Perceived Legibility and Aesthetic Pleasingness of Light and Ultralight Fonts

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

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

ABSTRACT

In the last five years the use of light and ultralight fonts has become the new norm in the user interface design of most digital products. Our previous research, based on performance measures, showed that lightweight fonts have negative effects on text legibility and cognitive load when performing word search tasks. However, objective data and subjective perceptions of legibility and aesthetic appeal of fonts do not always correspond. In this paper we present the results of a subjective evaluation study of four variations of the Helvetica Neue typeface (ultralight, light, normal and bold) presented with high vs. low contrast and positive vs. negative polarity. 63 subjects volunteered in a pairwise comparison survey aimed at evaluating the preferences of respondents with regard to perceived legibility and aesthetic pleasingness of sixteen combinations of font weight, contrast and polarity. The results suggest that users of different ages evaluate the legibility of ultralight font as being very low, but younger users may perceive ultralight fonts as more aesthetically appealing than do older users. Based on our study, we provide recommendations on the use of lightweight fonts in user interface design.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)** → Empirical studies in HCI

KEYWORDS

Perceived legibility, perceived pleasingness, ultralight fonts

ACM Reference format:

T. Zlokazova and I. Burmistrov. 2017. Perceived Legibility and Aesthetic Pleasingness of Light and Ultralight Fonts. In *European Conference on Cognitive Ergonomics 2017 Proceedings*, 4 pages.
https://doi.org/10.1145/3121283.3121296

ECCE 2017, September 19-22, 2017, Umeå, Sweden

ACM ISBN 978-1-4503-5256-7/17/09…$15.00
https://doi.org/10.1145/3121283.3121296

1 INTRODUCTION

In the past five years the use of light and ultralight fonts (often combined with low text-to-background contrast and negative text-to-background polarity) has become the new norm in the user interface design of most digital products. This trend contradicts the knowledge accumulated during more than a century of empirical legibility research, and there are a number of compelling reasons to believe that software industry leaders have consciously sacrificed the legibility of text in favour of the aesthetic appeal and innovative look of their products:

1) for decades, experts in typography have been convinced of the superiority of regular and bold fonts over their lighter variants, to the extent that lightweight fonts have almost never been included in legibility studies [9];
2) obvious congruency in the treatment of typography by industry-leading digital vendors (eg. Microsoft, Apple, Google and Sony) was not supported by any published research, so it can be supposed that their design teams are simply "aping" each other's work [2];
3) typography-related controversies regularly shake the industry (some examples: in 2012, Microsoft received a massive negative feedback from Visual Studio users because of their switch to uppercase fonts in the software main menu; in 2013, due to user complaints, the ultralight font introduced by Apple as a system font in the initial release of iOS 7 was replaced by a more legible light font; three years later, Apple was under fire again, now because of its replacement of light fonts with bold ones in iOS 10; in 2016, Amazon was criticized for "anorexic" fonts introduced in Kindle 5.7.2 and, as a result, had to rollback the update).

One might almost have the impression that software vendors' top management is enforcing a new typographic style without conducting usability testing or asking users about their attitudes to this new trend in type design.

At the current stage of evolution of user interfaces their design became a fashion-driven practice [7] and design decisions are often made on a "cool / not cool" basis [4]. In particular, lightweight fonts are currently considered being cool.

Our previous experimental research, based on performance measures and oculomotor indicators of cognitive load, showed that lightweight fonts have negative effects on text legibility and cognitive load when performing word search tasks [1]. However, objective performance data and subjective perception of legibility and user preferences do not always correspond [5].

In the present research, echoing a classic study of Tinker and Paterson [8], we asked users to evaluate the subjectively perceived legibility and aesthetic appeal of ultralight and light fonts in comparison with their thicker counterparts – fonts of normal and bold weight. Of particular interest were possible age differences in font perception since it is known that younger users may perceive fashionable user interface design more positively than older users [6].

2 MATERIALS AND METHOD

2.1 Study design

The research used a repeated measures design. Within-subjects factors were:

1) font weight with four levels: Ultralight, Light, Normal and Bold;
2) background color with two levels: White and Black (i.e. positive and negative text-to-background polarity);
3) contrast between text and background with two levels: High – black text on white background or white text on black background, and Low – gray (50% gray, RGB=128:128:128) text on white or black background.

Dependent variables were participants' subjective judgments on text "legibility" and "pleasingness". Following the approach used in [8], legibility was defined in the survey instruction as "ease and speed of reading", and no specific definition was provided for "pleasingness".

2.2 Stimuli

The stimuli were images of a text, 200×300 pixels in size. All images presented the same text – a fragment of a famous folk tale. Helvetica Neue typeface (font size 6pt) was chosen for the typographic treatment of stimuli due to its popularity in modern user interfaces.

In accordance with varied factors (4 font weights × 2 text-to-background polarities × 2 contrasts), 16 stimuli were prepared.

2.3 Task and Procedure

The research consisted of two series:

a) evaluation of text legibility,
b) evaluation of text pleasingness.

For text evaluation in each series the pairwise comparison method was used. The task for participants was to choose the more legible / pleasant text in each pair of stimuli. A sample task is presented in Fig. 1.

The number of text pairs to compare was calculated as follows: $(16^2-16)/2=120$. So, in each series participants compared 120 pairs of text images. The order of the series (first legibility evaluation, then pleasingness evaluation, or the other way round), the order of stimuli presentation within series, as well as the stimuli position in a pair (left or right) were randomized.

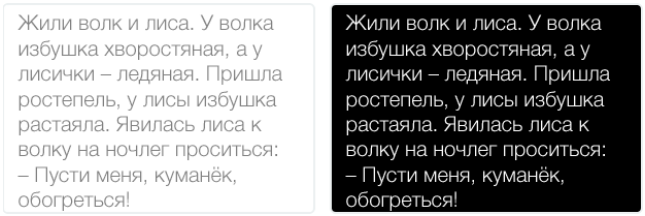

Figure 1: Sample stimuli for pairwise comparison survey

The online survey tool SurveyGizmo was used to conduct the research.

2.4 Participants

An invitation to participate in the survey was sent via e-mail and social networks. The overall sample included 63 volunteers who completed the survey in full. The participants were aged 19–68 years (mean age was 39.0), 21 male and 42 female. (A preliminary data analysis did not reveal any gender-related differences, so we did not include the gender factor in further statistical tests.)

3 RESULTS AND DISCUSSION

3.1 Data analysis

To calculate preference scores we followed the procedure described in [3]. The preference score for each stimulus (or research condition) was calculated as a sum of its selections in each series. The preference level for a stimulus ranged from 0 to 15 – the bigger the score, the higher the stimulus preference. Mean values of legibility and pleasingness scores for different conditions are presented in Fig. 2.

To evaluate the effects of font weight, text-to-background polarity and contrast on the dependent measures the 4×2×2 ANCOVA with repeated measures was used. The respondent's age was added as a covariate, as it might affect participants' pleasingness and legibility scores. To specify the effects of the main factors the paired-samples Wilcoxon signed-rank test was used.

3.2 Legibility

ANCOVA and Wilcoxon test results for legibility scores are shown in Table 1 and Table 2.

Table 1: Legibility: ANCOVA results

Factors	*df*	*F*	*Sig.*
Weight	1.79, 109.02	460.894	0.000
Polarity	1.61	54.862	0.000
Contrast	1.61	379.496	0.000
Weight×Polarity	2.57, 156.87	4.210	0.010
Weight×Contrast	2.22, 135.53	25.966	0.000
Weight×Polarity×Contrast	2.71, 165.37	4.750	0.005

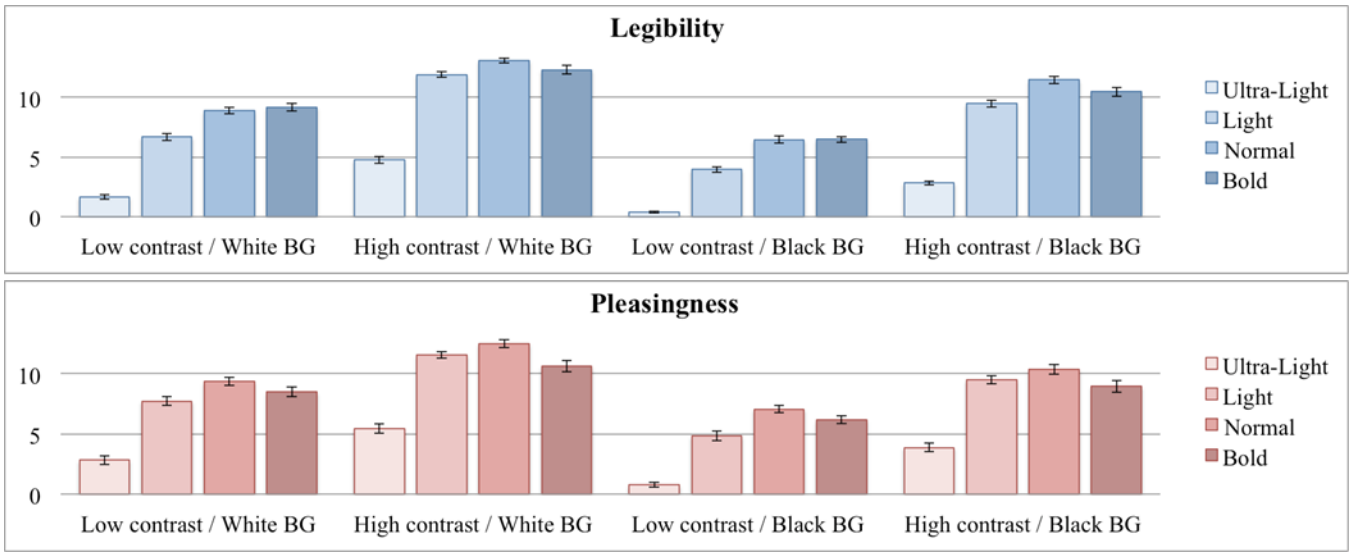

Figure 2: Legibility and pleasingness scores

Table 2: Legibility: Wilcoxon signed-rank test for white vs black background and high vs low contrast

	Ultralight	Light	Normal	Bold
Ultralight		**-6.899*** ** -6.929***	**-6.880*** ** -6.926***	**-6.593*** ** -6.912***
Light	**-6.932*** ** -6.936***		**-4.458*** ** -6.290***	**n/s** -4.964***
Normal	**-6.915*** ** -6.921***	**-5.508*** ** -6.177***		**n/s** n/s
Bold	**-6.843*** ** -6.937***	**-2.413*** ** -5.861***	**-2.494*** n/s	

* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; n/s – non-significant

The results showed that texts with higher font weight, positive text-to-background polarity and high contrast were in general evaluated as more legible. Ultralight font was judged as less legible in all research conditions, while normal and bold fonts received the highest preference scores with respect to ease and speed of reading.

Significant two and three-way interactions of font weight, contrast and polarity were observed. In particular, the lower the font weight, the more significant the decrease in its legibility scores when moving from high to low contrast: scores drop more than 3 times lower for the ultralight font, nearly 2 times lower for the light font and 1.5 times lower for normal and bold fonts. In addition, normal and bold fonts were judged equally legible and significantly more legible than thinner fonts under low contrast conditions. But with high contrast bold font lost a significant degree of legibility against normal font.

The degree of differences between legibility scores for the low and high contrast texts becomes greater when moving from positive to negative text-to-background polarity. On a white background low-contrast text maintains relative legibility and does not differ too much from high-contrast text, while on a black background the decrease in legibility scores of low-contrast texts is more significant. In particular, ultralight and light font under the condition of low contrast and negative polarity have much lower legibility rating than their high-contrast versions. Age did not have significant effect on the subjective ratings of legibility.

3.3 Pleasingness

ANCOVA and Wilcoxon test results for pleasingness scores are shown in Table 3 and Table 4.

Table 3: Pleasingness: ANCOVA results

Factors	*df*	*F*	*Sig.*
Weight	1.68, 102.54	142.999	0.000
Weight×Age	1.68, 102.54	6.967	0.003
Polarity	1.61	43.280	0.000
Contrast	1.61	121.843	0.000
Weight×Contrast	1.97, 120.25	8.300	0.000
Weight×Contrast×Age	1.97, 120.25	5.843	0.004
Polarity×Contrast	1.61	4.860	0.031

For pleasingness scores we obtained significant effects of main factors (font weight, polarity and contrast), as well as their interactions (font weight × contrast), that were similar to those described above for legibility. In addition, the results show a more significant decrease in pleasingness scores for bold font (compared to normal), especially in high-contrast conditions. Also, lighter fonts more often had higher pleasingness scores than legibility scores, while the most thick and high-contrast texts were more often judged as significantly more legible than pleasant.

Table 4: Pleasingness: Wilcoxon signed-rank test for white vs black background and high vs low contrast

	Ultralight	Light	Normal	Bold
Ultralight		**-6.605*** ** -6.870***	**-6.325*** ** -6.792***	**-5.155*** ** -6.070***
Light	**-6.632*** ** -6.591***		**-3.343*** ** -4.892***	**n/s** -2.021*
Normal	**-6.288*** ** -6.860***	**-2.320*** -5.259***		**-4.120*** ** n/s
Bold	**-5.464*** ** -6.538***	**n/s** -2.739**	**-3.510*** ** n/s	

* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; n/s – non-significant

The significant interaction of polarity and contrast corresponds with the legibility results: the pleasingness of low-contrast texts decreases more significantly when moving from positive to negative polarity conditions.

In addition, a correlation analysis (Spearman test) was conducted into legibility and pleasingness scores for each research condition. The results showed that in almost all conditions legibility and pleasingness scores have a high level of correlation ($p \leq 0.001$), which corresponds to the results obtained in the classic work [8]. Significant interactions of font weight and contrast with age covariate were revealed. The younger participants rated lighter fonts as more pleasant and thicker fonts as less pleasant compared to older ones. These differences between age groups were more evident for normal and bold fonts under high-contrast conditions.

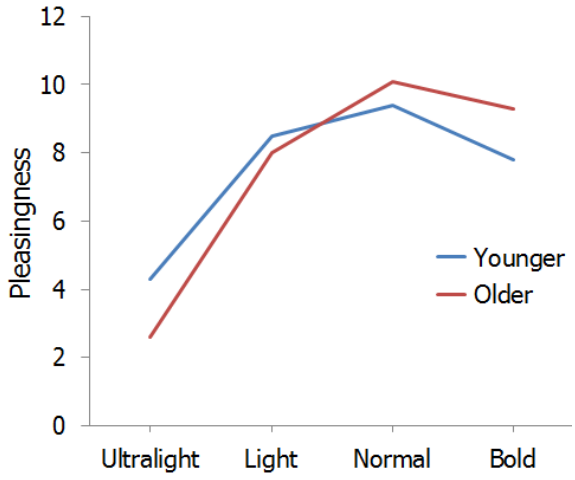

Figure 3: Mean pleasingness scores for younger and older users

To look deeply into age differences we compared two contrasting age groups – younger (19–30 years old, N=19) and older (45–68 years old, N=23). For legibility, no significant differences between the two groups were revealed, but pleasingness rates (Fig. 3) had an age-related effect in the case of ultralight font: younger participants rated ultralight font significantly higher (4.3) than the older group (2.6).

Older participants rated pleasingness of ultralight (2.6) and light (8.0) fonts significantly lower than normal (10.1) and bold (9.3), and perceived the latter two fonts as equally pleasing. Younger group rated ultralight font (4.3) as less pleasing than other font variations, and light font (8.5) as having generally the same pleasing scores as normal (9.4) and bold (7.8).

4 CONCLUSIONS

Our research has shown that ultralight and light fonts are in general less preferable than thicker fonts from a subjective point of view on both legibility and pleasingness. We found that such factors as text polarity and contrast also influence the subjective perception of text both independently and also in interaction with font weight factor. We also demonstrated that while there were no significant differences in the subjective perception of the fonts *legibility*, the users' age may influence the perception of *pleasingness* of ultralight fonts. A possible explanation may be the fashion-proneness of younger users, who are keen to follow the design trend imposed by software vendors.

The results of our study provide a basis for suggesting guidelines on the use of lighter fonts in user interface design: (a) negative polarity and low contrast conditions should be avoided for lighter fonts as there is a prominent negative effect on text legibility and pleasingness scores; (b) ultralight fonts should be avoided under any of the investigated conditions.

ACKNOWLEDGMENTS

We thank Andrei Filimonov for his invaluable help in data analysis. The research was supported by a grant from the Russian Foundation for Basic Research #17-06-00652.

REFERENCES

[1] Ivan Burmistrov, Tatiana Zlokazova, Iuliia Ishmuratova, and Maria Semenova. 2016. Legibility of Light and Ultra-Light Fonts: Eyetracking Study. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. ACM, New York, NY, Article 110. DOI: https://doi.org/10.1145/2971485.2996745

[2] Ian C. Dyer. 2014. An Examination of Typographic Standards and Their Relevance to Contemporary User-Centred Web and Application Design. In *Proceedings of 6th International Conference on Cross-Cultural Design (CCD 2014)*. Springer, Cham, 140-149. DOI: https://dx.doi.org/10.1007/978-3-319-07308-8_14

[3] Jerzy Grobelny and Rafał Michalski. 2015. The Role of Background Color, Interletter Spacing, and Font Size on Preferences in the Digital Presentation of a Product. *Computers in Human Behavior* 43, 85-100. DOI: https://dx.doi.org/10.1016/j.chb.2014.10.036

[4] Philip Kortum. 2017. Has Human Factors and Usability Lost Its Mojo?. *Journal of Usability Studies* 12, 2, 50-55. http://uxpajournal.org/wp-content/uploads/pdf/JUS_Kortum_Feb2017.pdf

[5] Jari Laarni. 2002. Searching for Optimal Methods of Presenting Dynamic Text on Different Types of Screens. In *Proceedings of NordiCHI '02*. ACM, New York, NY, 219-222. DOI: https://dx.doi.org/10.1145/572020.572050

[6] Kate Meyer. 2016. *Young Adults Appreciate Flat Design More Than Their Parents Do*, https://www.nngroup.com/articles/young-adults-flat-design/

[7] Yue Pan and Erik Stolterman. 2015. What if HCI Became a Fashion-Driven Discipline?. *interactions* 22, 6, 50-53. DOI: http://dx.doi.org/10.1145/2828430

[8] Miles A. Tinker and Donald G. Paterson. 1942. Reader Preferences and Typography. *Journal of Applied Psychology* 26, 1, 38-40. DOI: https://dx.doi.org/10.1037/h0061105

[9] Dirk Wendt. 1994. Legibility. In *Font Technology: Methods and Tools*, Peter Karow (Ed.). Springer, Berlin, 271-306. DOI: https://doi.org/10.1007/978-3-