Acta Psychologica 199 (2019) 102904

Contents lists available at ScienceDirect

Acta Psychologica

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

Smaller visual angles show greater benefit of letter boldness than larger visual angles

Sofie Beier*, Chiron A.T. Oderkerk

The Royal Danish Academy of Fine Arts, School of Design, Denmark

ARTICLE INFO

Keywords:
Stroke boldness
Visual angle
Font
Letter recognition
Legibility

ABSTRACT

Research has shown that fonts viewed at a smaller visual angle benefit from greater letter boldness. Since small and large visual angles operate on different spatial frequencies, we examined whether the effect was dependent on font size. By applying a paradigm of single-letter exposure across two experiments, we showed that fonts of thinner letter strokes and of extreme boldness decreased recognition for all tested font sizes, and that there was a positive effect of boldness at small visual angles which did not occur at large visual angles. The paper provides evidence that bolder fonts are less effective at improving recognition at larger visual angles, and that over a scale of font weights there is a drop-off at the lightest and the heaviest extremes at all tested font sizes.

1. Introduction

The act of reading words involves several levels of parallel processing, including the identification of letter components, whole letters, and words. These processes are believed to involve a range of reciprocal feedback loops (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; McClelland & Rumelhart, 1981; Perry, Ziegler, & Zorzi, 2014). Letter recognition is hence central to word recognition and most real life reading situations.

To determine the most essential features for letter recognition, some researchers have employed methodologies involving the degrading or removing parts of the stimuli. Fiset et al. (2008) used the 'bubbles technique' to show that the letter stroke terminations of the Arial font are the most important for letter recognition. In a lexical decision experiment, Rosa, Perea, and Enneson (2016) found that removing midsegments of letter strokes was more detrimental to reading than removing the junctions or the letter stroke terminations in the Minion font; similar findings were shown for the Courier font in a single-letter recognition task (Petit & Grainger, 2002). In contrast, Lanthier, Risko, Stolz, and Besner (2009) demonstrated that removing the junctions of the Arial font was more damaging than removing midsegments.

In spite of the diverse range of results in determining the most essential letter features for identification, there appears to be a consensus within cognitive psychology that feature detection is a vital aspect of letter recognition (Finkbeiner & Coltheart, 2009; Pelli, Burns, Farell, & Moore-Page, 2006; Sanocki & Dyson, 2012). The methodologies of the above-mentioned experiments were employed to study components within one font of regular weight. As extensive research has demonstrated that font style has an effect on letter recognition (Beier & Larson, 2010; Beier, Starrfelt, & Sand, 2017; Pelli et al., 2006; Pušnik, Podlesek, & Možina, 2016), it is likely that the specific choice of font may have significantly influenced the results.

In this paper, we intend to show that letter boldness alone can influence letter recognition within the same font family. Furthermore, due to the influence of spatial frequencies, we predicted that the effect of letter boldness differs between different font sizes.

1.1. Spatial frequency

The effect that a test font has on a participant's reading greatly depends on test methodology (Beier & Larson, 2010). One of many reasons suggested for this is that fonts presented at larger or smaller visual angles draw on different spatial frequency channels. In the perceptual system, the spatial-frequency tuning of the visual neurons varies in relation to both the size of the stimulus and the luminance contrast (Alexander, Xie, & Derlacki, 1994; Chung, Legge, & Tjan, 2002); this mechanism is found to be largely similar in both foveal and peripheral vision (Chung et al., 2002; Chung & Tjan, 2009). Majaj, Pelli, Kurshan, and Palomares (2002) demonstrated that observers employed only one channel at a time, the choice of which was dependent on the

* Corresponding author.
E-mail address: sbe@kadk.dk (S. Beier).

https://doi.org/10.1016/j.actpsy.2019.102904
Received 14 December 2018; Received in revised form 24 July 2019; Accepted 4 August 2019

stimulus and could not be selected by the observer. In other words, small visual angles[1] require the use of lower spatial frequencies, which results in letters being perceived as blurred images. The finer details and edges that are perceived at higher spatial frequencies, are therefore not available at small visual angles (Fig. 1). To facilitate greater letter recognition at small visual angles, the focus should be on those visual features that are visible when the observer makes use of the relevant channels. One such feature is the distribution of letter boldness.

1.2. Letter boldness

The different shapes of the individual letter parts of the alphabet can be viewed as different distributions of black and white surface areas — this relationship between black and white changes with letter boldness (Fig. 2). Extremely bold fonts cover a bigger black surface area, which 'eats up' the surrounding white area, while the black surface area of extremely light fonts is much smaller, which results in a bigger white surface area inside and around the letter (Noordzij, 2005). This difference in the distribution of black and white surface area changes the shapes of letter features between fonts of different boldness.

Our hypothesis is that the influence of boldness on letter identification varies depending on the visual angle. In this study, we aim to identify the ideal letter-stroke boldness at different visual angles.

Bold fonts play a central role in headlines and for text emphasis, as the darker surface area makes them stand out from the page relative to regular weights (Bateman, Gutwin, & Nacenta, 2008; Dyson & Beier, 2016). Furthermore, bolder fonts can also facilitate font legibility in certain reading scenarios. Measuring visual acuity, Sheedy, Subbaram, Zimmerman, and Hayes (2005) showed that Franklin Gothic Book was less legible than Franklin Gothic Medium, Demi, and Heavy at small font sizes. This indicates that small font sizes require a minimum stroke weight to maintain their readability (Fig. 3). These findings matched earlier findings by Kuntz and Sleight (1950). In a study involving contrast threshold and visual acuity of numerals that had stroke width/height (SW/H) ratios ranging from 1:4.2 to 1:16.0, Kuntz and Sleight (1950) found that the font that delivered the best performance was the one that had a SW/H ratio of 1:6.0 (this ratio equals Franklin Gothic Medium in Fig. 3). In a follow-up experiment, the researchers tested a smaller range of ratios and here demonstrated similar performances for fonts in ratios of 1:4.0 to 1:6.0.

There are also indications that fonts of bold weights facilitate reading in low-luminance conditions. Using a visual search task that involved eye-tracking, Burmistrov, Zlokazova, Ishmuratova, and Semenova (2016) found a general disadvantage of lighter-weight fonts, showing longer search time, increased fixation duration, and lower saccadic amplitude when testing different boldnesses of Helvetica Neue. The study further demonstrated a positive effect of the bold font on fixation duration, albeit only when there was a low-luminance contrast between foreground and background. These results were in line with earlier findings by Luckiesh and Moss (1940), who measured contrast threshold and found that the light-weight version of the font family Memphis was inferior to all the heavier weights tested.

In reading scenarios that involved regular text sizes and high-luminance contrast, however, letter boldness has not been found to have any real effect in either lexical decision tasks (Dobres, Reimer, & Chahine, 2016; Dyson & Beier, 2016) or reading speed tasks (Bernard, Kumar, Junge, & Chung, 2013; Tinker, 1964).

A look at the studies above suggests that the effect of letter boldness on legibility is dependent on the reading situation and the level of the spatial frequency. Bold fonts have been demonstrated to provide an advantage at small visual angles and low-luminance contrasts, which suggests that low-frequency channels benefit from bold fonts. Spatial frequency tuning also results in the perceptual phenomenon that the same font weight appears lighter in small sizes than in large sizes (Fig. 4). As edges and details disappear in small sizes, the stroke becomes less defined and hence appears lighter. Since the early days of printing, typographers have been aware of this effect and have used optical scaling of their fonts to make letters of small sizes heavier, wider, and with lower stroke contrast than in larger font sizes (Ahrens & Mugikura, 2014).

Fig. 1. Spatial frequency. The two images are identical. Left: at larger visual angles the higher-frequency channels show details and edges, and thus, the viewer sees the sharp 'F' more than the blurred 'D'. Right: at small visual angles, the lower-frequency channels show letter weight and proportions, and thus, the viewer sees the blurred 'D' more clearly than the sharp 'F'. One can further experiment with viewing the large-size image from a larger reading distance. This will make the letter 'D' stand out instead of the 'F'.

However, the different processing of frequency channels might indicate that those features that enhance letter recognition at small visual angles could be diminished when replicated at large visual angles. In the present study, we tested the hypothesis that letter recognition benefits more from boldness at small visual angles than at large visual angles. Thus, the goal of the experiments was to examine the varying effect of letter-stroke at larger and smaller visual angles.

2. Experiment 1

2.1. Material and methods

2.1.1. Participants

The experiment was advertised through a participant recruitment website (Forsoegsperson.dk). A total of 21 participants aged from 19 to 52 years (M_{age} 26.9 years, $SD = 7.5$ years, 15 women) took part. Participants received a gift card of DKK 150 upon completion of the experiment. All reported normal or corrected-to-normal vision. Written informed consent was obtained from each participant after the experiment was explained. The research followed the tenets of the Declaration of Helsinki and The Danish Code of Conduct for Research Integrity.

2.1.2. Test material

Research has shown that reading rate is independent of whether the test fonts are widely used or new to the participant, as long as the fonts have common letter shapes (Beier & Larson, 2013). The five test fonts originate in the font family Ovink and were designed for this experiment. This made it possible to choose the weights that fit our

[1] Following this, a sign showing 120-point type viewed at 4 m' distance would have the same visual angle as a sign showing 12-point type viewed at a normal reading distance of 40 cm.

Fig. 2. The bolder the letter, the smaller the letter counter which is the area not covered by the letter stroke (marked in red). Demonstrated in the font family Avenir Next. (For interpretation of the references to colour in this figure legend, the reader is referred to the digital version of this article.)

Fig. 3. Sheedy et al. (2005) tested four weights of the font family Franklin Gothic and found Franklin Gothic Book (far left) to be inferior to all the other weights in a visual acuity experiment. The stroke width/height ratio (SW/H) from left: Franklin Gothic Book (1:9.0), Gothic Medium (1:6.0), Demi (1:4.7) and Heavy (1:3.5).

Fig. 4. Top row: all letters are set in the font weight Ovink Regular. Bottom row: The smallest letter is set in Ovink Bold, the following is set in Ovink Medium, while the three larger ones are set in Ovink Regular. In the top row, the weight appears increasingly lighter in smaller sizes, while the weight in the bottom row appears more even across the different sizes.

experimental aims and avoid dependency on font weights designed by others. To generate the interpolation of the intermediate fonts in the Glyphs software, the two extreme fonts were assigned weight values of 1 (Light) and 5 (Bold), with the three intermediate fonts assigned weight values of 2 (Regular), 3 (Medium), and 4 (Semi Bold). Thus, the boldness increased linearly across the five test fonts.

As is customary in professional font design, the heavy weights are perceptually adjusted in the junctions, so that when a round shape meets a stem, the letter stroke thins, and for letters like ‘a’ and ‘e’, which have many details in small spaces, the middle part is thinner than the rest of the letter. This results in a different kind of letter contrast in the bolder fonts compared with the lighter fonts; however, it also ensures that the findings of the experiment can be directly transferred into real-life usage (Fig. 5). This way of adding a perceptually comparable amount of weight to the stroke follows the tradition of sans serifs fonts. The tradition of serif fonts, which adds weight by increasing the stroke contrast in such a way that the main bulk of the weight is placed at vertical strokes (Noordzij, 2005), is left for future investigations.

We tested the following stroke width/height ratios: Light (1:20.0), Regular (1:10.0), Medium (1:6.4), Semi Bold (1:4.7), Bold (1:3.8). Stroke width was measured from the horizontal width of the lowercase stem, and height was measured from the baseline to the top of the ascending lowercase letters. Thus, the letter ‘h’ with a stroke width of 20 units and a height of 200 units has a ratio of 1:10.

2.1.3. *Apparatus*

The stimuli were displayed on a 12.3-inch LCD monitor in a dimly lit room (refresh rate = 60hz, resolution = 3000 × 2000). Experiments were created using the software OpenSesame 3.2. (Mathôt, Schreij, & Theeuwes, 2012). Stimuli were presented as black text (#000000) on a light grey background (#DADADA). However, stimuli were presented without anti-aliasing.[2] Therefore, in order to ensure that the resolution of the stimuli was comparable for all three sizes participants were seated far away from the monitor. The distance between the participant and the monitor was determined at the beginning of the experiment, such that the distance between the participant's eyes and the monitor was 200 cm when the participant was seated in their preferred position with their back to the chair.

2.1.4. *Procedure*

Fonts of different boldness vary in the amount of inter-letter spacing so that bolder fonts have smaller inter-letter spacing compared to lighter fonts. To eliminate an unwanted variable of inter-letter spacing, which would have shown in letter-string presentations, we presented the stimuli as single lowercase letters.

We applied a method of short exposure at 3.15° left or right side of the fixation circle. We tested three size conditions (measured from the top to the bottom of the descender): Small 0.08° (equals x-height[3] of 0.06 cm at a reading distance of 40 cm = 3.5 points), Medium 0.14° (equals x-height of 0.01 cm at a reading distance of 40 cm = 6 points), and Large 0.20°(equals x-height of 0.14 cm at a reading distance of 40 cm = 9 points), and five weight conditions: Light (1), Regular (2), Medium (3), Semi Bold (4), and Bold (5).

In each trial, the stimulus consisted of one of sixteen lowercase letters (a, d, e, f, g, h, k, m, n, o, p, r, s, t, u, y), presented individually for a short exposure time. Stimuli were presented on the left or right side of the fixation circle, 496 ms after the initiation of a trial.

[2] Anti-aliasing is a technique that smooths the edges of strokes so they do not appear jagged.

[3] The x-height is the height of the lowercase ‘x’.

Fig. 5. The sixteen letters tested in experiment 1, set in the five test fonts. From the top: Ovink Light (1), Ovink Regular (2), Ovink Medium (3), Ovink Semi Bold (4), and Ovink Bold (5). The numbers to the right are the stroke width/height ratio.

Participants were instructed to maintain fixation on the fixation circle while the stimulus was being presented. In order to remove any possible after-image effects, the stimuli were followed by a mask for 496 ms in the form of a Gaussian noise patch of variable size. Following a stimulus presentation, participants were asked to name the stimulus letter if they were able, after which it was recorded by the experimenter.

In each block, each of the five fonts was shown for each of the three sizes on both the left and the right side of the fixation circle for a total of 60 trials per block. Over the course of an experimental session, each participant engaged in one practice block and eight test blocks.

In order to ensure performance comparability between participants, the stimulus exposure duration was determined for each participant separately using a staircase procedure adapted from the accelerated stochastic approximation by Kesten (1958) (Treutwein, 1995). The trial outline for this staircase procedure was similar to the later testing session, although participants were only presented with Medium-sized stimulus letters of Medium (3) boldness over a series of trials. After an initial exposure duration of 160 ms on the first trial, the exposure duration during the following trials would increase or decrease by 64 ms, depending on the accuracy of the participant's report in the preceding trial. This meant that the exposure duration would continue to increase by steps of 64 ms if the participant continually failed to report the correct letter. When the exposure duration was long enough for the participant to make their first correct reports, the step size by which the exposure duration changed between trials would decrease after a predetermined number of reversals of accuracy (i.e., a correct report followed by an incorrect report, or vice versa). Step sizes decreased to 48 ms after seven reversals, to 32 ms after 13 reversals, and to 16 ms after 18 reversals. The staircase procedure was terminated after 25 reversals, at which time the final exposure duration to be used during that participant's following test session was the average of the exposure durations of the final six reversals.

2.2. Results

Using a 3 (size condition: Small, Medium, and Large) × 5 (weight condition: Light, Regular, Medium, Semi Bold, and Bold) repeated-measures ANOVA on mean accuracy, we found large main effects of size condition, $F(2, 40) = 410.31$, $p < .001$, $\omega^2 = 0.81$, and of weight condition, $F(4, 80) = 49.06$, $p < .001$, $\omega^2 = 0.21$. However, the size of the stimuli influenced the accuracy of the responses differently across the five weight conditions, resulting in a small significant interaction effect, $F(8, 160) = 2.55$, $p = .012$, $\omega^2 = 0.02$.

Using 30 planned comparison, corrected for multiple comparisons using the Bonferroni method, showed that the interaction resulted from a significant difference between the mean accuracy in the Regular (2) and Semi Bold (4) weight, $t(20) = 5.48$, $p = .001$, $dz = 1.20$, which only occurred in Medium-sized stimuli. Conversely, there was no effect between Regular (2) and Semi Bold (4) weight conditions at the Small, $t(20) = 2.49$, $p = .648$, $dz = 0.54$, or Large sizes, $t(20) = 1.80$, $p = .999$, $dz = 0.39$.

Accuracy in the Light (1) weight condition was impaired in all sizes. Specifically, in the Small size mean accuracy in the Light (1) weight was significantly lower than Medium (3), $t(20) = 4.98$, $p = .002$, $dz = 1.09$, Semi Bold (4), $t(20) = 5.12$, $p = .002$, $dz = 1.19$, and Bold (5), $t(20) = 5.40$, $p = .001$, $dz = 1.18$, though the difference between mean accuracy of Light (1) and Regular (2) did not reach significance, $t(20) = 3.12$, $p = .164$, $dz = 0.68$; in the Medium size mean accuracy for Light (1) was significantly lower than Regular (2), $t(20) = 4.93$, $p = .002$, $dz = 1.08$, Medium (3), $t(20) = 6.83$, $p < .001$, $dz = 1.49$, Semi Bold (4), $t(20) = 8.59$, $p < .001$, $dz = 1.86$, and Bold (5), $t(20) = 6.52$, $p < .001$, $dz = 1.42$; and in the Large size mean accuracy for Light (1) was near-significantly lower than Regular (2), $t(20) = 3.49$, $p = .070$, $dz = 0.76$, and significantly lower than Medium (3), $t(20) = 4.61$, $p = .005$, $dz = 1.01$, Semi Bold (4), $t(20) = 6.13$, $p < .001$, $dz = 1.34$, and Bold (5), $t(20) = 3.81$, $p = .033$, $dz = 0.83$. No other comparisons reached significance (all p's > 0.117) (Fig. 6).

2.3. Discussion experiment 1

Supporting our hypothesis that the effect of boldness was dependent on size, Experiment 1 showed a small significant interaction effect between size and weight, which indicated that boldness had a stronger effect on the Medium font sizes, which did not occur in the Small or the Large font size. Furthermore, we found that recognition improved in nearly all weight conditions, relative to the lightest weight. Measuring reading speed with rapid serial visual presentation of words, Bernard et al. (2013) found that performance dropped when reading the boldest font of their experiment. We, however, did not replicate this in Experiment 1, as our boldest font (Ovink Bold (5)) resulted in generally good performances. The boldest font of Bernard et al. (2013) was much bolder than any of our fonts. By adding an extreme bold font to our Experiment 2, we were interested in seeing if the experimental paradigm of single letter recognition, could similarly result in a performance drop with extreme letter boldness.

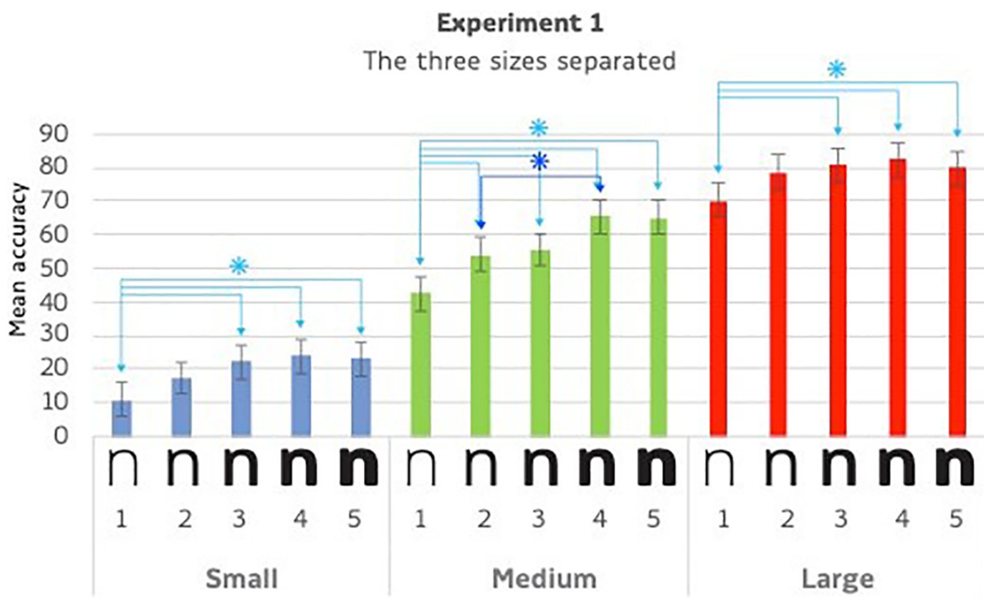

Fig. 6. Mean accuracy of the responses across size and weight conditions. The blue bars represent Small size conditions, the green bars represent Medium size conditions, and the red bars represent Large size conditions. Numbers on the x-axis represent the weight conditions Light (1), Regular (2), Medium (3), Semi Bold (4), and Bold (5). Comparisons marked with * were significantly different. (For interpretation of the references to colour in this figure legend, the reader is referred to the digital version of this article.)

3. Experiment 2

We were interested in replicating the interaction between Regular (2) and Semi Bold (4) at Medium and Large sizes, as well as seeing if, like the lightest condition, there was a drop in performance with very heavy weight.

Firstly, we predicted that the new font, Ovink Ultra Black (6), would impede performance in regardless of stimulus size. Secondly, we hoped to replicate the size and weight interaction from Experiment 1. Namely, we predicted that in the Medium sized letters the increased boldness of Ovink Semi Bold (4) would facilitate greater letter recognition relative to Ovink Regular (2), only in the Medium sized letters, while there would be no such significant difference between the same two weights for the Large letters.

At the time of testing, presenting stimulus letters with anti-aliasing in OpenSesame 3.2 was dependent on the x-height of letters on the monitor being no smaller than 0.73 cm; this was not the case for any of the sizes in Experiment 1. In order to ensure anti-aliasing, participants in Experiment 2 were seated yet further from the monitor, such that the smallest size included in Experiment 2, the Medium font size, had the same visual angle as in Experiment 1 of 0.14° and an x-height of 0.73 cm.

3.1. Material and methods

3.1.1. Participants

Participant recruitment was the same as in Experiment 1. 15 participants (M_{age} 27.33 years, $SD = 5.09$ years, 11 women) took part. Each received a DKK 150 gift card in remuneration upon completion of the experiment. All reported normal or corrected-to-normal vision.

3.1.2. Test material

The fonts Ovink Regular (2) and Ovink Semi Bold (4) are identical to the same fonts in Experiment 1 (Fig. 7). The weight of the new font Ovink Ultra Black (6) was chosen based on the premises of including a font that is as heavy as possible without the letter counters closing up.

3.1.3. Apparatus

Stimuli in Experiment 2 were displayed on a backlit 17-inch IBM/Sony CRT monitor (refresh rate = 85 hz, resolution = 1024 × 768) in a darkened room. Distance between the participant and the monitor was maintained through the use of a chin rest.

3.1.4. Procedure

With a few exceptions, Experiment 2 was identical to Experiment 1. Though the distance between the participant and the monitor was increased to 300 cm, the sizes of the Medium and Large stimuli were kept at the same visual angles as in Experiment 1, at 0.14° and 0.20°, respectively. Given the increased distance and the limited width of the monitor, stimuli in Experiment 2 were presented 2.80° left or right of the fixation circle. A new weight condition - Ultra-Black (6) - was added, while Light (1), Medium (3) and Bold (5) weights were excluded. Stimulus letters were, therefore, presented in either Regular (2), Semi Bold (4), or Ultra Black (6), at Medium or Large sizes. Contrary to Experiment 1, participants were first given a chance to learn the task in the practice block, before the stimulus exposure duration was calibrated to their performance in the staircase block. The stimulus exposure duration during the practice block and the first trial of the staircase block was set to 180 ms. The step sizes with which the exposure duration increased or decreased during the staircase block was changed to match the refresh rate of the monitor. The initial step size was set to 48 ms, after which it decreased to 36 ms after seven reversals, to 24 ms after 13 reversals, and to 12 ms after 18 reversals. As in Experiment 1, all stimuli in the staircase block were presented at Medium visual angle and at Medium (3) boldness. Lastly, participants recorded their own unspeeded responses on a keyboard.

3.2. Result

Using a 2 (size condition: Medium, and Large) × 3 (weight condition: Regular (2), Semi Bold (4), and Ultra Black (6)) repeated-measures ANOVA on mean accuracy, we found large main effects of size, $F(1, 14) = 157.78$, $p < .001$, $\omega^2 = 0.45$, and of weight, $F(2, 28) = 136.54$, $p < .001$, $\omega^2 = 0.46$, as well as a small but significant interaction effect of size and weight, $F(2, 28) = 5.38$, $p = .011$, $\omega^2 = 0.02$.

This interaction appears to result from the effect of weight, which only facilitated recognition of the Medium sized stimuli. Specifically, planned comparisons, corrected for multiple comparisons using the Bonferroni method, showed a significantly lower mean accuracy in the Regular (2) than in the Semi Bold (4) weight condition of the Medium-sized stimuli, $t(14) = 4.80$, $p = .002$, $dz = 1.24$, while this same comparison did not reach significance for the Large stimuli, $t(14) = 1.24$, $p > .999$, $dz = 0.32$.

Conversely, excessive weight appeared detrimental to performance regardless of size. Mean accuracy of the Ultra Black (6) font was

Fig. 7. The sixteen letters tested in Experiment 2, set in the three test fonts. From the top: Ovink Regular (2), Ovink Semi Bold (4), and Ovink Ultra BLack (6). The numbers to the right are the stroke width/height ratio.

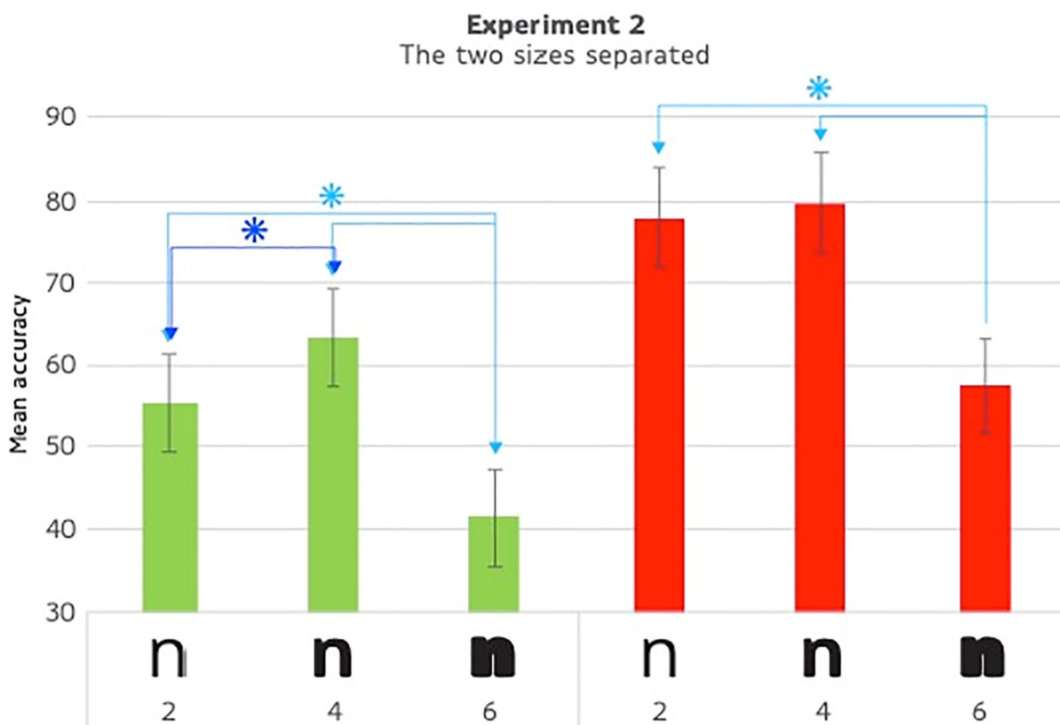

Fig. 8. Mean accuracy of the responses across size and weight conditions. The green bars represent Medium size conditions, and the red bars represent Large size conditions. Numbers on the x-axis represent the weight conditions Regular (2), Semi Bold (4), and Ultra Black (6). (For interpretation of the references to colour in this figure legend, the reader is referred to the digital version of this article.)

significantly lower than the Regular (2) font in both the Medium size, $t(14) = 6.02$, $p < .001$, $dz = 1.56$, and the Large size, $t(14) = 10.87$, $p < .001$, $dz = 2.81$. Similarly, mean accuracy of the Ultra Black (6) font was significantly lower than the Semi Bold (4) font in the Medium size, $t(14) = 11.26$, $p < .001$, $dz = 2.91$, as well as the Large size, $t(14) = 14.91$, $p < .001$, $dz = 3.85$ (Fig. 8).

3.3. Discussions experiment 2

As in Experiment 1, Experiment 2 showed that letter recognition was significantly lower at the Regular (2) weight compared to the Semi Bold (4) weight, though only in the Medium size. We further found a decline in recognition performance in the Ultra Black (6) weight compared to all other fonts at both font sizes. The implications of this will be discussed in the following.

4. General discussion

Our goal was to study the influence of font size on the effect of boldness on letter recognition. The data showed that boldness influences letter recognition in different ways for small and large sizes, and that extreme weights caused lower letter recognition.

4.1. Results differ between font sizes

We found that participants were significantly better at recognising bolder fonts – such as Ovink Regular (2) compared to Ovink Semi Bold (4) – although this was only true for the Medium-sized letters in Experiment 1. We then replicated this finding in Experiment 2 when comparing Ovink Regular (2) and Ovink Semi Bold (4) at Medium and Large sizes. Based on the theory of spatial frequency processing, which holds that small visual angles are sensitive to boldness and proportions with letters appearing blurred, while large visual angles are sensitive to details and edges, we hypothesised that small visual angles would show a greater benefit of letter boldness than large visual angles. In line with our predictions, our results showed that letter recognition of fonts viewed at a small visual angle (Medium size) benefitted more from boldness than fonts viewed at a large visual angle (Large size).

Our findings are in line with previous studies on visual acuity and luminance contrast (Burmistrov et al., 2016; Kuntz & Sleight, 1950; Luckiesh & Moss, 1940; Sheedy et al., 2005), demonstrating that boldness can enhance legibility. Specifically, our data supports the work of Sheedy et al. (2005), who found that Franklin Gothic Book (which has a similar stroke-width ratio to Ovink Regular (2)) had to be read at a larger visual angle than Franklin Gothic Heavy (which has a similar stroke-width ratio to Ovink Bold (5)). Sheedy et al. (2005) further found Franklin Gothic Book to be inferior to all heavier weights tested.

4.2. Extreme weights impair recognition

At all tested visual angles in Experiment 1, Ovink Light (1) resulted in a lower recognition rate. This finding that very thin letter strokes impeded performance in all sizes suggests that thin lines not only blur out in small point sizes but also cause edges and details to become insufficiently visible in large sizes. In extension of this, the heaviest weight tested, Ovink Ultra Black (6), was inferior to all other test fonts at both visual angles in Experiment 2. This follows earlier findings by Bernard et al. (2013), who demonstrated that extremely heavy font weights had a negative effect on reading speed, as there is a limit to how much weight a stroke can carry before the inside of the letter counter fills out completely. Our collective data suggests that for all font sizes, there is an optimal level of boldness with drop-offs at both extreme ends, as for both small and large sizes there is evidence for a significant drop in performances at the two extremes of Ovink Light (1) and Ovink Ultra Black (6).

4.3. Visual cues and letter recognition

Prior research aimed at identifying the most important letter components for recognition only tested one font of regular weight. Some identified the midsegment of the letter stroke to be the most important feature (Petit & Grainger, 2002; Rosa et al., 2016), while others identified the junctions (Lanthier et al., 2009) or the stroke terminations (Fiset et al., 2008) to be most important. However, visual representation of letters are not generic. A given letter will always be visualized in a specific font style and boldness. Findings from one font cannot necessarily translate into the reading of other fonts. By testing different letter weights within one font family, we demonstrated both that letter recognition is enhanced by letter boldness, and that this effect is dependent on size. It could be that our finding that letter boldness enhanced recognition on small visual angles resulted from boldness enhancing the visibility of all the letter components and thus enhancing important visual cues needed for letter recognition.

The experiment adds to the existing body of knowledge by demonstrating that the positive effect of letter boldness on recognition can be found in bold weights if the size of the letters is small.

4.4. Letter boldness and older age

The literature indicates that while sensitivity to low spatial frequencies remains relatively constant throughout adulthood, healthy ageing adults may suffer a loss of sensitivity in the higher and middle spatial frequency regions (Derefeldt, Lennerstrand, & Lundh, 1979; Owsley, Sekuler, & Siemsen, 1983; Wright & Drasdo, 1985). Older readers may, therefore, struggle to identify the finer details of letters, while the recognition of the overall letter proportions remains intact. This suggests that findings concerning letter boldness are especially relevant for this age group. As we did not include older participants in the present investigation, it is likely that a replication of the experiments with an ageing pool of participants would yield an even greater effect of weight in the Small and Medium sizes.

4.5. Reading situations

Our findings relate to single lowercase letter recognition. The way our results translate into the reading of letter strings and words depends on the inter-letter spacing. As fonts read at small visual angles are affected by a phenomenon known as crowding, where neighbouring letters appear to merge (Hess, Dakin, & Kapoor, 2000), it is possible that the tradition of adding a small amount of inter-letter spacing in bold fonts will be counter-productive as narrow letter spacing is known to induce letter crowding (Bouma, 1970).

The results from our experiments could suggest that letter recognition of small font sizes will benefit from having text set in bold fonts, while for letter recognition of larger font sizes the text can be set in both regular and bold weights. To put this finding into a real-life context, the Medium font size of 0.14° typically equals that used for setting text for footnotes (6 point at a reading distance of 40 cm), while the Large font size of 0.20°, lies within contemporary newspaper and book font sizes (9 point at a reading distance of 40 cm, Legge & Bigelow, 2011). As the critical print size for normal vision readers is 0.20°, which is the smallest font size before reading speed will rapidly decline (Legge, 2006), and as studies of font boldness at large font sizes found no effect of bold fonts similar in weight to our Ovink Semi Bold (4) when compared to regular weights (Dobres et al., 2016; Dyson & Beier, 2016), we would expect that in any font bigger than 9 point text sizes read at normal reading distances will similarly fail to enhance reading performance through increased boldness alone.

The letters presented in the small size were so small that it would require participants with a visual acuity of under 0.0 logMAR (3.5 point at a reading distance of 40 cm), which does not represent any real-life reading situation. We did, however, include this font size in Experiment 1 as we expected the advantages of boldness to be the strongest here, although this did not turn out to be the case.

Considering the many reading situations involving small visual angles, our present findings provide evidence that under such reading conditions, bolder weights facilitate letter recognition, and that both light and ultra-black font weights should be avoided in any case where letter recognition is a priority.

5. Conclusion

In all font sizes, the light and the ultra-black fonts were inferior to all the fonts in the middle of the scale. The bolder weights in the middle of the scale enhanced recognition in the Medium font size, while failing to do so in the Large font size. We therefore suggest that, while boldness enhances letter recognition at small visual angles for the tested sans serif font family, it fails to do so at large visual angles, as these are perceived via higher-frequency channels and consequently are more affected by letter details and edges than by letter stroke weight and proportions.

Acknowledgement

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

References

Ahrens, T., & Mugikura, S. (2014). *Size-specific adjustments to type designs: An investigation of the principles guiding the design of optical sizes.* Just Another Foundry.

Alexander, K. R., Xie, W., & Derlacki, D. J. (1994). Spatial-frequency characteristics of letter identification. *JOSA A, 11*(9), 2375–2382.

Bateman, S., Gutwin, C., & Nacenta, M. (2008). *Seeing things in the clouds: The effect of visual features on tag cloud selections. Paper presented at the proceedings of the nineteenth ACM conference on hypertext and hypermedia*.

Beier, S., & Larson, K. (2010). Design improvements for frequently misrecognized letters. *Information Design Journal, 18*(2), 118–137.

Beier, S., & Larson, K. (2013). How does typeface familiarity affect reading performance and reader preference? *Information Design Journal, 20*(1), 16–31.

Beier, S., Starrfelt, R., & Sand, K. (2017). Legibility implications of expressive display typefaces. *Visible Language* (pp. 112–133). .

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. *Vision Research, 84*, 33–42.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226*, 177–178.

Burmistrov, I., Zlokazova, T., Ishmuratova, I., & Semenova, M. (2016). *Legibility of light and ultra-light fonts: Eyetracking study. Paper presented at the proceedings of the 9th Nordic conference on human-computer interaction*.
Chung, S. T. L., Legge, G. E., & Tjan, B. S. (2002). Spatial-frequency characteristics of letter identification in central and peripheral vision. *Vision Research, 42*(18), 2137–2152.
Chung, S. T. L., & Tjan, B. S. (2009). Spatial-frequency and contrast properties of reading in central and peripheral vision. *Journal of Vision, 9*(9), 1–19.
Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204–256.
Derefeldt, G., Lennerstrand, G., & Lundh, B. (1979). Age variations in normal human contrast sensitivity. *Acta Ophthalmologica, 57*(4), 679–690.
Dobres, J., Reimer, B., & Chahine, N. (2016). The effect of font weight and rendering system on glance-based text legibility. *Paper presented at the proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications*.
Dyson, M. C., & Beier, S. (2016). Investigating typographic differentiation: Italics are more subtle than bold for emphasis. *Information Design Journal, 22*(1), 3–18.
Finkbeiner, M., & Coltheart, M. (2009). Letter recognition: From perception to representation. *Cognitive Neuropsychology, 26*(1), 1–6.
Fiset, D., Blais, C., Ethier-Majcher, C., Arguin, M., Bub, D., & Gosselin, F. (2008). Features for identification of uppercase and lowercase letters. *Psychological Science, 19*(11), 1161–1168.
Hess, R. F., Dakin, S. C., & Kapoor, N. (2000). The foveal “crowding”effect: Physics or physiology? *Vision Research, 40*(4), 365–370.
Kesten, H. (1958). Accelerated stochastic approximation. *The Annals of Mathematical Statistics, 29*(1), 41–59.
Kuntz, J. E., & Sleight, R. B. (1950). Legibility of numerals: The optimal ratio of height to width of stroke. *The American Journal of Psychology, 63*(4), 567–575.
Lanthier, S. N., Risko, E. F., Stolz, J. A., & Besner, D. (2009). Not all visual features are created equal: Early processing n letter and word recognition. *Psychonomic Bulletin & Review, 16*(1), 67–73.
Legge, G. E. (2006). *Psychophysics of reading in normal and low vision.* CRC Press.
Legge, G. E., & Bigelow, C. A. (2011). Does print size matter for reading? A review of findings from vision science and typography. *Journal of Vision, 11*(5), 1–22.
Luckiesh, M., & Moss, F. K. (1940). Boldness as a factor in type-design and typography. *Journal of Applied Psychology, 24*(2), 170–183.
Majaj, N. J., Pelli, D. G., Kurshan, P., & Palomares, M. (2002). The role of spatial frequency channels in letter identification. *Vision Research, 42*(9), 1165–1184.
Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324.
McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. an account of basic findings. *Psychological Review, 88*(5), 375–407.
Noordzij, G. (2005). *The stroke.* London: Hyphen Press.
Owsley, C., Sekuler, R., & Siemsen, D. (1983). Contrast sensitivity throughout adulthood. *Vision Research, 23*(7), 689–699.
Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research, 46*(28), 4646–4674.
Perry, C., Ziegler, J. C., & Zorzi, M. (2014). When silent letters say more than a thousand words: An implementation and evaluation of CDP++ in French. *Journal of Memory and Language, 72*, 98–115.
Petit, J.-P., & Grainger, J. (2002). Masked partial priming of letter perception. *Visual Cognition, 9*(3), 337–353.
Pušnik, N., Podlesek, A., & Možina, K. (2016). Typeface comparison − does the x-height of lower-case letters increased to the size of upper-case letters speed up recognition? *International Journal of Industrial Ergonomics, 54*, 164–169.
Rosa, E., Perea, M., & Enneson, P. (2016). The role of letter features in visual-word recognition: Evidence from a delayed segment technique. *Acta Psychologica, 169*, 133–142.
Sanocki, T., & Dyson, M. C. (2012). Letter processing and font information during reading: Beyond distinctiveness, where vision meets design. *Attention, Perception, & Psychophysics, 74*(1), 132–145.
Sheedy, J. E., Subbaram, M. V., Zimmerman, A. B., & Hayes, J. R. (2005). Text legibility and the letter superiority effect. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 47*(4), 797–815.
Tinker, M. A. (1964). *Legibility of print.* Iowa State University Press.
Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research, 35*(17), 2503–2522.
Wright, C. E., & Drasdo, N. (1985). The influence of age on the spatial and temporal contrast sensitivity function. *Documenta Ophthalmologica, 59*(4), 385–395.