Legibility of Light and Ultra-Light Fonts: Eyetracking Study

Ivan Burmistrov
interUX Usability Engineering Studio
Tallinn, Estonia
ivan@interux.com

Maria Semenova
Lomonosov Moscow State University
Moscow, Russia
154mmi223@gmail.com

Tatiana Zlokazova
Lomonosov Moscow State University
Moscow, Russia
t.zlokazova@gmail.com

Iuliia Ishmuratova
Lomonosov Moscow State University
Moscow, Russia
ishmuratova08@gmail.com

NordiCHI '16, October 23-27, 2016, Gothenburg, Sweden
© 2016 ACM. ISBN 978-1-4503-4763-1/16/10…$15.00
DOI: http://dx.doi.org/10.1145/2971485.2996745

Abstract
The use of light and ultra-light fonts has become an omnipresent trend in the design of modern user interfaces. Although this trend has been criticized by a number of usability experts, no empirical research exists to date on the legibility of these fonts. We present the results of a preliminary eyetracking study showing that light and ultra-light fonts are less legible than their regular and bold counterparts in two variations of text-background contrast (low vs high) and two variations of text-background polarity (positive vs negative). Oculomotor indicators like mean fixation duration and saccade amplitude show that light and ultra-light fonts also induce higher cognitive load. Our study suggests avoiding light and ultra-light fonts for body text.

Author Keywords
Font weight; light fonts; ultra-light fonts; legibility; text-background contrast; text-background polarity, eyetracking; cognitive load.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction

The recent cardinal changes in the visual style of user interfaces commonly referred to as "modern", "flat" or "minimalistic" design have led, in particular, to the omnipresent use of light and ultra-light fonts and low text-to-background contrast in all digital products, including operating systems, websites and mobile apps. This change was initiated by the leading software producers like Microsoft, Apple and Google and was enthusiastically received by the graphic design community. The software industry leaders' striving to make their products look aesthetically appealing and innovative by consciously sacrificing basic ergonomics and usability principles coincided with the ignorance of these principles among many graphic designers, who often treat textual content simply as a kind of visual ornamentation [2]. The result was that low-weight low-contrast fonts became the new norm in contemporary user interfaces.

Although this trend has been specifically criticized by a number of usability experts [5, 6, 11], at the current stage of evolution of user interfaces when their design has become a fashion-driven practice [7], these claims often fall on deaf ears.

Nevertheless, we believe that expert opinion must be supported by experimental research and in the present study (to the best of our knowledge, this is the first empirical research on light and ultra-light fonts) we report the preliminary results of our eyetracking experiment aimed primarily at measuring the legibility of light/ultra-light fonts vs normal/bold ones. This is supplemented by an analysis of two other factors usually studied in legibility research, namely text-background contrast and text-background positive vs negative polarity (i.e. dark text on light background vs light text on dark background) since the latter also appears to be a trend in modern user interface design.

Experiment

Participants

Twenty four volunteer participants (12 male and 12 female; aged between 18 and 37, average – 24.7) took part in the experiment. All participants were experienced internet and text editor users and had normal or corrected-to-normal visual acuity.

Experimental design

The experiment was a repeated measures design. Within-subjects factors were:

- font weight with four levels: Ultra-Light, Light, Normal and Bold;
- background color with two levels: White and Black (i.e. positive and negative text-background polarity);
- contrast between text and background with two levels: High – black typeface on white background or white typeface on black background, and Low – gray typeface (50% gray, R:128, G:128, B:128) on white or black background.

Each combination of independent variables consisted of three tasks with a different position of the target word within one of three paragraphs.

Dependent variables were time of visual search and oculomotor indicators – fixation duration and saccade amplitude. These oculomotor parameters are considered in the literature as indicators of the

Figure 1: Sample image showing the layout of text on a stimulus screen.

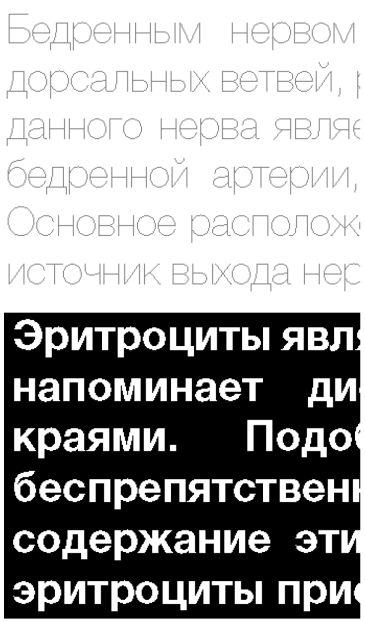

Figure 2: Example cut-outs from actual stimuli: Ultra-Light, Gray on White (top) and Bold, White on Black (down).

cognitive load [3, 8, 10], mean fixation durations are also considered as an indicator of text readability [9].

The average means of the dependent variables were calculated for each combination of independent variables (font weight, background color and contrast).

Task and stimuli

The task for participants was to search for a target word on a text page. The search task method has been used in legibility research since 1960s and its use is motivated by a disregard for text comprehension factors unrelated to the visual qualities of the text [1]. Also, this task has high ecological validity because it corresponds to real usage situations as users frequently use Google to find links to webpages relevant to their search terms and then visually search for specific words on a corresponding webpage.

The texts were taken from an online guide to human anatomy and physiology. For all texts the Helvetica Neue typeface of 12pt size was used. We chose this typeface because it is very popular in modern user interfaces (in particular, Helvetica Neue was a system typeface in versions 4 to 8 of the Apple iOS mobile operating system), and one expert opinion suggests that specifically Helvetica Neue Light and Ultra-Light fonts can be tedious to read [4].

Each page comprised three paragraphs of 6 lines each (**Figure 1**, **Figure 2**). This allowed all of the subjects to perform the search tasks within a reasonable timeframe. The target word (e.g. "movement") always consisted of 8 letters and was placed randomly in the first, second or third paragraph.

Procedure

The experiment was divided into two experimental series: (1) positive polarity, where black or gray texts appeared on a white background, and (2) negative polarity, where white or gray texts appeared on a black background. Each participant took part in both series. The order of series (first positive, then negative, or the other way round) was counterbalanced.

In each series participants performed 24 word search tasks. Random presentation of stimuli assigned to different experimental conditions was used. In each search task, first the target word was displayed in the center of a computer screen. Participants were instructed to press the "space" button when they memorized the target word. Then the text page was displayed. Participants were asked to fixate their gaze on the target word when they found it, and then press the "space" button to proceed to the next task. The detection of the target word was controlled by the experimenter on a separate monitor. The on-screen picture and subject's gaze position were recorded, allowing search accuracy to be re-checked after the experiment. The participants were given instructions before the experiment and two training tasks before each series.

Apparatus

A 23 inch LCD monitor with 1920*1080 pixel resolution was used to display stimuli. The viewing distance was 75 cm due to use of a chin rest to stabilize the participant's head. Eye movements were recorded with an infrared video-based eye tracker (SMI iView-X Hi-Speed 1250) at a sampling rate of 500 Hz and an instrument spatial resolution of 0.01°. BeGaze 3.6 software was used for data analysis.

Results

To evaluate the effects of the font weight, the text-background polarity and the contrast level on the dependent measures the 4x2x2 ANOVA with repeated measures was used. To specify these effects the paired-samples t-test was used. Mean values of search time, fixation duration and saccadic amplitude for different text-background contrast and polarity conditions are presented in **Figure 3**.

Search time

The effect of font weight on **search time** was significant ($F(3,21)=6.86$, $p=0.002$). Search time was lower when thicker fonts were used (Normal and Bold) in comparison with thinner ones (Ultra-Light and Light). In addition, with Normal font search time was the lowest, and with Light font – the highest. Pairwise comparison (t-test) showed that search time when Normal font was used was significantly less than when Light font was used in all experimental conditions ($p<0.01$ for black background and high contrast; $p<0.05$ – for all other combinations of factors). In terms of search time, Normal font appeared better than Ultra-Light against a black background and under all contrast conditions ($p<0.01$), and better than Bold against a black background with low contrast ($p<0.05$).

There was also a significant effect of text-background polarity ($F(1,23)=5.39$, $p=0.030$). In general, search in texts written on a black background was slower than in texts on a white background.

No significant interactions between different independent measures were observed.

Oculomotor indicators

A significant effect of font weight on **duration of fixations** was observed ($F(3,21)=5.66$, $p=0.005$). The results confirmed that the thicker fonts are better for performing a visual search. When the font was thinner there was an increase in the average fixation duration. This indicates a greater cognitive load and a reduction in text readability. Contrast level also had a significant effect ($F(1,23)=13.70$, $p=0.001$) – the average fixation duration appeared to be higher when contrast level was low. Pairwise comparison showed differences in fixation duration only in conditions of positive text-background polarity. In particular, a higher fixation duration was observed with Ultra-Light font, compared to Normal ($p<0.01$) and Bold ($p<0.05$) fonts in high contrast condition. Also, Bold had an advantage over all other fonts when contrast was low ($p<0.01$).

A highly significant effect of font weight on the **saccadic amplitude** ($F(3,21)=21.96$, $p<0.001$) was observed, which also confirmed the advantage of thicker fonts. With thinner fonts (Ultra-Light and Light) we detected lower saccadic amplitude, reflecting a higher level of cognitive load. The lowest saccadic amplitude was obtained with Light font. The paired-sample t-test for the positive text-background polarity condition in general showed the advantage of Normal and Bold fonts over the thinner ones ($p<0.05$ for high contrast, $p<0.01$ for low contrast). With negative text-background polarity we observed the advantage of Normal and Bold fonts over Light font ($p<0.05$) when there was high contrast, and also the advantage of Bold font over Ultra-Light and Light fonts ($p<0.01$) in low contrast conditions.

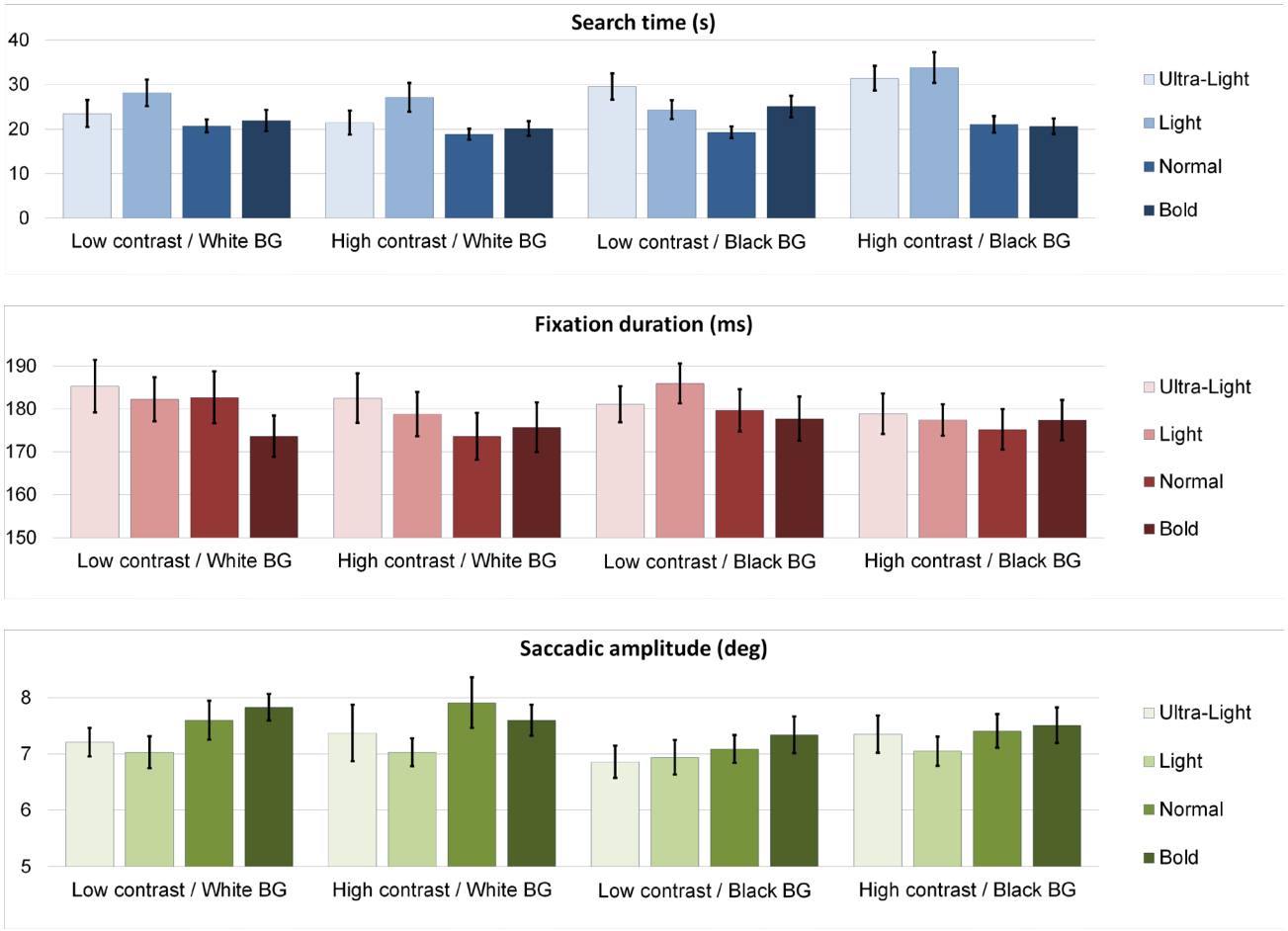

Figure 3: Mean values of search time, fixation duration and saccadic amplitude for different contrast and background (BG).

Conclusion and future work

The results of our preliminary study show that light and ultra-light fonts are less legible than their regular and bold counterparts in two variations of text-background contrast (low vs high) and two variations of text-background polarity (positive vs negative). Oculomotor indicators like mean fixation duration and saccade amplitude show that light and ultra-light fonts also induce higher cognitive load.

Our future research will include more participants of different ages, more text-background contrast gradations (for example, we have not tested gray backgrounds), and more text size variations because the dependence between legibility and font size is probably non-linear. We also plan to extend our methodology by capturing subjective user experience.

However, our study recommends avoiding light and ultra-light fonts for *body* text.

Acknowledgements

We thank Dr. Boris B. Velichkovsky, Zahhar Kirillov and Andrei Filimonov for their invaluable help in analyzing eyetracking data. The research was supported by a grant from the Russian Foundation for Basic Research #14-06-00371.

References

[1] Beier, S. *Typeface Legibility: Towards Defining Familiarity*. PhD Thesis. The Royal College of Art, London, 2009.

[2] Contrast Rebellion. http://contrastrebellion.com.

[3] Crosby, M.E., Iding, M.K., and Chin, D.N. Visual search and background complexity: Does the forest hide the trees? *Proc. UM 2001*. Springer (2001), 225-227.

[4] Itkonen, M. *Typography and Readability*. The Finnish Centre for Easy to Read, 2006.

[5] Massey, B. Landing page templates that maximize conversions. http://conversionsciences.com/blog/landing-page-templates-that-maximize-conversions/.

[6] Norman, D., and Tognazzini, B. How Apple is giving design a bad name. http://fastcodesign.com/3053406/how-apple-is-giving-design-a-bad-name.

[7] Pan, Y., and Stolterman, E. What if HCI became a fashion-driven discipline? *interactions 22*, 6 (2015), 50-53.

[8] Pomplun, M., Reingold, E.M., and Shen, J. Investigating the visual span in comparative search: The effects of task difficulty and divided attention. *Cognition 81*, 2 (2001), B57-B67.

[9] Rello, L., Pielot, M., and Marcos, M.-C. Make it big!: The effect of font size and line spacing on online readability. *Proc. CHI '16*. ACM Press (2016), 3637-3648.

[10] Renshaw, J.A., Finlay, J.E., Tyfa, D., and Ward, R.D. Designing for visual influence: An eye tracking study of the usability of graphical management information. *Proc. INTERACT '03*. IOS Press (2003), 144-151.

[11] Sherwin, K. Low-contrast text is not the answer. http://www.nngroup.com/articles/low-contrast/.