The effect of serifs and stroke contrast on low vision reading

Minakata, Katsumi; Eckmann-Hansen, Christina ; Larsen, Michael; Bek, Toke; Beier, Sofie

Published in:
Acta Psychologica

DOI:
10.1016/j.actpsy.2022.103810

Publication date:
2023

Document Version:
Publisher's PDF, also known as Version of record

Acta Psychologica 232 (2023) 103810

Contents lists available at ScienceDirect

Acta Psychologica

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

The effect of serifs and stroke contrast on low vision reading☆

Katsumi Minakata [a], Christina Eckmann-Hansen [b,c], Michael Larsen [b,c], Toke Bek [d], Sofie Beier [a,*]

[a] Centre for Visibility Design, Royal Danish Academy – Architecture, Design, Conservation, Denmark
[b] Department of Ophthalmology, Rigshospitalet, Glostrup, Denmark
[c] Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[d] Department of Ophthalmology, Aarhus University Hospital, Denmark

ARTICLE INFO

Keywords:
Sans serif
Serif
Font
Low vision
Reading
Word recognition

ABSTRACT

Purpose: Patients with low vision are generally recommended to use the same fonts as individuals with normal vision. However, we are yet to fully understand whether stroke width and serifs (small ornamentations at stroke endings) can increase readability. This study's purpose was to characterize the interaction between two factors (end-of-stroke and stroke width) in a well-defined and homogenous group of patients with low vision.

Methods: Font legibility was assessed by measuring word identification performance of 19 patients with low vision (autosomal dominant optic atrophy [ADOA] with a best-corrected average visual acuity 20/110) and a two-interval, forced-choice task was implemented. Word stimuli were presented with four different fonts designed to isolate the stylistic features of serif and stroke width.

Results: Font-size threshold and sensitivity data revealed that using a single measure (i.e., font-size threshold) is insufficient for detecting significant effects but triangulation is possible when combined with signal detection theory. Specifically, low stroke contrast (smaller variation in stroke width) yielded significantly lower thresholds and higher sensitivity when a font contained serifs (331 points; $d' = 1.47$) relative to no serifs (345 points; $d' = 1.15$), $E(\mu_{sans,\ low} - \mu_{serif,\ low}) = -14$ points, 95 % Cr. I. = [−24, −5], $P(\delta > 0) = 0.99$ and $E(\mu_{serif,\ low} - \mu_{sans,\ low}) = 0.32$, 95 % Cr. I. = [0.16, 0.49], $P(\delta > 0) = 0.99$.

Conclusion: In people with low visual acuity caused by ADOA, the combination of serifs and a uniform stroke width resulted in better text legibility than other combinations of uniform/variable stroke widths and presence/absence of serifs.

1. Introduction

Subnormal visual acuity reduces the ability to read, and the compensation induced by magnification aids is at the expense of overview (Brown et al., 2014; Szpiro et al., 2016; Tunold et al., 2019). Consequently, there is a need for maximizing the readability of the fonts used to produce text. It is well documented that different fonts have different reading thresholds for both people with normal (Beier et al., 2018; Beier & Oderkerk, 2022, 2019a, 2019b, 2021; Bernard et al., 2016; Dobres et al., 2017; Sawyer et al., 2017) and low vision (Beier et al., 2021; Mansfield et al., 1996; Tarita-Nistor et al., 2013; Xiong et al., 2018). Font styles vary with respect to a number of stylistic features, such as width and spacing of letters, stroke contrast and the lines at the end of strokes known as serifs (see Fig. 1). Serifs originate in the Roman Capital Letters and have been part of letter design since the early days of printing (Catich, 1968). The continuous popularity of serif fonts suggests that they may facilitate letter recognition, line tracing and reading (Frutiger, 1989; McLean, 1980; Unger, 2007). The phenomenon of visual crowding, where neighboring elements tend to perceptually interfere (Bouma, 1970), is known as a perceptual bottleneck for object recognition (Levi, 2008; Pelli & Tillman, 2008). As visual crowding causes migration of features (Coates et al., 2019), it could be speculated that the presence of serifs enhances the visibility of the letter stroke endings (Unger, 2007) and by that facilitate a more accurate integration of letter features.

Visual disorders have been shown to be accompanied with preferences for specific font styles. Thus, reading performance of patients with age-related macular degeneration (AMD) is facilitated by a wider shape and spacing between letters (Beier et al., 2021; Tarita-Nistor et al., 2013; Xiong et al., 2018), which is rarely taken into account by printed media

☆ The research on which this paper is based was partly supported by a grant from the Independent Research Fund Denmark [grant number DFF – 7013-00039]
* Corresponding author.
E-mail address: sbe@kglakademi.dk (S. Beier).

https://doi.org/10.1016/j.actpsy.2022.103810
Received 4 July 2022; Received in revised form 11 November 2022; Accepted 9 December 2022

(Galvin, 2014; HHS Accessibility, 2020; Kitchel, 2011; VisionAware, 2020). When it comes to effects of serifs, the matter is less conclusive. Interested in the influence of serifs on four participants of normal vision and two of low vision with age-related maculopathy, Arditi and Cho (2005) created new fonts with and without serifs. They tested reading speed and visual acuity, and only found significant results on visual acuity showing that their serif fonts could be read at smaller sizes than their sans serif font. The authors speculated this being related to the greater spacing of the serif fonts. Others have compared different font styles, some with serifs and others without. For these studies, serifs was not a defining factor on reading acuity, critical print size and reading speed of AMD patients (Tarita-Nistor et al., 2013; Xiong et al., 2018), nor was it a factor on speed of reading the text aloud for cataract and glaucoma patients (Rubin et al., 2006). Others have failed to find differences between boldness and regular weights for participants of central vision loss and AMD (Beier et al., 2021; Chung & Bernard, 2018). The thickness of the stroke and the use of serifs are defining features of font styles. Thus, the generally used sans serif fonts such as Arial, Helvetica and Verdana are designed with little or no variation in the thickness of the stroke (low stroke contrast) whereas serif fonts such as Times Roman, Georgia and Cambria traditionally have greater variation in the stroke thickness (high stroke contrast (Beier, 2017); see Fig. 1). However, the significance of serifs and stroke contrast for the readability of text fonts in visually impaired patients is yet to be fully understood.

The purpose of the present study was to compare the font feature combinations serif/sans serif and low/high stroke contrast. We have previously found that in persons with normal vision, the most legible types were those with low stroke contrast and no serif (Minakata & Beier, 2022). In this study, we have made a comparable evaluation in persons with *OPA1*-mutation with autosomal dominant optic atrophy which causes slow, childhood-onset subnormal visual acuity and a cecocentral scotoma in both eyes.

2. Method

2.1. Participants

Nineteen patients with *OPA1* autosomal dominant optic atrophy were recruited from the Department of Ophthalmology at the Rigshospitalet and the Department of Ophthalmology, Aarhus University Hospital (Rönnbäck & Larsen, 2014). Inclusion criteria were visual acuity between 20/63 and 20/200 for the eye with best visual acuity, and age between 15 and 75 years. The study included 11 females and 8 males with an average age of 39 years (range 17–72) (see Table 1) and an average Snellen visual acuity of 20/110. Danish was the primary language of all participants. The study followed the tenets of the Declaration of Helsinki and all persons gave their written informed consent to participate.

Table 1
Characteristics of the clinical population.

ID	Age	Gender	ETDRS letters	Snellen
1	51	F	R 53	R 20/80
			L 41	L 20/160
2	21	F	R 64	R 20/50
			L 57	L 20/63
3	39	F	R 49	R 20/100
			L 52	L 20/100
4	30	M	R 49	R 20/100
			L 55	L 20/80
5	72	F	R 45	R 20/125
			L 45	L 20/125
6	40	M	R 54	R 20/80
			L 51	L 20/100
7	63	M	R 30	R 20/250
			L 55	L 20/80
8	49	F	R 54	R 20/80
			L 59	L 20/63
9	17	F	R 58	R 20/63
			L 59	L 20/63
10	61	F	R 59	R 20/63
			L 41	L 20/160
11	35	M	R 38	R 20/160
			L 38	L 20/160
12	50	F	R 47	R 20/160
			L 49	L 20/125
13	38	M	R 52	R 20/100
			L 44	L 20/125
14	31	F	R 38	R 20/160
			L 39	L 20/160
15	23	F	R 52	R 20/100
			L 58	L 20/63
16	34	M	R 51	R 20/100
			L 61	L 20/63
17	24	M	R 49	R 20/100
			L 53	L 20/80
18	28	M	R 42	R 20/160
			L 39	L 20/160
19	32	F	R 47	R 20/125
			L 43	L 20/125

F = Female; M = Male; ETDRS = Early Treatment Diabetic Retinopathy Study; L = Left; R = Right.

2.2. Procedure

The procedure, font stimuli and equipment, were identical to that of Experiment 1 in Minakata and Beier (2022).

The test objects were sequences of letters varying between three and six letters in the font Helvetica that either represented a meaningful word or a pseudoword constructed by changing one of the central letters in a real word, so that the new pseudoword would still be pronounceable. Participants were exposed to one pair of words: a real word and a pseudoword. The two words were separated in time by a short delay. Throughout the threshold estimation procedure, the font Helvetica was used as the baseline font. Participants were informed that the real word

Fig. 1. Stylistic elements of font styles highlighted for the sans serif font Helvetica (top) and the serif fonts Times Roman (middle) and Courier (bottom).

would randomly appear either first or second and that their task is to indicate whether the real word came first or last.

The participants were also instructed that the word-recognition task would be presented under two different contexts (first font size threshold estimation and then the experiment proper). The trial sequence is shown in Fig. 2. During each trial the word and a pseudoword were presented in random order, each followed by the presentation of random noise for 50 ms corresponding to the text field to eliminate after images secondary to neural or visual persistence (Sperling, 1965). The random noise presentations were followed by a black screen which lasted 500 ms after the first presentation and after the second presentation lasted until the patient had pressed one of two designated keys on the keyboard to indicate whether the meaningful word had been presented first or last. The key-press responses were recorded via a keyboard that featured a "left arrow" key to indicate a "first interval" decision and a "right arrow" key that indicated a "second interval" response. The interval to the following trial was set to 1000 ms.

Next, each participants' font-size threshold was estimated with QUEST strategy which is an adaptive Bayesian procedure for threshold estimation (Watson, 2017; Watson & Pelli, 1983). The algorithm requires four input parameters: (1) A psychometric function which was chosen a cumulative normal distribution; (2) A slope which was set to 2; (3) The lapse rate which is the probability of an incorrect response when a supra-threshold stimulus is presented, which was set to 0.019 (i.e., 2 % of trials are assumed to be missed due to inattention, etc.); and (4) A guess rate set to 0.50 determined by the 2IFC task structure (½ + ½ / 2 = ½). The QUEST estimation yielded a distribution of responses to the threshold estimate from which the mean and the standard deviation was noted. This QUEST procedure was repeated three times and the three estimations of the mean and standard deviation of the threshold were averaged to result in an overall mean (μ) and standard deviation (σ), from which the range of the visible font sizes for the patient was calculated: ($\mu - [2 \times \sigma]$; $\mu + [4 \times \sigma]$). From the minimum to the maximum of this interval nine evenly-spaced steps of font sizes were defined to be used for experiment proper.

During the second phase (experiment proper), the same word-recognition task was used and instead of the QUEST algorithm, the method of constant stimuli (MOCS) was applied. The MOCS is a psychophysical method wherein stimuli pairs are serially, constantly, and randomly presented. The four experimental font conditions were presented, and the experiment was divided into 10 equally-sized blocks (72 trials each). At the beginning of each block, participants were allowed to take a break and each block took about 15 min to complete. Upon completion of the fifth and tenth block, the QUEST procedure (used to estimate the font-size threshold) was repeated. These additional thresholds were collected to statistically control for learning- or practice-effects, as needed.

2.3. Stimuli

The four test fonts took the outset in the font Ovink Regular. Using the software Glyphs, this style was modified to a ratio of thick and thin strokes of 3/2.4 to result in low contrast (right column, Fig. 3), and of 3/0.8 to result in high stroke contrast following a vertical stroke modulation of a pointed nib pen (left column, Fig. 3). To ensure a perceptual equal distribution of "blackness" across fonts, the stem weights of the serif fonts are 27 % heavier than the stem weight of the sans serif fonts. The fonts were also modified by the addition of serifs of sharp triangular brackets (lower row). In both serif fonts, the stroke thickness of the serifs followed the stroke contrast of the sans serif fonts. The four fonts were also tested in the previous experiment involving normal vision participants (Minakata & Beier, 2022).

2.4. Data analysis

2.4.1. Calculation of threshold and sensitivity

For each of the nine font sizes (tailored to each participant's performance) the percentage of correct responses to whether the meaningful word was presented first was calculated, and these percentages were fitted to a cumulative normal distribution as a function of font size. The probability of correct answer of 75 % point obtained from the curve, a chance/guessing rate of 0.5 and an inattention/lapse-rate of 0.019 were entered into the Psignifit Toolbox's maximum likelihood fitting procedure (Prins & Kingdom, 2018; Schütt et al., 2016) in order to estimate the font size threshold (α alpha). In signal detection theory, the hit rate (H: correct detection of word in interval 1) and the false alarm rate (F: incorrect detection of word in interval 2) are taken into account to get measures of sensitivity and response bias. The normalized distribution of the probability of correct word detection in interval 1 was z(H) and z(F) of incorrect detection in interval 2 (Macmillan & Creelman, 1991). The sensitivity of each test person to distinguish between meaningful words (signal) and pseudowords (noise) was calculated as $d' = z(H) - z(F)$ with the criterion level for a correct/false answer set to: $\beta = 0.5 \times [z(H) + z(F)]$. The calculations were based on ten trials for each font size.

2.4.2. Confirmation of results by posterior probability

To confirm the results, Bayesian hierarchical linear models were used to estimate the posterior probability of both the font-size threshold (*alpha*) and sensitivity (d') with stroke contrast and serif as independent factors. The estimation was performed in the Stan modeling language (Carpenter et al., 2017) in R and the package *brms* (Bürkner, 2017; Stan Development Team, 2017; Stan Modeling Language, 2017), which was needed to get a model that converged given our small sample size. This approach considered maximal random effect structures (Barr et al.,

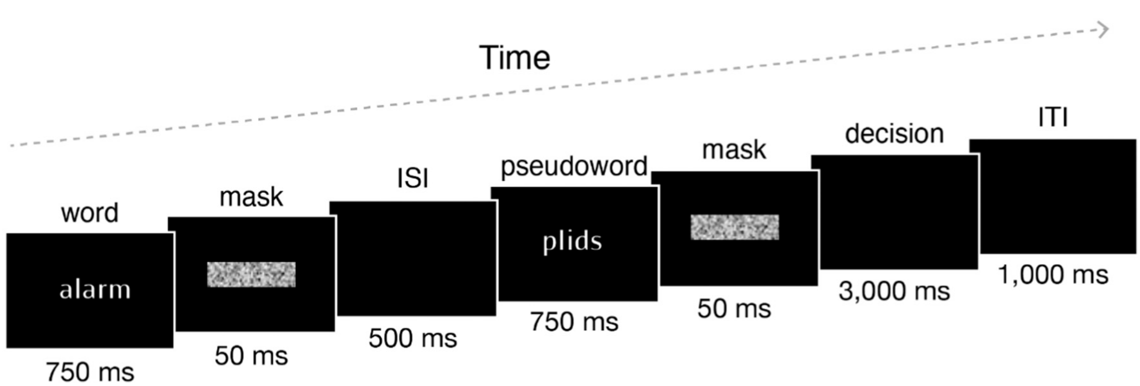

Fig. 2. *The experimental procedure.* The test session had two phases; threshold estimation (first font size threshold estimation and experiment proper. Participants were exposed to one pair of words set in the stimuli fonts: a real word and a pseudoword in random order. The two words were separated in time by a short delay. The meaningful word "alarm" is presented before the pseudoword "plids". ISI = inter-stimulus interval; ITI = inter-trial interval. For a thorough description of methodology see Minakata & Beier (Minakata & Beier, 2022).

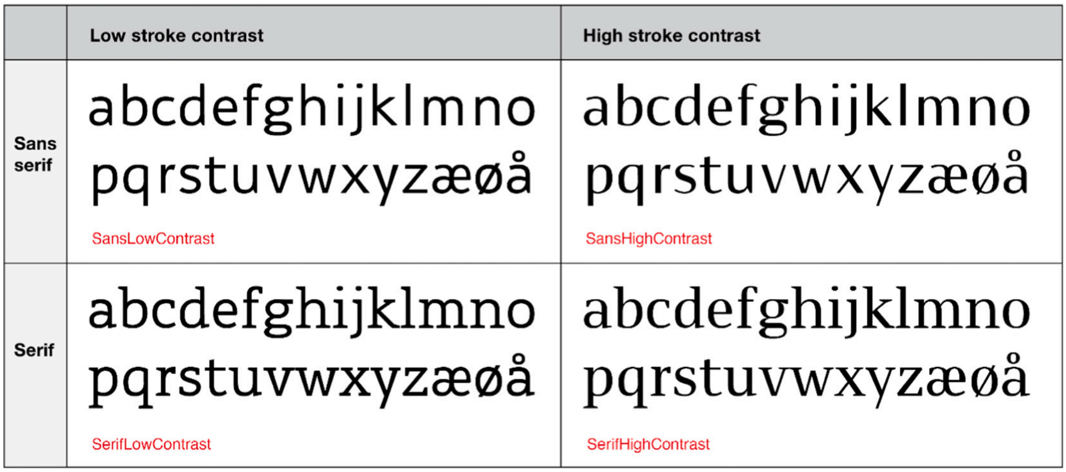

Fig. 3. The four font conditions, designed to control for the presence or absence of serif and for low stroke contrast/high stroke contrast.

2013), and that the predictors of interest and their interactions could vary for each participant. Two hierarchical levels were used as follows:

Level 1:

$$\text{Font}-\text{size Threshold}_{ij} = \beta_{0j} + \beta_{1j}(\text{Serif Type}) + \beta_{2j}(\text{Stroke Contrast}) + \beta_{3j}(\text{Serif Type})^{*}(\text{Stroke Contrast}) + \text{R}_{ij}$$

Level 2:

$$\beta_{0j} = \gamma_{00} + \text{U}_{0j}$$

$$\beta_{1j} = \gamma_{10}(\text{Serif Type}) + \text{U}_{1j}$$

$$\beta_{2j} = \gamma_{20}(\text{Stroke Contrast}) + \text{U}_{2j}$$

$$\beta_{3j} = \gamma_{30}(\text{Serif Type})^{*}(\text{Stroke Contrast}) + \text{U}_{3j}$$

Full Equation:

$$\begin{aligned}\text{Font}-\text{size Threshold}_{ij} &= \gamma_{00}(\text{Intercept}) + \gamma_{10}(\text{Serif Type}_{ij}) \\ &+ \gamma_{20}(\text{Stroke Contrast}_{ij}) + \gamma_{30}(\text{Serif Type}_{ij})^{*}(\text{Stroke Contrast}_{ij}) \\ &+ \text{U}_{0j}(\text{Intercept}) + \text{U}_{1j}(\text{Serif Type}_{ij}) + \text{U}_{2j}(\text{Stroke Contrast}_{ij}) \\ &+ \text{U}_{3j}(\text{Serif Type}_{ij})^{*}(\text{Stroke Contrast}_{ij}) + \text{R}_{ij}\end{aligned}$$

Let $\text{Font}-\text{size Threshold}_{ij}$ denote the i^{th} observation in the j^{th} participant. That is, $i =$ trial level, $j =$ participant level, $\text{U} =$ level $-$ two error, $\text{R} =$ population $-$ level error.

The serif, stroke contrast factors, and their interaction, were treated as categorical fixed effects to obtain their respective omnibus estimates. This was followed by treating each of different combinations of stroke contrast and serif and the interaction between the two as categorical reference cells and the other cells as test cells. The serif type = sans and stroke contrast = low and their interaction were given Student's *t*-distribution priors (*alpha*: $\nu = 3$, $\mu = 315.2$, $\sigma = 95.5$, & $\gamma = 2$, 0.10; *d*': $\nu = 3$, $\mu = 1.3$, $\sigma = 2.5$, & $\gamma = 2$, 0.10). We used the *brms* package's default priors for standard deviations of random effects (a Student's *t*-distribution with: $\nu = 3$, $\mu = 0$ & $\sigma = 95$; $\nu = 3$, $\mu = 0$ & $\sigma = 2.5$), as well as for correlation coefficients in interaction models (LKJ $\eta = 1$).

Six sampling chains with 10,000 iterations (i.e., more than sufficient for estimating the resultant posterior) were run with a warm-up period of 5000 iterations for each chain, thereby yielding 30,000 samples for each parameter tuple. For the marginal means and differences between them, we report the expected values under the posterior distribution and their 95 % credible intervals (Cr. I.s). For marginal mean differences, we also report the posterior probability that a difference δ is bigger than zero. If a hypothesis states that $\delta > 0$, then it would be considered *strong evidence* for this hypothesis would be if zero is not included in the 95 % Cr. I. of δ and the posterior $P(\delta > 0)$ is close to one (by a reasonably clear margin). To extract the estimated marginal means from the posterior distribution of the fitted models we used the "emmeans" R package (Russell, 2021).

The probability of direction (pd) was used to determine whether the non-significant post hoc comparisons were equivalent (Makowski et al., 2019). A low pd. is related to no direction (no effect) and a high pd. means there is a direction (positive effect). The "estimate_contrasts" function from the R package "model based" (Makowski et al., 2020) was applied to the brms model fit, which yielded the differences, 95 % credible intervals, pd., and percentage in the region of practical equivalence (ROPE). That is, an approximation of an alpha-level in the Bayesian-statistics framework is called the ROPE. Thus, analogously, 6 % in the ROPE is equal to $p = .06$.

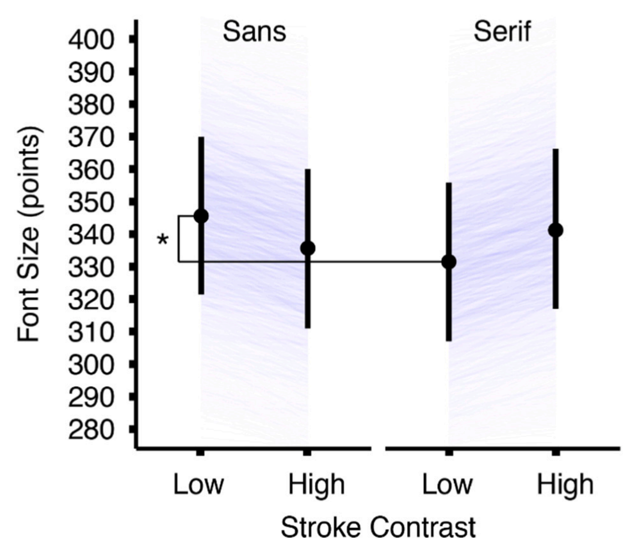

Fig. 4. *Font-size threshold as a function of serif type and stroke contrast.* Vertical Bars represent the 95 % Credible Intervals around the estimated marginal means, which are represented by black circles. Blue areas represent the posterior distribution. Note the only significant effect was the simple effect between sans serif with low stroke contrast and serif with low stroke contrast. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Results

3.1. Font-size threshold (α)

The feature combination with the lowest font size threshold, and hence the one that was easiest to read in small print, was low stroke contrast combined with serifs (Fig. 4). When features were analysed in isolation, the font size threshold was higher for sans-serifs ($E(\mu_{sans})$ = 340, 95 % Cr. I. = [287, 391]) than for serifs ($E(\mu_{serif})$ = 336, 95 % Cr. I. = [284, 388]), but the difference was not significant ($E(\mu_{sans} - \mu_{serif})$ = 4, 95 % Cr. I. = [−3,13], $P(\delta > 0) = 0.82$). The font-size threshold was marginally higher for high stroke contrast($E(\mu_{high})$ = 338, 95 % Cr. I. = [288, 392]) than for low stroke contrast ($E(\mu_{low})$ = 338, 95 % Cr. I. = [287, 389]), but the difference was not significant ($E(\mu_{low} - \mu_{high})$ = − 0.13, 95 % Cr. I. = [−7, 7], $P(\delta > 0) = 0.51$). In terms of the interaction between serif and stroke contrast, however, the serif font condition with low stroke contrast yielded the lowest font-size threshold ($E(\mu_{serif,\ low})$ = 331, 95 % Cr. I. = [279, 381]), followed by the sans-serif font condition with high stroke contrast ($E(\mu_{sans,\ high})$ = 335, 95 % Cr. I. = [284, 386]), then the serif font condition with high stroke contrast yielded a mean of ($E(\mu_{serif,\ high})$ = 340, 95 % Cr. I. = [287, 391]), finally, the sans-serif condition with low stroke contrast resulted in a mean of ($E(\mu_{sans,\ low})$ = 345, 95 % Cr. I. = [297, 397]; see Fig. 4). The serif type by stroke contrast interaction coefficient for the serif with high stroke contrast was >0 ($E(\mu_{serif,\ high})$ = 20, 95 % Cr. I. = [7, 33], $P(\delta > 0) = 0.99$). For means of the font-size thresholds and sensitivity for each participant, refer to Appendix 1.

At the level of the sans serif font condition, the low stroke contrast condition yielded a higher font-size threshold relative to the high stroke contrast condition ($E(\mu_{sans,\ low} - \mu_{sans,\ high})$ = 10, 95 % Cr. I. = [0, 19], $P(\delta > 0) = 0.96$). At the level of the serif font condition, the opposite pattern was found the low stroke contrast condition yielded lower sensitivity relative to the high stroke contrast condition. However, the magnitude of the difference did not provide compelling evidence ($E(\mu_{serif,\ low} - \mu_{serif,\ high})$ = 10, 95 % Cr. I. = [0, 20], $P(\delta > 0) = 0.95$). There was ample evidence, however, for the difference between serif and sans-serif, when analysed at the level of the low stroke contrast condition ($E(\mu_{sans,\ low} - \mu_{serif,\ low})$ = −14, 95 % Cr. I. = [−24, −5], $P(\delta > 0) = 0.99$). For Bayesian pairwise comparisons see Table 2.

3.2. Sensitivity (d prime [d'])

In contrast to the font-size threshold results, the combination of features that lead to the best performance, in terms of sensitivity, was low stroke contrast with serifs. When analysing the features in isolation, the serif font condition produced higher sensitivity ($E(\mu_{serif})$ = 1.36, 95 % Cr. I. = [0.96, 1.75]) than the sans-serif font condition ($E(\mu_{sans})$ = 1.24, 95 % Cr. I. = [0.83, 1.62]). There was no compelling evidence for this difference ($E(\mu_{sans} - \mu_{serif})$ = −0.12, 95 % Cr. I. = [−0.23, 0], $P(\delta > 0) = 0.98$). The low stroke-contrast font condition produced lower sensitivity ($E(\mu_{low})$ = 1.30, 95 % Cr. I. = [0.94, 1.74]) than the high stroke-contrast font condition ($E(\mu_{high})$ = 1.31, 95 % Cr. I. = [0.92, 1.71]). There was, however, no compelling evidence for this difference ($E(\mu_{low} - \mu_{high})$ = 0.02, 95 % Cr. I. = [0.06, −0.08], $P(\delta > 0) = 0.62$). These non-significant main effects were qualified by the significant serif type by stroke contrast interaction.

In terms of the interaction between stroke contrast and serif, the serif font condition with low stroke contrast yielded the highest sensitivity ($E(\mu_{serif,\ low})$ = 1.47, 95 % Cr. I. = [1.08, 1.88]), followed by the sans font condition with high stroke contrast ($E(\mu_{sans,\ high})$ = 1.33, 95 % Cr. I. = [0.91, 1.71]), then the serif font condition with high stroke contrast yielded a mean of ($E(\mu_{serif,\ high})$ = 1.24, 95 % Cr. I. = [0.83, 1.63]), finally, the sans-serif condition with low stroke contrast resulted in a mean of ($E(\mu_{sans,\ low})$ = 1.15, 95 % Cr. I. = [0.75, 1.54]; see Fig. 5 & Table 4).

At the level of the sans-serif font condition, the low stroke contrast condition ($E(\mu_{sans,\ low})$ = 1.15) yielded lower sensitivity relative to the high stroke contrast condition ($E(\mu_{sans,\ high})$ = 1.33). There was compelling evidence for this difference between low and high stroke contrast ($E(\mu_{sans,\ low} - \mu_{sans,\ high})$ = −0.19, 95 % Cr. I. = [−0.32, −0.02], $P(\delta > 0) = 0.98$). At the level of the serif font condition, the opposite pattern was found, the low stroke contrast condition ($E(\mu_{serif,\ low})$ = 1.47) yielded higher sensitivity relative to the high stroke contrast condition ($E(\mu_{serif,\ high})$ = 1.24). The difference value provided compelling evidence ($E(\mu_{serif,\ low} - \mu_{serif,\ high})$ = 0.22, 95 % Cr. I. = [0.07,

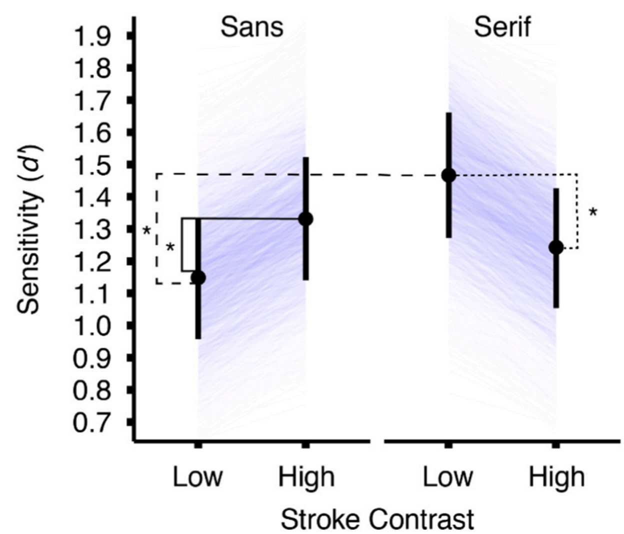

Fig. 5. Sensitivity as a function of serif type and stroke contrast. Asterisks (*) represent significant pairwise comparison, $p < .05$. Vertical Bars represent the 95 % Credible Intervals around the estimated marginal means, which are represented by black circles. Blue areas represent the posterior distribution. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Pairwise comparisons for font-size threshold as a function of serif type and stroke contrast. Cr. I. = credible interval; pd. = probability of direction; ROPE = region of practical equivalence. Grey rows represent statistically equivalent conditions, italicised font represents non-significant simple effects, and bold font represents significant simple effects.

Font-size Threshold Pairwise Comparisons					
Level 1	Level 2	Difference	95% Cr. I.	pd	% in ROPE
Sans, Low	**Serif, Low**	**14**	**(3, 26)**	**99.22**	**0**
Sans, Low	Sans, High	*10*	*(-2, 22)*	*95*	*.31*
Sans, Low	Serif, High	5	(-7, 16)	78.33	1.01
Serif, Low	Sans, High	-4	(-16, 7)	77.42	1.12
Serif, Low	*Serif, High*	*-10*	*(-21, 2)*	*95.25*	*0.38*
Sans, High	*Serif, High*	*-5*	*(-17, 6)*	*81.82*	*0.83*

Table 3
Pairwise comparisons for sensitivity as a function of serif type and stroke contrast. Cr. I. = credible interval; pd. = probability of direction; ROPE = region of practical equivalence. Italicised font represents non-significant simple effects, and bold font represents significant simple effects.

Sensitivity pairwise comparisons					
Level 1	Level 2	Difference	95 % Cr. I.	pd	% in ROPE
Sans, Low	**Serif, Low**	**−0.32**	**(−0.49, −0.16)**	**100**	**0**
Sans, Low	**Sans, High**	**−0.19**	**(−0.32, −0.02)**	**98.43**	**13.40**
Sans, Low	*Serif, High*	*−0.10*	*(−0.27, 0.07)*	*87.10*	*51.18*
Serif, Low	*Sans, High*	*0.13*	*(−0.03, 0.32)*	*94.30*	*32.66*
Serif, Low	**Serif, High**	**0.22**	**(0.07, 0.39)**	**99.43**	**3.82**
Sans, High	*Serif, High*	*0.09*	*(0.08, 0.25)*	*85.63*	*54.93*

Table 4
Sensitivity as a function of serif type and stroke contrast. Values in italics represent the estimated marginal means of the 2-by-2 factorial design and bold-values represent a frequentist equivalent of an alpha value <5 %.

		Serif type		Mean Difference	% in ROPE
		Sans	Serif		
Stroke Contrast	Low	*1.15*	*1.47*	0.32	**0**
	High	*1.33*	*1.24*	0.09	55
Mean Difference		0.18	0.22		
% in ROPE		13	**4**		

0.39], $P(\delta > 0) = 0.99$). At the level of the low stroke-contrast condition, the serif condition ($E(\mu_{serif,\ low}) = 1.47$) yielded higher sensitivity relative to the sans condition ($E(\mu_{sans,\ low}) = 1.15$). The difference value provided compelling evidence ($E(\mu_{serif,\ low} - \mu_{sans,\ low}) = 0.32$, 95 % Cr. I. = [0.16, 0.49], $P(\delta > 0) = 0.99$). For Bayesian pairwise comparisons see Table 3.

4. Discussion

Using the same test stimuli and procedure as a previous experiment (Minakata & Beier, 2022), the primary purpose of the study was to determine whether our findings, from persons with normal vision, could be replicated in patients with low vision. The secondary objective was to examine whether the typographic characteristics of: 1) presence or absence of serifs and 2) high or low levels of stroke contrast interact to determine legibility. In the previous experiment, the spectral power was obtained for each font condition by applying a fast fourier transformation of the four images that each contained a single font's alphabet. That is, an image of the alphabet was created with a font size of the grand mean of the font-size threshold for the four font conditions and was then transformed to cycles per degree, by taking the distance between the stimulus and the observer into account. The resultant spatial frequency magnitude (in decibels) was reorganized/binned into each frequency band of interest (e.g., 10–15, 15–20 Hz, etc.). We found that fonts of similar spectral power elicited similar results (see Fig. 6). Thus, for the present experiment, we expected that the performance results of our low-vision participant group would similarly fit with our previous spectral power findings. However, considering that low- and normal vision populations vary greatly across many visual parameters, this expected data pattern was only a tentative hypothesis.

We measured reading of four fonts with a word identification task, where the fonts varied in letter stroke contrast and in the presence or absence of serifs (see Fig. 3). Participants were visually impaired patients with autosomal dominant optic atrophy. Our sensitivity and font-size threshold measures showed better performance, which was reliant on our test fonts' spectral composition. However, the pattern was reversed from our previous normal-vision group's findings. While the normal-vision participants showed lower sensitivity with the fonts SerifLowContrast and SansHighContrast, our low-vision participants showed higher sensitivity with these same fonts.

As our low-vision participants all had autosomal dominant optic atrophy, it strengthened our experiment's internal validity by limiting the large variability of visual acuity from different visual disorders. Visual function in low-vision patients is known to vary between different diagnoses, most significantly between patients with intact central visual fields versus central visual field loss and between excessively blurred versus clear ocular media (Legge et al., 1985). However, this advantage can also be construed in the opposite way because our low-vision sample is not as heterogenous and may not transfer to a more general low-vision population. Autosomal dominant optic atrophy is characterized by central foveal visual defects, which results in impaired contrast sensitivity for higher spatial frequencies. Thus, their visual system has a limited bandwidth that is insensitive to images and text that contain more spectral power at higher spatial frequencies.

Visual stimuli presented at small visual angles tend to produce lower task performance (e.g., lower letter recognition rate) when the stimuli are mainly composed of high spatial frequencies (Majaj et al., 2002). In our spectral power analysis (see Fig. 6), all four conditions resulted in a frequency range between 0.01 and 55 cycles per image Hertz (Hz). The sans serif, high stroke contrast and the serif, low stroke contrast fonts all yielded higher power for low spatial frequency bands (0–10 & 10–20 Hz) and their first 2 frequency-band peaks fall below 20 Hz. The opposite was true for the sans serif with low stroke contrast font and the serif with high stroke contrast font, such that their first two frequency-band peaks were shifted to the right, and therefore, outside of the lower spatial frequencies that are important for low-vision reading. Thus, there is a possibility that the spatial frequency composition of the test fonts can be an alternative explanation for our results.

While normal-vision participants draw on the whole spectrum of spatial frequency information (Beckmann et al., 1991), participants with low vision usually have limited contrast sensitivity for higher spatial frequencies. The font conditions with more spectral power at low spatial frequencies would therefore provide better performance relative to the font conditions with more spectral power at high spatial frequencies. Our data follow this pattern (see. Fig. 6).

Our previous experiment was identical to our current experiment (except that our low-vision participants required larger font sizes in order to obtain identical word-recognition performance) and provided us with the opposite result. Higher spatial frequency information in images is represented as sharper and clearer edges, which were possibly not as well perceived by our low-vision sample of participants. By employing the same methods as the previous experiment, we showed that the optimal typographic features of fonts for normal vision participants are not always identical to the optimal typographic features of fonts for low-vision participants.

As our experimental paradigm allowed us to isolate two typographic variables instead of just one, we have been able to identify an interaction between typographic variables; stroke contrast and serifs. The interaction showed that, for low-vision readers, there is no evidence for an effect of serifs when they are experimentally isolated. Furthermore, there is no evidence for an isolated effect of stroke contrast. Yet, the effect emerged when the two variables were combined. Although the notion that typographic variables can interact is a phenomenon that has been long-observed by typographers (Beier, 2016, 2021), it has received little to no attention in the research literature. This point alone highlights the importance of our results because it emphasizes that one should not create design guidelines that are solely grounded on single-factor studies, that only manipulate a single typographic variable. Instead, one should also refer to experiments with more complicated designs such as the fully-crossed 2-factor experimental design we have implemented. More experiments need to be conducted on a combination of typographic variables.

Fig. 6. A visualization of the spectral power as a function of cycles per image (Hz) for the four font condition's alphabets. These differences relate to the spatial frequency information provided by the addition of serifs. (A) and (B) marks the two lowest peaks of cycles per image on all four fonts. Please note that for the two good performing test fonts for low vision participants, both peaks are below 10 cycles per image.

We suggest that the current tradition of advising to use specific fonts for low vision reading, should be reconsidered. Instead, recommendations should be based on font characteristics. This makes it possible for one to choose from the many fonts available on the market, rather than being restricted to a list of few, preselected, font options. Our findings suggest that serif fonts for low-vision readers should have low stroke contrast while sans serif fonts should have high stroke contrast.

While we found significant differences in performance between test fonts, others have failed to do so (Rubin et al., 2006). We argue that our methodology is more sensitive to smaller performance differences, compared to studies that measure reading speed when text is read aloud, which often show no effect (Beier & Oderkerk, 2019a; Tarita-Nistor et al., 2013; Xiong et al., 2018).

A limitation of the study was that only one font style was tested. Although we isolated our typographical variables, it is very likely that the same variables could influence the results in a different way, if they were used and isolated within a completely different font style. Second, we showed that typographic variables indeed interact and can lead to different results. It is, therefore, likely that other typographic variables could affect the results. Inter-letter spacing and letter width are already known to improve low-vision visual acuity (Beier et al., 2021; Tarita-Nistor et al., 2013; Xiong et al., 2018). Font style can differentially affect the task performance of normal- and low-vision participants; typographical variables can interact and lead to unanticipated results.

5. Conclusion

Using a sensitivity measure (d-prime) on a word identification task, we found that low-vision participants with autosomal dominant optic atrophy had smaller font-size thresholds when a sans-serif font was set with low stroke contrast compared to when it was set with high stroke contrast. Moreover, the opposite case was true for the serif font conditions, where better performance was found when words were set with low stroke contrast when compared with words that were set with high stroke contrast. Looking at the variables the other way around, both the sensitivity measure and a font-size threshold measure found that low stroke contrast fonts were read at significantly smaller sizes when words were set with serifs relative to words set without serifs (sans serif). Our findings demonstrating that typographic variables interact for low vision readers. Paradigms that manage to isolate such variables and measure the nature of their interaction should be preferred over a more traditional approach of comparing fonts of different origin.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Data availability

Data will be made available on request.

Appendix 1. Font-size threshold and Sensitivity as a function of Serif and Stroke Contrast by participant

ID	Serif Condition	Stroke Contrast Condition	Font-size Threshold	Sensitivity
	Sans	Low	305	1.33
	Sans	High	328	1.12
	Serif	Low	313	1.78
1	Serif	High	325	1.14
	Sans	Low	229	2.51
	Sans	High	227	1.77
	Serif	Low	220	2.06
2	Serif	High	216	1.35
	Sans	Low	321	1.13
	Sans	High	312	1.17
	Serif	Low	322	1.30
3	Serif	High	330	1.02
	Sans	Low	250	2.51
	Sans	High	259	3.34
	Serif	Low	265	3.31
4	Serif	High	248	2.67
	Sans	Low	361	0.61
	Sans	High	357	0.92
	Serif	Low	379	0.80
5	Serif	High	383	0.68
	Sans	Low	219	2.32
	Sans	High	233	3.04
	Serif	Low	222	2.76
6	Serif	High	234	2.33
	Sans	Low	343	0.99
	Sans	High	336	1.03
	Serif	Low	316	1.36
7	Serif	High	315	1.28
	Sans	Low	308	1.31
	Sans	High	340	1.01
	Serif	Low	314	1.45
8	Serif	High	332	1.01
	Sans	Low	312	0.80
	Sans	High	275	1.35
	Serif	Low	255	2.28
9	Serif	High	271	1.28
	Sans	Low	442	0.11
	Sans	High	440	0.17
	Serif	Low	452	0.05
10	Serif	High	467	0.43
	Sans	Low	571	0.13
	Sans	High	621	0.35
	Serif	Low	552	0.10
11	Serif	High	530	0.35
	Sans	Low	292	1.36
	Sans	High	246	1.73
	Serif	Low	244	1.84
12	Serif	High	269	1.64
	Sans	Low	353	0.80
	Sans	High	325	1.29
	Serif	Low	332	1.20
13	Serif	High	348	0.88
	Sans	Low	649	0.10
	Sans	High	616	0.00
	Serif	Low	606	0.18
14	Serif	High	670	0.10
	Sans	Low	241	1.89
	Sans	High	236	1.99
	Serif	Low	195	2.24
15	Serif	High	219	2.31
	Sans	Low	307	1.25
	Sans	High	273	1.40
	Serif	Low	282	1.81
16	Serif	High	266	1.88
	Sans	Low	276	1.56
	Sans	High	235	2.14
	Serif	Low	239	2.14
17	Serif	High	235	2.03

(*continued on next page*)

(*continued*)

ID	Serif Condition	Stroke Contrast Condition	Font-size Threshold	Sensitivity
	Sans	Low	450	0.63
	Sans	High	432	0.70
	Serif	Low	437	0.74
18	Serif	High	463	0.66
	Sans	Low	376	0.74
	Sans	High	346	1.04
	Serif	Low	380	0.78
19	Serif	High	392	0.64

References

Arditi, A., & Cho, J. (2005). Serifs and font legibility. *Vision Research, 45*(23), 2926–2933. https://doi.org/10.1016/j.visres.2005.06.013

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Beckmann, P. J., Legge, G. E., & Luebker, A. (1991). *7.5: Reading: Letters, words, and their spatial-frequency content*.

Beier, S. (2016). Letterform research: An academic orphan. *Visible Language, 50*(2), 64.

Beier, S. (2017). *Type Tricks: Your personal guide to type design*. BIS Publishers.

Beier, S. (2021). *Type Tricks: Layout design*. BIS Publishers.

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

Beier, S., & Oderkerk, C. A. (2022). Closed letter counters impair recognition. *Applied Ergonomics, 101*, Article 103709. https://doi.org/10.1016/j.apergo.2022.103709

Beier, S., & Oderkerk, C. A. T. (2019a). The effect of age and font on reading ability. *Visible Language, 53*(3), 51–69. https://doi.org/10.34314/vl.v53i3.4654

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

Beier, S., & Oderkerk, C. A. T. (2021). High letter stroke contrast impairs letter recognition of bold fonts. *Applied Ergonomics, 97*.

Beier, S., Oderkerk, C., Bay, B., & Larsen, M. (2021). Increased letter spacing and greater letter width improve reading acuity in low vision readers. *Information Design Journal, 26*(1), 1–16. https://doi.org/10.1075/idj.19033.bei

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), Article e0152506. https://doi.org/10.1371/journal.pone.0152506

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

Brown, J. C., Goldstein, J. E., Chan, T. L., Massof, R., Ramulu, P., & Low Vision Research Network Study Group. (2014). Characterizing functional complaints in patients seeking outpatient low-vision services in the United States. *Ophthalmology, 121*(8), 1655–1662.

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1–28.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Grantee Submission, 76*(1), 1–32.

Catich, E. M. (1968). *The origin of the serif: Brush writing & Roman letters*. Catfish Press.

Chung, S. T., & Bernard, J.-B. (2018). Bolder print does not increase reading speed in people with central vision loss. *Vision Research, 153*, 98–104. https://doi.org/10.1016/j.visres.2018.10.012

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

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. *Transportation Research Record: Journal of the Transportation Research Board, 2624*, 1–8. https://doi.org/10.3141/2624-01

Frutiger, A. (1989). *Signs and symbols: Their design and meaning*. Van Nostrand Reinhold Company.

Galvin, K. (2014). Online and print inclusive design and legibility considerations. https://www.visionaustralia.org/services/digital-access/blog/12-03-2014/online-and-print-inclusive-design-and-legibility-considerations.

HHS Accessibility. (2020). *Microsoft Word accessibility reference*. Department of Health & Human Services Office of the Secretary Accessibility Program. https://www.hhs.gov/sites/default/files/os-a11y-word-reference.pdf.

Kitchel, J. E. (2011). *APH guidelines for print document design*. APH. https://www.aph.org/aph-guidelines-for-print-document-design/.

Legge, G. E., Rubin, G. S., Pelli, D. G., & Schleske, M. M. (1985). Psychophysics of reading—II.Low vision. *Vision Research, 25*(2), 253–265.

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research, 48*(5), 635–654. https://doi.org/10.1016/j.visres.2007.12.009

Macmillan, N., & Creelman, C. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

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.

Makowski, D., Ben-Shachar, M. S., Chen, S., & Lüdecke, D. (2019). Indices of effect existence and significance in the Bayesian framework. *Frontiers in Psychology, 10*, 2767.

Makowski, D., Lüdecke, D., & Ben-Shachar, M. (2020). *Modelbased: Estimation of model-based predictions, contrasts and means*.

Mansfield, J. S., Legge, G. E., & Bane, M. C. (1996). Psychophysics of reading. XV: Font effects in normal and low vision. *Investigative Ophthalmology & Visual Science, 37*(8), 1492–1501.

McLean, R. (1980). *The Thames and Hudson manual of typography*. Thames and Hudson.

Minakata, K., & Beier, S. (2022). The dispute about sans serif versus serif fonts: An interaction between the variables of serif and stroke contrast. *Acta Psychologica, 228* (103623). https://doi.org/10.1016/j.actpsy.2022.103623

Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience, 11*(10), 1129.

Prins, N., & Kingdom, F. (2018). Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes toolbox. *Frontiers in Psychology, 9*, 1250.

Rönnbäck, C., & Larsen, M. (2014). Macular sensitivity and fixation patterns in patients with autosomal dominant optic atrophy. *Investigative Ophthalmology & Visual Science, 55*(13), 6211-6211.

Rubin, G. S., Feely, M., Perera, S., Ekstrom, K., & Williamson, E. (2006). The effect of font and line width on reading speed in people with mild to moderate vision loss. *Ophthalmic and Physiological Optics, 26*(6), 545–554.

Russell, V. L. (2021). emmeans: Estimated marginal means, aka least-squares means (R package version 1.6.1). https://CRAN.R-project.org/package=emmeans.

Sawyer, B. D., Dobres, J., Chahine, N., & Reimer, B. (2017). In *, 61. The cost of cool: Typographic style legibility in reading at a glance* (pp. 833–837). https://doi.org/10.1177/1541931213601698

Schütt, H. H., Harmeling, S., Macke, J. H., & Wichmann, F. A. (2016). Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Research, 122*, 105–123.

Sperling, G. (1965). Temporal and spatial visual masking.I. Masking by impulse flashes. *JOSA, 55*(5), 541–559.

Stan Development Team. (2017). Stan: A C++ library for probability and sampling (version 2.14.0). URL http://mc-stan.org/.

Stan Modeling Language. (2017). User's guide and reference manual. http://mc-stan.org/manual.html.

Szpiro, S. F. A., Hashash, S., Zhao, Y., & Azenkot, S. (2016). In *How people with low vision access computing devices: Understanding challenges and opportunities* (pp. 171–180).

Tarita-Nistor, L., Lam, D., Brent, M. H., Steinbach, M. J., & González, E. G. (2013). Courier: A better font for reading with age-related macular degeneration. *Canadian Journal of Ophthalmology/Journal Canadien d'Ophtalmologie, 48*(1), 56–62. https://doi.org/10.1016/j.jcjo.2012.09.017

Tunold, S., Radianti, J., Gjøsæter, T., & Chen, W. (2019). In *Perceivability of map information for disaster situations for people with low vision* (pp. 342–352).

Unger, G. (2007). *While you're reading*. Mark Batty Publisher.

VisionAware. (2020). Tips for making print more readable. http://documents.jcahpo.org/documents/webinars/visionaware_pdfs/AFB_VisionAware_PrintReadable.pdf.

Watson, A. B. (2017). QUEST+: A general multidimensional Bayesian adaptive psychometric method. *Journal of Vision, 17*(3), 10. https://doi.org/10.1167/17.3.10

Watson, A. B., & Pelli, D. G. (1983). Quest: A Bayesian adaptive psychometric method. *Perception & Psychophysics, 33*(2), 113–120. https://doi.org/10.3758/BF03202828

Xiong, Y.-Z., Lorsung, E. A., Mansfield, J. S., Bigelow, C., & Legge, G. E. (2018). Fonts designed for macular degeneration: Impact on reading. *Investigative Ophthalmology & Visual Science, 59*(10), 4182–4189. https://doi.org/10.1167/iovs.18-24334