Design Guidelines for readability on a computer screen with focus on typography

Oskar Mothander

Department of Applied Physics and Electronics
Umeå University, Sweden
dit06@cs.umu.se

Abstract. This paper explains the background of typeface problems and how old problems have affected the view of when to use a Sans-Serif typeface versus using a Serif typeface on a digital screen. The study for this paper has proven that Serif typefaces are better for readability, especially when it comes to users understanding the text they read, on both paper and on digital screens.

1 Introduction

The art of writing is at least a thousand years old and the art of printed writing is a about five hundred years old [1]. We have on the other hand only used computers for reading and writing for the last few decades, therefore it is far from as optimized and researched as much as the two previously mentioned.

The importance of readability applies to all areas where users are presented with text. It is not only important for school purposes and learning, most situations benefit from users reading with ease. A web-page selling products will obviously sell better if users better comprehend the intended message or comprehend the message with less effort. In any situation where the intention is that users should understand what is written; readability is important!

In the past screens have lacked sufficient resolution, computers and software have lacked support for anti-aliasing and typefaces have not been adjusted to be presented on a screen. These issues created problems especially with Serif fontfaces for which the simple solution was to recommend user instead use Sans-Serif fontfaces. Today these problems have almost vanished. One example is that iPhone 4 is claimed to have a resolution that is practically the maximum of what the average human can distinguish. That is to say, any higher resolution would not improve (or deteriorate) the experience at the normal viewing distance.

The hypothesis of this paper is that with modern computer systems reading on a screen is not much different to reading on paper. Research has been done that does in fact supports this [2]. Therefore the same guidelines should be used for both. Most of the guidelines for text on screen has been determined because of problems with presenting the user with the same visual experience as on paper. The main focus of the paper is to bust the myth that you should always use a Sans-Serif typeface for screens. The hypothesis is actually the opposite, that Serifs increase readability.

2   Background

In this article typeface is used as a wide term for the style of letters and glyphs on any medium. Fontface is used to refer to the digital representation of the typeface. Body of copy text is normal text in paragraphs, that is to say not the headlines and similar. The term readability is used to describe the users ease in reading text. Good readability supports fast reading and good comprehension of the text read.

In classic typography (on paper) Serif-typefaces where developed to be used in body of copy text for readability, whilst Sans-Serif typefaces usually where used for headlines and stylishness [3, 4]. The design of the bottom of letters in Serif typefaces have proven to aid horizontal eye tracking while reading text (Schultz:1997). Schultz also points out that variations in the top of the letters also help users to distinguish between letters. While most users do not actually read words letter by letter, but instead see the pattern of words, capitalization of words decrease the difference in patterns between words and even letters; since the top of each letter always extends to the same height (example: ALWAYS). Therefore capitalizing words decrease the readability and should be avoided or at least very sparsely used [5].

2.1   Aliasing and Anti-aliasing

In signal processing aliasing refers to when an curved object is poorly represented in the digital world. Letters that have curved lines can be disturbed by aliasing. In the digital world where you have pixels, a line is represented as pixels where a pixel is either filled or not (usually black or white). When a line curves, there might become a gap because the pixels are square. Since the pixels are not directly above or next to each other the line does not "fill" as the analog variant would. For a user the line appears as "hairy" and discontinuous. See figure 2.1.

Anti-aliasing solves this problem by inserting semi-transparent (usually gray) pixels in these gaps. For the user the line then appears as continuous [6, 7].

Fig. 1. Times New Roman aliased versus anti-aliased [6]

3   Review of studies

A fair amount of font-face studies that have been performed are out of date with modern screens and software. Additionally many studies rely on users not having

anti-aliasing support or not having it installed [8], which in fact practically all modern computer systems have [9, 10]. Many books and tutorials today (and especially in the past) tell developers to use a Sans-Serif typeface for computer screens. David Benyon (etal.) for example states in their book "Designing Interactive Systems" that a Sans-Serif font is the best for readability on a screen [11]. The reason they give is that Serif fonts "can appear hairy onscreen" which can be a result of aliasing [9]. The authors however also point out that these issues can be mitigated with anti-aliasing [11].

Studies preformed by Wheildon 2005 show a significant increase in comprehension when using a Serif typeface instead of a sans-Serif [4]. Wheildon goes on to give this recommendation:

> "Use Serif typefaces for body text (no smaller than 8 points and no bigger than 12 points)" [3]

In the previously mentioned book by David Benyon, they point out that

> "type sizes less than 10 pixels are difficult to read in most situations, and if you must use small type, it's usually best to go with a sans-Serif typeface without anti-aliasing" [11].

They give however no reference as to how this conclusion was arrived at.

4   User Study

A user study has been preformed for this article by studying Internet users efficiency in a reading comprehension test with different fontfaces.

4.1   Method

Existing reading comprehension tests were used that had three selected parts from Mark Twain novels [12]. Each text had four questions. A web-page was built to support two fontfaces. The answers where logged in a database and the tests where timed in order to measure efficiency. The web-page automatically selected the fontface which had less submitted results. If both fontfaces had an equal amount of test subjects, the fontface was selected randomly between Georgia (serif) and Arial (sans-serif).

Users were aware that the test was timed but did not know the expected time the test would take. The users were not informed of the expected time as this could influence users to rush through the test instead of taking the time they needed for the test. The authorshypothesis was that a statistically significant time difference would show between the two user groups.

Half of the users performed the reading comprehension test with Arial as chosen fontface. Arial was chosen simply because it is a common fontface on the web today and a sans-Serif typeface [8]. The other half had Georgia as fontface which is a Serif fontface especially designed for screens [8]. Both user groups had the same text and questions.

Users also stated their native language as this could have an impact on the results. The test was written in English, most users had Swedish as native language but some of the user had English. Since other research studies [8] had been unable to show any significant difference in readability the test was made to be quite long with the intention of bringing forth even small differences. In the authors opinion the font-size used was quite small (14 pixels) to make the test harder for users. In theory the Serifs should then help the users keep track of where they are in the text and ease the understanding of words. A small research study in 2009 indicates however that the average font size on the web to be 13 pixels [13]. The hypothesis was that if the test was too simple (too short lines or too little amount of words) Serifs would make no difference.

4.2   Measurements

Since making readability tests are quite time consuming and there are a lot of factors that you could consider, the test was made as simple as possible. The test was focused on determining if there was any significant difference in reading efficiency between a Sans-Serif fontface and a Serif fontface. To ensure the test differed only in this perspective the two pages were made identical except for the fontface. Because fontfaces have different default sizes (in relative measurements) the web-page had an absolute font-size in pixels set to make the letters identical in height. This was done to avoid sources of errors where users benefited from larger sized letters. The study assumed that most users had a laptop screen 13 to 15 inches as the test was preformed mostly by friends of the author.

5   Results

A total of 14 users preformed the test, 7 user for each fontface. The average reading time for the Georgia test group was 8 minutes and 2 seconds with a standard deviation of 3 minutes and 54 seconds. For Arial the average time was 6 minutes and 56 seconds with a standard deviation of 2 minutes and 12 seconds. The shortest time spent on the test from a user was 3 minutes and 18 seconds, the user had Georgia as fontface. The highest was 15 minutes and 41 seconds, the user had Georgia as fontface. Through a Welch t-test [14, 15] no significant difference in time could be claimed.

Table 1. Time based calculations [16, 17]

	Georgia	Arial
Mean time:	481.8571	416.4286
St. deviation:	233.8610	131.7800
T-Test:	$t = 0.6449$ df $= 6.076$	

In the understanding part of the test, the Arial test group had an average of 5 correct answers with a standard deviation of 2. The Georgia test group had an average of 7 correct answers with a standard deviation of about 1.7. The lowest score from a user was 2 out of 12 correct answers, the user had Arial as fontface. The highest score was 9 out of 12 correct answers, the two users with this result both had Georgia as fontface. Through a Welch t-test it could be claimed that Georgia was in fact significantly better for understanding with 90% certainty.

Table 2. Understanding based calculations [16, 17]

	Georgia	**Arial**
Mean corr answers:	7.0000	5.1429
St. deviation:	1.7728	2.0304
T-Test:	t = 1.8229 df = 18.70	

6   Discussion

At a first glance calculations showed no significant difference between the two fontfaces. The hypothesis was that users would spend more time on a less readable text to comprehend the text. Instead users spent, statistically, an equal amount of time and the comprehension varied between the two groups.

The readability test has been intentionally left open so more users can submit the test and the statistics will be recalculated automatically by the script that provides data for this article. Hopefully it can later show a higher certainty (than the now 90%) that Serif is more easily understood than Sans-Serif. It is possible that a time difference will show up as well. Depending on how important the users feel the test is it is possible they might rather give up before understanding everything, just because they know the test is anonymous and will not affect them in a bad way. The test was found quite hard by all interviewed users and the fact that not a single user got all answers correct also supported this. However, making a test that is not anonymous might discourage users from preforming the test at all. If they preform poorly others will be aware of their failure.

Some suggestions for further studies in the area of this article are for example; studying whether fontfaces constructed especially for screens have an actual impact on readability when you compare two Serif fontfaces.

Only one user selected English as native language therefore it was impossible to say if this had any significant impact on readability. Although since the total result from all users claimed significance, the language was not of importance for the study. The intention of the study was not to show that native language affected performance. The purpose of including preferred language in the test was to study native (English) speaking people if the people with English as

second language failed to show any significant results. It is highly likely however that native English speaking subjects would be a preferred test group in order to show higher significance in the tests. Because of the small number of willing test subjects and the lack of readability tests in Swedish (the native language for most users) this was however not possible.

7   Conclusion

With todays technology Serif typefaces are in fact more readable for long texts no matter if the text is on a screen or on paper. In this report the user study has proved this with 90% certainty statistically. Sans-Serif typefaces should be used as they were used in the past on paper; for design purposes only, examples of this situation are in headlines and other short text parts. Previous research also indicates that typefaces smaller than 10 pixels are difficult to read in any situation [11]. Another factor to consider is that capitalizing of words decrease the readability and should be avoided or at least very sparsely used. [5]

It should however be said that the research preformed here depends on users not using very old systems that lack support for anti-aliasing and use poor screens. The fontface Georgia, which was used as the Serif fontface in the readability study, is constructed especially for readability on a screen. A very common and popular Serif typeface is Times New Roman which was not made especially for screens (but for printing). All research (see background and results) done for this paper indicate however that similar results are expected with these Serif fontfaces. That is to say; they should preform better in readability than Sans-Serif fontfaces on both paper and screen. It is unclear if the construction of typefaces especially for screens has a significant impact on readability, whereas Serif versus Sans-Serif turns out to be a big factor when it comes to readability. Research done for this paper gives the hypothesis of this paper, that Serifs increase readability, scientific support.

References

[1] Needham, P.: The First Four Printed Bibles in the Scheide Library: Exhibited in Firestone Library, Princeton University, 31 May 2002. Princeton University Library (2002)

[2] Gould, J.D., Alfaro, L., Finn, R., Haupt, B., Minuto, A.: Why reading was slower from crt displays than from paper. In: Proceedings of the SIGCHI/GI conference on Human factors in computing systems and graphics interface. CHI '87, New York, NY, USA, ACM (1987) 7–11

[3] Smith, M.P.: Keep it simple stupid: Using basic design maxims for effective communication (1998) http://www.stcsig.org/mgt/docs/readability-kiss.pdf, accessed 2010-11-28.

[4] etal, C.W.: Type and Layout - Are you communicating or just making pretty shapes. THE WORSLEY PRESS (2005) http://www.eyelevelbooks.co.uk/pdfs/TypeBookextract.pdf, accessed 2010-11-28.

[5] Schultz, T.W.: Why johnny won't read - producing readable printed matter. 1997 ASEE/IEEE Frontiers in Education Conference (1997)

[6] Author, U.: Aliasing http://en.wikipedia.org/wiki/Aliasing, accessed 2010-12-23.

[7] Author, U.: Anti-aliasing http://en.wikipedia.org/wiki/Anti-aliasing, accessed 2010-12-23.

[8] Bernard, M., Mills, M.: So, what size and type of font should i use on my website? Usability News, Vol. 2 Issue 2 (2000) http://www.surl.org/usabilitynews/22/font.asp, accessed 2010-11-07.

[9] Author, U.: Cleartype: frequently asked questions (2010) http://windows.microsoft.com/en-US/windows-vista/ClearType-frequently-asked-questions, accessed 2010-11-28.

[10] Author, U.: Advanced typography with mac os x tiger - using and managing fonts (2005) http://images.apple.com/pro/pdf/L311277A_FontTT_v4.pdf, accessed 2010-11-28.

[11] Benyon, D., Turner, P., Turner, S.: Designing interactive systems: People, activities, contexts, technologies. Addison-Wesley Longman (2005)

[12] Jackson, R.: Reading comprehension quiz (2009) http://www.teacherneedhelp.com/readingquizzes/, accessed 2010-12-03.

[13] Martin, M.: Typographic design patterns and best practices (2009) http://www.smashingmagazine.com/2009/08/20/typographic-design-survey-best-practices-from-the-best-blogs/, accessed 2011-01-14.

[14] Dr. Fuller, P.: Welch t test for independent samples http://frank.mtsu.edu/~dkfuller/notes302/welcht.pdf, accessed 2011-01-06.

[15] : t-test table http://changingminds.org/explanations/research/analysis/t-test_table.htm, accessed 2011-01-06.

[16] Mothander, O.: Readtest statistics (live) http://www.oskarmothander.se/readtest/stats.php, accessed 2011-01-06.

[17] Mothander, O.: A cognitive test in reading comprehension and efficiency http://www.oskarmothander.se/readtest/, accessed 2011-01-06.