Applied Ergonomics 97 (2021) 103499

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

High letter stroke contrast impairs letter recognition of bold fonts☆

Sofie Beier *, Chiron A.T. Oderkerk

Royal Danish Academy – Architecture, Design, Conservation, Center for Visibility Design, Denmark

ARTICLE INFO

Keywords:
Crowding
Font
Reading
Letter features
Visual acuity

ABSTRACT

To make graphical user interfaces look more fashionable, designers often make use of high-stroke-contrast fonts. We are yet to understand how these fonts affect reading. We examined the effect of letter-stroke contrast on three bold fonts, one with extreme contrast between thick and thin strokes, one with no contrast, and one in between. The fonts were designed for this experiment to enable control of font variables. Participants identified the middle letter in a lowercase letter trigram in each trial, briefly presented in the parafovea (at 2° left and right of fixation) and at the foveal fixation point. There was evidence for letter recognition impairment for the font with high stroke contrast compared to the fonts with low and medium stroke contrast, while there was no significant difference in performance between the medium- and low-stroke-contrast fonts. The results suggest that bold fonts with high stroke contrast should not be considered for designs where letter recognition is a priority.

1. Introduction

As digital platforms such as highway displays, in-vehicle interfaces and various mobile devices depend on fast reading, it is of increasing importance to investigate a font's glance-based legibility (Dobres et al., 2017; Sawyer et al., 2017).

Research shows that font styles come with certain semantic associations. This has been found in relation to both products (Doyle and Bottomley, 2004) and emotions (Brumberger, 2003; Juni and Gross, 2008) and suggests that an organisation striving to communicate its values in a visual language can do so through its choice of font. A font style that is widely used for the purpose of expressing specific values is the one known as 'Didone' or 'Modern' (Fig. 1). It is a style with high contrast between thinner and thicker letter parts that originated during the period of neoclassicism (for a thorough review of the development of the Didone style, see Beier, 2017). Today, Didone style fonts are often used in association with high-end fashion and luxury brands.

However, the visual expression of graphical interface designs has to maintain a fine balance between the need for visualising the artistic feel of an organisation's visual identity and support for the perceptual and cognitive functions of reading and page navigation.

Multiple experiments within design and vision research have demonstrated that font style can affect both letter and word identification. Examples include serifs at vertical extremes improving distance letter recognition (Beier and Dyson, 2014), small-size sans serif resulting in faster reading speed (Morris et al., 2002), simple letter shapes causing faster recognition of trigrams (Beier et al., 2018), and greater letter differentiation improving letter recognition (Beier and Larson, 2010; Bernard et al., 2016). One typographical feature that is yet to be investigated is the impact on the perception of bold fonts of high stroke contrast on letter recognition.

1.1. The two-stage model of visual processing

Reading is a complex operation in which words and sentences are recognised based on parallel operations of lower-level identification of letters and of higher-level syntactic and semantic processes (Coltheart et al., 2001; Pelli and Tillman, 2007). The main focus of the present paper is on crowded letter identification.

There is a general consensus within vision science that supports a two-stage model of visual letter processing. The first stage concerns independent feature detection, while the second concerns the integration of features into an overall perception of the letter (Levi, 2008; Pelli et al., 2003, 2006). Letter recognition is a central part of reading (Coltheart et al., 2001; Pelli et al., 2003), and has led to multiple experimental investigations drawing on letter identification paradigms to better inform our understanding of how we read (Bernard et al., 2016; Coates et al., 2019; Morin Duchesne et al., 2014; Ngiam et al., 2018; Pelli et al., 2006). In the search to identify the letter parts that are essential for letter and word identification, several studies have looked into removing parts

☆ This work was supported by the Danish Council for Independent Research [grant number DFF – 7013-00039].
* Corresponding author.
E-mail address: sbe@kglakademi.dk (S. Beier).

https://doi.org/10.1016/j.apergo.2021.103499
Received 13 October 2020; Received in revised form 11 May 2021; Accepted 7 June 2021

Fig. 1. In the Vox font classification system, fonts of the neoclassical era are categorised as Didone. One example is Linotype Didot Bold (left), which was inspired by the work of Firmin Didot (1764–1836). Another example is ITC Bodoni Six (right), which was inspired by the work of Giambattista Bodoni (1740–1813).

of the letter stimuli (Fiset et al., 2008; Lanthier et al., 2009; Petit and Grainger, 2002; Rosa et al., 2016). These studies tap into the first stage of visual processing of feature detection. Although there is little consensus about which of the removed parts impair recognition the most, studies that compared stimuli with and without removed letter parts found that eliminating letter features generally had a negative effect on identification (Lanthier et al., 2009; Rosa et al., 2016).

Letter crowding is the phenomenon popularly described as the presence of neighbouring letters severely impairing the recognition of a target letter (Bouma, 1970). It is believed to be one of the most fundamental bottlenecks for visual word recognition and reading rate (Levi, 2008; Pelli et al., 2007; Pelli and Tillman, 2008). While crowding can be found in the fovea at very small font sizes (Coates et al., 2018), it is often found in the periphery, as the magnitude of the crowing effect increases with eccentricity (Chung and Legge, 2009; Yu et al., 2014). A new font, designed so as to reduce crowding, was found to facilitate both peripheral crowded letter recognition in trigram whole-report paradigms and peripheral word recognition in lexical decision paradigms for the same reason; specifically, in direct support of theories that describe letter recognition as a necessary step for word recognition, researchers found that it was the same letters that were vulnerable to confusion during the crowded letter recognition task and the word recognition task, and that reducing letter confusability due to crowding increased legibility of both letters and words in both paradigms (Bernard et al., 2016).

According to the two-stage model of visual processing, crowding does not affect the first stage of feature detection. At the second stage, the feature integration is disproportionally large, which may result in a mistaken integration of neighbouring letter parts (Pelli et al., 2004). Both letters and letter features may mistakenly migrate when affected by crowding. This phenomenon of neighbouring features grouping and being mistakenly identified as a non-existing third letter is known as 'illusory conjunction' (Pelli et al., 2004; Treisman and Schmidt, 1982). The feature transposition occurs even when whole-letter transposition does not take place (Coates et al., 2019). Such findings suggest that a font style that is vulnerable to letter part fragmentation will have a greater tendency to letter part migration and poor letter recognition. One font style that could be expected to cause letter part fragmentation is that of high stroke contrast.

1.2. *Fovea and parafovea*

In sentence reading, one simultaneously draws on information from different retinal locations (Clifton et al., 2016; Shepherd et al., 1986). The greater the number of characters recognised in the periphery, the faster the reading speed (Legge et al., 2007). The visual periphery is further essential for those who suffer from central (foveal) vision loss, which is a well-known result of age-related macular degeneration (AMD), and the most common vision impairment of people above the age of 60 in the developed world (Wong et al., 2014).

Fonts that facilitate peripheral recognition could, therefore, lead to better reading for both normal vision and AMD readers and that to investigate letter identification in relation to reading, one would need to include multiple retinal locations in the experiment.

1.3. *Letter stroke boldness*

In addition to the Regular font weight, well-equipped font families contain styles of both Italic and Bold weights, with many newer variable fonts including several levels of boldness, ranging from Medium to Ultra Black. When the same amount of blackness is added throughout a letter stroke, it can result in the poor distribution of the black and white surface areas, which may impair reading performance (Bernard et al., 2013). To avoid this, the designer will focus on producing the optimal distribution of black and white within each letter, which typically results in letters such as 'a', 'e', and 's' having greater contrast between the thin and thick letter parts compared to both their Regular counterparts and the other letters within the font (Unger, 2018, p. 112). This is done to avoid having the white surface area inside of letters perceptually disappear.

Due to the darker surface area, bold fonts can work well for text emphasis (Bateman et al., 2008) and for reading with low luminance contrast between text and background (Burmistrov et al., 2016). Texts presented at acuity limit are more easily identified when set in bold fonts compared to regular-weight fonts (Beier and Oderkerk, 2019; Kuntz and Sleight, 1950; Sheedy et al., 2005), while both Light (s/w ratios of 1:20) and Ultra Black (s/w ratios of 1:2.5) weights impair recognition at all font sizes (Beier and Oderkerk, 2019). Experiments involving lexical decision tasks displaying high luminance contrast stimuli above the visual acuity limit find no benefit of greater letter-stroke boldness (Dobres et al., 2016; Dyson and Beier, 2016). This suggests that the effect of boldness is closely related to the reading situation.

Except for Sheedy et al. (2005), who looked into the collective effect of boldness in six different font families, the remaining experiments on boldness mentioned above have employed test fonts of low stroke contrast, with no examples of boldness involving high stroke contrast.

1.4. *Variations of boldness*

Serif fonts traditionally have a relatively high stroke contrast, while the tradition of sans serif fonts is to have a low stroke contrast (Fig. 2). Although Miles A. Tinker never directly investigated the effect of letter stroke contrast, the prolific legibility researcher raised a concern with both extremes. He argued that the 'tendency to use hairlines to form a part of certain letters should be strongly condemned' (Tinker, 1964, p. 36), as he expected that a thin crossbar in 'e' would produce more frequent misreadings for 'c'. For the low stroke contrast, he found that 'an unduly thick horizontal stroke infringes on the enclosed space and therefore reduces legibility' (Tinker, 1964, p. 36). If we follow Tinker's line of thought, fonts of low stroke contrast may cause the white areas within the letter to become too small for identification; this was later demonstrated to be true in relation to reading speed by Bernard et al. (2013), while for fonts of thin hairlines, the thin part of the letter risks disappearing. There are multiple examples of typography scholars expressing the historical dislike of fonts with hairline strokes, arguing that they lack good legibility due to the heavy contrast between thick

Fig. 2. As highlighted in the letter 'e', the bold weight of the font family Helvetica has a lower contrast between thin and thick strokes than the bold weight of the font family Times New Roman.

and thin strokes (Dwiggins, 1947; Gill, 2001; Goudy, 1963). However, the negative effect of thin hairline strokes is yet to be empirically verified, as most prior investigations involving letter stroke boldness have focused on low-stroke-contrast fonts. Following this, we hypothesised that bold fonts of very high stroke contrast may impair letter recognition. If the hypothesis is unsupported, it would suggest that high contrast does not impact letter recognition and that it does not cause letter parts to break up and migrate. We aimed to identify the effect of the different styles of adding weight to bold fonts and how this influences letter recognition.

2. Experiment

2.1. Material and methods

2.1.1. Participants

A total of 24 participants took part in this study, their ages ranging from 20 to 35 (M_{age} = 25.9 years, *SD* = 4.61 years, 16 women). Two participants were excluded because they were unable to complete the experiment. All self-reported normal or corrected-to-normal vision. The experiment was advertised through the participant recruitment website Forsoegsperson.dk. Participants received a gift card of DKK 300 (about USD 50) for participation. We obtained written informed consent from all participants after explaining the experiment both verbally and in writing. The research followed the tenets of the Declaration of Helsinki and The Danish Code of Conduct for Research Integrity.

2.1.2. Apparatus

The experiment was created using the software OpenSeame 3.2 (Mathôt et al., 2012) and presented on a backlit 17-inch IBM/Sony CRT monitor (refresh rate = 85hz, resolution = 1024 × 768) in a darkened room. The stimuli were presented as dark text (#000000) on a light background (#DADADA). The distance between the participant and the monitor was maintained by the use of a chinrest with a forehead strap.

2.1.3. Test material

The three test fonts designed for this experiment all originate in the font family KarloTest, we chose this font family in part because we had permission to alter it and in part because the width and letter skeleton have conventional proportions. All fonts had identical vertical letter axes, meaning that for the letter ‘o’, the thinnest part was always at the top and bottom of the letter; they all had identical vertical and horizontal proportions, meaning that they took up the same amount of space on the line of text; and they have all been adjusted to have the same amount of perceptual boldness (the thicker parts of the low-stroke-contrast font are thinner than the thicker parts of the High-stroke-contrast font). The main variation between the three test fonts was the level of letter stroke contrast, which affects the distribution of black and white surface areas. One font had high stroke contrast (High); another had low stroke contrast (Low); and the third had medium stroke contrast (Medium), which was the result of interpolation between the two extremes (Fig. 3). By isolating the typographical variable of letter stroke contrast from all other typographical variables, we ensured that findings would relate to this specific visual feature that we were interested in and not any other visual difference that typically exists between different font styles.

The ratio in the three fonts was measured based on the thinnest part of the ‘o’ and the height of the ‘l’. The stroke/height (s/h) ratio of the High font was 1/30; of the Medium font, it was 1/10; and of the Low font, it was 1/5.

Fig. 3. The three text fonts designed for the experiment. The fonts have identical letter proportions; the only variation between the fonts is the nature of the letter stroke boldness.

2.1.4. Stimuli and task

The experimental paradigm was based on the methodology by Beier et al. (2018). Participants took part in a partial report trigram recognition task, where they were shown a string of three letters and were asked to report only the middle letter. The experimental session was split up into two near-identical sessions that differed on the retinal location of the stimulus. In the parafoveal session, stimuli were presented at 2° left or right side of the fixation circle while the participant was seated 200 cm from the monitor. In the foveal session, stimuli were presented at central fixation while the participant was seated 350 cm from the monitor. In order to compare foveal and parafoveal eccentricities without incurring floor or ceiling effects, task difficulty was controlled through the use of an adaptive staircase procedure, described below, which determines the stimulus x-height for the foveal and parafoveal sessions separately.

A central fixation cross (120px by 120px, 0.63° at 200 cm and 0.36° at 350 cm) was presented at the start of a trial for a variable duration of 1300 ms, with a jitter of ±300 ms. This was immediately followed by the trigram stimulus, shown for 200 ms and consisting of a string of three letters, all of which were presented in one of the three font conditions that had either High, Medium, or Low stroke contrast. The central letter of the trigram appeared at 2° left or right of the fixation circle, in the parafoveal session or at central fixation, in the foveal session. The letters that made up each trigram were chosen from 16 lowercase letters, broadly representing the different letter structures of the alphabet (a, d, e, f, g, h, k, m, n, o, p, r, s, t, u, and y), without replacement, such that every letter occurred in every position of the trigram equally often. A backwards mask, consisting of a rectangular noise patch of variable size that covered all three letters of the trigram followed the stimulus for 500 ms, after which participants could report central trigram letter, if they were able, using the keyboard, and then continue to the next trial by pressing the space key. As participants had been instructed to fixate on the fixation cross in the centre of the screen, they were told that they could report gaze shifts away from the fixation cross by reporting any of the numbers on the keyboard instead of the stimulus letter they might have seen. These trials would later be discarded. After a participant ended the trial, they were immediately presented with feedback for 500 ms. The feedback informed them whether they were correct (Correct) or wrong (Wrong) or showed a dash if they either reported an eye movement or failed to report a letter.

In every block, each of the three fonts was presented in random order 16 times, for a total of 48 trials per block. The foveal and parafoveal sessions each consisted of 6 test blocks of 288 test trials, one practice block of 48 trials, and the staircase procedure described below.

2.1.5. Adaptive staircase procedure

Task difficulty was controlled by determining the x-height of the stimulus letters using a staircase procedure adapted from the accelerated stochastic approximation (Kesten, 1958; Treutwein, 1995). The staircase procedure took place at the beginning of the foveal and parafoveal sessions and employed the same trigram recognition task and trial outline as the subsequent practice and test blocks, with the exception that stimulus letters in the staircase procedure were presented only in the font of Medium stroke contrast and that the x-height of the stimulus letters changed between trials, depending on the accuracy of the participant’s responses.

The task thus grew more difficult after every correct response, as the x-height decreased, and easier after every incorrect response as the x-height grew larger. In order to make finer adjustments to the x-height, the size with which the x-height changed between trials, also referred to as the step size, decreased after a shift in response category occurred (i. e., from correct to incorrect or vice versa). This meant that the x-height would continue to increase with constant step size, as long as the participant continued to fail to report any target stimuli correctly.

In order to allow participants to grow accustomed to the experimental paradigm, there was no change in x-height for the first eight trials of the staircase procedure, during which stimuli were presented at an x-height of 0.16° when presented in the fovea at a distance of 350 cm and 0.28° in the parafovea at 200 cm distance. After the first eight trials, the x-height was determined through,

$$x_{n+1} = x_n - \frac{c}{m_{shift}}(z_n - 0.50), \quad 1$$

where n is the current trial number (excluding the first eight trials), x_n is the x-height of the current trial, x_{n+1} is the x-height of the subsequent trial, m_{shift} is the number of shifts in response category that occurred after the first eight trials (from correct to incorrect or vice versa), z_n is equal to 1 if the response for the current trial is correct and 0 if the response in the current trial is incorrect, c is the initial step size of 0.11° at 350 cm distance or 0.19° at 200 cm distance. The final response accuracy was based on the methodology previously used by Beier et al. (2018), such that the staircase was terminated after 19 reversals, which yielded a response accuracy of approximately 50% (foveal x-height average: 0.10° (33.81 pixels); STD: 0.03° (9.56 pixels); range: 0.07°–0.21° (24–70 pixels); parafoveal x-height average: 0.23° (43.50 pixels); STD: 0.05° (10.5 pixels); range: 0.14°–0.37° (26–71 pixels)).

3. Results

Trials in which participants reported eye movements towards the stimulus were excluded from analysis (1.31% of parafoveal trials). Using a 3 (font condition: Low, Medium, High) x 2 (location condition: 0°and 2°) repeated-measures ANOVA on mean accuracy, we found a large main effect of font condition, $F(2, 42) = 42.76$, $p < .001$, $\eta p^2 = 0.073$. We found no significant effect of location, $F(1, 21) = 2.51$, $p = .128$, $\eta p^2 = 030$, nor did we find a significant interaction between font and location, $F(2, 42) = 2.32$, $p = .110$, $\eta p^2 = 0.07$.

Post hoc paired t-tests, corrected for multiple comparisons using the Bonferroni correction, showed that mean accuracy for the font of High stroke contrast was significantly lower than the Medium font, $t(21) = 8.42$, $p < .001$, $d = 1.795$, decreasing by 3.12%, and the Low font, $t(21) = 8.72$, $p < .001$, $d = 1.859$, decreasing by 15.23% (Fig. 4). The Medium font, however, was not significantly different from the Low font, $t(21) = 1.66$, $p = .334$, $d = 0.354$.

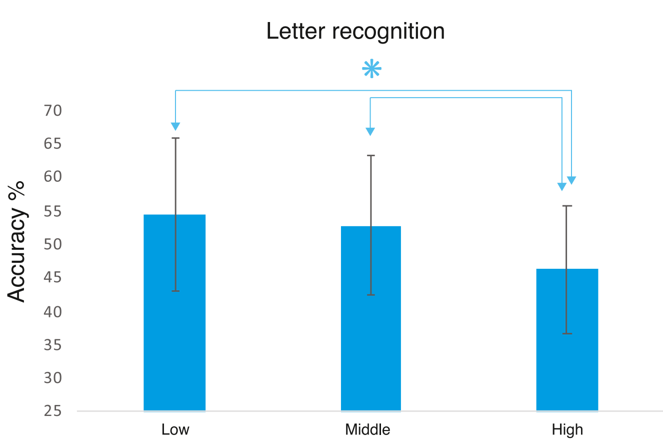

Fig. 4. Mean recognition for the tested fonts showing the fovea and parafovea presentations collectively. Error bars represent standard deviation. The stars highlight significant differences.

4. Discussion

By measuring the effect of font contrast on letter identification of the middle letter within a three-letter string, we found that glance-based crowded letter recognition is negatively affected by the font of High stroke contrast, which resulted in inferior performances compared to fonts of Medium and Low stroke contrast. There was no further significant difference between the performances of the Low and Medium contrast fonts. The findings support the hypothesis in demonstrating that high stroke contrast impairs letter recognition. In order to compare the effects of stroke contrast in the fovea and the parafovea, we controlled task difficulty across retinal locations by employing two adaptive staircase procedures that determined stimulus x-heights separately. For this reason, there was no effect of retinal locations, as the resulting size of parafoveal stimuli was just over two times larger, on average, than the foveal stimuli. Likewise, we did not find that the effects of stroke contrast differed between the fovea and the parafovea.

We speculate that the thin hairline strokes caused the poor performance of the font of high stroke contrast and that it resulted in negative effects on recognition at both stages of the two-stage model of visual processing. At the first stage, the thin strokes of the High-stroke-contrast font were less visible to our participants, which made it difficult to identify all letters features. As only detected features are available at the second stage of visual processing (Pelli et al., 2006), the undetected letter features of the High-stroke-contrast font would cause difficulties for the remaining detected features to connect into letters. Due to the effect of crowding, unconnected features would have a greater risk of mistakenly integrating with neighbouring letter features.

Some of the specific reading situations that would directly benefit from great letter recognition are password setting, reading of street, and road signs with unfamiliar place names and letter and number coding in complex wayfinding and navigation systems. In addition to this, all general reading situations could indirectly benefit. Reading is known to be based on the three processes of letter, word, and sentence identification. In a measure of reading speed where the three processes were isolated from each other, it was shown that the process of letter identification was the strongest factor, accounting for 62% of the reading rate, with holistic word recognition and contextual sentence processing accounting for respectively 16% and 22% of the total reading rate (Pelli and Tillman, 2007). Improving low-level letter identification, and thus the reading speed for letters, would improve reading speed in general. That being said, great letter identification is not inherently identical to great word identification. The process of drawing on the phonological and lexical levels, themselves activated at the word level (Rumelhart and McClelland, 1982), benefits from letters within words being predictable and formed of familiar word units (Sanocki and Dyson, 2012). As the stimuli letters of our experiment were both predictable and familiar in shape, we expect no negative effects on the word units, which suggests that the findings are also relevant for the word and sentence reading.

The most popular approach in the measurement of font style effects on reading is to compare performances using fonts of different font families. As multiple variables vary between different font families, we argue against this methodology, because it fails to isolate important variables within the fonts. Any difference between such fonts can only be attributed to those specific fonts and cannot be transferred to other fonts in general. The experimental paradigm of our experiment aimed to isolate one typographic variable with minimal interferences from other variables. Although the only claim we can make with certainty is that the measured effect is valid in relation to the tested font family, it follows from the chosen approach that the findings may possibly be generalized to other fonts as well.

5. Conclusion

The data showed that while there was no significant difference between bold fonts of low letter stroke contrast and medium letter stroke contrast, in comparison, a bold font of high letter-stroke contrast impaired crowded letter recognition. These results follow previous efforts in showing that font characteristics affect reading and suggest that in reading situations where time is a limiting factor and in a reading environment of smaller font sizes, choosing high-stroke-contrast fonts can have a negative impact on the functionality of the interface design. While it may be valid to create eye-catching interfaces using sophisticated font styles with thin hairlines to enhance certain semantic associations, this approach needs to be balanced against the function of operating the device, which ideally should work without negative interferences from stylistic font choices.

By demonstrating performance differences between different styles of font boldness, the result informs developers of the value of avoiding the use of high-stroke-contrast fonts when letter recognition is of importance.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the Danish Council for Independent Research [grant number DFF – 7013-00039].

References

Bateman, S., Gutwin, C., Nacenta, M., 2008. Seeing things in the clouds: the effect of visual features on tag cloud selections. In: Proceedings of the Nineteenth ACM Conference on Hypertext and Hypermedia, pp. 193–202. https://doi.org/10.1145/1379092.1379130.

Beier, S., 2017. Type Tricks: Your Personal Guide to Type Design. BIS Publishers.

Beier, S., Bernard, J.-B., Castet, E., 2018. Numeral legibility and visual complexity. In: *DRS Design Research Society*. https://doi.org/10.21606/drs.2018.246, 2018: Limerick.

Beier, S., Dyson, M.C., 2014. The influence of serifs on 'h' and 'i': useful knowledge from design-led scientific research. Visible Lang. 47 (3), 74–95.

Beier, S., Larson, K., 2010. Design improvements for frequently misrecognized letters. Inf. Des. J. 18 (2), 118–137. https://doi.org/10.1075/idj.18.2.03bei.

Beier, S., Oderkerk, C.A.T., 2019. Smaller visual angles show greater benefit of letter boldness than larger visual angles. Acta Psychol. 199, 102904. https://doi.org/10.1016/j.actpsy.2019.102904.

Bernard, J.-B., Aguilar, C., Castet, E., 2016. A new font, specifically designed for peripheral vision, improves peripheral letter and word recognition, but not eye-mediated reading performance. PloS One 11 (4), e0152506. https://doi.org/10.1371/journal.pone.0152506.

Bernard, J.-B., Kumar, G., Junge, J., Chung, S.T., 2013. The effect of letter-stroke boldness on reading speed in central and peripheral vision. Vis. Res. 84, 33–42. https://doi.org/10.1016/j.visres.2013.03.005.

Bouma, H., 1970. Interaction effects in parafoveal letter recognition. Nature 226, 177–178. https://doi.org/10.1038/226177a0.

Brumberger, E.R., 2003. The rhetoric of typography: the persona of typeface and text. Tech. Commun. 50 (2), 206–223. https://doi.org/10.1109/47.7818.

Burmistrov, I., Zlokazova, T., Ishmuratova, I., Semenova, M., 2016. Legibility of light and ultra-light fonts: eye tracking study. In: Proceedings of the 9th Nordic Conference on Human-Computer Interaction, pp. 1–6. https://doi.org/10.1145/2971485.2996745.

Chung, S.T., Legge, G.E., 2009. Precision of position signals for letters. Vis. Res. 49 (15), 1948–1960. https://doi.org/10.1016/j.visres.2009.05.004.

Clifton, C., Ferreira, F., Henderson, J.M., Inhoff, A.W., Liversedge, S.P., Reichle, E.D., Schotter, E.R., 2016. Eye movements in reading and information processing: keith Rayner's 40 year legacy. J. Mem. Lang. 86, 1–19. https://doi.org/10.1016/j.jml.2015.07.004.

Coates, D.R., Bernard, J.-B., Chung, S.T.L., 2019. Feature contingencies when reading letter strings. Vis. Res. 156, 84–95. https://doi.org/10.1016/j.visres.2019.01.005.

Coates, D.R., Levi, D.M., Touch, P., Sabesan, R., 2018. Foveal crowding resolved. Sci. Rep. 8 (1) https://doi.org/10.1038/s41598-018-27480-4.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., Ziegler, J., 2001. DRC: a dual route cascaded model of visual word recognition and reading aloud. Psychol. Rev. 108 (1), 204–256. https://doi.org/10.1037/0033-295x.108.1.204.

Dobres, J., Chrysler, S.T., Wolfe, B., Chahine, N., Reimer, B., 2017. Empirical assessment of the legibility of the highway Gothic and Clearview signage fonts. Transport. Res. Rec.: J. Transport. Res. Board 2624, 1–8.

Dobres, J., Reimer, B., Chahine, N., 2016. In: The Effect of Font Weight and Rendering System on Glance-Based Text Legibility, vols. 91–96. https://doi.org/10.3141/2624-01.

Doyle, J.R., Bottomley, P.A., 2004. Font appropriateness and brand choice. J. Bus. Res. 57 (8), 873–880. https://doi.org/10.1016/s0148-2963(02)00487-3.

Dwiggins, W.A., 1947. Mss. By WAD: Being a Collection of the Writings of Dwiggins on Various Subjects Some Critical, Some Philosophical, some Whimsical. The Typophiles, p. 49.

Dyson, M.C., Beier, S., 2016. Investigating typographic differentiation: italics are more subtle than bold for emphasis. Inf. Des. J. 22 (1), 3–18. https://doi.org/10.1075/idj.22.1.02dys.

Fiset, D., Blais, C., Ethier-Majcher, C., Arguin, M., Bub, D., Gosselin, F., 2008. Features for identification of uppercase and lowercase letters. Psychol. Sci. 19 (11), 1161–1168. https://doi.org/10.1111/j.1467-9280.2008.02218.x.

Gill, E., 2001. An Essay on Typography, 1931. Lund Humphries.

Goudy, F.W., 1963. The Alphabet and Elements of Lettering. Dover Publications.

Juni, S., Gross, J.S., 2008. Emotional and persuasive perception of fonts. Percept. Mot. Skills 106 (1), 35–42. https://doi.org/10.2466/pms.106.1.35-42.

Kesten, H., 1958. Accelerated stochastic approximation. Ann. Math. Stat. 29 (1), 41–59. https://doi.org/10.1214/aoms/1177706705.

Kuntz, J.E., Sleight, R.B., 1950. Legibility of numerals: the optimal ratio of height to width of stroke. Am. J. Psychol. 63 (4), 567–575. https://doi.org/10.2307/1418871.

Lanthier, S.N., Risko, E.F., Stolz, J.A., Besner, D., 2009. Not all visual features are created equal: early processing in letter and word recognition. Psychon. Bull. Rev. 16 (1), 67–73. https://doi.org/10.3758/pbr.16.1.67.

Legge, G.E., Cheung, S.-H., Yu, D., Chung, S.T., Lee, H.-W., Owens, D.P., 2007. The case for the visual span as a sensory bottleneck in reading. J. Vis. 7 (2) https://doi.org/10.1167/7.2.9, 9–9.

Levi, D.M., 2008. Crowding—an essential bottleneck for object recognition: a mini-review. Vis. Res. 48 (5), 635–654. https://doi.org/10.1016/j.visres.2007.12.009.

Mathôt, S., Schreij, D., Theeuwes, J., 2012. OpenSesame: an open-source, graphical experiment builder for the social sciences. Behav. Res. Methods 44 (2), 314–324. https://doi.org/10.3758/s13428-011-0168-7.

Morin Duchesne, X., Gosselin, F., Fiset, D., Dupuis-Roy, N., 2014. Paper features: a neglected source of information. J. Vis. 14 (13), 1–7. https://doi.org/10.1167/14.13.11.

Morris, R.A., Aquilante, K., Yager, D., Bigelow, C., 2002. P-13: serifs slow RSVP reading at very small sizes, but don't matter at larger sizes. In: SID Symposium Digest of Technical Papers, vol. 33. Blackwell Publishing Ltd, Oxford, UK, pp. 244–247. https://doi.org/10.1889/1.1830242. No. 1.

Ngiam, W.X., Khaw, K.L., Holcombe, A.O., Goodbourn, P.T., 2018. Visual working memory for letters varies with familiarity but not complexity. Journal of Experimental Psychology. Learning, Memory, and Cognition. https://doi.org/10.31234/osf.io/cxkw5.

Pelli, D.G., Burns, C.W., Farell, B., Moore-Page, D.C., 2006. Feature detection and letter identification. Vis. Res. 46 (28), 4646–4674. https://doi.org/10.1163/156856897x00366.

Pelli, D.G., Farell, B., Moore, D.C., 2003. The remarkable inefficiency of word recognition. Nature 423 (6941), 752–756. https://doi.org/10.1038/nature01516.

Pelli, D.G., Palomares, M., Majaj, N.J., 2004. Crowding is unlike ordinary masking: distinguishing feature integration from detection. J. Vis. 4 (12), 1136–1169. https://doi.org/10.1167/4.12.12.

Pelli, D.G., Tillman, K.A., 2007. Parts, wholes, and context in reading: a triple dissociation. PloS One 2 (8), e680. https://doi.org/10.1371/journal.pone.0000680.

Pelli, D.G., Tillman, K.A., 2008. The uncrowded window of object recognition. Nat. Neurosci. 11 (10), 1129.

Pelli, D.G., Tillman, K.A., Freeman, J., Su, M., Berger, T.D., Majaj, N.J., 2007. Crowding and eccentricity determine reading rate. J. Vis. 7 (2), 1–36.

Petit, J.-P., Grainger, J., 2002. Masked partial priming of letter perception. Vis. Cognit. 9 (3), 337–353. https://doi.org/10.1080/13506280042000207.

Rosa, E., Perea, M., Enneson, P., 2016. The role of letter features in visual-word recognition: evidence from a delayed segment technique. Acta Psychol. 169, 133–142. https://doi.org/10.1016/j.actpsy.2016.05.016.

Rumelhart, D.E., McClelland, J.L., 1982. An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. Psychol. Rev. 89 (1), 60. https://doi.org/10.1037/0033-295x.89.1.60.

Sanocki, T., Dyson, M.C., 2012. Letter processing and font information during reading: beyond distinctiveness, where vision meets design. Atten. Percept. Psychophys. 74 (1), 132–145. https://doi.org/10.3758/s13414-011-0220-9.

Sawyer, B.D., Dobres, J., Chahine, N., Reimer, B., 2017. The cost of cool: typographic style legibility in reading at a glance. Proc. Hum. Factors Ergon. Soc. Ann. Meeting 61, 833–837. https://doi.org/10.1177/1541931213601698.

Sheedy, J.E., Subbaram, M.V., Zimmerman, A.B., Hayes, J.R., 2005. Text legibility and the letter superiority effect. Hum. Factors: J. Human Factors Ergon. Soc. 47 (4), 797–815. https://doi.org/10.1518/001872005775570998.

Shepherd, M., Findlay, J.M., Hockey, R.J., 1986. The relationship between eye movements and spatial attention. Quart. J. Exper. Psychol. Sect. A 38 (3), 475–491.

Tinker, M.A., 1964. Legibility of Print. Iowa State University Press.

Treisman, A., Schmidt, H., 1982. Illusory conjunctions in the perception of objects. Cognit. Psychol. 14 (1), 107–141. https://doi.org/10.1093/acprof:osobl/9780199734337.003.0019.

Treutwein, B., 1995. Adaptive psychophysical procedures. Vis. Res. 35 (17), 2503–2522. https://doi.org/10.1016/0042-6989(95)00016-x.

Unger, G., 2018. Theory of Type Design. nai010.
Wong, W.L., Su, X., Li, X., Cheung, C.M.G., Klein, R., Cheng, C.-Y., Wong, T.Y., 2014. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. Lancet Glob. Health 2 (2), e106–e116. https://doi.org/10.1016/s2214-109x(13)70145-1.
Yu, D., Legge, G.E., Wagoner, G., Chung, S.T., 2014. Sensory factors limiting horizontal and vertical visual span for letter recognition. J. Vis. 14 (6) https://doi.org/10.1167/14.6.3, 3–3.