# Understanding Gaze and Scrolling Strategies in Text Consumption Tasks


**Jayson Turner**
Lancaster University
Lancaster, LA1 4WA, UK
jaysont@acm.org

**Shamsi Iqbal**
Microsoft Research
Redmond, WA 98052, USA
shamsi@microsoft.com

**Susan Dumais**
Microsoft Research
Redmond, WA 98052, UAS
sdumais@microsoft.com





*Ubicomp/ISWC'15 Adjunct,* September 07-11, 2015, Osaka, Japan
© 2015 ACM. ISBN 978-1-4503-3575-1/15/09…$15.00
DOI: http://dx.doi.org/10.1145/2800835.2804331



## Abstract
Scrolling is an integral part of our everyday computing experience, and many techniques and devices have been developed to enhance scrolling. We have conducted an 18 participant user study to understand how users' gaze position and scrolling strategies are coordinated. Our data showed that people scrolled within preferred reading regions of the screen. Scrolling patterns were also influenced by document structure, with most scrolling occurring at the intersection between paragraphs, in order to bring new text into the preferred reading region. We conclude with the implications of our findings with a new gaze-enhanced scrolling technique aimed to support behaviors evident in manual scrolling.



## Author Keywords
Gaze-enhanced scrolling; eye-based interaction; scrolling strategies; reading behavior; document structure;

## ACM Classification Keywords
H.5.2. Information interfaces and presentation (e.g. HCI): User interfaces – Input devices and strategies; Evaluation/methodology;

## Introduction

Eye gaze coupled with navigation form the building blocks of how one reads digital documents. When reading digital documents, eye gaze shows the loci of user attention, and navigation is controlled explicitly by the user, usually by scrolling with the mouse wheel, keyboard, or touch input. As documents typically extend beyond the constraints of current displays in terms of length (and occasionally width), scrolling enables parts of the whole to be brought into focus in a controlled manner. Users are able to read comfortably, controlling the rate and quantity of new information delivered with each scroll.

The research we present aims to move toward a form of scrolling controlled solely by the eyes, namely by understanding how gaze and manual scrolling are coordinated during normal text consumption.

Automatic scrolling is important in application contexts where manual interaction needs to be reduced, or catering to users' desire to scroll without explicit execution. Automatic scrolling is also desirable in situations where other forms of manual input are not available, on public or shared displays [16] on mobile devices [5], or in situations where the hands are busy.

We ask: *Can we identify common gaze and manual scrolling strategies in text consumption tasks?* To address this question, we collect data in reading and skimming tasks to analyze three factors of users' scrolling behavior: gaze location in the viewport (part of the document currently in view), gaze location within the document (relative to document structure), and distance scrolled. We then discuss how these findings could be used to develop a new gaze-enhanced scrolling technique.

## Related work

Works by Buscher et al. have shown how gaze can be used to enhance reading experience. *EyeBook* [1] and *Text 2.0* [2] demonstrated gaze triggered reading annotations, changing accompanying images and sounds based on the context of words being read. Other functionality included scrolling and page turning.

In 1991 Jacob et al. conducted fundamental studies into the use of the eyes as input for human-computer interaction [7]. The work suggested several applications for the eyes including gaze-activated scrolling. Work by Kumar and Winograd developed and evaluated techniques that enable implicit and explicit gaze-enhanced scrolling [10]. The work explored several approaches using fixed regions that, when gazed upon, would trigger scrolling. Users were able to explicitly activate/deactivate automatic scrolling using a dedicated key press.

Several techniques in Kumar's work were based on the hypothesis that users tend to read within the central third (1/3) of the viewport. Work by Buscher et al. found that users have individually preferred reading regions [4]. During reading, gaze is concentrated around the middle of the screen and the vertical spread of gaze is greater toward the left screen border. Additionally, they found variation in the height and position of preferred reading regions, with individuals tending to prefer either the top, middle, or bottom of the screen. A small correlation was found between scrolling distance and the distribution of gaze, suggesting that frequent scrolling entailed reduced spread. In our analysis we consider the regions and scroll distance measures highlighted by this work.

Sharmin et al. [14] conducted an initial study to find preferred reading regions of users. A new automatic scrolling technique was developed to support users' preferred regions. Preferred regions were determined by the average range of reading-related fixations between scroll events, and an iterative approach was used to remove outliers. The automatic scrolling technique they developed initiates smooth scrolling when gaze crosses the lower bound of the average range, scroll speed is increased when gaze is one standard deviation from the lower bound. This behavior ensures new text is scrolled in a continuous smooth manner into the preferred reading region. No significant effects of manual vs. automatic scrolling were observed. A significant effect was found for font sizes, with larger font sizes resulting in shorter preferred regions.

Their work was later extended to analyze gaze patterns during automated scrolling. The analysis focused on the smooth scrolling approach implemented previously. It was shown that users read in a more regular pattern than observed in normal reading [12].

Our observations are contrary to Sharmin et al. [14]. We found that users did not naturally scroll in a smooth continuous manner, but instead several lines at a time. Additionally, we found that people typically initiate scrolling at paragraph and section boundaries. By taking both preferred reading regions and document structure into account we show that users' manual scrolling strategies relate to these factors (i.e., paragraph and heading structure).

*Manual scrolling strategies and reading behaviors*
There have been several investigations around scrolling strategies and reading behaviors. Braganza et al. conducted a comparison of scrolling behavior in vertical vs. horizontal textual layouts [3]. Users felt more comfortable in a horizontal layout over vertical. They found that 46% of users scrolled a few lines at a time in a vertical layout, 13% scrolled by page. In a horizontal layout, 64% scrolled one or two columns at a time and 50% scrolled by page. They found fixations most likely to be in the bottom region of the screen and that users generally scrolled small portions of text. Interestingly, Oquist et al. found that scrolling line-by-line on mobile devices caused slower reading behavior when compared to paging or marquee style scrolled text [11]. Hornbæk et al. also revealed that when reading electronic documents, users do not always read linearly but in fact skip large blocks of text and regress over previously read text [6]. Kelly et al. investigated if reading time, scrolling and clicks can reveal the relevance of a document to a particular topic. Scrolling had no significant indication to relevance [8].

To summarize, the research presented in this paper extends previous work by characterizing people's natural scroll behavior and corresponding gaze patterns during three text consumption tasks (reading, skimming and visual search). We characterize the last fixation before scrolling in terms of screen viewport properties. We also examine structural properties of the document and the distance scrolled which, to our knowledge, have not been previously addressed.

## Data Collection

The goal of this investigation was to determine gaze patterns, if any, that preceded a scroll while consuming text. Users typically consume text by either reading for comprehension or skimming for 'gist'. The former refers to the lexical processing of an entire document to develop a complete understanding of its content. The latter refers in essence to speed reading, the process of searching sentences, often skipping large chunks of irrelevant information, for key words that give insight to the meaning of a text. On average, reading speed is ~250 words per minute (wpm) and skimming speed is ~650 wpm [13]. For our data collection, we record participants' gaze and scroll data during reading and skimming tasks. We also collected data on a third task involving visual search, but we do not analyze or report those results here. However, we mention this condition to describe the details of our experimental design and data collection.

*Data Collection*

We collected data from 18 participants (12 M, 6 F) aged between 24 and 55 (M = 30, SD = 9.2). All participants were employees of a large technology company, and had normal or corrected-to-normal vision, either via glasses or contact lenses. Eight of the 18 participants were non-native English speakers, but all were fluent in the English language. Participants were compensated for their participation.

*Tasks*

Three tasks were used in the experiment (with repeated measures), reading, skimming, and visual search (not presented in this paper). For the reading task, participants were instructed to read a given document of around 500 words at their normal rate to fully comprehend the presented text, to the best of their ability. They were not given any explicit time constraints, but we expected that it would take approximately 2 minutes to read each article based on average reading speeds of 250 wpm. For the skimming task, participants were given 60 seconds to comprehend as much of the text as possible. Similar sized documents were used here too (~500 words). Enforcing a time limit in the skimming condition ensured that participants would not be able to carefully read the entire text and would instead need to skim the text. At the end of each trial, participants answered two comprehension questions based on the text. The comprehension test was used to encourage participants to genuinely comprehend the text and use a typical reading pattern.

*Reading Material*

The experiment consisted of three conditions, corresponding to the three types of tasks. Each condition had 1 training trial and 10 experimental trials. We therefore compiled a collection of 33 documents for use as reading material. We selected pages from the English version of Wikipedia that were popular, written at a broadly accessible level, and cover a range of topical categories. We selected pages from the 5000 most viewed Wikipedia articles during the week 24th June 2013, as reported by Wikipedia [15]. We further filtered pages to those that were classified at a seventh-grade (12-13 years of age) level of reading comprehension, as described in [9]. From this collection, we chose 3 documents at random for each of the following 10 topics: films, medical, music, mythology, nature, objects people, places, sports and technology, and randomly assigned them to each of the 3 categories. For training trials we chose 3 random documents with unrelated topics. Documents were formatted to contain only the text of the main article, with all navigation, images, and hypertext removed.

We hypothesized that readers would use salient features such as section headings for orientation within a text. To ensure the inclusion of headings, articles were formatted to include only 1 paragraph and 1 heading from each section of the original text, up to the

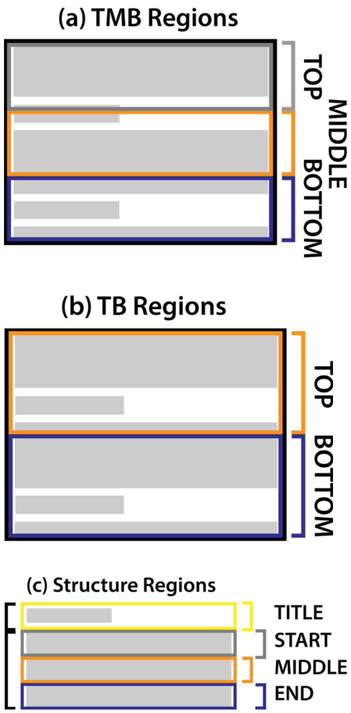


Figure 1: (a) Top Middle Bottom ROIs (b) Top Bottom ROIs (c) Structural ROIs

enforced word limit. This resulted in structured documents that contained from 3-9 section headings and paragraphs. The font size of each document was increased to 32px (2 em) point to improve overall accuracy in linking gaze data to regions of interest locations. The larger font also meant that we would observe more scrolling which is desirable. (We examine more general layouts in a preliminary experiment described later.)

*Experimental Setup*

A Tobii TX300 remote eye-tracking system was used to record gaze data. The system was situated below a 23-inch display (1920 x 1080 pixels (px)). A seat was placed 60-70 cm from the display. Gaze data was recorded at a rate of 300 Hertz (Hz) using Tobii's Studio recording suite. Stimuli were displayed using a separate application. Visually, the application encompassed the entire display with a dark-grey background to remove visual distractions. Stimuli were shown centered on the display with a web browser control 1024 x 768 px in size. The application also recorded scroll events reported from the web browser control. All events were logged and time stamped to microsecond precision (Tobii timestamps). Timestamps were synchronized between the two applications using Tobii's clock synchronization API.

## Procedure

The experiment followed a within subjects design. For each task condition (reading, skimming, visual search), participants completed 1 training trial, during which the task was explained, and 10 experimental trials. A calibration was performed at the beginning of each condition to reduce drift in gaze tracking accuracy over time. Participants were instructed to navigate through each document using only the scroll wheel on the mouse. The experiment took ~1 hour to complete.

We used a Latin-square design to determine the order in which the tasks were presented. For each trial, a document from our collection was presented; the presentation order of documents was randomized for each condition. At the end of each trial, participants were given two multiple-choice questions to assess their comprehension of the article

*Measures*

Two types of data were recorded – gaze data from the eye tracker, and interaction data from the reader application. Gaze data included fixations and saccades. Interaction data included scroll events and keystrokes.

Scroll data was recorded as a stream of timestamps. Since the flicking of a scroll wheel can generate multiple data points in close succession, the first step in our analysis was to identify distinct scroll events. A new scroll event was determined by the time between successive scroll wheel interactions. Following thresholds used in Braganza et al. we used a one second gap between successive scroll data to determine a new scroll event [3].

For each trial, we computed the total number of scroll events and reading time. For each distinct scroll event, we considered 3 dependent measures: (1) Scroll distance, (2) Gaze location within a viewport, (3) Gaze location within the document structure. The viewport is determined by the x, y coordinates on the screen; and the location within the content is based on structural properties of the article (section title, and the start/middle/end of paragraphs).

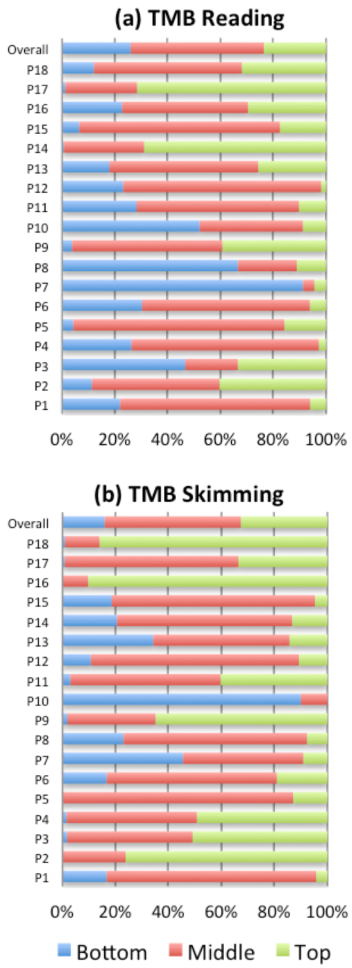


Figure 3: Proportion of fixations before a scroll across viewports for (a) Reading and (b) Skimming for TMB regions.

## Results

In the results that follow, we only report on events where participants scrolled downwards through text.

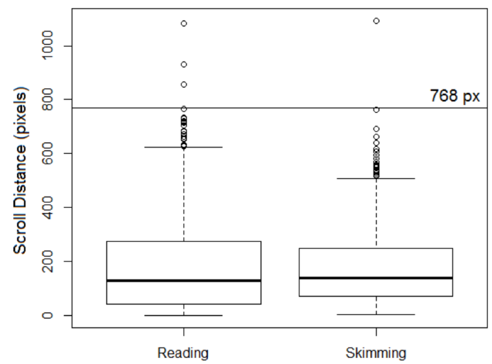


Figure 2: Distance scrolled in reading and skimming tasks. Events greater than 768px indicate a full viewport scroll.

*Reading Speed and Scroll Frequency*

Participants read on average 246 wpm (SD = 60 wpm) in the reading condition. Due to the time constraint enforced in skimming, we did not calculate reading speed for this condition. Across all users, there were 1091 scroll events for reading (180 trials) and 1090 scroll events for skimming (180 trials) in total. Regardless of whether participants were reading or skimming, they scrolled downward 6 times per article on average. This corresponds to a rate of 3.12 scrolls per minute (spm) for reading and 6.93 spm for skimming.

*Scroll Distance*

The box plot shown in Figure 2 shows the scroll distance data through their quartiles. The first, second (median) and third quartiles are shown in the box. The whiskers denote the value that is 1.5 times the interquartile range (IQR), and outliers that fall beyond this range are shown as individual points. One line of text on screen was 36 px in height; approximately 22 lines were visible on each page on the screen (including section titles).

During reading, the mean distance scrolled was 182 px/5.05 lines, SD = 166 px/4.16 lines, with the median being 128 px/3.5 lines. The first quartile is 43 px and the third quartile 276 px. 2% of events are outliers. Only 4 scroll events covered a distance that was greater than the full screen (768 px). For skimming, participants scrolled a mean of 174 px/4.83 lines, SD = 137 px/3.80 lines, the median being 137.5 px/3.8 lines. The first quartile is 71 px and the third is 248 px, with 3% of events outliers beyond 514 px. Only one scroll event was greater than the full screen. Although the distribution of scroll distances is somewhat larger for reading than skimming, a paired t-test showed no statistical significance in scroll distance between the two conditions ($t(1089) = 1.2786$, $p = 0.2013$).

*Viewport Analysis*

In the context of this work, we refer to the viewport as the region in which content can be viewed. Buscher et al. [4] and Sharmin et al. [14] have previously demonstrated that users have preferred on-screen regions in which they read, defined loosely as top, middle and bottom regions of a page. Our investigation aimed to understand where users fixate prior to scrolling, as opposed to the full distribution (heatmap) of their gaze throughout an entire reading session. We felt that this measure provides the best insight into what participants were looking at when scrolling. We classified participants point of gaze into on-screen regions of interest using two different granularities: (1)

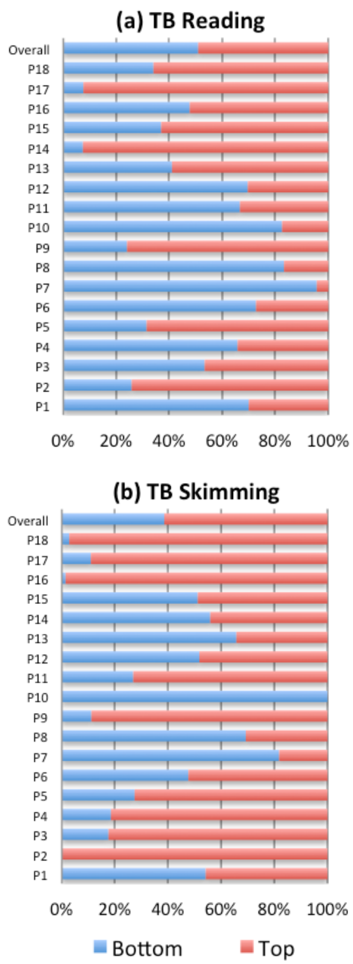


Figure 4: Proportion of fixations before a scroll across viewports for (a) Reading and (b) Skimming for TMB regions.

Top Middle Bottom (TMB), divided the viewport in to 3 vertically equal regions of interest (ROIs), top, middle, and bottom, each 256 px in height; (2) Top Bottom (TB), divided the viewport in to 2 vertically equal regions of interest (ROIs), top and bottom, each 384 px in height (see Figure 1 (a) and (b)).

Figure 3 shows the percentage of each participant's scroll events relating to TMB regions for both reading and skimming, the last column shows the distribution for all participants combined. Figure 4 shows scrolls events relating to TB regions. In reading for TMB we found overall that 50% of participants scroll events occurred when fixating the middle region, 26% for the top region and 24% for the bottom region. For TB regions we observed a near-even split, 51% for the top region and 49% for the bottom region. Although the aggregate data shows an even split of fixations in the top and bottom regions before a scroll event, few individual participants mirror this distribution. Only 2 participants (P3 and P16) have at least 40% of their fixations in each of the two regions. Most tend to gaze predominantly in one region or the other before scrolling, e.g., more than 80% of the fixations for P7 and P8 occur in the top region and more than 90% of fixations for P14 and P17 occur in fixating more often in the top region before scrolling, 8 were in the bottom region, and 2 were evenly split. We found 2 of 18 participants for TMB and 3 for TB that were significantly different than the aggregate using Chi-squared tests ($p < .005$). In skimming (see Figure 3) for TMB we found overall that 51% scrolled in the middle region, 33% in the top region, and 16% in the bottom region. For TB regions, 61% of scrolls occurred in the top and 39% occurred in the bottom. When skimming, gaze tends to be somewhat more prevalent in the top regions than it is for reading. Chi-squared tests found 6 participants in TMB and 3 for TB that significantly deviated from the aggregate ($p < .005$).

*Document Structure Analysis*

In addition to viewport regions, we were also interested to understand if participants' scroll events were correlated with the structure or layout of the document. We chose 4 distinct regions of interest that are represented throughout the text. The first was the section title, the bounding box of each heading HTML element, spanning the full width of the viewport. The remaining 3 were defined by splitting each paragraph HTML element into 3 vertically equal regions, start, middle, and end (see Figure 2(c)).

Results from the analysis of structure regions of interest are shown in Figure 5. We found that during reading 18% of scroll events occurred while fixating in the section heading regions, 33% in the start of paragraph regions, 17% in the middle, and 32% at the end of each paragraph. For skimming overall, 16% scrolled at section headings, 32% at paragraph start, 23% in the middle, and 29% at the end. Individually participants generally do not deviate much from the aggregate result. Chi-squared tests found 1 reading and 1 skimming participant that significantly deviated from the aggregate ($p < .005$).

## Summary

Our data collection uncovered several interesting points about the behaviors of our sample. First, there was no statistically significant difference between reading and skimming conditions in the distance scrolled. Except for a few outliers, participants did not scroll beyond the vertical bounds of the screen. Specifically, for reading, 50% of scrolls were between 43 px and 276 px, and skimming between 71 px and 248 px. This indicates

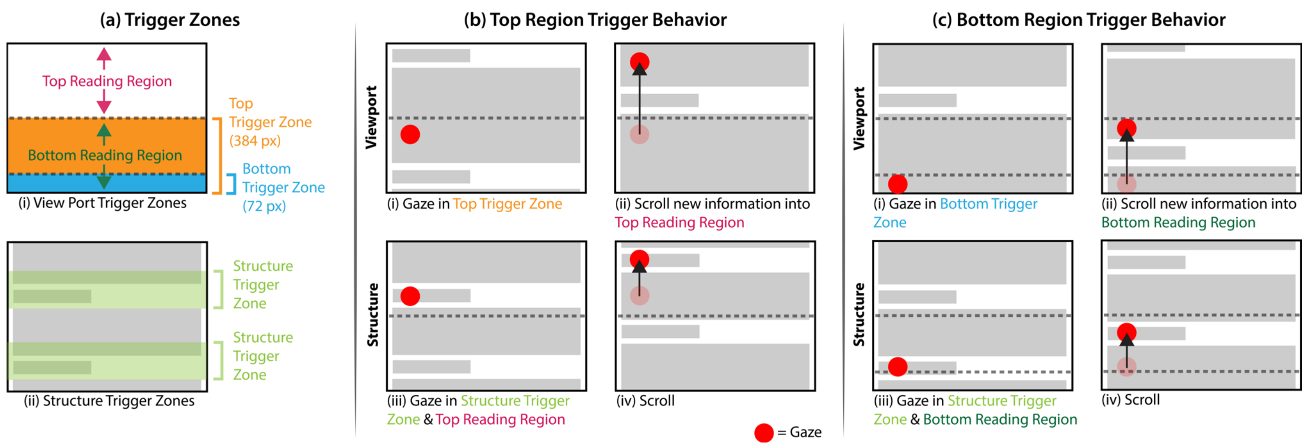


Figure 6: (a) Viewport and Structure trigger zones. (b) Top region trigger behavior. (c) Bottom region trigger behavior.

that participants scroll in 1-7 line chunks, leaving a large proportion of previously consumed text still in view after scrolling. We believe that participants used this residual text to orientate when new context is introduced. Furthermore, participants very rarely scroll a page at a time – only 5 (of 2181) scroll events were greater than the extent of the screen (768 px).

When looking at participants' gaze in relation to document structure we found that the majority of scroll events, for reading, occurred when participants' gaze was either at the start or end of a paragraph. The middle of paragraphs were given least attention when scrolling. Section headings are vertically smaller (36 px) than paragraph regions (~170 – 849 px).

We can conclude that participants generally scrolled at the intersection between paragraphs. For skimming, events are more evenly distributed across regions, both overall and for individual participants. Although the majority of scroll events are still at the beginning and end of paragraphs, there are now more scrolls in the middle of paragraphs. The larger spread in the distribution of events suggests that skimming depends less on structure, but the majority of scrolls are still at paragraph boundaries. Based on these findings we introduce the notion of document structure (in combination with preferred reading region) for automatic scrolling. Specifically, we use the intersection between paragraphs as a trigger for scrolling in our implementation of a new gaze-enhanced scrolling technique.

Two aspects of our experimental settings are worth noting. First, our stimuli included only one paragraph per section heading, meaning there were no instances where paragraphs were consecutive. Thus we cannot isolate the effects of title vs. paragraph boundaries as landmarks for scrolling; we leave this for future work. Second, when reaching the end of the document,

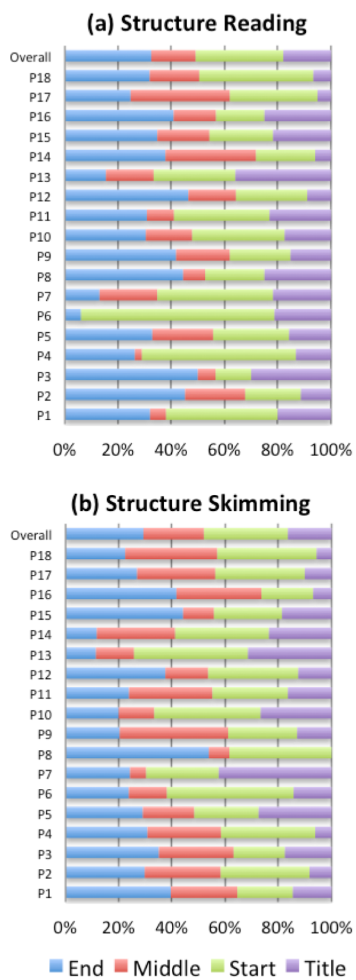


Figure 5: Proportions of fixations before a scroll for different regions (title, paragraph).

participants were forced to read to the end of the viewport. While this is the default behavior in most browsers and readers, in order to support a user's preferred reading region consistently, all content should be accessible within their preferred region. One solution to this is to allow the document to continue scrolling into the preferred region, forming whitespace at the end of the viewport.

## Region&Structure Gaze-enhanced Scrolling

As a result of our analysis we have designed a novel gaze-enhanced scrolling technique. The technique is designed to: (1) support a person's preferred reading region (2) take into account document layout, and (3) scroll less than a page in order to maintain some context. In our implementation we support two preferred reading regions, top and bottom, as outlined in our previous analysis. The preferred region can be set as a parameter before interaction begins.

Our technique requires only gaze for input. Scroll events are triggered when gaze has remained within specific bounds for a set duration, i.e. 250 ms. These bounds refer to invisible trigger zones within the viewport and document. The trigger zones are based on features of the viewport and document layout. Figure 6(a)(i) shows the reading regions and viewport trigger zones for the two variants of our technique (top and bottom). The top reading region covers the upper 384 px, (50%) of the viewport vertically. The bottom reading region covers the lower 384 px (50%) of the viewport. The viewport trigger zones are at the bottom of the respective reading regions. Figure 6(a)(ii) shows how trigger zones are laid out with respect to document structure. The trigger zone covers the intersection between paragraphs – specifically the area including the end of a paragraph, the following section title, up to the next paragraph start.

Certain conditions need to be met to fire a scroll event by gaze in a trigger zone. Content or structural zones can only trigger a scroll event if they lie within the preferred reading region. If no structure trigger zone exists within the preferred reading region, or if gaze skips over all relevant structure triggers, the region trigger zone will fire. If gaze reaches the trigger zone at the bottom of the reading region, a scroll event fires. Once a scroll event is fired, the system must determine the distance to scroll. The distance is calculated as the difference between the current gaze point and the upper bound of the preferred reading region, minus 36 px (1 line). This calculation ensures the new information being scrolled in is comparable in quantity to what has just been read, while keeping 1 line from the previous context. Figure 6(b) and (c) demonstrate the flow of interaction for both top and bottom variants of our technique.

At present this technique is still under development. As an extension of this work we aim to conduct an empirical study to compare our technique with that of Kumar and Winograd [10].

## Conclusion

Our initial data collection identified three important variables that pertain to gaze in reading and scrolling patterns: preferred reading region, document structure, and scroll distance. We found that there are individual differences in preferred reading regions and scrolling behavior between users, and that most scrolling events occurred in alignment with paragraph boundaries. This finding motivated us to develop a new gaze-enhanced scrolling technique with the future aim of verifying the reliability of these features.

## References


1. Biedert, R., Buscher, G., and Dengel, A. The eyebook - using eye tracking to enhance the reading experience. *Informatik-Spektrum 33, 3* (2010), 272–281.
2. Biedert, R., Buscher, G., Schwarz, S., Hees, J., and Dengel, A. Text 2.0. In *Proc. CHI'10 Extended Abstracts*, (2010), 4003–4008.
3. Braganza, C., Marriott, K., Moulder, P., Wybrow, M., and Dwyer, T. 2009. Scrolling behaviour with single- and multi-column layout. In *Proc. WWW'09* (2009), 831.
4. Buscher, G., Biedert, R., Heinesch, D., and Dengel, A. 2010. Eye tracking analysis of preferred reading regions on the screen. In *Proc. CHI'10 Extended Abstracts*, (2010)
5. Dybdal, M.L., Agustin, J.S., and Hansen, J.P. 2012. Gaze input for mobile devices by dwell and gestures. In *Proc. ETRA '12*, 225-228.
6. Hornbæk, K., and Frøkjær, E. 2003. Reading patterns and usability in visualizations of electronic documents. In *Proc. CHI '10,* (2003), 119–149.
7. Jacob, R. J. K. 1990. What you look at is what you get: eye movement-based interaction techniques. In *Proc. CHI'90*, 11–18.
8. Kelly, D., and Belkin, N. J. 2001. Reading time, scrolling and interaction: Exploring implicit sources of user preferences for relevance feedback. In *Proc. SIGIR'01* (2001), 408–409.
9. Kim, J., Collins-Thompson, K., Bennett, P. and Dumais, S.T. 2012. Characterizing web content, user interests, and search behavior by reading level and topic. In *Proc. WSDM'12*, (2012), 213-222.
10. Kumar, M., and Winograd, T. 2007b. Gaze-enhanced scrolling techniques. In *Proc. UIST'07*, (2007), 213–216.
11. Oquist, G., and Lundin, K. 2007. Eye movement study of reading text on a mobile phone using paging, scrolling, leading, and RSVP. In *Proc. MUM '07*, (2007),
12. Räihä, K.-J., and Sharmin, S. 2014. Gaze-contingent scrolling and reading patterns. In *Proceedings of NordiCHI'14*, (2014), 65-68.
13. Rayner, K. 1998. Eye movements in reading and information processing. *Psychological Bulletin, 124, 3*, (Nov 1998)
14. Sharmin, S., Spakov, O., and Räihä, K.-J. 2013. Reading on-screen text with gaze-based auto-scrolling. In *Proc. ETSA '13*, (2013), 24–31.
15. Wikipedia:User:West.Andrew.g/Popular_pages, http://en.wikipedia.org/wiki/User:West.andrew.g/Popular_pages
16. Zhang, Y., and Gellersen, H. 2013. SideWays: A gaze interface for spontaneous interaction with situated displays. In *Proc. CHI'13*, 851–860.