An Eye Tracking Study of How Font Size and Type Influence Online Reading

David Beymer
IBM Almaden Research Center
650 Harry Rd, San Jose, CA 95120
beymer@us.ibm.com

Daniel Russell
Google
1600 Amphitheatre Pkwy, Mountain View, CA
drussell@google.com

Peter Orton
IBM Center for Advanced Learning
20 Old Post Rd, Armonk, NY 10504
porton@us.ibm.com

ABSTRACT
In order to maximize online reading performance and comprehension, how should a designer choose typographical variables such as font size and font type? This paper presents an eye tracking study of how font size and font type affect online reading. In a between-subjects design, we collected data from 82 subjects reading stories formatted in a variety of point sizes, san serif, and serif fonts. Reading statistics such as reading speed were computed, and post-tests of comprehension were recorded. For smaller font sizes, fixation durations are significantly longer, resulting in slower reading – but not significantly slower. While there were no significant differences in serif vs. san serif fonts, serif reading was slightly faster. Significant eye tracking differences were found for demographic variables such as age group and whether English is the subject's first language.

Categories and Subject Descriptors
H.5.2. User Interfaces: Screen Design. H.1.2 User/Machine Systems: Human Factors

General Terms
Experimentation, Design, Human Factors.

Keywords
Eye tracking, typography, font size, font type, reading.

1. INTRODUCTION
From a design standpoint, how does one adjust the typography of a web page or online document for optimal reading? In this paper, we present studies that address two design issues: (1) what is the best font size for reading online, (2) which font type, serif or san serif, is best for reading. We are motivated by the design of online e-learning material, so retention of the material is as important a factor as speed and efficiency.

These typographic issues have been studied by advertisers, psychologists, ergonomists, and designers for over 100 years, focusing mostly on paper but now moving to address online presentation on computer screens. For example, in the font size issue, using too small a font makes the letters illegible, while too large a font needlessly wastes page space. Paterson and Tinker, who in the 1920s – 1940s studied a number of typographical issues [1], found that for paper, 10 pt text was read faster than 6, 8, 12, or 14 pt [2]. Measuring character size by visual angle, Legge, *et al* [3] found that reading speed was fairly constant for a range of character sizes, 0.3º-2º, but it deteriorates outside of this range. Looking at font sizes of 10, 12 and 14 pt on a computer screen, Bernard, *et al* [4] found that 12 pt was read the fastest. Darroch *et al* [16] investigated different font sizes on a handheld device for young and old subjects, but they found no effect of font size on reading time.

For font type, there is a lack of statistically significant differences in font studies to rule in favor of serif or san serif fonts. Those favoring serif fonts claim that the serif brackets and the contrasting use of thin and thick lines makes the letters and words more distinctive and hence easier on the eye. According to those favoring san serif, those same shifts in line width create an exaggerated contrast that impairs reading speed. In a study of 10 fonts types on paper, Paterson and Tinker [5] isolated 2 fonts as poor performers, but those fonts are not in use today (American typewriter, Old English). In a more recent study, Boyarski *et al* [6] tested modern computer fonts, explicitly comparing serif vs. san serif fonts for computer screens. There was a comprehension advantage for Georgia (serif) over Verdana (san serif), but no speed difference.

To understand the detailed structure of how people read text, psychologists and other researchers have turned to eye gaze tracking as a valuable analysis tool. In eye gaze tracking, a camera tracks and records where a subject's eye is looking; these gaze points are mapped to the text to follow the subjects' reading behavior.

Eye tracking analysis has revealed how the eye moves during the reading process – see Rayner and Pollatsek [7] for an excellent summary. The eye reads an individual line of text in discrete chunks by making a series of fixations and saccades. A *fixation* is a brief moment, around 250 ms, where the eye is stopped on a word or word group, and the brain processes the visual information. A *saccade* is a fast eye movement, usually forward in the text around 8-12 characters, to position the eye on the next section of text. A *regression* is a backwards motion in the text, and it indicates confusion. A *return sweep* is the eye motion from the end of one line of text to the beginning of the next. The *perceptual span* refers to the size of the visual window processed at each fixation.

Accepted to HCI 2008
Liverpool, UK—Sept, 2008
Published by the British Computer Society

Paterson and Tinker used eye gaze tracking to study the typographical issues of font size and font type [8, 9]. For font size, they compared 10 pt (the optimal size, according to their previous reading speed study) against 6 pt and 14 pt, and they replicated their previous speed findings. The 6 pt font was slower due to increased fixation duration, which was probably caused by reduced character visibility. For 14 pt, there were more fixations and thus probably a smaller perceptual span. The perceptual span result, however, is contradicted in a later study by Morrison & Rayner [10], who show that saccade length depends on character spacing, not visual angle – that is, saccades should scale up for larger fonts. For font type, Paterson & Tinker compared Scotch Roman with Old English to further examine a large reading speed difference from [5]. Old English was slower to read because of an increased number of fixations and an increased regression rate.

In this paper, we use eye tracking to investigate how font size and type influence online reading. All the Paterson and Tinker studies were done on paper – will we see the same results on computer screens, and using modern eye tracking equipment? Will saccade length vary linearly with font size, as suggested by [10]? We will use modern computer fonts for the study of font type, as opposed to Scotch Roman and Old English.

2. EXPERIMENT

Using a between-subjects design, we collected data from 82 subjects for each typographical issue, distributed as shown in Table 1. A one-page story was assigned to each typographic issue and formatted appropriately (e.g. small, medium, and large fonts) for each subject group. The stories were taken from a science news web site with universally appealing stories written at an 8th grade reading level; the content was selected to go beyond common knowledge to allow for testing of retention.

Stories are formatted on a single page to avoid scrolling. The text color is black against a white background, and the fonts are anti-aliased. Within paragraph structure, line breaks are constrained to occur at the same location across formatting conditions, preserving line structure for each story. The resulting layouts for font size are different magnifications of the same page (see Figure 1), which avoids confounding font size with paragraph formatting. For the font size task, the font type is Verdana. For the font type task, the font size is 12 pt. (For reference, 12 pt on our experiment monitor has a font cap height of 3.2 mm.)

Subjects participating in the experiment were employees of a major computer company, and we recruited them at two company cafeterias, offering them a cafeteria voucher to reward their participation. While N = 82 for each task, subjects were drawn from a larger pool of 132 subject recordings. Of these recordings, a base set of 114 had usable eye tracking data. We had a good distribution of ages and gender in the base set: 74

Typographic Issue	Conditions		N	
1) Font Size	A) small B) medium C) large	10 pt 12 pt 14 pt	28 27 27	(82 total)
2) Font Type	A) san serif B) serif	Helvetica Georgia	41 41	(82 total)

Table 1. The font sizes and types used in our reading study.

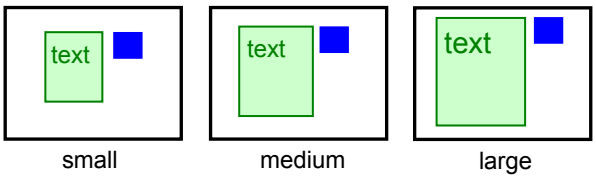

Figure 1. Page layouts for font size task, where the blue box is an accompanying picture.

male, 40 female, 2 subjects in their 20s, 63 in their 30s, 28 in their 40s, 18 in their 50s, and 3 above 60.

During the experiment, subjects sit at a distance of around 60-70 cm from our Tobii 1750 eye tracker, and the system first calibrates the eye tracker for the subject. Subjects are given instructions to read the stories for comprehension, followed by a questionnaire asking for name, first language, and a self-estimate of web usage. Task stories are then presented in a random order, with a 3-question, multiple-choice post-test of retention after each story.

To collect and analyze the experimental data, we use WebGazeAnalyzer (WGA) [11, 12], a tool that records eye gaze in the context of a web browsing session. WGA's web browser is instrumented to record all URLs visited and HTML content, so at analysis time, WGA can automatically map eye gaze to web page text and graphics. By designing our experiment as a series of web pages, we can use WGA to compute reading statistics and thus address our font size and type issues.

3. RESULTS

Our main evaluation criteria are eye tracking reading statistics and post-test retention scores, as these give us a performance measure for the subject's comprehension task. Our eye tracking statistics emphasize speed and re-reading of the material, and include: 1) 1st-pass reading speed (defined as 1st-pass gaze duration / characters read), 2) regression rate, 3) time in return sweeps, 4) fraction of the material re-read, 5) saccade length, and 6) fixation duration.

In our analysis, we found that the greatest source of variation in the data is from the subjects' first language – whether or not it is English. Thus, in reporting on our font issues, we will mention the post hoc analysis of native-English subjects as well as the original subject pool. Following the font issues, we analyze the data in terms of the demographic variables of first language and age.

3.1 Font Size

Table 2 shows eye gaze reading statistics (plus retention) for task 1 on font size. In 1st-pass reading speed, there is a slight trend for reading the larger fonts faster, but this trend is not significant. For example, the 14 pt font is read 12.6% faster than the 10 pt font, but this is not significant, $F(1,53) = 2.845$, $p < 0.1$. We found this lack of influence of font size on reading speed surprising itself, as we were expecting the larger fonts to be easier and faster to read. Furthermore, the regression rate, saccade length, fraction of the material re-read, and retention in the post test are very similar across the font size conditions.

The significant eye tracking differences are in fixation duration and return sweeps. In Table 2, fixation duration shows a nice linear trend of roughly 10 ms/point size that is a "penalty" for the smaller fonts. The 10 pt font induces significantly longer

fixation durations as compared to the 14 pt font, $F(1,53) = 25.6$, $p < 0.00001$.

Reading Statistic	Font Size		
	Small	Medium	Large
1st-pass speed (char/sec)	41.1 (9.5)	44.6 (13)	46.3 (13)
Regression rate (reg/sec)	0.39 (.20)	0.40 (.19)	0.38 (.16)
Total sweep time (sec)	3.48 (1.3)	4.22 (1.2)	4.66 (1.1)
Fraction re-read (%)	30.6 (15)	30.0 (14)	28.2 (15)
Saccade length (char)	11.0 (2.3)	10.8 (2.8)	10.4 (2.7)
Fixation duration (ms)	281 (36)	261 (47)	239 (22)
Retention (% correct)	89.2 (16)	90.1 (18)	88.9 (16)

Table 2. Reading statistics under changes in font size. Row measures with significant differences are shaded, and standard deviations are in parentheses.

On the other hand, subjects given the 14 pt font spend 34% more time in return sweeps than those given 10 pt, $F(1,53) = 12.16$, $p < 0.001$. In order to maintain line formatting across conditions, line length gets longer for the larger fonts, making these return sweeps more difficult for the reader. Beymer, *et al* [12] found this same effect when studying the effect of line length on reading. Figure 2 (top), shows a plot of sweep time vs. font size, and the middle and bottom show histograms of the sweep times for the 10 pt and 14 pt conditions. In the peak near 50 ms, the eye is making a single saccade from the end of one line to the beginning of the next, while the peak at 200 ms involves an additional correction fixation and saccade [12]. The histogram for the 10 pt size is dominated by the single saccade near 50 ms, whereas the 14 pt size has a much larger fraction of sweeps clustered around 200 ms.

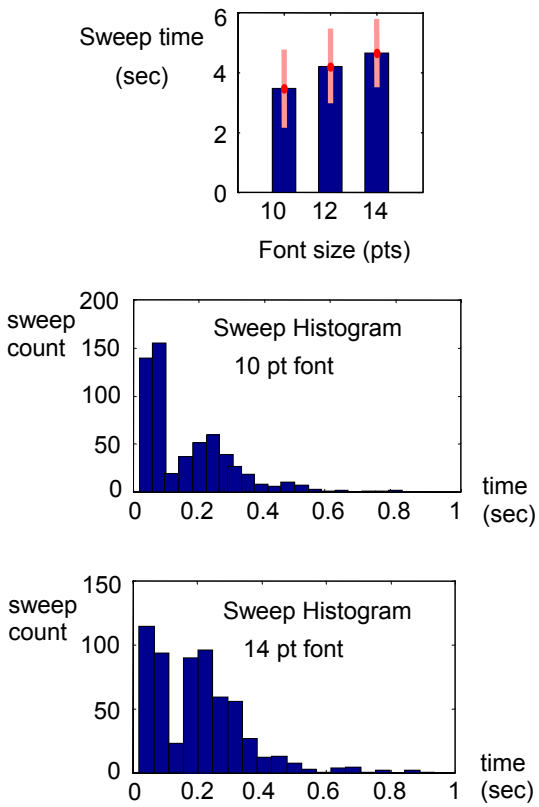

Figure 2. Top, the time spent in return sweeps increases significantly with font size (light bars show std dev). Middle and bottom, histograms of return sweep times for the font size conditions.

The exact same trends and the same significant differences are present in both the original 82 subjects (includes both native-English and non-native-English subjects) as well as the subgroup of native-English subjects.

3.2 Font Type

Using the overall speed metric, the serif font, Georgia, was read 7.9% faster than the san serif font, Helvetica, although this difference is not significant, $F(1,80) = 1.73$, $p < 0.2$. Overall, there are no statistically significant reading or retention differences between the two font types. (This is true for both the original 82 subjects and the native-English subject pool.) Visually, the Georgia and Helvetica versions of the task 2 story are quite similar, so this result is not too surprising.

3.3 Reading Statistic by Demographic

While our data analysis has so far focused on the experimental variables of font size and type, we also collected demographic data from the subjects. In this section, we present a post hoc analysis of reading behavior across age groups and the subject's first language. When investigating a particular demographic, we include reading data from both font tasks.

For subject age, the results from past studies on the effects of aging suggest that reading speed may be lower among the older subjects [13]. To investigate this, we compared a group of 27 subjects in their 30s with a group of 15 subjects in their 50s, where English is the subjects' first language. (The 20s and 60s groups had sample sizes that were too small.) The 1st-pass reading speed for the two subgroups were very close, 48.6 char/sec and 48.1 char/sec, respectively, for the 30s and 50s, an insignificant difference. The only reading statistic that showed a significant age difference was re-reading, with the 30s group re-reading 32% of the material, compared to the 50s group at only 24% ($F(1,40) = 5.69$, $p < 0.03$). This re-reading caused the younger group to spend significantly more overall time on the story URL page ($F(1,40) = 7.135$, $p < 0.02$). Retention was the same between the two groups, so it may be said that the

Reading Statistics – 30s subject group, font tasks	Is English the subject's first language?	
	yes	no
1st-pass speed (char/sec)	48.6 (12)	39.4 (9.5)
Regression rate (reg/sec)	0.43 (.15)	0.44 (.20)
Fraction re-read (%)	31.5 (9.5)	41.9 (16)
Saccade length (char)	11.4 (2.3)	9.9 (2.0)
Fixation duration (ms)	248 (33)	269 (41)
Retention (% correct)	84.0 (15)	88.2 (14)

Table 3. Reading statistics broken down by whether English is the subject's first language. Row measures with significant differences are shaded, and standard deviations are in parentheses.

older group has a higher time-normalized retention score.

To analyze online reading by first language, we divided the subject pool into two groups: (1) subjects in their 30s reporting English as their first language (27 subjects), and (2) subjects in their 30s reporting some other language (34 subjects). Table 3 shows a number of significant differences between the two groups, with the English first group reading faster and re-reading less than the non-English first group. (For example, the 1st-pass speed of the English first group is 23% faster than the non-English first group, a significant difference, $F(1,59) = 10.85$, $p < 0.002$.) This speed difference can be explained by non-English first group having significantly shorter saccades and significantly longer fixation durations. There is no difference in retention or regressions, so the non-English first subjects simply have a time handicap compared to the English first group. These results align well with a previous eye tracking study for ESL [14]; for non-native English speakers, they also found increased fixation duration, decreased speed, but no differences in regressions.

4. DISCUSSION

In this paper, we investigated how the typographical issues of font size and font type impact online reading. For font size, the slight reading speed advantage for larger fonts, while not significant, is probably caused by significantly longer fixations for smaller fonts. Might there be some underlying cognitive explanation for why smaller characters take longer to process on each fixation? This effect of font size on fixation duration agrees with Tinker [9] as well as recent eye tracking work by Yen and Radach [15].

On the other hand, the constancy of saccade length across font size is consistent with Morrison and Rayner's claim that saccade length should scale with point size. Tinker claimed that it would get smaller (smaller perceptual span), and Rayner claimed that it would stay the same (saccade length scales with character size). While saccade length in Table 3 drops just 5% from 10 to 14 pt, this is against the background of a 45% increase in pixel size, so our data favors Rayner's interpretation.

With regards to design, the lack of a significant difference in speed across font sizes may tempt a designer to use a smaller font in order to cram material on one page. However, while we did not quantitatively estimate subject font size preference, the reaction to the small 10 pt font was fairly negative. This is especially true since the eye tracker constrains the subject from getting closer to the monitor. We feel that a combined measure of speed and preference would probably select the 12 pt font.

For font type, our study of serif vs. san serif yielded no significant differences in the eye tracking or retention metrics. There was a 7.9% advantage in 1st-pass speed for the serif font, but the difference was not significant.

Post hoc analysis comparing subject age groups did not find reading speed reductions in older subjects. However, existing studies on reading in older adults typically use an older population sample (70s) than our sample (50s), which probably explains this difference. Post hoc analysis of the subjects' first language yielded large speed differences favoring native-English subjects. This difference can be explained by significantly longer fixations and shorter saccades in non-native English subjects.

5. REFERENCES

[1] Tinker, M. Legibility of Print. Iowa State University Press, Ames, Iowa (1963)

[2] Paterson, D.G., Tinker, M.A. Studies of Typographical Factors Influencing Speed of Reading: II. Size of Type. *Journal of Applied Psychology*, 13:2 (1929), 120-130.

[3] Legge, G.E., Pelli, D.G., Rubin, G.S., Schleske, M.M. Psychophysics of Reading: I. Normal Vision, *Vision Research*, 25:2 (1985), 239-252.

[4] Bernard, M., Lida, B., Riley, S., Hackler, T., Janzen, K. Comparison of Popular Online Fonts: Which Size and Type is Best? *Usability News*, 4.1, (2002)

[5] Paterson, D.G., Tinker, M.A. Studies of Typographical Factors Influencing Speed of Reading: X. Style of Type Face. *Journal of Applied Psychology*, 16:6 (1932), 605-613.

[6] Boyarski, D., Neuwirth, C., Forlizzi, J., Regli, S.H. A Study of Fonts Designed for Screen Display. *Proc. CHI 1998*. 87-94.

[7] Rayner, K. and Pollatsek, A. The Psychology of Reading. Lawrence Erlbaum Associates, Hillsdale, NJ (1989)

[8] Paterson, D.G., Tinker, M.A. The Effect of Typography upon the Perceptual Span in Reading. *American Journal of Psychology*, 60 (1947), 388-396.

[9] Tinker, M.A., Paterson, D.G. The Effect of Typographical Variations upon Eye Movement in Reading. *Journal of Educational Research*, 49 (1955), 171-184.

[10] Morrison, R.E., Rayner, K. Saccade Size in Reading Depends upon Character Spaces and Not Visual Angle. *Perception and Psychophysics*, 30:4 (1981), 395-396.

[11] Beymer, D. and Russell, D. WebGazeAnalyzer: A System for Capturing and Analyzing Web Reading Behavior Using Eye Gaze. In *CHI 2005 Extended Abstracts*. 1913-1916.

[12] Beymer, D., Russell, D., Orton, P. Wide vs. Narrow Paragraphs: An Eye Tracking Analysis. In *Interact 2005*. 741-752.

[13] Kemper, S. and McDowd, J. Eye movements of Young and Older Adults while Reading with Distraction. *Psychology and Aging*, 21:1 (2006), 32-39.

[14] Oller, J.W. Assessing Competence in ESL: Reading. TESOL Quarterly, 6:4 (1972), 313-323.

[15] Yen, M.-H. and Radach, R. Saccades and binocular coordination in reading: Effects of viewing distance and font size, In *14th European Conference on Eye Movements* (2007)

[16] Darroch, I., Goodman, J., Brewster, S.A., and Gray, P.D. The effect of age and font size on reading text on handheld computers, In *Interact 2005*, 253-266.