects (including an interaction term), whereas participants were treated as random effects. P-values were obtained by likelihood ratio tests of the full model against the model without the fixed effect in question. Significant main effects were followed up by post-hoc t-tests.

To examine the spatial congruence between target locations and gaze positions, horizontal and vertical root-mean-squared errors (RMSEs) were calculated between the given target location and the observed gaze position in every trial for every participant and condition. First, linear mixed models (fixed effects: condition, target location; random effect: participant) were fitted to the horizontal/vertical RMSEs averaged across blocks to determine whether RMSEs varied as a function of target location or condition. Second, t-tests were performed to test whether RMSEs were greater than zero, thus indicating a significant difference between target and gaze position.

To investigate whether the target location that was fixated by the participant in a trial can be correctly identified based on the (two-dimensional) gaze data in that trial, a k-nearest neighbors algorithm (k = 3) was trained on the data of block 1 and 2 and tested on the data of block 3 for every participant and condition. The classifier's performance was evaluated by calculating confusion matrices summarizing the key statistics, e.g. accuracy scores. Resulting accuracy scores were further compared between the "hand-held" and "mounted" condition as well as against chance level by calculating paired t-tests across participants.

3 RESULTS

3.1 Horizontal and vertical gaze positions varied with target location

The linear mixed effects analyses of the horizontal and vertical gaze positions revealed a significant effect of target location (horizontal: $\chi^2(1) = 35.53$, $p < .001$; vertical: $\chi^2(1) = 78.79$, $p < .001$), thus indicating that fixating target positions spaced by 10.0 and 9.0 mm also resulted in distinguishable gaze positions. The main effect "target location" was further explored by post-hoc paired t-tests comparing gaze locations to adjacent targets. Results showed significant differences for all horizontal (t's > 6.68, p's < .003) and vertical (t's < -4.93, p's < .008) locations. The mean difference of gaze positions to adjacent targets equaled 10.2 mm in the horizontal, and 8.8 mm in the vertical direction.

Condition had an effect only on vertical gaze positions ($\chi^2(1) = 35.85$, $p < .001$) as indicated by gaze positions that were on average 15.1 mm higher (standard error, SE = 2.03) in the "hand-held" rather than the "mounted" condition. No interaction effects were found. Figure 2 shows the horizontal and vertical gaze positions as a function of target location for individual participants in the two conditions.

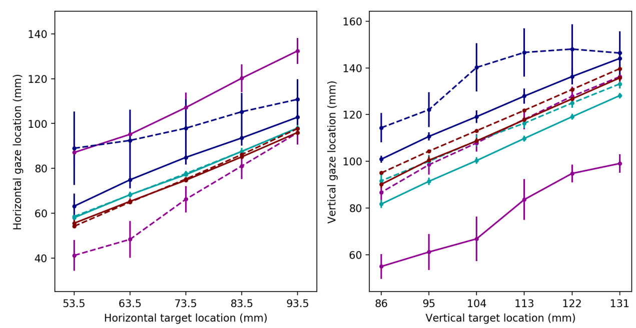

Figure 2: Gaze locations as a function of target locations averaged across blocks. Tick marks on the x-axis correspond to the target locations. Individual lines denote participants; solid lines: "mounted" condition; dashed lines: "hand-held" condition.

3.2 Gaze locations deviated from target locations

Horizontal and vertical RMSEs did not significantly vary across conditions and targets, nor was there an interaction effect (horizontal: χ^2's < 1.05, p's > .593; vertical:χ^2's < .354, p's > .552). Thus, to determine whether there was a significant offset between gaze and target, RMSEs were averaged across targets and conditions. The t-test against zero yielded a significant difference (horizontal: $t(3) = 21.82$, $p < .001$; vertical: $t(3) = 15.39$, $p < .001$) as gaze and target deviated on average by 8.3 mm in the horizontal and 11.2 mm in the vertical dimension. This deviation closely resembles the inter-target distance (10.0 and 9.0 mm), thus, on average, gaze position coincided with the adjacent rather than the actual target (note that only one target was visible at a time).

3.3 Decoding accuracies for target locations

Accuracy scores varied considerably between participants, ranging from 30 to 93 % in the "mounted" and from 8 to 100 % in the "hand-held" condition. Table 1 shows a summary of the results of the classification analysis. Accuracy scores did not differ between conditions (t(3) = 1.59, p = .210). Overall, the accuracy scores averaged across conditions exceeded chance performance (1/30 = 0.03) as indicated by a t-test (t(3) = 2.96, p = .030).

4 DISCUSSION AND CONCLUSION

The present proof-of-concept study investigated the accuracy/validity of gaze positions when participants fixated targets on a mobile phone. Results showed that gaze deviated from the target positions (RMSEs > 0), thus revealing a bias that prevents, or at least urges caution, to superimpose measured gaze uncorrected on the presented stimuli. Differences between gaze positions still reflected the pattern of target locations (main effect target location and t-tests) and individual targets could be classified above chance level implying that a correction by a simple (linear) transformation may be possible.

The mounted and the hand-held condition did not differ in terms of RMSEs and classification accuracies. Anecdotal observations showed that participants mostly kept their head and hand in the same position, even in the "hand-held" condition, although they were not explicitly told so. In addition, the screen of the laptop on which the phone was clamped in the "mounted" condition was adjusted to resemble the orientation and distance of a hand-held smartphone, thus rendering both conditions similar by design.

Fiducial markers were used to define a surface around the smartphone to facilitate offline analyses. Although the frame on which the markers were placed did not obstruct grasping the mobile phone, the setup can be improved. An alternative solution has recently been proposed by MacInnes et al. ([MacInnes et al. 2018]). They suggest a feature matching algorithm that is based on the characteristics of a natural surface (rather than defining a surface explicitly by markers) and which is applied automatically on every frame recorded by the world camera during offline analyses.

For future studies, we plan to present a similar task and target pattern before and/or after conducting the actual experiment to estimate and then account for the bias between gaze and target locations. Data and analyses presented in this manuscript as well as additional data acquired after submission, are available at http://dx.doi.org/10.23668/psycharchives.2384 and http://dx.doi.org/10.23668/psycharchives.2383, respectively.

ACKNOWLEDGMENTS

To Immo Schütz and Mathias Klinghammer, for proofreading this manuscript and helpful comments.

REFERENCES

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2014. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823* (2014).

Evgenia Dimitriadou, Kurt Hornik, Friedrich Leisch, David Meyer, and Andreas Weingessel. 2008. *Misc functions of the Department of Statistics (e1071), TU Wien.*

Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: Adjunct publication.* ACM, 1151–1160.

Florian Krause and Oliver Lindemann. 2014. Expyriment: A Python library for cognitive and neuroscientific experiments. *Behavior Research Methods* 46, 2 (2014), 416–428.

Dachuan Liu, Bo Dong, Xing Gao, and Haining Wang. 2015. Exploiting eye tracking for smartphone authentication. In *International Conference on Applied Cryptography and Network Security.* Springer, 457–477.

Jeff J MacInnes, Shariq Iqbal, John Pearson, and Elizabeth N Johnson. 2018. Wearable Eye-tracking for Research: Automated dynamic gaze mapping and accuracy/precision comparisons across devices. *bioRxiv* (2018), 299925.

Lucas Paletta, Helmut Neuschmied, Michael Schwarz, Gerald Lodron, Martin Pszeida, Stefan Ladstätter, and Patrick Luley. 2014. Smartphone eye tracking toolbox: accurate gaze recovery on mobile displays. In *Proceedings of the Symposium on Eye Tracking Research and Applications.* ACM, 367–68.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Zofija Tupikovskaja-Omovie and David Tyler. 2018. Mobile consumer shopping journey in fashion retail: eye tracking mobile apps and websites. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications.* ACM, 89.

Zofija Tupikovskaja-Omovie, DJ Tyler, Sam Dhanapala, and Steve Hayes. 2015. Mobile App versus Website: A Comparative Eye-Tracking Case Study of Topshop. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering* 9 (2015).

H Wickham. 2018. *Francois R. dplyr: A Grammar of Data Manipulation. R package version 0.4. 3. 2015.*

Chris Keefer Williams, Allan Engelhardt, Tony Cooper, Zachary Mayer, Andrew Ziem, Luca Scrucca, Yuan Tang, Can Candan, Tyler Hunt, and Maintainer Max Kuhn. 2018. *Package 'caret'.*