Pagination versus Scrolling in Mobile Web Search

Jaewon Kim[†], Paul Thomas[§†], Ramesh Sankaranarayana[†], Tom Gedeon[†], Hwan-Jin Yoon[¶]
[†]Research School of Computer Science, [¶]Statistical Consulting Unit
The Australian National University
Canberra, Australia
{jaewon.kim, ramesh.sankaranarayana, tom.gedeon, hwan-jin.yoon}@anu.edu.au
[§]Microsoft
Canberra, Australia
pathom@microsoft.com

ABSTRACT

Vertical scrolling is the standard method of exploring search results pages. For touch-enabled mobile devices that are not equipped with a mouse or keyboard, we adopt other methods of controlling the viewport with the aim of investigating user interaction. From the intuition that people are used to reading books by turning pages horizontally, we conducted a user experiment to investigate the effects of *horizontal* and *vertical control types* (*pagination* versus *scrolling*) on a touch-enabled mobile phone. Our findings suggest that participants using pagination were more likely to find relevant documents, especially those over the fold; spent more time attending to relevant results; and were faster to click while spending less time on the search result pages overall. We also found that the main reason for the difference in search speed is the time taken for the scroll itself. We conclude that search engines need to provide different viewport controls to allow better search experiences on touch-enabled mobile devices.

Keywords

Scroll effect; pagination; mobile device; user study

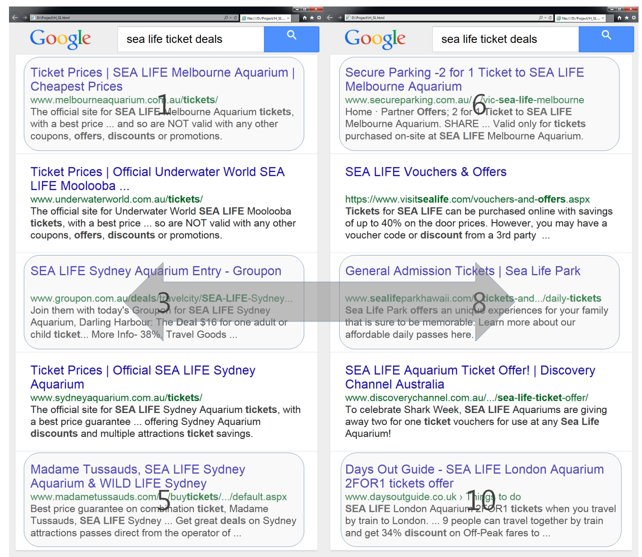

Figure 1: Example of two subsequent pages of a SERP with the horizontal control type.

1. INTRODUCTION

In the interest of enhancing the user search experience, a number of studies have investigated user interactions with elements of the search engine results pages (SERPs) on desktop monitors by evaluating search performance, behaviour, and user satisfaction. Researchers have studied user interaction under the standard conditions provided by search engines [9, 12, 20], manipulated elements of SERPs (e.g., length of snippets and rank order) [5, 10, 12, 16], and classified task types according to the search goals [3, 20]. The results of these studies have led to better SERP interfaces, and a number of search engines have incorporated some of these suggestions into their design. Although several studies have indicated that the use of a scroll function is closely related to search performance and behaviour [5, 9, 12, 16], people must still scroll through SERPs to see beyond the page folds.

In recent years, web search has become one of the most important activities conducted on mobile devices [1]. However, relatively few studies have considered user interaction when conducting web search on touch-enabled mobile devices [11, 17, 18, 19]. Compared to scrolling on desktop monitors, screens for smartphones need to be scrolled more because they have less visible space, and the effect of scrolling may be more important because users need to cover part of the screen with their fingers while scrolling.

To initiate a scroll event on a touch screen, users drag their fingers vertically to produce a similar effect to spinning a mouse wheel, holding and dragging the scroll bar, or using the page up/down keys on desktops. To the best of our knowledge, vertical scrolling is the only option provided by all search engines to control SERPs on touch-enabled mobile devices. However, people are familiar with turning

CIKM'16 , October 24-28, 2016, Indianapolis, IN, USA
© 2016 ACM. ISBN 978-1-4503-4073-1/16/10. . . $15.00
DOI: http://dx.doi.org/10.1145/2983323.2983720

pages horizontally when reading a book, either on paper or via an eReader, with discarding unwanted applications or web pages on smartphones [29], with the start screens for smartphones (pages of icons or tiles), and with controlling applications which provide a switching interface using horizontal swiping gestures (e.g., weather, stock price and shopping). Therefore, horizontal swiping for mobile web search could also have been considered by interface designers, and we wondered what would happen, in terms of search performance and behaviour, if users could swipe horizontally (paginate), as shown in Figure 1. Although most SERPs have buttons for moving to the next page, this function is somewhat different from a horizontal swiping interface. With new HTML and CSS standards, it is now relatively simple to load multiple SERPs, and the pre-loaded pages can be hidden before being called up. This means that there is no additional loading time to display the same number of search results as with a vertical scroll function.

In this study, we investigate the effect of two control types (horizontal pagination and vertical scrolling) for mobile web search, with relevant results at different positions, as shown in Figure 1. We form a series of tasks that have only one target result among ten items to examine the effect of the control type.

Under horizontal swiping, users bring a set of new links over the page fold in one swipe without scrolling. This might lead to users spending more time on each SERP, and paying attention to more results could allow them to carefully consider the most relevant link.

To examine the above inference, we formulated the following four hypotheses:

- H1. When the relevant result is located after the page fold, users take longer to make their first decision and complete the task.
- H2. Users exhibit better search speed with the vertical control type.
- H3. Users exhibit higher search accuracy with the horizontal control type.
- H4. Users are more satisfied with the vertical control type.

Although people are familiar with turning the pages of a book horizontally, and swiping horizontally between screens in weather apps for example, they are used to scrolling with a vertical control type on touch-enabled devices, as this is the only mechanism provided by search engines.

In the following sections, we survey previous studies regarding user interaction in web search on both desktop and mobile devices, and describe the design and procedure of our user study. We then present our findings regarding the effects of both control types and the position of relevant results. We discuss the observed effects, and describe some limitations of our study. Finally, we conclude and suggest ideas for future work.

2. BACKGROUND

To better explain our experiments, we explore two general lines of background knowledge from prior research. The first concerns *reading and scanning SERPs*, and the second addresses *web search on small screens*.

2.1 Reading and scanning SERPs

Although this investigation considers touch-enabled mobile devices, it is important to understand the previous studies of user interaction on desktop monitors, which have various implications for general search behaviour on SERPs.

First, several studies examined how users interact with SERPs in their standard condition. Granka et al. [9] investigated how users scan SERPs until their first click. Their findings indicated that users tend to spend most of their reading time on the first- and second-ranked results, and to view higher ranks above the selected link, although different search behaviour was observed near the page break. A study by Joachims et al. [12] produced similar findings in reading and clicking patterns, and additionally indicated that users tend to scan the search results in a top-to-bottom pattern. Lorigo et al. [20] defined a compressed sequence and a minimal scanpath using the fixation sequence, and described three patterns of user evaluation on SERPs (i.e., complete, linear, and strictly linear). One of their results suggests that users do not generally trust the rank order of the search results, which is the opposite of the result from Joachims et al. [12].

Second, several studies have investigated the effects of different SERP elements (title, URL, snippets) on search behaviour by manipulating the contents and rank order. Joachims et al. [12] investigated user interaction with three different manipulated rank orders (i.e., normal, swapped, and reversed). Their findings show that a user's click decisions are affected by the relevance of search results. A similar study was conducted by Guan and Cutrell [10] to investigate the effects of *target position*: the rank position of the relevant result. They suggested that the search speed and accuracy were much lower when the relevant links were located at lower ranks. Cutrell and Guan [5] also described the effects of changes in snippet length. They generated three snippet lengths, and observed that search speed and accuracy were significantly affected by the amount of information in the snippet. Recently, Kelly and Azzopardi [16] investigated the effects on search behaviour and user experience of the number of results on a SERP. By showing the participants in their experiments a different number of results (three, six, or ten), they examined the resulting different click patterns and found that participants shown three results viewed more SERPs than those shown ten results.

Third, Broder [3] classified task types according to the search goals. He developed a taxonomy of *informational*, *navigational*, and *transactional* web searches that aim to find specific information, reach a particular website, and perform some web-mediated activity, respectively. Lorigo et al. [20] confirmed the effects of the informational and navigational task types. They suggested that the time for task completion is affected by the task type, although there was no significant task-type effect on the scanning behaviour and search accuracy. These two task types have been broadly adopted, and we use one of them here.

2.2 Web search on small screens

Relatively few studies have investigated user interaction in web search on small screens. Jones et al. [13] evaluated users' abilities on mobile phones, PDAs, and desktop interfaces. They found that both the search speed and accuracy were worse on smaller screens. Kim et al. [17] compared user performance and behaviour on SERPs with large and small screens (for a conventional monitor and a mobile device, respectively). Their results indicate that there is no significant difference in time to first click and search accuracy. However, with the small screen, users had difficulty extracting information, exhibited less eye movement, and were slower to complete tasks. Guo et al. [11] investigated touch interactions (i.e., gestures, zooming, swiping, and inactivity) during web search on a mobile device by comparing interactions on a desktop computer with a mouse and keyboard. Their findings indicate that touch behaviours on mobile devices are significantly correlated with the document relevance. Lagun et al. [19] studied the effect of relevance in answer-like results (e.g., today's weather) on a mobile device. Their results indicated that users were less satisfied, spending more time below the answer-like results and scrolling more when the answer-like results were irrelevant. They also found that the second links received more user attention than the first links. This differs from search behaviour on desktop screens, which tend to produce a top-to-bottom scanning pattern. Raptis et al. [24] evaluated the perceived usability, task completion times (efficiency), and task completion rates (effectiveness) on three different sizes of mobile device screens. Their findings suggested that users on the medium and large screens required less time to complete tasks, but there was no significant screen size effect on the perceived usability and task completion rate. Recently, Kim et al. [18] studied search performance and behaviour, and user satisfaction on three small screens: early smartphones, recent smartphones and phablets. They found that subjects recorded a similar search speed and accuracy across the screens. However, users exhibited different search behaviour: less eye movement within a few top links on the phablet screen, fast reading with some hesitation in making a decision on the screen of recent smartphones, and frequent use of scrolling with lower user satisfaction on the screen size of early smartphones. Finally, they suggested some better presentation designs for search results on each screen size.

3. USER STUDY

In this section, we describe the experimental design and procedure, and introduce the participants, tasks, and apparatus.

3.1 Participants

We recruited 27 subjects from both inside and outside of the university campus. The data from three participants were excluded because of low eye gaze calibration accuracy, leaving us 24 participants (14 male) aged 22–41 years (mean: 28.3, standard deviation: 5.0). Two-thirds were studying computer science, and the remaining third had varied backgrounds: accounting, business, law, biochemistry, mathematics, and international studies. Most subjects (22) classified themselves as heavy users of web search engines, usually submitting multiple queries every day. These subjects considered themselves good at using search engines, although one participant thought that he/she was neither good nor bad at searching. Over half of the subjects (17 of 24) regarded themselves as good or expert at using mobile devices; two subjects were not familiar with using mobile phones, and the remaining five claimed to be neither good nor bad at using mobile devices. The subjects participated voluntarily in the experiment, and were compensated for attending with a meal voucher.

3.2 Tasks

Each participant completed twelve tasks, which were based on the tasks used in a study by Dumais et al. [7], and modified for the local participants. Samples of the task descriptions and initial queries are listed in Table 1. We varied the task category, including categories such as tourism, sports, science, weather, and transport. All initial SERPs were extracted from the Google mobile search engine. Each task had 10 search results (five results before the page fold), as shown in Figure 1, and we ensured that each task showed the same elements (title, snippet, and URL) by excluding advertisements, related links, and stars for popularity as in previous studies [7, 16, 17, 18]. All of the tasks were informational in that they required a particular piece of information to be found [20]. To control some variability, we excluded navigational tasks, which are much easier to complete [10, 20], and testing other task types remains as a future study. The tasks were very simple, but were made more difficult by changing the target position. Inspired by Guan and Cutrell [10], we ensured that each task had six variant SERPs, each with a single target result at a different position: ranks 1, 3, and 5 before the fold, and ranks 6, 8, and 10 after (see Figure 1). This let us investigate the effect of target position.

Table 1: Examples of task descriptions and queries.

Task description	Initial task query
You are interested in membership of Questacon. What are the benefits?	Questacon membership benefits
Which two countries played the third match in the 1996 Cricket world cup?	Cricket world cup 1996
How many demerit points can you collect before your driving license is suspended in ACT?	demerit point ACT suspend

3.3 Design and procedure

We used a within-subject design (two control types × six target positions). Participants were assigned a total of 12 tasks across two task sets (tasks 1–6: from task set 1; tasks 7–12: from task set 2), i.e., six tasks from each task set with each control type. Under this condition, each task in a task set was shown with one target result among the six positions, so only two of twelve tasks had the same target position. To minimize the carry-over effect in this design, we randomized the task order in each task set. Furthermore, the orders of task set and control type were counter-balanced, and every task was presented with all six target positions across the participants.

Before starting the experiment, we attempted to relax the participants by emphasizing that they were not being tested. They were then given a consent form explaining the privacy of individual data and their rights, such as terminating their participation at any time or withdrawing their data. After signing the form, subjects were instructed about the number of tasks, the control types, and the experimental procedure. To familiarize themselves with both methods of controlling the viewport, participants conducted two sample tasks with each control type, although they were already very familiar with the vertical control type. We then calibrated their gaze recordings using a 9-point procedure provided by the

Figure 2: An Example of the experimental environment.

eye-tracker software, and presented the task list page. When the participants clicked the first task on the list, a description and initial query were shown. Once they clearly understood the descriptions and the query, they pressed the start button, and the initial SERPs were presented. When the participants announced the correct answer, the task was considered complete. There was no time limit on reaching the correct answer. After each task, they were asked to score the difficulty of the task and usability of the control type. This cycle was repeated until all twelve tasks had been completed. The task/task set order was controlled by Javascript in the cached HTML files. After completing all tasks, participants were asked to fill out a post-experiment questionnaire that asked about their age, preferred scrolling type, and familiarity with search engines and mobile devices. The runtime from sitting on the chair to leaving the laboratory was about 35–40 min.

3.4 Apparatus

All tasks were displayed with Internet Explorer 9 on an iPhone 6 Plus (5.5 inches, which is one of the most popular screen sizes [22]) connected as a secondary monitor by a USB cable using the Twomon software [6], as shown in Figure 2. We obtained eye gaze data using Facelab 5 [27], which records gaze data at 60 Hz, and used the Eyeworks software [8] for our analysis. With the mobile device, the SERPs were displayed at full resolution (1920×1080), and the zoom level was adjusted in Internet Explorer to display five links on the initial screen, similar to the Google mobile search engine with most browsers.

We implemented the pagination function using the smoothscroll jQuery function [15]. The touchable screen recognized a horizontal swipe if the distance between the initial point pushed by the finger and the point at which the finger left the screen was over 100 pixels along the x-axis (about 0.63 cm).

4. RESULTS

We analyzed the gaze data from 288 tasks: 144 tasks for each control type, 48 tasks for each target position. As mentioned earlier, our focus is on the effect of the control type and the target position. First, we divided each SERP into the *front* and *back pages* (before and after the page fold) given by the horizontal swiping interface (see Figure 1), and analyzed the target position effect by aggregating target positions 1, 3, 5 as *target-front* and 6, 8, 10 as *target-back*. For further analysis, we investigated the effect of each target position with both control types. To test the hypotheses, we measured search speed, search accuracy, and user satisfaction. We then examined other data such as time consumed for scroll actions and user attention in the area of interest (AOI).

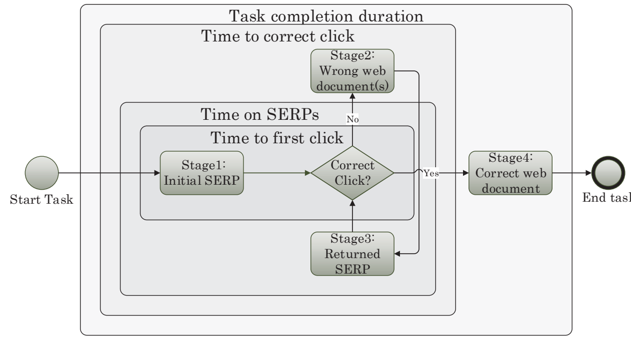

Figure 3: Definitions of four kinds of search speed.

For data analysis, we adopted analysis of variance (ANOVA) [4] for continuous data such as search speed and fixation duration. To maintain the normality assumption, we used a log-transformation $log(x+1)$, so that 0 maps to 0. We employed generalized linear models (GLMs) [21] and generalized linear mixed models (GLMMs) [2] with the binomial distribution and a logit link function for binary data, and a Poisson distribution and logarithm link function for count data. To consider the individual difference in varying expertise at using mobile devices, we adopted a block structure (subject) for ANOVA, and we used a GLMM instead of a GLM if there was a random effect between subjects (σ_s^2). For the score data from the 7-point Likert scale, we used a linear mixed model (LMM) [30]. Analysis was conducted using the GenStat statistical package [28] and R [23].

4.1 Search performance and user satisfaction

Participants' time on each task could be divided into the four stages shown in Figure 3. If a user clicked (tapped) the relevant link at the first attempt, the time spent on the initial SERP (stage 1) was the first stage, and the time spent on the correct web document (stage 4) was the last stage needed to reach the correct answer. In addition, we investigated instances in which users made an incorrect choice on the SERP. Therefore, we considered the time spent on the wrong web document(s) (stage 2), and the time spent re-reading SERP(s) (stage 3) while searching for a different choice.

Using these different stages, we could measure four search speeds: *Time to first click* denotes the time taken to make the first decision, *Time on SERPs* combines stages 1 and 3 to give the total time spent on SERPs, *Time to correct click* combines stages 1–3 to investigate the time taken to make the correct decision, and *Task completion duration* is the total time required to complete a task.

Time to correct click excludes the time spent on stage 4 (correct web document). Although participants reached the same web document at the end of each task, some pages

Table 2: Search performance and user satisfaction.

		Target-front		Target-back		*p*-value		
		H	V	H	V	Target position	Control type	Interaction
Search speed	Time to first click [s]	25.83	33.10	30.97	32.89	**	0.175	0.187
	Task completion duration [s]	58.00	63.69	63.39	72.85	**	0.057	0.699
	Time to correct click [s]	36.09	40.89	42.02	53.13	***	*	0.535
	Time on SERPs [s]	31.13	37.40	36.02	43.85	***	*	0.870
Search accuracy	Correct answer rate [%]	79.20	81.94	77.78	58.33	*	0.105	*
User satisfaction	7-point Likert scale	5.60	5.74	5.49	5.44	*	0.601	0.332

Note: H and V denote horizontal and vertical control types, respectively. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3: Search performance, user satisfaction, and scroll rate with each target position.

	p-value		
	Position	Type	Interaction
Time to first click	**	0.172	0.610
Task completion duration	**	0.056	0.432
Time to correct click	***	*	0.322
Time on SERPs	***	*	0.593
Search accuracy	0.105	0.199	0.071
User satisfaction	*	0.599	0.342
Scroll rate	***	0.381	0.181

Note: Position and Type denote target position and control type, respectively. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

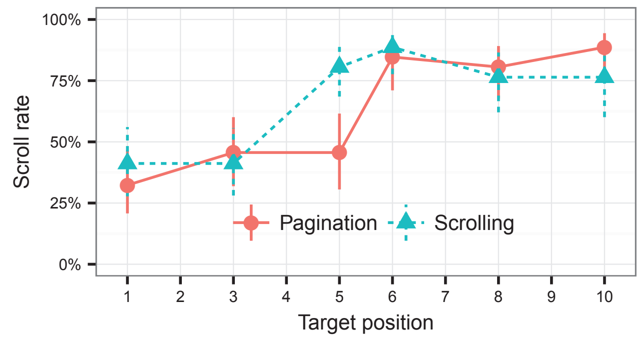

Figure 4: Scroll rate with each target position.

were better suited to desktop screens, which is beyond the search engine's control.

We applied ANOVA to the search speed measurements. First, only the target position significantly affected time to first click ($F_{(1,261)} = 7.39$, $p < 0.01$) and task completion duration ($F_{(1,261)} = 10.13$, $p < 0.01$). Table 2 indicates that subjects spent more time finding targets at lower ranks, where it was necessary to perform a scroll (mean time to first click: 29.46 s vs. 31.93 s, and task completion duration: 60.84 s vs. 68.12 s).

For the remaining two search speeds, we found significant effects due to target position and control type on time to correct click ($F_{(1,261)} = 20.67$, $p < 0.001$, and $F_{(1,261)} = 4.81$, $p < 0.05$, respectively) and time on SERPs ($F_{(1,261)} = 18.01$, $p < 0.001$, and $F_{(1,261)} = 5.05$, $p < 0.05$, respectively). With vertical scrolling, participants needed more time to identify the correct result (about 7.96 s longer) and spent longer on the SERP (7.05 s).

The search accuracy was assigned a value of 1 if a participant clicked the target result and found the answer at the first attempt; otherwise, the search accuracy was set to 0. Using a GLM with a binomial distribution, significant differences were observed in search accuracy according to target position ($\chi^2 = 5.99$, df = 1, $p < 0.05$) and the interaction of target position and control type ($\chi^2 = 3.83$, df = 1, $p < 0.05$). Table 2 displays a small difference (2.74%) in the mean values for both control types when the target was on the front of the page, but this difference was much larger when the target was on the back page (19.45%). With these results, we can say that horizontal pagination results in better search accuracy when the correct answer is located on the back page. It seems that pagination helps compensate when rankings are poor, but does not degrade performance when rankings are good.

We employed an LMM to examine user satisfaction. User satisfaction exhibited a significant difference due to target position ($\sigma_s^2 = 0.530$, $\chi^2 = 4.69$, df = 1, $p < 0.05$). This means that users were more satisfied (about 0.21 points higher) when a relevant link was located on the front of the page. However, participants expressed no preference for either control type, despite being more familiar with vertical scrolling on current search engines.

We also investigated the effect of each (of six) target position and both control types using the same statistical methods. Across all search performance results, user satisfaction and scroll rate, we found very similar results compared to those obtained when the target positions were divided into target-front and -back, as shown in Table 3.

However, when we looked at control type effects by each target position, there was an interesting result for scroll rates. Figure 4 displays the scroll rates for the six target positions. The scroll rates with both control types on each target position were not different except for one target position—when the correct answer was at rank 5, subjects recorded only about 46% scroll rate using the horizontal swiping, whereas that with the vertical control type was as high as 81%. The line in Figure 4 for vertical scroll suddenly increases from target position 5 (from about 40% to 80%), whereas the other line for horizontal swiping starts to soar from target position 6. This suggests that participants performed a scroll event to either look ahead a few links beyond rank 5, or used scrolling to position rank 5 in the middle of the screen with the vertical control type. This is addressed further when we examine the fixation duration with target position 5 in Section 4.3.

Across the results for search performance and user satisfaction, there were clear target position effects, and we found several task type effects and interactions of the variables.

Table 4: Scroll actions and search speed excluding the scroll duration.

		Target-front		Target-back		p-value		
		H	V	H	V	Target position	Control type	Interaction
Scroll	Rate [%]	41.67	54.17	83.33	79.17	***	0.405	0.129
	Finger actions	0.82	3.57	0.74	3.96	0.577	***	0.512
Search speed excluding scrolling	Time to first click [s]	25.42	30.22	30.60	29.79	**	0.850	0.114
	Task completion duration [s]	57.56	60.64	62.99	68.67	**	0.317	0.840
	Time to correct click [s]	35.65	37.84	41.63	48.95	***	0.242	0.683
	Time on SERPs [s]	30.69	34.35	35.63	39.67	***	0.591	0.612

Note: Finger action and scroll duration measured until first click.
Note: H and V denote horizontal and vertical control types, respectively. ** $p < 0.01$, *** $p < 0.001$.

Consistent with H1, when the target was on the front of the page, users tended to require less time, show better accuracy, and exhibit higher satisfaction. However, participants were not faster or more satisfied with vertical scrolling, and in fact were faster to find the target and spent less time on SERPs when given the horizontal controls. Therefore, our results do not support H2 or H4. In addition, consistent with H3, users recorded better search accuracy with pagination when the target was on the back page.

4.2 Scroll action and duration

Most people are more familiar with vertical scrolling in mobile web search, which is generally the only type provided by search engines. The results for search performance and user satisfaction indicated that horizontal pagination was similar or better for mobile web search. To determine why, we looked more closely at the most salient difference: the scrolling action itself.

First, using a GLMM with a binomial distribution, only the target position effect was observed on scroll rate ($\sigma_s^2 = 0.494$, $\chi^2 = 36.07$, df = 1, $p < 0.001$), and we found no control type effect (Table 4). We can explain the target position effect as follows. When the correct link was on the front page, neither control type required the subject to scroll to reach the relevant answer: overall scroll rate on the front page was less than 50%. However, the scroll rate does not seem to explain the difference in speed between the control types. Therefore, we decided to investigate how much effort the user made to use the scroll function.

In our experiment, minimal effort was required (only one flip) to see a whole SERP with pagination. Even if users wanted to see a few links on the back page, once swiped, they could see all of the results 5–10. Besides the time taken to change the cached pages (maximum 0.5 s), users could read the SERP immediately after swiping, whereas it may have been difficult to read while scrolling vertically. Thus, it is interesting to examine how vertical scrolling affected the search. Unfortunately, we were unable to find a previous study regarding scroll duration.

We needed to define the *continuous scroll action* from the scroll events to calculate the duration. First, the duration of a continuous scroll action was measured by considering whether the scroll event happened continuously over 100 ms. Second, we considered users to have stopped scrolling if no scroll event has been recorded for 100 ms, because, by definition, they could have made a fixation during this time (see Section 4.3). Finally, we neglected scroll distances of less than 30 pixels (the height of one snippet line) over 1 s to exclude slow scrolling in which the SERPs were being read and slight movements of the finger while holding the screen causes minor further scrolling. There was a clear limitation in this calculation, as certain participants might be able to fixate while scrolling relatively fast, although we excluded the case of reading while slowly scrolling. However, this could be considered as an effect of scrolling that made it more difficult to read the search results.

Using the above definition, we compared how many times the users performed a continuous scroll action with vertical scrolling against the number of swiping events with the horizontal control type. In addition, we investigated how long the participants spent scrolling vertically for each task. Table 4 lists the number of finger actions required to scroll with both control types until the first decision was made. Applying a GLMM with a Poisson distribution, we found a significant difference in the number of scrolling motions on the first SERP according to control type ($\sigma_s^2 = 0.262$, $\chi^2 = 97.09$, df = 1, $p < 0.001$). This result indicates that participants conducted about five times more finger actions with the vertical control type.

Finally, we recalculated the contents of Table 2 for the four search speeds by removing the scroll duration (mean 3.05 s and 4.18 s with target-front and -back) with the vertical scroll type. We recalculated the search speed under horizontal swiping by assuming that participants could not read SERPs while changing pages (0.5 s per scroll event). Target position effects can still be found across the four measurements. However, the control type effects on the time to correct click and time on SERPs have disappeared, the mean values of both control types become closer, and the p-values of the control type effects are now out of the significance level.

Thus, we can be sure that the time needed for vertical scrolling is by itself the main factor in users spending more time on SERPs. It is not productive time, such as reading for example; but simply time needed to manage the interface.

4.3 User attention

While the search performance and user satisfaction results represent how efficiently users search and their impression of the search experience, fixation duration is believed to represent the degree to which users try to extract information from SERPs and areas of interest (AOIs) [14, 25].

In this experiment, we recorded fixations of over 100 ms within a region of 70 pixels diameter using algorithms in the Eyeworks software. We determined the duration and distance thresholds by considering the screen resolution (1920 × 1080), the distance between the user and the eye-tracker, and the calibration accuracy [26]. To investigate

Table 5: Fixation duration: relations between control types and others.

		Target-front		Target-back		p-value		
		H	V	H	V	Target position	Control type	Interaction
AOI-front [s]		8.91	11.73	8.42	10.47	0.345	**	0.749
AOI-back [s]		3.37	3.42	7.12	4.52	***	0.105	*
p-value	AOI-pages	***		***				
	Control type	0.197		0.096				
	Interaction	0.507		***				

Note: H and V denote horizontal and vertical control types, respectively.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

the effect on user attention with each control type, we applied ANOVA and assigned two AOIs (*AOI-pages*): contents on the front page (*AOI-front*), and the remaining content (*AOI-back*).

First, we analyzed the effects of AOI-pages, target position, and control type on fixation duration. Using a three-way ANOVA, we found no significant control type effect or interaction of the three variables, but there were significant differences in user attention due to AOI-pages ($F_{(1,545)} = 246.46$, $p < 0.001$), target position ($F_{(1,545)} = 21.45$, $p < 0.001$), interaction of AOI-pages and target position ($F_{(1,545)} = 30.73$, $p < 0.001$), interaction of AOI-pages and control type ($F_{(1,545)} = 7.20$, $p < 0.01$), and interaction of target position and control type ($F_{(1,545)} = 4.12$, $p < 0.05$).

Although we confirmed the effects of target position, AOI-pages, and their interactions, it was difficult to explain the implication of these interactions. For further analysis, we investigated the effects of control type and target position and control type and AOI-pages, because the control type had no significant effect but did interact with the other factors.

First, Table 5 describes the attention paid to AOI-pages (AOI-front and -back), which can explain the interaction of control type and target position. When investigating the fixation duration on AOI-front, there was a significant control type effect ($F_{(1,261)} = 6.84$, $p < 0.01$) on fixation duration, whereas no difference according to target position was found. In addition to this, we could see a target position effect ($F_{(1,261)} = 33.95$, $p < 0.001$) and an interaction between control type and target position ($F_{(1,261)} = 4.62$, $p < 0.05$) for AOI-back.

These results indicate that subjects spent more time (about 2.4 s) extracting information from AOI-front with the vertical control type. On AOI-back, participants spent a similar amount of time reading the target on the front page with both scrolling types, but paid more attention to targets on the back page while using the horizontal swiping function (about 2.6 s).

The interaction of target position and control type on AOI-back may be explained by the fact that all links on the back page were displayed with one action under the horizontal control type. This allowed subjects to read more links on the back page, whereas the vertical control type required continued scrolling to see links on the back page, which moved downwards with each vertical scrolling action. To investigate this inference, further analysis of the fixation duration per link with minimal scanpath values [17] is required.

Second, we investigated the effect of AOI-front and -back on fixation duration to analyze the interaction of AOI-pages and control type as shown in Table 5. When we considered a target on the front page, a significant AOI-pages effect could be observed in fixation duration ($F_{(1,261)} = 215.70$, $p < 0.001$), whereas no difference due to control type was found. Considering tasks where the target was on the back page, we found significant effects on fixation duration due to AOI-pages ($F_{(1,261)} = 60.01$, $p < 0.001$) and the interaction of AOI-pages and control type ($F_{(1,261)} = 11.29$, $p < 0.001$).

Table 6: Fixation duration: relation between AOIs and control types.

	p-value		
Target positions	AOI	Control type	Interaction
Rank 1	***	***	0.366
Rank 3	***	0.171	0.584
Rank 5	***	0.607	**
Rank 6	***	0.815	***
Rank 8	***	0.430	*
Rank 10	***	0.454	**

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

This indicates that when the target result was located on the front page, participants spent more time reading the front page with both control types: 5.54 s and 8.31 s more with horizontal swiping and vertical scrolling, respectively. When the target was on the back page, participants using the scrolling interface still spent 5.95 s longer on the front page (2.36 s less than before), but participants using horizontal swiping spent only 1.30 s longer on the front page than the back (4.24 s less). That is, when the target document was on the back page, people using the swiping interface shifted their attention much more than people using the scrolling interface. This result might be related to the difference in search accuracy on the back pages (77.78% with pagination and 58.33% with vertical scrolling). Although further investigation is required, the longer reading time with the horizontal control type seems to lead to better search accuracy on the back page.

We also assigned ten AOIs, one to each link in the SERP, to investigate which results participants focused on. In particular, we were interested in the effect of target position and AOI with each control type.

When we considered the effects of AOI, control type and target position on fixation duration, the effects were very similar to the results with target-front and -back with AOI-pages: we could see effects due to target position, AOI and interactions between these two variables.

To investigate these interactions, we analyzed the fixation duration of each AOI along the target positions using ANOVA (see Table 6 and Figure 5). First, we examined

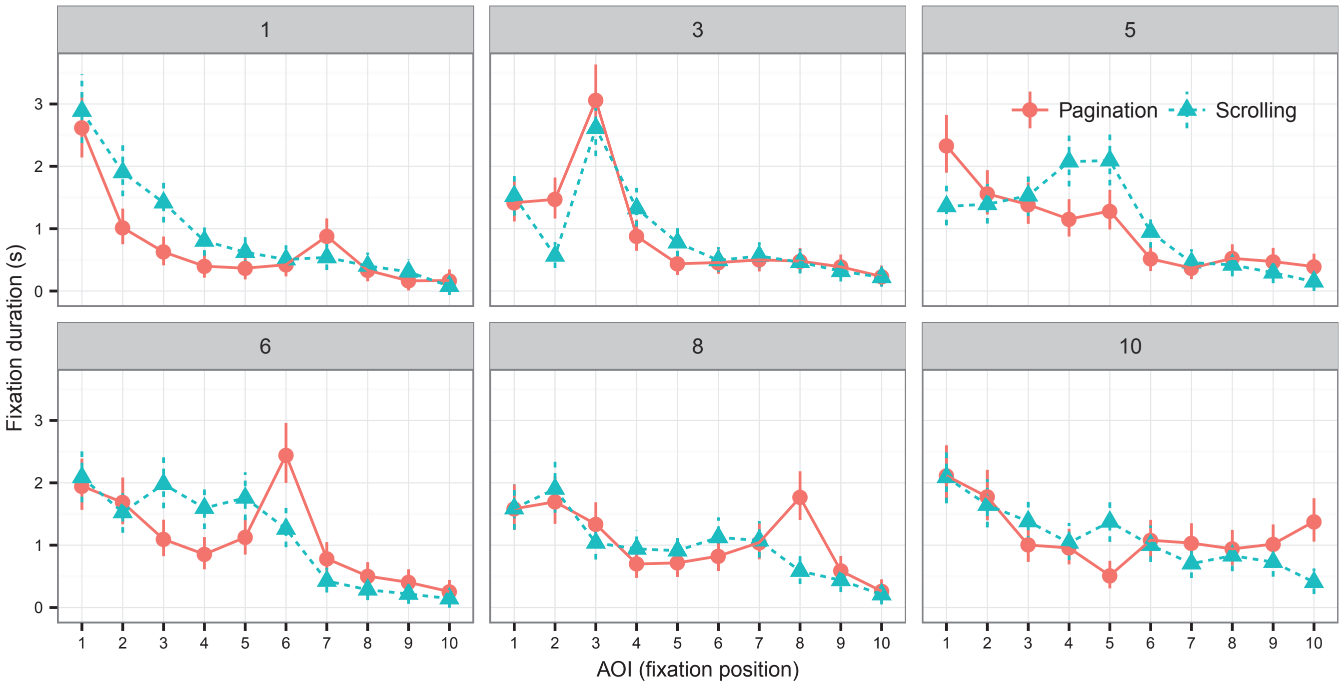

Figure 5: Fixation duration on each AOI along target positions [s]: numbers on each y-axis are values after back-transformation.

the effects for target positions 1, 3, and 5, located on the front pages. With target position 1, there were significant AOI and control type effects ($F_{(1,437)} = 25.97$, $p < 0.001$ and $F_{(1,437)} = 11.95$, $p < 0.001$, respectively) on fixation duration. Fixation duration with pagination and vertical scrolling followed a fairly similar pattern of decline across AOIs. The fixation duration on each AOI was normally higher with vertical scrolling, which causes the control type effect, and the main difference could be seen on AOIs 2, 3. When the target position was at rank 3, fixation duration was significantly affected by AOI ($F_{(1,437)} = 30.77$, $p < 0.001$). AOI 3 recorded the longest read times with both control types, which would appear to be due to the target result.

With target position 5, users exhibited significant differences in user attention due to AOI and the interaction of AOI and control type ($F_{(1,437)} = 21.27$, $p < 0.001$ and $F_{(1,437)} = 2.69$, $p < 0.01$, respectively). This interaction was apparently caused by the reversed pattern of reading effort between AOI 1 and AOIs 4, 5. Subjects spent more time reading AOI 1, but less on AOIs 4 and 5 with the horizontal control type. This can be connected to the scroll rate discussed in Section 4.1: participants only exhibited higher scroll rates for target position 5. Although further analysis is needed to confirm this, with the higher fixation duration in AOI 5, it seems that users employed the vertical scroll function to reposition rank 5 to the upper side (or maybe the middle) of the screen, and then focused their attention on the link.

Across target positions 6, 8, and 10, we found significant effects on fixation duration due to AOI ($F_{(1,437)} = 24.62$, $p < 0.001$, $F_{(1,437)} = 11.54$, $p < 0.001$, and $F_{(1,437)} = 6.11$, $p < 0.001$) and due to the interactions of AOI and control type ($F_{(1,437)} = 3.40$, $p < 0.001$, $F_{(1,437)} = 2.17$, $p < 0.05$, $F_{(1,437)} = 2.95$, $p < 0.01$) on user attention. With target position 6, the reason for this interaction may come from the reversed pattern for AOIs 3, 4, and AOI 6. Users took more time to read AOI 6, which contained the correct answer, when using the horizontal swiping interface, whereas subjects using the vertical control type did not exhibit any enhanced reading effort in this AOI. This might be explained by AOI 6 being the top link on the second page when using pagination. With the relevant result at rank 8, subjects using horizontal swiping concentrated more on AOI 8, which contained the correct answer. Similar to the result for target position 6, we could not find a relation between the target position and fixation duration on the AOI with the vertical control type. Finally, for target position 10, subjects displayed similar results to those for target positions 6 and 8. The main reason for this interaction seems to come from AOI 10. Subjects took more time reading AOI 5, but spent less time on AOI 10 with the vertical control type. This may be the reason for the different search accuracy (75% with horizontal swiping and 50% with vertical scrolling) with target position 10.

When the target results were on the front page, the fixation displayed a similar pattern with both control types: decreasing along the sequence of AOIs and increasing on the AOIs containing the correct answer. However, whereas targets on the back page received more attention than others with the horizontal swiping interface, there as no such effect with vertical scrolling. This may indicate that the targets were simply not recognised as useful when scrolling. This, in turn, may explain why the paginated interface led to higher accuracy when targets were over the fold.

4.4 User preference

At the end of each session, we asked participants about their preferred control type, the reason for this preference, and their thoughts regarding the use of horizontal swiping on their phones.

Twelve of the 24 participants felt that the vertical scrolling was more convenient than horizontal pagination, and one reason from seven of the 12 was familiarity.

"That [the vertical scrolling] is common in most phones including mine"

"Phones provide vertical browsers, familiar"

The other reason from the remainder (5) of this 12 is that the vertical control type was easier/more convenient for web searches.

"I prefer to continuous scrolling to pagewise [one page at a time]"

One-third (8) replied that they preferred the pagination to the vertical scrolling, and they considered it to be more convenient and easier to see all of the results.

"I just needed one click to see two pages — it [the horizontal swiping] was fast"

"It [pagination] was improved visibility"

The remainder (4) considered both scroll types to have similar usability. If horizontal swipes were provided by commercial search engines, a quarter of participants said that they would definitely use it instead of vertical scrolling, and the majority (16 of 24) were willing to try using pagination as the main scroll type.

5. DISCUSSION

In this section, we aggregate the previous discussions by focusing on the effects of both control types and consider the limitations of this study.

First, in dividing SERPs into front and back pages, we found interesting effects of control type on search performance and no difference in user satisfaction. Although it was their first time using pagination on SERPs with a touch-enabled mobile device, participants exhibited similar or better search performance across four speed measurements, and they were more accurate when relevant results were located after the page fold. Participants also expressed similar satisfaction with both control types. Considering each target position, we found an interesting search behaviour: participants scrolled at different rates, across the two interfaces, when the correct answer was at link 5. Similar to the results of previous studies [9, 12], special scanning behaviour occurred at the page break. Although this may have several implications and further analysis with fixation sequence (scanpath) [7, 17, 20] is required to be conclusive, one possible reason is that participants scrolled vertically to read a few links after the page fold or to locate the link from the bottom to the middle/top of the screen with higher user attention when a correct answer was in that AOI.

Second, since we have controlled for task and other effects, differences in speed were most likely due to control type: something about vertical scrolling made users slower. To investigate this, we measured the vertical scroll duration by defining a continuous scroll action and re-investigated the search speed by excluding the scroll duration. From these results, we could confirm that the time consumed when using vertical scrolling is the main reason for the difference in search speed.

Third, a major effect on user attention was that participants with the horizontal control type spent more time extracting information over the page fold when the correct answer was located on the back page. We inferred that the higher search accuracy with horizontal swiping when the relevant links were on the back page may be due to the higher user attention on lower ranks. We left this for a further study. With 10 AOIs, subjects using horizontal swiping exhibited a similar pattern with each target position: a decline in fixation duration along the rank order, but strong attention on AOIs containing the relevant links. However, we could find no such effect with the vertical control type when the relevant links were on the back page. Thus pagination appears to encourage users to notice results further down the list.

5.1 Limitations

We acknowledge that this study had several limitations. First, user interaction on mobile web search may differ for various screen sizes [17]. We considered the recent trend in screen size, and adopted one of the most popular smartphone sizes for this study [22]. However, it is possible that the results may be different with other screen sizes. Second, although we recruited participants with various backgrounds and a wide range of ages, the results cannot cover all mobile users' individual web search behaviours. Third, participants were seated while searching, and were not free to move the mobile device as it was held fixed in a holder as shown in Figure 2. This restricted the participants' freedom of movement with the smartphone, which could have affected their behaviour. Finally, the results may be affected by the task type. Because we adopted only informational tasks to avoid complex interactions of many independent variables, the results may be different with other tasks.

6. CONCLUSIONS AND FUTURE WORK

We conducted an experiment to determine the effect of horizontal and vertical control types when conducting web searches on mobile devices.

Despite participants having greater familiarity with vertical scrolling during search, our results suggested that *horizontal swiping gave similar or better search speed.* The main reason was the effect of the scroll action itself when using vertical scrolling. In addition, *participants using horizontal swiping tended to pay more attention to links beyond the fold* when a correct answer was located there, *and with better search accuracy*. Furthermore, users using pagination exhibited special attention to the positions of the relevant link, whereas this trend was not observed with the vertical control type when the correct answer was beyond the page-break. Finally, although half of the participants expressed a preference for vertical scrolling, many stated that this was because of familiarity, and most subjects were willing to try horizontal swiping if provided.

Considering the limitations regarding users' individual differences and the lab conditions of this study, we cannot conclusively say that the horizontal control type is better than vertical scrolling. We do suggest, at least, that it is worthwhile for search engines to provide both scrolling types to enhance the user search experience.

This study left further analysis and experimentation for future work: e.g., understanding the high scroll rate when a relevant answer is just above the page fold, and the relationship between search accuracy and user attention. We will run further analyses with fixation sequence measures such as the minimal scanpath, compressed sequences [7, 20], and trackback [17] to investigate the implications of the results from this study, and will examine the correlation between search performance and user attention. In addition, we plan to study the control type effect with navigational tasks and various screen sizes of mobile phones.

7. REFERENCES

[1] Adwords. Building for the next moment, 2015. Retrieved April 03, 2016 from http://adwords.blogspot.com.au/2015/05/building-for-next-moment.html.

[2] N. E. Breslow and D. G. Clayton. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421):9–25, 1993.

[3] A. Broder. A taxonomy of web search. *ACM SIGIR Forum*, 36(2):3–10, 2002.

[4] W. G. Cochran and G. M. Cox. Experimental designs. 1957.

[5] E. Cutrell and Z. Guan. What are you looking for?: An eye-tracking study of information usage in web search. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 407–416. ACM, 2007.

[6] Devguru. Twomon USB. Retrieved from http://devguru.co.kr/easynlight-en/.

[7] S. Dumais, G. Buscher, and E. Cutrell. Individual differences in gaze patterns for web search. In *Proceedings of the third symposium on Information interaction in context*, pages 185–194. ACM, 2010.

[8] Eyetracking. Eyeworks. Retrieved from http://www.eyetracking.com/Software/EyeWorks.

[9] L. A. Granka, T. Joachims, and G. Gay. Eye-tracking analysis of user behavior in WWW search. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 478–479. ACM, 2004.

[10] Z. Guan and E. Cutrell. An eye tracking study of the effect of target rank on web search. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 417–420. ACM, 2007.

[11] Q. Guo, H. Jin, D. Lagun, S. Yuan, and E. Agichtein. Mining touch interaction data on mobile devices to predict web search result relevance. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 153–162. ACM, 2013.

[12] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay. Accurately interpreting clickthrough data as implicit feedback. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 154–161. ACM, 2005.

[13] M. Jones, G. Buchanan, and H. Thimbleby. Improving web search on small screen devices. *Interacting with Computers*, 15(4):479–495, 2003.

[14] M. A. Just and P. A. Carpenter. Eye fixations and cognitive processes. *Cognitive Psychology*, 8(4):441–480, 1976.

[15] Karl Swedberg. Smooth scroll. Retrieved from https://plugins.jquery.com/smooth-scroll/.

[16] D. Kelly and L. Azzopardi. How many results per page? A study of SERP size, search behavior and user experience. In *Proceedings of the 38th International ACM SIGIR Conference*, pages 183–192. ACM, 2015.

[17] J. Kim, P. Thomas, R. Sankaranarayana, T. Gedeon, and H.-J. Yoon. Eye-tracking analysis of user behavior and performance in web search on large and small screens. *Journal of the Association for Information Science and Technology*, 66(3):526–544, 2015.

[18] J. Kim, P. Thomas, R. Sankaranarayana, T. Gedeon, and H.-J. Yoon. Understanding eye movements on mobile devices for better presentation of search results. *Journal of the Association for Information Science and Technology*, 2016.

[19] D. Lagun, C.-H. Hsieh, D. Webster, and V. Navalpakkam. Towards better measurement of attention and satisfaction in mobile search. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 113–122. ACM, 2014.

[20] L. Lorigo, B. Pan, H. Hembrooke, T. Joachims, L. Granka, and G. Gay. The influence of task and gender on search and evaluation behavior using Google. *Information Processing & Management*, 42(4):1123–1131, 2006.

[21] P. McCullagh and J. A. Nelder. *Generalized linear models*. Chapman & Hall/CRC, 2nd edition, 1989.

[22] Mobile Marketing. Majority of mobile web browsing on smartphones bigger than 5 inches, 2015. Retrieved from http://mobilemarketingmagazine.com/54708-2/.

[23] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. ISBN 3-900051-07-0.

[24] D. Raptis, N. Tselios, J. Kjeldskov, and M. B. Skov. Does size matter?: Investigating the impact of mobile phone screen size on users' perceived usability, effectiveness and efficiency. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*, pages 127–136. ACM, 2013.

[25] K. Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3):372, 1998.

[26] D. D. Salvucci and J. H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 71–78. ACM, 2000.

[27] Seeingmachines. Facelab5. Retrieved from http://www.seeingmachines.com/.

[28] VSN International. Genstat for Windows 17th edition., 2014. VSN International, Hemel Hempstead, UK. Web page: www.vsni.co.uk.

[29] A. Warr and E. H. Chi. Swipe vs. scroll: web page switching on mobile browsers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2171–2174. ACM, 2013.

[30] B. T. West, K. B. Welch, and A. T. Galecki. *Linear mixed models: A practical guide using statistical software*. CRC Press, 2014.