Reading and Estimating Gaze on Smart Phones

Ralf Biedert*
Andreas Dengel†
German Research Center for
Artificial Intelligence (DFKI)

Georg Buscher‡
Microsoft Bing

Arman Vartan§
Technical University Kaiserslautern

Abstract

While lots of reading happens on mobile devices, little research has been performed on how the reading-interaction actually takes place. Therefore we describe our findings on a study conducted with 18 users which were asked to read a number of texts while their touch and gaze data was being recorded. We found three reader types and identified their preferred alignment of text on the screen. Based on our findings we are able to computationally estimate the reading area with an approximate .81 precision and .89 recall. Our computed reading speed estimate has an average 10.9% wpm error in contrast to the measured speed, and combining both techniques we can pinpoint the reading location at a given time with an overall word error of 9.26 words, or about three lines of text on our device.

CR Categories: H.5.2 [Information Systems]: User Interfaces—Evaluation/methodology;

Keywords: touch, smart phone, reading, eye tracking

1 Introduction

Recently the sales of digital books began to surpass the number of paper sales [Miller and Bosman 2011] and probably this trend will not reverse. While reading and interacting with web sites or PDFs on desktop PCs has become very common, the trend to mobile reading devices is quite new. Tablets, cell phones and dedicated ebook readers appear to become the device-of-choice, and a variety of them manifested during the last months and years. With them arrive new interaction paradigms [Shneiderman 1991], and traditional keyboard-mouse interaction is more and more replaced by small, touch-sensitive fullscreen reading devices. At the same time, traditional eye tracking and reading research [Rayner 1998] has shown that by taking into account gaze and interaction data and putting them into relation with the displayed content, there is a considerable potential for improving human-computer interaction techniques [Biedert et al. 2010b][Buscher 2010]. There have been interaction [Drewes et al. 2007] and reading [Oquist and Lundin 2007] studies on traditional cell phones and we believe on top of that touch interface with smoothly scrollable screens offer some unique insights and interaction possibilities. For this paper our goals are therefore twofold. First we want to explore how the reading interaction usually takes place on these devices, especially in terms of eye movements and touch behavior. Second we want to investigate to what extent we can approximate the reader's current focus of attention (i.e., the read text) by analyzing the available inputs. Such an approximation can enable us to provide eyeBook [Biedert et al. 2010a] like ambient reading effects on existing hardware without the need for expensive eye tracking equipment. If properly estimated interaction data is being considered for many users, it also could give valuable insights on possible problematic passages within the text [Biedert et al. 2012].

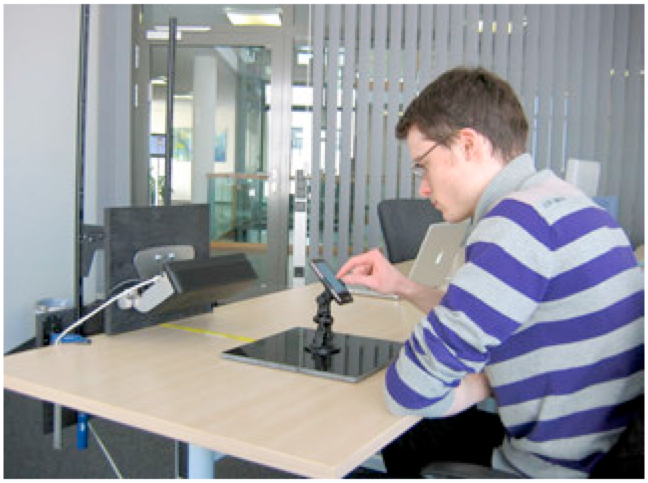

Figure 1: *A user reading texts on a mobile phone while being eye tracked. While he reads, the current touch and gaze information is being recorded as well as the screen's content for a later analysis. Notice that the tracking device is flipped and all calibration is performed directly on the smart phone.*

2 Setup

In order to correlate eye tracking data with scroll and touch information we need to integrate an eye tracking device into our interaction scenario. For the purpose of our study we use a Nexus One as the actual presentation device, which contains a 480x800px display with a screen diagonal of 94mm. It also has a capacitive touch sensor, a built-in HTML rendering engine and WiFi networking facilities, and we rely on all of them for the creation of our experimental application. For the purpose of our study the devices is fixed on a table and used in portrait mode.

As the tracking device we use a Tobii X120 unit which we integrated in a novel setup, allowing us to calibrate and record eye tracking data without the need of external cameras. For that we mount the unit head-down (compare Figure 1) to ensure that it can properly track the user's eye which when looking at the device which is placed *below* the tracker. Furthermore, the incoming gaze data needs then to be post-processed to match the flipped order of axes.

3 Experiment

Using the setup described in the previous section we conducted a user study to investigate how people actually interact with mobile devices. We drafted 18 students from the local university, 13 male, 5 female, all aged between 18 and 28 years, most of them were

*e-mail: ralf.biedert@dfki.de
†e-mail: andreas.dengel@dfki.de
‡e-mail: georgbu@microsoft.com
§e-mail: arman.vartan@gmail.com

ETRA 2012, Santa Barbara, CA, March 28 – 30, 2012.
© 2012 ACM 978-1-4503-1225-7/12/0003 $10.00

computer science students and all of them had German as their native language. We also surveyed the user for previous smartphone experience and 8 users reported they had used or do use such devices, and all were given time to familiarize with the device.

The participants were told to participate in a reading comprehension study. We asked them to read three HTML documents on the topics of cosmetics (D_1), biology (D_2) and gardening (D_3), excerpts from 'excellent' German Wikipedia articles, and answer a number of questions afterwards. The texts were about 530 words each, with different paragraph lengths on average (9, 12 and 15 screen lines per paragraph, respectively). The whole device could display approximately 15 lines of text and a single line contained approximately 5 words.

After the instructions were given the users familiarized themselves with the device. Eventually the eye tracker was calibrated and the actual experiment started with the documents presented in random order.

4 Evaluation

Using the data acquired in the experiment we start by analyzing a number of principal questions and eventually investigate how well the true reading position can be estimated by heuristics based on touching and scrolling behavior.

4.1 General Definitions

On the display of the device, a document D_i is presented to the reader. More specifically, at each moment $t \in T$ during the document interaction time T only a part of the document $p(t)$ is visible on the screen. There are two major ways to represent the viewport (the visible area of the document): a pixel-based variant that maps to two y-coordinates (y and $y + 800$), and a character-based variant that maps to two character offsets (o_1 and o_2, the offsets that were fully visible in the upper left and lower right part of the screen at t), counted from the start of the text. With the help of p we segment our recordings into two classes, namely *reading phases* and a *scrolling phases*. As reading we consider phases $r_i \subseteq T$ with $p'(r_i) = \{0\}$ (i.e., the first derivation of p) and $\Delta r_i > 2s$, i.e., where the content of the screen stood still for more than 2 seconds (a value empirically found when analyzing the recordings). We consider each block of time that does not form a reading phase as a scrolling phase s_i, and for each document interaction timeline we now have a partition into the set R of all reading phases and a set S of all scrolling phases. In addition to the scroll movements we also recorded gaze data $g(t)$ and touch data $f(t)$, and similar to p their values can be interpreted as raw screen position or character offsets in the document.

4.2 General Behavior and Gaze Distribution

We start by describing the overall distribution of reading areas. Since previous research has indicated that favored reading areas might exist on desktop screens [Buscher et al. 2010] the main question in this part is whether they also exist on small screens where only a very limited number of lines can be displayed at a time.

Visually inspecting the overall scrolling behavior (compare Figure 2 for a general overview) we noticed three general classes of readers. Four of our readers employed a reading pattern that is mostly page-wise, i.e., they read one page more or less completely and then scroll so that all the screen's content is replaced with new text. In contrast, four others exhibited mainly line by line reading behavior in which they tend to focus on a single or very few lines on the screen. They scroll almost constantly to keep new information *flowing* into that preferred area. The majority of our users (10) however preferred mostly blockwise scrolling, in which they changed only parts of the screen with each scrolling phase.

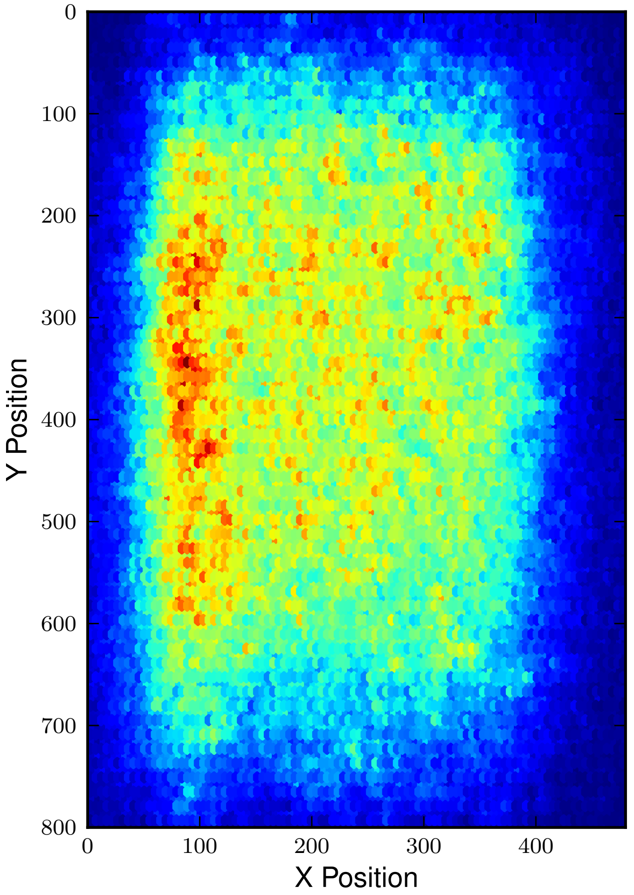

Figure 2: *Heat map with the general distribution of gaze data for all users and all documents. Although no top or bottom bar were present not the full height of the screen was generally being used.*

It should be noted that these modes are not strict. Individually also a mixing of different behaviors can be observed, such as when a paragraph does not fit into the entire screen it is read on a line-by-line basis, and when another can be fit again, the reading pattern changes back to a blockwise mode. When looking only at the mostly non-fullscreen readers and investigating their average gaze distribution over the whole document, preferred reading regions can also be observed here. While the average upper placement position was ranging between 8% and 15% screen height we could notice that with an increase of average paragraph length the reading bounds shifted outwards. It appears that this is caused by the general preference of these readers to align paragraphs so that they can be read in whole.

4.3 Correlation of Touch and Gaze

We also investigate to what extent gaze and touch behavior correlate. The main question is whether the touch-down or touch-up position, i.e., the vertical position where the finger touched or left the screen for a swipe, related directly to the reading bounds. From a coarse view the test group can be separated in a small group of 3 people out of 18 which used a dedicated portion of the screen for short and more frequent scrolling. The remaining users usually used the full length of the screen. In addition there is the obvious finding that all right handed persons used the right side of the screen, while all left handed persons used the left side of the screen. Overall each participant had an individual swipe pattern which was uniquely distinguishable from the other patterns.

The first analysis indicates that the majority performed their touch movements within the middle 80% of their individual reading area with a noticeable scatter increasing from the touch-up ($\mu = 448px$) and touch-down ($\mu = 152px$) region. For further investigation the Pearson correlation coefficient between the measurement parameters were calculated resulting in a very low correlation of $r = .094$ for the touch-down position and the bottom bound of the reading area. The same applied to the touch-up position and the top bound of the reading area $r = .164$.

4.4 Locating the Reading Area

In order to estimate the actual reading position we first have to determine which part of the screen we actually consider for its computation. We base our algorithm on two key observations: First, the non-fullscreen users mostly seem to ignore the upper and lower parts of the screen in general. Second, users instead tend to align paragraphs so that they can be read in total. We combine both observations into our area estimator $a(r_i)$. Given a reading slice r_i we analyze the document content within $p(r_i)$ and locate the paragraph with the start location most proximate to $y = 120px$ (which is equal to 15% screen height). If the entire paragraph fits onto the screen, we assume its bounds to be the reading area, otherwise we assume the remaining part of the screen to be the reading area.

In order to evaluate this heuristic, we excluded the data of full screen readers (for which the bounds are obviously known, and which can be detected automatically since their $p(r_i)$ results are non-overlapping) and line-wise readers (which we deemed to be too difficult and which can probably be detected automatically by an analysis of their $\Delta r_i : \Delta s_{i+1}$ ratios) we considered the remaining blockwise readers that had sufficiently good eye tracking (7 total[1]) data. For these we compared the computed regions with the actually measured gaze regions as defined by the highest and lowest fixation in $g(r_i)$. The results of this analysis can be seen in Figure 3 in which we plotted the precision and recall values of the area of the true reading regions compared against the estimated reading regions. Overall we achieve a precision of .81 and a recall of .89 for the true reading regions. A precision of 1.0 means that all of the estimated reading area was actually read, a recall of 1.0 that all of the area that really was used for reading was also being detected as such.

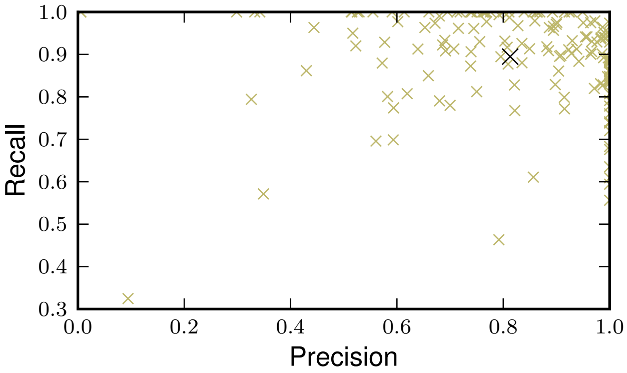

Figure 3: *Precision-recall map of the computed reading areas, in contrast to the real reading areas for all users and all documents.*

4.5 Estimating the Reading Speed

The main assumption in the computation of the reading speed is that it is approximately constant within one r_i slice. While in reality the fixation and saccade pattern is influenced by many factors, the actual slices are our atomic observation unit and therefore we take the a priori assumption that the time spent in each part of it is evenly distributed. Thus, for the estimation of the reading speed of given slice r_i we considered the amount of text available within $a(r_i)$ and the time Δr_i this text was available. This is simply the number of words within that area divided by the time taken for the area, $v^*(r_i) = a(r_i) : \Delta r_i$

For estimating reading speed, we measured how many words were presented between the first and last fixations of a given reading slice and also the time the slice was displayed. Comparing both readings we can observe an average error rate of about 17.6% in terms of words per minute. Overall we measured real average reading speeds in the range from 174 *wpm* to 272 *wpm* per session[2], and the errors reported for the individual users varied quite widely. However, when comparing the relative errors our algorithm produced between the three documents, we could not find a significant difference ($p = .55$) on a Kruskal-Wallis test. In general the computation appears to have the slight tendency to underestimate the true reading speed.

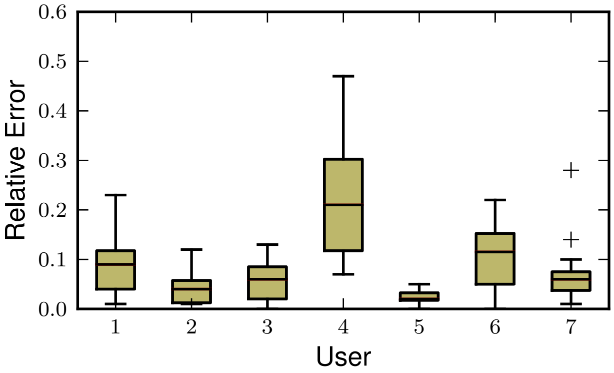

Figure 4: *Relative errors when estimating the current speed per user. While the inter-document differences were not significant, the differences between users vary quite widely.*

In terms of realtime computation the presented numbers can only be computed for the previous slice since they require knowledge for how long the slice has been presented. Thus we estimate the most likely current speed to be the moving average of all previously observed slices, hence $v_i = \varnothing\{ v^*(r_j) \}$ for all $j < i$. Using v_i instead of $v^*(r_i)$ for a comparison against the actual reading speed yields an even better estimate and achieved a 10.9% total average wpm error, compare Figure 4.

4.6 Pinpointing the Reading Position

Based on the estimation of the reading bounds and the approximation of the current reading speed, we eventually try to estimate the most likely reading location l for a given time. Taking the currently displayed reading segment r_i we measure the time Δt elapsed since its start and compute an offset by multiplying it

[1] Apparently our number of dropouts (3 of effectively 10) is approximately in the range of the numbers reported in [Oquist and Lundin 2007], which was 4 of 16. We considered data to be good when it could be matched to the text in most cases.

[2] Our recorded reading speeds are above the reported speeds of 178 *wpm* mentioned in [Oquist and Lundin 2007] for scrolling interfaces which can probably be attributed to the differences in content (sci-fi prose vs. encyclopedic articles), the language (Swedish vs. German) and the way the interfaces are operated and sized (small key-operated cell phone vs. relatively large touch operated smart phone)

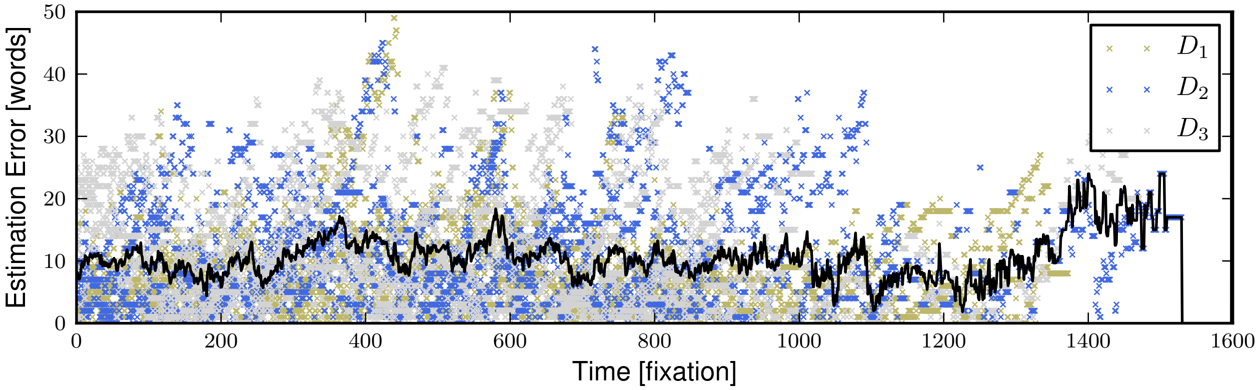

Figure 5: *Visualization of the data on all documents and all block-wise readers we were able to extract proper reading positions from. The x-axis reflects the nth fixation, i.e., the time. The y-axis shows how many words the computed position was away from the real position. The black line reflects the average error over all users and documents. It can be seen that, except for some outlines within two exceptionally long document passes, there is no obvious trend for the accuracy to degrade over time. If there were a general user drift, the average should increase within the latter parts.*

with the average speed as described in the previous chapter, i.e., $l(r_i + \Delta t) = a(r_i) + v_i \Delta t$.

To evaluate the quality of this approach we compare the computed position l with our actual eye tracking data g. We consider all fixation events that occurred during reading slices r_i and for each fixation we compare the computed word offset against the true word offset in the document. The overall average word error we achieve is 9.26 words, with a standard deviation of 8.90 words. Investigating the overall error over time, compare Figure 5, there appears to be no obvious trend for the algorithm to degrade in performance over time. While individual measures can be far off (which can be caused, for example, by *accidental* saccades to the other end of the text) the average error after the *n*th fixation of runtime remains at around 10 words. We also investigated the individual users and documents. The lowest average error we observed was 6.49 words for one user and one document, the highest average error 16.02 words, with standard deviations ranging from 4.39 to 10.67 words. Expressed in lines, our prediction was, on average, about 3 lines off from the true reading position.

5 Conclusion & Outlook

We implemented a novel setup and investigated how text interaction and reading are performed on a mobile touch screen device. Inviting 18 users we had them read a number of documents and recorded their gaze and touch behavior. We categorize our readers into three types (full screen, linewise and blockwise) and find that blockwise readers tend to align the read paragraph so that it fits on screen in its entirety. Measuring their scrolling speed and modeling our findings in an algorithm we are able to pinpoint their reading position with an average accuracy of 10 words. There are also a number of open questions. For example it is unclear how stable our classification into three reader groups actually is, whether it changes over sessions, days or weeks, and if it changes or converges with smartphone or touchscreen experience. Likewise the fit-to-screen strategy could be verified in an independent experiment. Lastly we can imagine that, although we could not find any obvious relation in reading and touch positions in the scope of our experiment (beyond as a means for scrolling), there are some relations waiting to be uncovered.

References

Biedert, R., Buscher, G., and Dengel, A. 2010. The eyeBook - Using eye tracking to enhance the reading experience. *Informatik-Spektrum 33*, 3 (June), 272–281.

Biedert, R., Buscher, G., Schwarz, S., Hees, J., and Dengel, A. 2010. Text 2.0. *CHI EA '10: Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems* (Apr.), 4003–4008.

Biedert, R., Hosseiny, M., Georg, B., and Dengel, A. 2012. Towards Robust Gaze-Based Objective Quality Measures for Text. In *Seventh ACM Symposium on Eye Tracking Research & Applications (ETRA)*, German Research Center for Artificial Intelligence.

Buscher, G., Biedert, R., Heinesch, D., and Dengel, A. 2010. Eye tracking analysis of preferred reading regions on the screen. In *CHI EA '10: Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, New York, NY, USA, 3307–3312.

Buscher, G. 2010. *Attention-Based Information Retrieval*. PhD thesis, University Kaiserslautern, Kaiserslautern.

Drewes, H., De Luca, A., and Schmidt, A. 2007. Eye-gaze interaction for mobile phones. *Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology*, 364–371.

Miller, C. C., and Bosman, J. 2011. E-Books Outsell Print Books at Amazon. *New York Times* (May), B2.

Oquist, G., and Lundin, K. 2007. Eye movement study of reading text on a mobile phone using paging, scrolling, leading, and RSVP. *Proceedings of the 6th international conference on Mobile and ubiquitous multimedia*, 176–183.

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

Shneiderman, B. 1991. Touch screens now offer compelling uses. *Software, IEEE 8*, 2, 93–94.