Reading and Learning Smartfonts

Danielle Bragg
University of Washington
Seattle, WA
dkbragg@cs.washington.edu

Shiri Azenkot
Cornell Tech
New York, NY
shiri.azenkot@cornell.edu

Adam Tauman Kalai
Microsoft Research
Cambridge, MA
adam.kalai@microsoft.com

ABSTRACT

As small displays on devices like smartwatches become increasingly common, many people have difficulty reading the text on these displays. Vision conditions like presbyopia that result in blurry near vision make reading small text particularly hard. We design multiple different scripts for displaying English text, legible at small sizes even when blurry, for small screens such as smartphones and smartwatches. These "smartfonts" redesign visual character presentations to improve the reading experience. Like cursive, Grade 1 Braille, and ordinary fonts, they preserve orthography and spelling. They have the potential to enable people to read more text comfortably on small screens, e.g., without reading glasses. To simulate presbyopia, we blur images and evaluate their legibility using paid crowdsourcing. We also evaluate the difficulty of learning to read smartfonts and observe a learnability/legibility tradeoff. Our most learnable smartfont can be read at roughly half the speed of Latin after two thousand practice sentences. It is also legible smaller than half the size of traditional Latin (i.e. "English") when blurry.

Author Keywords

Fonts; Reading; Learning; Scripts; Low-vision; Accessibility.

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation (e.g. HCI): User Interfaces

INTRODUCTION

Small text is inaccessible to many people with blurry vision. Nearly everyone's vision declines with age, resulting in presbyopia, the inevitable and irreversible decrease in the eyes' ability to focus [38, 1, 19]. Requiring people to put on reading glasses every time they look at a personal computing device prohibits seamless interaction with pervasive computing interfaces, yet, to our knowledge, this barrier has been ignored in the research literature. Even with corrected vision, out-of-focus text appears blurry, e.g., text on a smartphone navigation system when the user focuses on the path ahead. People with visual impairments, whose vision cannot be fully corrected with glasses, also have difficulty accessing text on larger screens. About 285 million people are visually impaired, 246 million of whom have low vision.[1] To improve the usability of devices with small displays, we argue that it is important to make the text on these displays more readable.

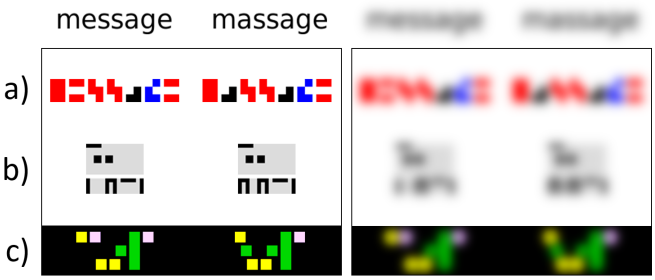


Figure 1: The words *message* and *massage* clear (left) and blurred (right) in our smartfonts: (a) Tricolor (b) Logobet (c) Polkabet (on black). Words are sized to have equal areas.

To address this usability challenge, we propose a novel approach: *smartfonts*, new scripts that improve legibility for various reading conditions. The Latin script that the Romans inscribed in stone millenia ago is not necessarily optimal for modern uses and technologies. Our research questions are: (a) what are effective script designs to improve these ancient twenty-six letters for display on small screens, and (b) how difficult would they be to learn? Regardless of the answers, smartfonts, like the Dvorak keyboard and Esperanto language, can help us better understand existing writing systems.

Smartfonts aim to improve the reading experience by replacing traditional letter shapes with more easily distinguished characters. We focus on English, but similar ideas can be applied to other languages. Smartfonts, demonstrated in Figure 1, comprise distinct renderings of the twenty-six letters so users can read text, letter for letter, without changes in orthography. In particular, they do *not* involve spelling changes or shortenings such as reading without vowels, though these tactics could be used in combination with smartfonts. Software can render text in smartfonts as easily as existing fonts. For instance, we have modified the firmware of a smartwatch to display everything in smartfonts (see Figure 2). Importantly, smartfonts could be used by any individual to display any digital text *without large-scale adoption*. Secondary potential benefits include increased privacy (e.g., for personal messages that may pop up), improved reading speed, aesthetics, personalization, and comfort (e.g., reduced fatigue).

[1] according to the World Health Organization http://www.who.int/mediacentre/factsheets/fs282/en/

UIST 2016, October 16 - 19, 2016, Tokyo, Japan

ACM 978-1-4503-4189-9/16/10...$15.00
DOI: http://dx.doi.org/10.1145/2984511.2984554

We designed five smartfonts to improve legibility of small, blurry text commonly viewed on small screens, and optimized our designs by iteratively evaluating them with crowd workers. Our smartfonts employ blocks (not merely strokes) of color to preserve character distinguishability under blurry reading conditions. Blurring replaces each pixel with a weighted mixture of surrounding pixels. Blocks of color survive largely unchanged because nearby pixels already share the same color. A diverse color palette also helps differentiate between confusable characters. These techniques produced smartfonts rendered with high fidelity at very small sizes, some perfectly renderable at only six pixels per letter.

We evaluated smartfont legibility with crowd workers who were asked to identify small, blurry strings. Because the general public does not know how to read smartfonts, evaluating their legibility is difficult. However, if a reader can make out an unfamiliar smartfont more clearly than a familiar font, it is likely that the smartfont is more legible, and the reader's experience will only improve with practice. Thus, our evaluation uses crowdsourcing to compare the identifiability of random strings in our smartfonts and in Latin characters. Blurry vision was simulated for the crowd by applying a Gaussian blur to the text, and size was varied. Our data suggests that it is possible to design smartfonts that, compared to the traditional Latin A-Z, are more legible when blurry, or equivalently can be displayed at smaller sizes with equal clarity. In particular, the smartfont in Figure 1a is legible smaller than half the size of Latin text when blurry, *without training*, by crowd members. This increased legibility could help people read smartphones or smartwatches at a glance, even without reading glasses.

We evaluated smartfont learnability through a website that supports practice and tracks user progress. Practice was provided through flashcards and fun yes/no practice questions in the smartfont. We found that our smartfonts, to varying degrees, can be read fluently with a reasonable amount of practice. We also found a learnability/legibility trade-off: certain scripts, especially ones that resemble the Latin alphabet, are easier to learn but perform worse with blur. Our Tricolor script offers a reasonable compromise in that it is relatively easy for many people to learn but also improves legibility.

Our key contributions are: (a) introducing the concept of smartfonts that radically redesign characters to improve certain aspects of the reading experience, (b) demonstrating how one can design and optimize (based on data) smartfonts for learnability and legibility under specific reading conditions, in our case, small blurry text, and (c) providing a methodology to evaluate legibility without teaching people to read fluently.

RELATED WORK

Our smartfont designs and evaluations were informed by knowledge about the reading process, character design research, and technological solutions to improve reading.

The Reading Process

Psycholinguistics provides a basis for understanding the reading process. When people read, their eyes dart from one fixed position to the next in jumps called "saccades." For the majority of the time, an experienced reader's eyes are stationary. The region from which a person's eyes can gather information during a fixation is referred to as the "perceptual span." There are many competing models of how people convert visual text into meaning. Word identification lies at the heart of many of these models. For example, dual-route (i.e., dual-process) models (e.g., [13]) propose that there are two ways that people recall word meanings: 1) by sounding out the word's phonemes and registering the word by its sound or 2) by converting the word's visual representation directly to vocabulary. In competition with dual-route models are single-route models, which are modeled after neural networks, and propose that lexicographical, phonological, and orthographic units exist in the brain; in between lie hidden layers of computation that refine over time with experience. Grounded by the importance of word identification in psycholinguistics, we use string identification as the basis for our evaluation of smartfont legibility.

Figure 2: A smartwatch displaying an SMS. The sender is oblivious to the fact that the SMS is read in a smartfont.

Individual letter identification is another important part of reading in many models. For example, in the phoneme-based route of dual-route models, sequential letter identification is thought to play an important role in sounding out words [45]. The number of letters that fit in the perceptual span has been linked to reading speed [30], providing further evidence of the importance of individual letter identification. The reading process is still largely not understood, and both high-level cognitive processes and low-level physical mechanisms involved are active research areas [43]. The posited importance of letter identification in psycholinguistic reading models informs our use of letter identification tasks in designing smartfonts.

Alphabet Character Design

Traditional alphabets have evolved to support both the visual reading and manual writing processes. Some scripts, such as the Korean alphabet Hangul, are *featural* meaning the shapes of the letters encode phonological features of the sounds they represent. *Mature* scripts, say those that have been in use for over 350 years, have been found to have many fewer mirror-image letter pairs, such as the lower-case Latin pair b/d, than younger scripts [51]. Unlike traditional alphabets, our character sets are designed for display on screens and thus are freed from the constraint that they be easily written by hand. Removing this writing constraint allows us to optimize the reading experience beyond the experience afforded by traditional alphabets. We are not the first to create novel scripts; motivated by various factors, artists and hobbyists have constructed creative scripts,[2] though we are unaware of any rigorous studies of their legibility or learnability, which we provide.

[2] A set of constructed scripts can be found at http://omniglot.com.

Smartfonts are conceptually similar to Braille, a tactile writing system for people who are blind or have very low vision. In Grade 1 Braille, each English letter is represented by a collection of raised dots in a 2x3 grid. Because Braille completely redesigns character shapes, it faced fierce opponents who thought it was too radical, that it would be unhelpful or detrimental to the blind community, and even attempted to ban it [21]. Despite initial resistance, Braille was eventually accepted. It has since greatly benefited many people, and is linked to higher rates of employment, education, and financial stability.[41] Like Braille, smartfonts completely redesign the written form of each English letter to improve legibility. There is evidence that tactile sensing of letters is very similar to visual sensing of letters subject to a low-pass spatial filter (i.e. blurry letters) [35]. Just as Braille's 2x3 structure is more legible than embossed characters to the finger [34], we expect Braille's 2x3 structure to be more easily discernible to readers with blurry vision. Consequently, we adopt Braille's 2x3 structure in some of our smartfonts. Our hope is that smartfonts, too, will benefit users by making text more accessible.

Font design has been shown to strongly impact the reading experience. Various letter shape properties, including stroke width or boldness (e.g. [6]) and serifs (e.g. [3]), have been shown to impact legibility. Contrast level in both luminance [32] and color [31] can impact both readability and aesthetic appeal. Certain text/background color combinations are known to be more readable and pleasing than others [25, 22, 42]. Prior studies on color-grapheme synethesia, where people have strong associations between letters and colors (see, e.g., [12]), have shown that reading books with colored letters suffices to passively learn and create strong perceptual associations between letters and colors. In her dissertation, Bessemans explored font design for children who are low-vision and just learning to read [9]. Custom font designs can also improve reading for people with dyslexia [40] and pilots in the cockpit [48]. While typography research informs our work, we are not addressing the question of traditional legibility. Typography refers to the stylization of existing written letters resulting in fonts with varied legibility and personality; we propose redesigning visual letter forms from scratch to create entirely new character sets, which we name "smartfonts."

Evaluating Fonts

Existing techniques for evaluating text readability and legibility are not readily applicable to smartfonts. These evaluation methods rely on participants' ability to read the script being tested, but reading smartfonts requires training. Prior studies typically ask participants to read text and then complete a task based on that text. For example, participants might read paragraphs of text in different fonts and then answer basic comprehension tests (e.g. [25]). Reading time and comprehension level serve as metrics for readability. An alternate setup consists of presenting a paragraph of text with individual word substitutions (e.g. [8, 14, 7]). The number of word substitutions detected measures readability or legibility. Other tests involve showing a single word or pseudoword, and asking the person whether the word they saw was a real word (e.g. [18]). Accuracy in distinguishing words from non-words in relation to word display time determines legibility. These tests would dramatically favor traditional Latin characters due to the participants' experience in reading them.

A motivating starting point for our smartfont evaluation techniques is the work on human perception by Pelli et al. [39]. This work compares the "efficiency" of letter identification across traditional and made-up alphabets. Efficiency was measured by how well individual letters could be identified in the presence of random noise, which is different but possibly related to blur. They also found that a few thousand training examples sufficed to teach someone to identify unfamiliar letters fluently. Pelli's methods for evaluating character distinguishability and learnability inform our smartfont legibility and learability test designs.

Crowdsourcing has been shown to yield reliable results in perception studies, and has been used by many researchers. For instance, Demiralp et al. explored the use of crowdsourcing to evaluate the perceptual similarity of different shapes and colors, and developed perceptual kernels to quantify crowd-learned similarity [17]. They found crowdsourcing to be an inexpensive, rapid, and efficient means to gather data on human perception. Heer and Bostock demonstrated the viability of Mechanical Turk, a popular crowdsourcing platform, for evaluating visualization graphics by replicating previous results and running new studies that produced new insights [26]. In our work, we use Mechanical Turk to reach a wide pool of potential smartfont users and evaluate smartfonts.

Technological Approaches to Improve Reading

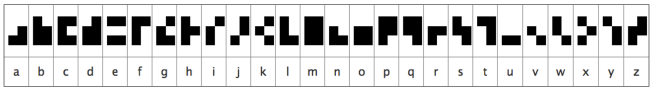
HCI techniques proposed to improve digital reading could be combined naturally with smartfonts, as they build off of existing letter forms. Such techniques include RSVP[3] [27], leading [24], Froggy [52], ClearType [20], and visual syntactic text formatting [49]. Interactive teaching techniques could also help people learn smartfonts, such as software that gradually teaches a language by introducing new words over time [46].

A number of factors affect the adoption of new technologies like smartfonts. Even with models like the Technology Acceptance Model [16], predicting adoption is difficult. A notable source of contention is the debate about the Dvorak keyboard's adoption "failure": Economists cited early studies claiming that it is 20-40% faster than the QWERTY keyboard, and thus the low adoption rate of the significantly "superior" Dvorak keyboard proves how difficult it is to change behaviors [15], while later studies found Dvorak to be only 2% faster [33]. Other input techniques garnered higher adoption on PDAs and smartphones [23, 53], highly influenced by user preference. Similarly, early smartfont adopters would likely be those who benefit most and find them easiest (or most enjoyable) to learn.

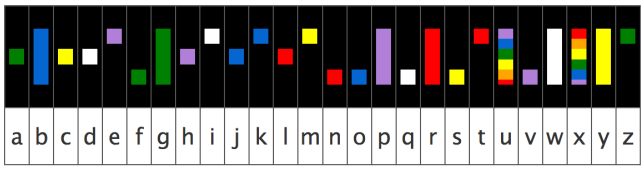
OUR SMARTFONTS

We designed three initial smartfonts to be easily legible at small sizes and out of focus. We leveraged three main techniques: 1) using blocky shapes known to be resilient to blur, 2) using color to distinguish between characters, and 3) radically reducing the space between adjacent characters. Visibraille

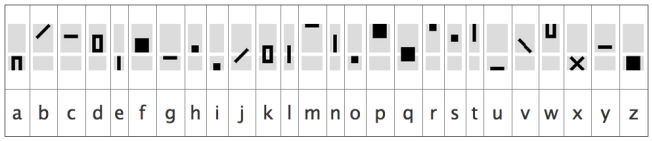
[3]RSVP has recently received attention due to http://spritzinc.com and its inclusion on the Microsoft Band smartwatch.



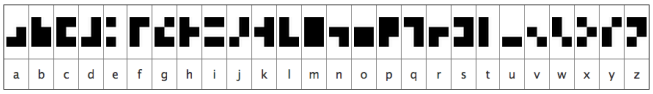
(a) Visibraille alphabet



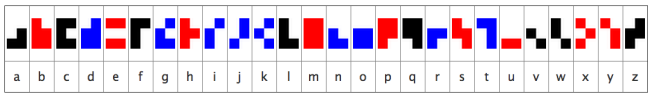
(b) Polkabet alphabet



(c) Logobet alphabet



(d) Visibraille 2 alphabet



(e) Tricolor alphabet

Figure 3: Our smartfonts designed to improve legibility of small, blurry text.

and Polkabet are designed to support blurry character distinguishability, while Logobet is designed to minimize text area.

Visibraille

Our smartfont Visibraille is modeled after Braille. Braille's 2x3 structure of dots has proven to be more tangible than embossed traditional letter forms [34]. Because fingers have low spatial resolution, what is "seen" with the fingers resembles what is seen in a blurry image [35]. Consequently, we expect Braille's 2x3 structure to be more easily discernible than traditional characters for blurry vision. This expectation is supported by empirical evidence showing that 2x3 characters are more visually distinguishable than traditional characters; out of a range of established and made-up alphabets, 2x3 characters have been shown to be most visually recognizable [39].

Visibraille, shown in Figure 3a, maps 26 2x3 blocks onto the 26 letters of the English alphabet. We selected the 3×2 blocks to be similar in shape to the English characters they represent. Because of its simple design and similarity to Latin characters, we expect this smartfont to be both legible and learnable.

Polkabet

Polkabet adds color to the design space to increase distinguishability. It is designed for a black background, which is common to personal devices like the smartwatch. Contrast in color and luminance help the eyes distinguish elements visually. They impact text readability (e.g. [32, 31]), and help data visualizations distinguish between data (e.g. [50]). To support distinguishability, we chose five colors, shown in Figure 3b, spaced out in hue and to have a strong luminance contrast with the black background. We then adjusted them based on participant feedback, since colors vary greatly across displays. Similarly, colors could be personalized and tailored for various types of color blindness. Rather than adding an additional color, we included two rainbows to minimize the number and confusabililty of colors.

Polkabet's characters are solid squares and rectangles. Perimetric complexity (the ratio of character perimeter to "ink" area) has been found to correlate strongly with people's efficiency at identifying characters, with less complex characters identified more efficiently [39]. Because solid squares and rectangles have low perimetric complexity, we expect these shapes to be highly discernible. Characters with low perimetric complexity are also resilient to blur. When an image is blurry or out of focus, each pixel appears to be a mixture of nearby pixels. Solid blocks are highly robust to this type of blurring because many nearby pixels share the same color.

Like Braille, Polkabet uses dot position to distinguish certain letters. While many letters are similar and hence possibly confusable, the similarities are vertical, motivated by the preference for vertically reflected letter pairs such as p-b over horizontally reflected letter pairs such as b-d across natural language [51].

Figure 4: Mnemonics for Polkabet's small square characters.

To facilitate learnability, we used mnemonics to match letters to colors, shown in Figure 4. If a character uses a small square of color, the reader can think of the associated item from Figure 4. The first letter of that item is the letter that the character represents. Squares of color at the top are associated with foods, middle squares with animals, and bottom squares with miscellaneous items. For example, suppose a reader forgets what a red square at the top of the line means. He/she would think, "This character uses a red square at the top, so think of a red food... Tomato! 'Tomato' starts with *t*, so that's a *t*!" Tall blocks of color represent the first letter of that color.[4]

Logobet

Logobet aims to minimize text area by reducing the spacing between letters, a typographic process called "kerning." In a proportional font, the space allotted to characters depends on their size. For example, an *m* is allotted more horizontal space than an *l*. However, reducing the spacing between pairs of

[4] X, which stands for "Rainbow", and U, which stands for "Upside-down Rainbow", are exceptions.

(a) The alphabet in Logobet characters without kerning

(b) The alphabet in Logobet with kerning

Figure 5: Example of aggressive kerning with the alphabet.

characters can be desirable. For example, a capital *T* allows a short subsequent letter, such as an *o*, to shift left under the *T*'s umbrella, as in the word *Tomato*. Because Logobet employs extreme kerning to condense strings, it visually resembles a character-based logography, such as Chinese.

Logobet's characters were chosen with vertical positioning, like Polkabet, to avoid left-right mirror pairs. Horizontal, vertical, and diagonal lines and circles and semi-circles are known to be easily distinguished [47]. Logobet's characters, illustrated in Figure 3c, consist of these shapes except that boxes were used in place of circles to avoid pixelation effects. Logobet maximizes kerning by allowing characters to shift entirely underneath preceding characters. Each character occupies a fraction of the row's height, and subsequent letters that are strictly lower slide underneath. Letters are ordered first top-to-bottom, then left-to-right, to ensure that every printed Logobet string corresponds to a unique letter sequence. For example, Figure 5 shows Logobet's kerning on the alphabet.

OPTIMIZATIONS

We explored optimizing color and shape over 2x3 characters. We present two optimized fonts: Visibraille 2, whose 2x3 blocks are chosen to be minimally confusable; and Tricolor, which adds color to Visibraille's Latin-esque 2x3 characters. We generated a crowdsourced confusion matrix of 2x3 shapes which determined Visibraille 2's shapes and Tricolor's colors.

2x3 Character Confusability Study

We determined the confusability of 2x3 characters[5] using paid crowdsourcing on Amazon Mechanical Turk.[6] Our study is modeled after a rich history of studies on character recognition and legibility, where a predominant technique is to collect confusion information on characters presented in conditions that obscure distinguishability (e.g., [2, 4, 10, 44, 28, 35]), and brings this tradition to the crowd. Our experimental setup mimicked a Snellen eye chart test, a standard eye exam test.

Workers were shown rows of characters at decreasing sizes, specified in Figure 6. Each row was shown separately, and contained 1-9 characters. Participants transcribed the character(s) using a virtual keyboard of 7 characters. The target characters were chosen randomly with replacement from the keyboard's characters, which were also chosen randomly.

In total, we collected 4022 evaluations from 548 people. Each character was shown 379-500 times (mean 442.1). Each pair

[5]Of the $2^6 = 64$ possible 2x3 characters, we considered a subset of 42 characters to reduce labor. In particular, two configurations were considered to have the same shape and likely to be confused if the block patterns were translations of one another.

[6]http://www.mturk.com

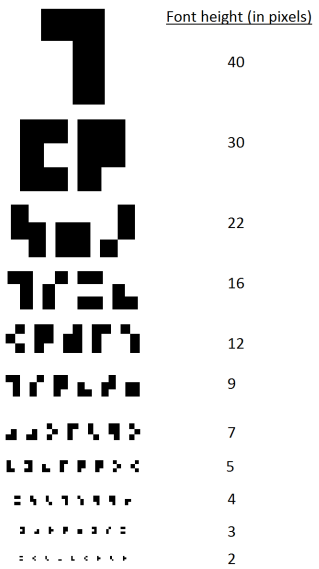


Figure 6: Our confusion matrix was generated by testing a series of rows of random characters, like a Snellen chart.

of characters appeared together on the virtual keyboard at least 33 times (mean 64.7). Since experiments were conducted remotely through web browsers, we did not control for display conditions or viewing factors such as retinal angle. However, this enables us to assess the confusability of our shapes "in the wild," across a wide variety of display types and people. To minimize pixelation artifacts, participants were instructed to keep their web browsers at the default 100% zoom.

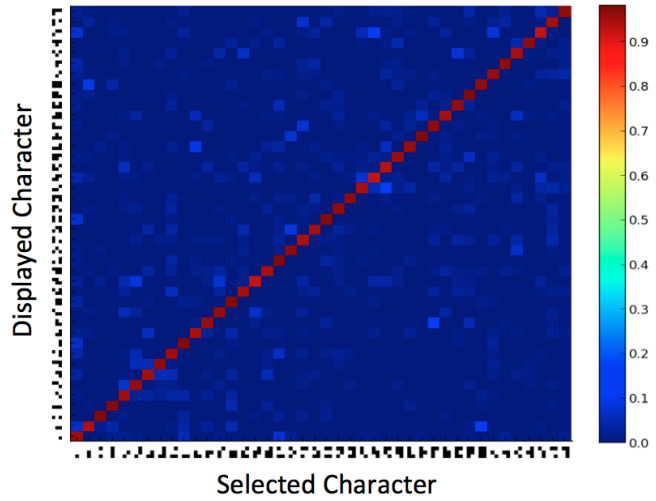


Figure 7: 2x3 character confusion matrix.

Our confusion matrix C, shown in Figure 7, consists of the confusability score c_{ij} between each pair of shapes i and j. Here, c_{ij} denotes the fraction of times shape j was transcribed when i was the target, out of the total times j was available as a transcription choice for i. The confusability of a shape with

itself c_{ii} is similarly defined to be the fraction of times that i was transcribed when i was shown.

Visibraille 2

Finding the 26 most empirically distinct (least "confusable") characters for an optimized smartfont was modeled as choosing the set S of size 26 so as to minimize $\sum_{i,j \in S, i \neq j} c_{ij}$. This problem is NP-hard, but we used a branch-and-bound search to quickly find the exact optimum among the $\binom{42}{26} \approx 10^{11}$ possible solutions. The set found by our branch-and-bound algorithm is visualized in Figure 8. The 26 selected letters, shown in black, minimize the sum of edge weights between nodes in the selected letters. We mapped these 26 shapes to the Latin A-Z to create Visibraille 2, as illustrated in Figure 3d. The mapping was chosen to ease learning by assigned shapes to Latin characters they resembled (upper- or lower-case[7]).

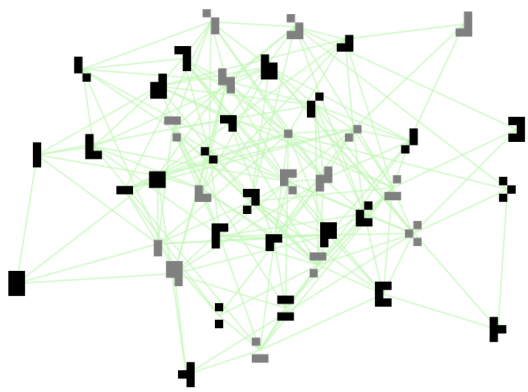


Figure 8: The 26 selected shapes (black) of the 42 considered. Edges denote highly confusable shapes. Although the confusion matrix is high-dimensional (as measured by eigenvalues), D3's force-directed graph layout [11] displays many confusable pairs near one another.

Tricolor

We also used the confusion matrix to identify commonly confused characters from our initial smartfont Visibraille. We color each pair of confused characters differently to introduce a new smartfont Tricolor. Because its characters closely resemble the Latin A-Z, we hypothesized that it would be easy to learn and remember. Because it uses both shape and color to distinguish between characters, we hypothesized that it would also be highly legible at small sizes and blurry.

To assign characters optimally to three distinct colors, we solve the following problem: partition the set of letters into three disjoint sets $S = S_1 \cup S_2 \cup S_3$ so as to minimize $\sum_{k=1}^{3} \sum_{i,j \in S_k, i \neq j} c_{ij}$. This NP-hard search over $\approx 4 \times 10^{11}$ partitions succumbed easily to exact optimization again using branch-and-bound search. The selection, visualized in Figure 9, minimizes intra-color edge weights (between characters of the same color), or equivalently maximizes inter-color edge weights.

[7]For simplicity, we focus on only twenty-six letters, while the Latin font has 52 letters if one includes both cases. An additional twenty-six letters could be added similarly, or like Braille, a single symbol could be incorporated to indicate case.

The Tricolor alphabet is shown in Figure 3e. It adopts the easily-learned Latin-esque shapes of Visibraille, and assigns 3 colors to help distinguish easily confused letters. Because pairs of "mirror images" were commonly confused in our study (in accordance with prior work [51]), our optimization assigned different colors to each such pair. For example, "a" and "n" are mirror images of one another and are colored black and blue respectively. Other highly confusable pairs, such as "l" and "n," are also colored differently.

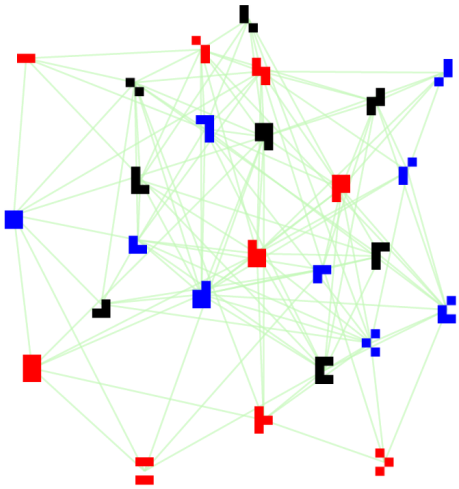


Figure 9: A D3 [11] force-directed layout of Tricolor attempts to locate similar pairs of letters near one another and also illustrates the colors chosen by our optimization algorithm.

LEGIBILITY EVALUATION

Evaluating the legibility of new smartfonts is difficult when nobody knows how to read them. Evaluating legibility compared to Latin is further complicated by the test population's lengthy experience reading and identifying Latin characters. Even with training, we cannot reasonably expect our test population to accumulate a comparable amount of experience with a new smartfont over the course of a study. Our evaluation method does not require training people to read smartfonts.

Experimental Setup

Our legibility experiments consist of showing participants a target string and asking them to select the matching string from a list, as shown in Figure 10. The targets were random strings of length five, roughly the average word length for English. We chose strings as visual stimuli based on the theory that words are preferable to both individual letters and sentences for evaluating visual acuity [5]. Individual letters are inappropriate since they cannot blend with neighboring characters as longer strings do, and longer text is not ideal because individual factors unrelated to visual clarity contribute to reading comprehension (like inference from surrounding words). Each question came with four possible answer choices. One of the answer choices was the same five-letter string as the target. The other three matched four out of five target characters, with one random replacement.

We used a within-subject design, with each participant answering questions for a single smartfont and for Latin. Latin text

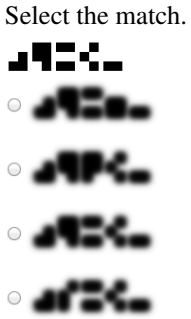


Figure 10: Sample legibility task with smartfont Visibraille.

was presented in Helvetica. The target image was presented at decreasing sizes, with three questions at each size for both fonts. Blur was manipulated with a Gaussian filter, which replaces each pixel with a weighted average of nearby pixels. A large radius creates highly distorted images and mimics severe presbyopia, while a small radius leaves images largely intact. The blur radius was fixed throughout each experiment. We first scaled text then applied blur. This simulates the experience of reading small, blurry text (e.g., a presbyope reading text on a smartphone), where the stimulus is fixed and clear and the eyes essentially apply a blurry filter.

Simulating blurry vision allowed us to crowdsource our evaluation through Amazon Mechanical Turk.[8] Instead of screening for specific blurry vision conditions, we blurred text and asked the crowd to "read" it. Low visual acuity and other problems early in visual processing are in some sense transformations on the stimulus. Whether blurring occurs on the screen or in the visual system, the perceptual effect is similar. Consequently, there is a precedent for blurring images to simulate blurry vision in reading studies (e.g., [28]).

We ran two experiments. Our first compares all five smartfonts at a fixed blur of 3.5. We had 154 participants (69 female, 81 male, 4 other). Ages ranged 18-72 (mean 35). 32 evaluated Polkabet; 30 evaluated Visibraille 2; 31 evaluated Visibraille; 33 evaluated Tricolor; and 28 evaluated Logobet. 69 were wearing glasses during the study, and 85 were not. Our second experiment compares Tricolor at three different blurs. It involved 104 participants (37 female, 64 male, 3 other). Ages ranged 20-69 (mean 36). Of these, 36 saw a blur of 2.5; 33 saw a blur of 3.5; and 35 saw a blur of 4.5. 41 participants wore glasses during the study, and 63 did not. Participants received $1.50. Workers quit and were paid when the size became too small, so few answered all questions. Workers had at least a HIT Approval Rate of 97% and 1000 approved HITs.

Legibility Results

It is not obvious how to compare the legibility of smartfonts, especially with experiments performed "in the wild" where users have varied screens and software. To address this, we compare, for each participant, the smallest size they can read Latin text to the smallest size they can read their smartfont. Since each participant viewed Latin and a single smartfont, this enabled us to compare, on an individual-by-individual basis, the smartfont and Latin letters under the same conditions.

Determining a metric for font size applicable to diverse scripts is challenging. Vision scientists typically use visual angle between the bottom of the text, the viewer's eye, and the top of the text, while typographers prefer the physical print size of characters [29]. Variance in character height and width within fonts further complicates defining size. To fairly compare different scripts, we use *text area* as our metric. Text area includes the white space between and around characters required to render the text that cannot be occupied by surrounding text.

We define a *Minimal Reading Area* (MRA) for font f, MRA_f, which is specific to the participant (and blur). Note that in our experiment we asked three questions for each font at each size. As we decrease size, we say the participant fails to read at the first size where they make a majority of errors (2 out of 3). We define the MRA to be the just-larger size used before failure. Although participants were asked to attempt reading at smaller sizes, this further data was not used in the analysis because it typically reflected random guessing. We also exclude data from participants who failed to read at the largest size, since they likely misunderstood the instructions or were guessing. It is convenient to consider the log-MRA, shown in Figure 11, since a constant difference in log-MRA reflects a constant factor change in legible size.

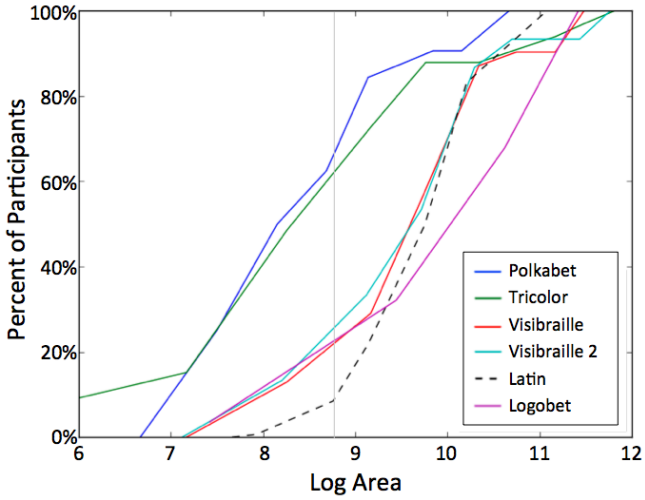


Figure 11: The (smoothed) empirical distribution function of log-MRA (at blur of radius 3.5 pixels) for the fonts. y is the fraction of participants whose log-MRA was larger than x.

To get some intuition for the meaning of text area, a Facebook post on a Chrome desktop[9] web browser today appears in a font whose full ascender-to-descender height is 13 pixels, which corresponds to a log area 8.75 (see the vertical line in Figure 11) in our experiments. With the interpolation in Figure 11, this suggests that only 8% of participants could read this size text, at our blur, in the Latin font while over 60% of the participants could read Polkabet and Tricolor fonts.

[8] http://www.mturk.com

[9] Most of our participants were using desktop, not mobile, browsers. See http://facebook.com and http://google.com/chrome.

Our definition of MRA focuses on font sizes large enough that users are most often correct, based on reading research which shows that users prefer and read faster at font sizes for which they can readily discern letters (e.g [14]).

To quantify each smartfont's performance by a single number bounded by a confidence interval, we define the *log-score* (LS), for each experiment to be the log of the ratio of the MRA for Latin to the MRA for each smartfont f, or equivalently,

$$\mathrm{LS}_{Latin,f} = \lg \frac{\mathrm{MRA}_{Latin}}{\mathrm{MRA}_f} = \lg(\mathrm{MRA}_{Latin}) - \lg(\mathrm{MRA}_f),$$

where lg denotes base 2 logarithm. A log-score of 0 means the participant read the smartfont at the same size as Latin; 1 means they read the smartfont at half the size; 2 means they read the smartfont 4 times smaller; etc. Note that our experiment is inherently one-sided: upper bounds on log-score do not bound the legibility of the smartfont *after training*.

A histogram of the log-scores for Tricolor is displayed in Figure 12. 26 of the 33 participants (79%) had positive log-scores, meaning that they "read" the smartfont at a smaller size than Latin, and 18 of the 33 (55%) had log-scores greater than 1, meaning they "read" the smartfont at least half as small as Latin. The sample mean log-score was 1.28. The wide variance in this histogram means that some users might benefit significantly more than others from adoption.

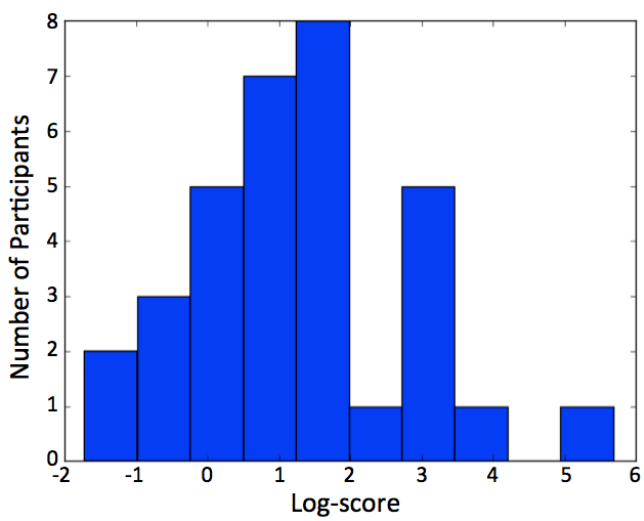


Figure 12: Histogram of log-scores for Tricolor with blur radius 3.5 pixels.

Table 1 displays confidence interval bounds for our smartfonts. For *simultaneous* 95% post-hoc confidence intervals for five smartfonts, we choose what would normally be 99% confidence intervals bounding each (the union bound on the 1% failure probability of each estimate then implies 95% confidence). Since our test is inherently one-sided, as mentioned, we use simultaneous one-sided confidence intervals, based on mean and standard deviation. Only Tricolor and Polkabet's confidence intervals are entirely positive, suggesting particularly strong legibility for these smartfonts.

Smartfont	CI lower-bound	Mean log-score
Polkabet	0.78	1.30
Tricolor	0.62	1.28
Visibraille	-0.23	0.14
Visibraille 2	-0.32	0.14
Logobet	-0.56	-1.03

Table 1: Mean and 95% simultaneous (one-sided) confidence interval lower-bounds.

To see how performance would vary as we change the blur parameter, we compared Tricolor versus Latin at three different blur radii. The results at radii 2.5 pixels, 3.5 pixels, and 4.5 pixels, were all greater than 0 with statistical significance, though the differences were not statistically significant. The mean log-scores of 1.17, 1.28, and 1.41, respectively, suggest a possible increasing trend.

LEARNABILITY EVALUATION

In order for our smartfonts to be usable, they must be learnable. To evaluate learnability, we designed an online learning system and tracked partcipants' progess. The learning site assigned each visitor a single smartfont. It provided a tutorial about the smartfont, yes/no practice questions in the smartfont, and flashcards for drilling the meaning of individual characters.

Learning Site Design

The site welcomed visitors with a brief smartfont tutorial. The tutorial presented 1) the mapping between the smartfont characters and Latin (i.e. "English") characters, 2) a description of the smartfont's organization, and 3) examples of words in the smartfont with their Latin equivalents. The site's welcome page provided the tutorial and a chart of the participant's performance over time for self-tracking. Participants could return to the welcome page at any time.

The site provided short yes/no practice questions to help participants learn their smartfont. The questions were generated via crowdsourcing, consisting of questions from MindPixel [37] and questions we gathered from Amazon Mechanical Turk workers. We screened the questions for inappropriate content. In total, we used 2739 questions: 1245 with an answer of "yes" and 1494 with an answer of "no." The questions were generally fun and entertaining. Examples include "Is the moon out at night?" and "Are you a celery?".

Figure 13: Sample practice question for smartfont Logobet. (*Do fish wear clothing? Yes/No*)

Practice questions were displayed with smartfont characters, as shown in Figure 13. After receiving an answer from the user, the site showed the question in Latin characters and gave feedback on correctness. While answering practice questions, participants could reference a cheatsheet of the mapping

between smartftont and Latin characters. The Polkabet cheatsheet also provided mnemonics. To view the cheatsheet, a user could click a link above the question and the cheatsheet would overlay the practice question. This design forced participants to use their memory when answering questions, rather than relying entirely on the cheatsheet to look up each character.

To further help participants memorize their smartfont, we made flashcards available throughout the study. Each flashcard presented a single smartfont character, and quizzed the participant on its Latin equivalent. Mistaken characters were repeated until the participant got them right.

Experimental Setup

We recruited 23 people to use our smartfont learning site through Amazon's Mechanical Turk platform.[10] Each participant was assigned randomly to a single font: 8 to Polkabet, 6 to Tricolor, and 9 to Logobet. Varying numbers for each font are due to participant dropout during the study. Participants chose how long they spent on our site. They typically spent 2-3 hours per day on our site over the course of about a week. Our study workers received $5 for their first 10 questions, $0.05 per question for the next 3333, and an extra $50 if they completed 3333 by the study end date. Workers had at least a HIT Approval Rate of 97% and 1000 approved HITs.

Participants set up accounts on the learning site so that they could log back in to continue learning and we could track their progress. Participants were compensated for the yes/no practice questions that they answered, but were free to use the flashcards, cheatsheet, or tutorial at any time. We recorded time and accuracy in answering the yes/no questions. One in every 10 yes/no questions was displayed in Latin characters for baseline comparison. We also recorded their use of the cheatsheet and flashcards throughout the study. Participants were free to provide open-ended feedback through a form on the site at any point during the study.

Learnability Results

We evaluated learning in terms of speed and accuracy in reading and answering the practice questions. To evaluate speed, we calculated the ratio of the time it took them to answer each smartfont practice question to the average time it took them to answer the Latin control questions. A value of 1 means that it took the same time to answer smartfont questions as Latin ones, a value of 2 means it took twice as long, and so on.

All participants held over 95% accuracy in answering the encoded questions, so they were not guessing. Practice did not increase accuracy. Average accuracies were: Polkabet 97.8%, Logobet 97.3%, Tricolor 98.2%, Latin 98.9%. The difference between each smartfont and Latin was statistically significant ($p<0.001$, Kruskal-Wallis). Accuracy and response time were weakly correlated ($r=-0.0030$, $p=0.2534$, Pearson).

Figure 14 shows the general trends across our smartfonts. Tricolor exhibits the easiest learning curve, followed by Logobet and then Polkabet. After $2,000$ questions, participants learning Tricolor were reading a median of 2.1 times slower than they did in Latin; participants learning Logobet were 5.2 times slower; and participants learning Polkabet were 6.7 times slower. We ran an unpaired t-test to determine whether the differences in response time across fonts was significant after 2000 questions. We found a statistically significant difference between each pair of fonts: Polkabet and Logobet ($t(8498) = 10.6623, p < 0.0001$), Polkabet and Tricolor ($t(7998) = 4.0640, p < 0.0001$), and Logobet and Tricolor ($t(8498) = 2.6588, p < 0.008$).

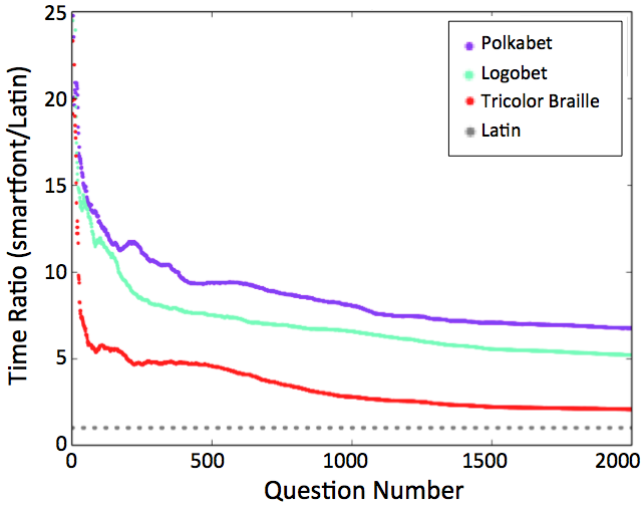


Figure 14: Smartfont response time, normalized by individual average Latin time. Each point is the median of a sliding window of averages across participants to remove outliers.

There was some variation in learning curves between individual participants. Two participants learned their smartfonts, Tricolor and Logobet, extremely quickly. They became as fast at reading the smartfont as they were at Latin after only around $1,000$ questions. Their quick learning curves show that while learning smartfonts might be somewhat challenging for most people, it is quite natural for some. Those with an affinity for learning smartfonts have a low barrier to start reading and benefiting from smartfonts. The variation in learning curves between individuals also suggests the benefit of personalization. Some may prefer to learn smartfonts that would challenge others, and these preferences may be individual.

Adding color to Tricolor's characters appeared to support learning. Since its characters' shapes are unique and resemble the Latin alphabet, one might be concerned that people learn to read it ignoring color. This did not seem to be the case. We found that participants remembered the colors of common words, not just the letter shapes. In a post-test three days after the site closed, seven readers of Tricolor were asked to identify the coloring of five common words (like "the") on multiple-choice questions with four choices (three random colorings of the same shaped letters). The aggregate accuracy was 28/35 (80%) over these four questions, strongly indicating that they had remembered at least some of the colors.

Participants used both the flashcard and cheatsheet as they learned our smartfonts. Participants learning Tricolor made more use of the learning resources than participants learning

[10] http://www.mturk.com

Polkabet or Logobet. It is possible that participants learning Tricolor relied more on the learning resources because their font tutorial did not include additional information beyond its mapping to Latin characters. We provided mnemonics for Polkabet, which likely helped Polkabet participants recall more characters independently, if slowly. Similarly, we provided a lengthy tutorial for Logobet detailing the character organization, that likely helped participants remember the character representations. Because of its relative simplicity, Tricolor had no mnemonics or details about font organization.

We gave participants the opportunity to provide open-ended feedback. Their responses indicated that they largely enjoyed learning and reading our fonts. The majority remarked that their experience was "fun." Several compared the reading smartfonts to solving puzzles. One even wrote, "someone should find a way to turn this into an Android game." Participants also noted their progress. One found it, "super hard in the beginning but on the last couple I actually was reading them as though I was seeing the letters." At the end of the experiment, one participant contacted us, asking if they could continue using our site to practice their font. Coupled with our learning curves, the participants' positive reflections suggest that people can enjoyably learn to read smartfonts fluently.

DISCUSSION AND FUTURE WORK

There are several limitations to the current work. First, we did not control for screen type, screen resolution and distance between the viewer and screen. Moving away from a controlled laboratory setting allowed us to use crowdsourcing for rapid experimentation. Reproducing our results in a lab with users with presbyopia would be beneficial. Second, we do not currently offer users who learn a smartfont the ability to use it on their devices, which could be crucial to adoption.

We would like to explore benefits of smartfonts beyond improved legibility. Smartfonts can provide *privacy* by visually encrypting text. "Substitution ciphers" which encrypt text by replacing each letter with a symbol, have been used by da Vinci in mirror-writing, by Union prisoners in the Civil War, and by children in games and journals. Privacy can be especially valuable on smartwatches, where embarrassing personal communications may appear without warning, visible to anyone sufficiently close. Smartfonts might also affect cost or durability. For instance, the seven-segment digit display common to digital alarm clocks and other electronics is cheaper and has fewer pieces that may fail than a high-resolution screen. Smartfonts could similarly improve printing or hardware costs.

We limited the design space to create each of our smartfonts. A limited design space makes it difficult to ensure that Latin text is always unambiguously recoverable from smartfont text, and that all Latin characters are covered by the smartfont. Possible solutions include 1) relaxing design space constraints, 2) using different design spaces for different character sets (e.g. letters vs. digits vs. punctuation), 3) creating compound characters and 4) adding indicator characters at the start of different modes (e.g. text vs. numbers). Braille uses 3) and 4) with a 63-character design space. Color in smartfonts is especially suited for messaging, where text is typically monocolor, but raises interesting questions in broader graphic design where color serves important functions. We plan to further explore the smartfont design space, including expanding smartfonts to characters beyond the alphabet.

In the future, smartfonts could be tailored to an individual's eyesight or display screen. Each person is unique, and a wide variety of vision conditions exist. We imagine a system that evaluates a person's vision and generates optimized smartfonts on-the-fly. Such a system would require learning a model of how vision relates to script readability. Just as many Southeast Asian scripts have rounded letters because straight lines would tear the palm leaves on which they were written [36], smartfonts could also be tailored to their display screens.

Smartfonts could also be generalized to other character systems besides Latin. For example, we can develop smartfonts for the Hebrew alphabet or Chinese characters. Some East Asian scripts are read top-to-bottom, so any smartfont involving kerning would need to support combining adjacent characters vertically. The size of character sets can also vary enormously. For example, there are over 50,000 Chinese characters. A smartfont for such a large character set would likely need to take advantage of language or character structure.

CONCLUSION

In this work, we introduced smartfonts, scripts that completely redesign the written alphabet with the purpose of improving the reading experience. We do not claim to have created the best smartfonts or even optimal smartfonts for reading blurry text, but we have hopefully demonstrated that it is possible to improve over the millenia-old letters in use today.

We also presented experimental designs for evaluating 1) the legibility and 2) the learnability of smartfonts under various reading conditions. We addressed a challenge in the design-evaluation loop through a novel experimental setup that allowed us to evaluate legibility under various reading conditions by users *without requiring fluency*. We evaluated learnability by teaching smartfonts through an online system that provided a tutorial, encoded simple yes/no questions, and tracked yes/no question response times.

Smartfonts have many potential benefits: improved legibility for various reading conditions, privacy, aesthetics, and intellectual challenge (though easier than learning a new script together with a new language). Allowing interested users to opt-in, smartfonts do not require alphabet reform. They do not require new hardware or software, and are deployable on existing platforms. As our experiments showed, people can learn them with a reasonable amount of practice.

As we move into an age of personalized electronics, screen sizes shrink and enabling people to read comfortably on small screens becomes increasingly important. Similarly, making it possible for people to read text that looks blurry is also important. As people age, their eyes lose the ability to focus, and glasses are not always convenient or available. Similarly, a variety of low-vision conditions exist and cannot be corrected with glasses. Our legibility experiments provide evidence that our smartfonts are indeed more legible for a range of small sizes and with varying amounts of blur.

Acknowledgements
The authors would like to thank Kevin Larson for helpful discussions.

REFERENCES

1. Alpern, M. Accommodation and the pupil. *The Eye 3* (1969).
2. Appelman, I. B., and Mayzner, M. S. Application of geometric models to letter recognition: Distance and density. *Journal of Experimental Psychology: General 111*, 1 (1982), 60.
3. Arditi, A., and Cho, J. Serifs and font legibility. *Vision Research 45*, 23 (2005), 2926–2933.
4. Ashby, F. G., and Perrin, N. A. Toward a unified theory of similarity and recognition. *Psychological Review 95*, 1 (1988), 124.
5. Bailey, I. L., and LOVIE, J. E. The design and use of a new near-vision chart. *Optometry and Vision Science 57*, 6 (1980), 378–387.
6. Bernard, J.-B., Kumar, G., Junge, J., and Chung, S. T. The effect of letter-stroke boldness on reading speed in central and peripheral vision. *Vision Research 84* (2013), 33–42.
7. Bernard, M., Liao, C. H., and Mills, M. The effects of font type and size on the legibility and reading time of online text by older adults. In *Ext. Abstracts CHI 2001*, ACM Press (2001), 175–176.
8. Bernard, M. L., Chaparro, B. S., Mills, M. M., and Halcomb, C. G. Comparing the effects of text size and format on the readibility of computer-displayed times new roman and arial text. *Intl. Journal of Human-Computer Studies 59*, 6 (2003), 823–835.
9. Bessemans, A., et al. *Letterontwerp voor kinderen met een visuele functiebeperking*. PhD thesis, Leiden University, 2012.
10. Blommaert, F. J., and Timmers, H. Letter recognition at low contrast levels: effects of letter size. *Perception 16*, 4 (1987), 421–432.
11. Bostock, M., Ogievetsky, V., and Heer, J. D^3 data-driven documents. *IEEE TVCG 17*, 12 (2011), 2301–2309.
12. Colizoli, O., Murre, J. M., and Rouw, R. Pseudo-synesthesia through reading books with colored letters. *PLoS One 7*, 6 (2012), e39799.
13. Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. Drc: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review 108*, 1 (2001), 204.
14. Darroch, I., Goodman, J., Brewster, S., and Gray, P. The effect of age and font size on reading text on handheld computers. In *Proc. INTERACT 2005*. Springer, 2005, 253–266.
15. David, P. A. Clio and the economics of qwerty. *The American Economic Review* (1985), 332–337.
16. Davis Jr, F. D. *A technology acceptance model for empirically testing new end-user information systems: Theory and results*. PhD thesis, MIT, 1986.
17. Demiralp, C. D., Bernstein, M. S., and Heer, J. Learning perceptual kernels for visualization design. *IEEE TVCG 20*, 12 (2014), 1933–1942.
18. Dobres, J., Reimer, B., Mehler, B., Chahine, N., and Gould, D. A pilot study measuring the relative legibility of five simplified chinese typefaces using psychophysical methods. In *Proc. AutomotiveUI 2014*, ACM (2014), 1–5.
19. Duane, A. Normal values of the accommodation at all ages. *Journal of the American Medical Association 59*, 12 (1912), 1010–1013.
20. Farrell, J., Xu, J., Larson, K., and Wandell, B. 47.2: Visual preference for cleartype technology. In *SID Symposium Digest of Technical Papers*, vol. 40, Wiley Online Library (2009), 702–705.
21. Freedman, R. *Out of darkness: the story of Louis Braille*. Houghton Mifflin Harcourt, 1999.
22. Garcia, M. L., and Caldera, C. I. The effect of color and typeface on the readability of on-line text. *Computers and Industrial Engineering 31*, 1 (1996), 519–524.
23. Goldberg, D., and Richardson, C. Touch-typing with a stylus. In *Proc. INTERCHI 1993*, ACM (1993), 80–87.
24. Granaas, M. M., McKay, T. D., Laham, R. D., Hurt, L. D., and Juola, J. F. Reading moving text on a CRT screen. *Human Factors 26*, 1 (1984), 97–104.
25. Hall, R. H., and Hanna, P. The impact of web page text-background colour combinations on readability, retention, aesthetics and behavioural intention. *Behaviour and Information Technology 23*, 3 (2004), 183–195.
26. Heer, J., and Bostock, M. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proc. CHI 2010*, ACM (2010), 203–212.
27. Just, M. A., Carpenter, P. A., and Woolley, J. D. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General 111*, 2 (1982), 228.
28. Kwon, M., and Legge, G. E. Higher-contrast requirements for recognizing low-pass–filtered letters. *Journal of Vision 13*, 1 (2013), 13–13.
29. Legge, G. E., and Bigelow, C. A. Does print size matter for reading? a review of findings from vision science and typography. *Journal of Vision 11*, 5 (2011), 8.
30. Legge, G. E., Mansfield, J. S., and Chung, S. T. Psychophysics of reading: XX. Linking letter recognition to reading speed in central and peripheral vision. *Vision Research 41*, 6 (2001), 725–743.
31. Legge, G. E., Parish, D. H., Luebker, A., and Wurm, L. H. Psychophysics of reading: XI. Comparing color contrast and luminance contrast. *Journal of the Optical Society of America A 7*, 10 (1990), 2002–2010.

32. Legge, G. E., Rubin, G. S., and Luebker, A. Psychophysics of reading: V. The role of contrast in normal vision. *Vision Research 27*, 7 (1987), 1165–1177.

33. Liebowitz, S. J., and Margolis, S. E. The fable of the keys. *Journal of Law and Economics* (1990), 1–25.

34. Loomis, J. M. On the tangibility of letters and braille. *Perception and Psychophysics 29*, 1 (1981), 37–46.

35. Loomis, J. M. A model of character recognition and legibility. *Journal of Experimental Psychology: Human Perception and Performance 16*, 1 (1990), 106.

36. Masica, C. *The Indo-Aryan Languages*. Cambridge Language Surveys. Cambridge University Press, 1993.

37. McKinstry, C., Dale, R., and Spivey, M. J. Action dynamics reveal parallel competition in decision making. *Psychological Science 19*, 1 (2008), 22–24.

38. Moses, R. A. Accommodation. *Adler's Physiology of the Eye* (1987), 291–310.

39. Pelli, D. G., Burns, C. W., Farell, B., and Moore-Page, D. C. Feature detection and letter identification. *Vision Research 46*, 28 (2006), 4646–4674.

40. Rello, L., and Baeza-Yates, R. Good fonts for dyslexia. In *Proc. ASSETS 2013*, ACM (2013), 14.

41. Ryles, R. The impact of braille reading skills on employment, income, education, and reading habits. *Journal of Visual Impairment and Blindness 90* (1996), 219–226.

42. Shieh, K.-K., and Lin, C.-C. Effects of screen type, ambient illumination, and color combination on VDT visual performance and subjective preference. *Intl. Journal of Industrial Ergonomics 26*, 5 (2000), 527–536.

43. Smith, F. *Understanding reading: A psycholinguistic analysis of reading and learning to read*. Routledge, 2012.

44. Townsend, J. T. A note on the identifiability of parallel and serial processes. *Perception & Psychophysics 10*, 3 (1971), 161–163.

45. Traxler, M. J. *Introduction to psycholinguistics: Understanding language science*. John Wiley & Sons, 2011.

46. Trusty, A., and Truong, K. N. Augmenting the web for second language vocabulary learning. In *Proc. CHI 2011*, ACM (2011), 3179–3188.

47. Unger, G. *While you're reading*. Mark Batty Pub, 2007.

48. Vinot, J.-L., and Athènes, S. Legible, are you sure?: an experimentation-based typographical design in safety-critical context. In *Proc. CHI 2012*, ACM (2012), 2287–2296.

49. Walker, S., Schloss, P., Fletcher, C. R., Vogel, C. A., and Walker, R. C. Visual-syntactic text formatting: A new method to enhance online reading. *Reading Online 8*, 6 (2005), 1096–1232.

50. Wang, L., Giesen, J., McDonnell, K. T., Zolliker, P., and Mueller, K. Color design for illustrative visualization. *IEEE TVCG 14*, 6 (2008), 1739–1754.

51. Wiebelt, A. Do symmetrical letter pairs affect readability?: A cross-linguistic examination of writing systems with specific reference to the runes. *Written Language and Literacy 7*, 2 (2004), 275–303.

52. Yu, C.-H., and Miller, R. C. Enhancing web page readability for non-native readers. In *Proc. CHI 2010*, ACM (2010), 2523–2532.

53. Zhai, S., and Kristensson, P.-O. Shorthand writing on stylus keyboard. In *Proc. CHI 2003*, ACM (2003), 97–104.