PseudoSloan: A perimetric-complexity and area-controlled font for vision and reading research

Vladimir Y. Vildavski

Wu Tsai Neurosciences Institute, Department of Psychology, Stanford University, California, USA

✉

Luca Lo Verde

Wu Tsai Neurosciences Institute, Department of Psychology, Stanford University, California, USA

✉

Gail Blumberg

Blumberg Communication Design, 309 Arden Road, Menlo Park, California, USA

✉

Joss Parsey

Freelance Designer, 3810 Page Mill Road, Los Altos Hills, California, USA

✉

Anthony M. Norcia

Wu Tsai Neurosciences Institute, Department of Psychology, Stanford University, California, USA

✉

Artificial orthographies have long been used in studies of verbal learning and reading. These orthographies, also known as pseudo or false fonts, are designed to match the letters of an existing alphabet on a range of visual features, isolating effects of orthography from those owing to lexical processing. In a parallel line of research, there has been much interest in the design of optotypes for measuring visual acuity that have good properties in terms of character complexity and graceful degradation under blur. Here we merge these two traditions by designing a fully scalable pseudofont, "PseudoSloan," that is based on the design rubric of the widely used Sloan optotypes. The font includes 26 Latin letters as well as two sets of letter-like symbols matching the Latin alphabet on a letter-by-letter basis. Quantitative matching of the pairs of Sloan and PseudoSloan glyphs is done on the basis of ink area and perimetric complexity. We provide the installable PseudoSloan font in TrueType and OpenType formats, plus a large number of PseudoSloan glyphs in .svg format that vary over wide ranges in their perimetric complexity and ink area (https://osf.io/qhj2b/).

Introduction

Written symbols are critical visual forms that mediate the reading of text, and their recognition is thus an essential part of modern life. There has been much interest in visual processing of letters in the case of alphabetic languages and of idiograms in non-alphabetic languages. Letters were first used to measure visual acuity by Kuechler in the mid-nineteenth century (cited by (Colenbrander, 2008). Kuechler's chart, reproduced in Colenbrander (2008), comprised single words per line of decreasing size. The words were rendered in a complex Fraktur typeface. Subsequently, Snellen introduced several innovations (Colenbrander, 2008): the use of specially constructed serifed "optotypes" designed for acuity measurement, the arrangement of letters rather than words into the familiar chart form, and calibration of letter size to units of visual angle (Colenbrander, 2008). Louise Sloan (Sloan, 1951) pointed out that the use of serifs may produce unfamiliar, difficult to recognize forms, subsequently introducing sans serif optotypes based on Snellen's 5×5 grid and equal black and white stroke widths (Sloan, 1959). Optotypes have become the preferred means of clinical acuity measurement as their recognition is a familiar task, at least to literate patients, and depends strongly on the quality of refraction, the most common source of decreased visual acuity.

Reading involves acuity, orthographic, and lexical processes. With the advent of functional neuroimaging, the reading literature has made important use of letter-like symbols—"false-fonts" or "pseudofonts"—as a means of studying the visual processing demands of reading separate from semantic content. Importantly, the space of letter-like symbols used in many writing systems is relatively limited as characters of many writing systems tend to have approximately three

strokes (Changizi & Shimojo, 2005) and the complexity of characters does not depend strongly on the number of characters used in the writing system (Changizi & Shimojo, 2005; Miton & Morin, 2021).

The design of these pseudofonts conserves essential features of the native font, while at the same time avoiding combinations of features that are graphemes in the alphabet under study. Pseudofont creation has been accomplished in a variety of ways, for example, "The false-font font characters were composed of compositions of contiguous curved and straight-line segments designed to be at once dissimilar from any of the alphabetical symbols, but equivalent in the content of primitive features" Petersen et al. (1990) or "The false fonts were created by recombination of letter elements (ascending, descending horizontal, and curved lines of the font type Arial" (Indefrey et al., 1997), or "Finally, false-font strings (category 1) were generated using a custom-designed pseudo-font with fixed character spacing, where each uppercase letter was replaced by an unfamiliar shape with an almost equal number of strokes and angles and an overall similar visual appearance (Vinckier et al., 2007). More recently, a rubric based on equating strokes, junctions terminators and symmetry between the pseudofont with letters of the Latin alphabet (Courier New font) has been used to create the Brussels Artificial Character Set (BACS) stimulus set (Vidal, Content, & Chetail, 2017), which is available as an installable .otf font https://osf.io/dj8qm/.

The previously described pseudofonts differ in their construction rules and thus the equivalence to letters of the alphabet to which they are to be compared. Moreover, the details of the construction rule may vary over different fonts used to render a particular alphabet. An important question is thus how to assess the equivalence of a given font and its corresponding pseudofont quantitatively. A key concept in both the acuity and reading literatures is the notion of legibility. Legibility depends on stimulus contrast, letter size, spacing, and stroke width, as well as character complexity (Bigelow, 2019; Loomis, 1990). Complexity has been quantified by a simple metric—perimetric complexity, which is an empirical measure proportional to the ratio of the squared perimeter of the shape and its area (Arnoult & Attneave, 1956; Pelli, Burns, Farell, & Moore-Page, 2006). Perimetric complexity provides a good description of the efficiency of character identification over a wide range of character types and fonts (Pelli et al., 2006) and we thus chose to use perimetric complexity, along with ink area, as a means to quantitively equate glyphs used to match letters of the Latin alphabet. Motivated by a desire to link the visual acuity and reading literatures through the use of optotypes that can be varied in size and specified complexity, here we describe the PseudoSloan font, its design principles and the corresponding downloadable .otf and .ttf fonts (https://osf.io/qhj2b).

Methods

Pseudofont construction

The Sloan letters (Sloan, 1959) are based on a 5×5 square grid, as illustrated in Figure 1. The thickness of the lines equals the thickness of the white spaces and also the thickness of the gap in the letter "C." The height and width of the optotypes is five times the thickness of the line. In addition to the 5×5 grid, there are two sizes of inscribed circles, portions of which are used to compose letters with curves, such as the letters B, C, and D. A font for the Sloan letters authored by Denis Pelli has been available for a number of years https://github.com/denispelli/Eye-Chart-Fonts.

The design language of the Sloan letters admits the creation of a large number of alternative glyphs. Using this design language, we created 26 Latin letters and a set of 264 candidate artificial glyphs by hand using Adobe Illustrator and OminiGraffle software. Using high-resolution vector representations in SVG format, we computed perimetric complexity and ink area of all Latin letters and candidate glyphs, and then selected two sets of artificial glyphs matching each Latin letter in both the complexity and the area simultaneously. When pseudofonts are used to measure neural responses, it is important to control the ink area, because this factor affects the luminance of the stimuli. Equating the mean luminance of the font and pseudofont stimuli ensures that luminance does not inadvertently drive the response.

Perimetric complexity (Watson, 2012) can be defined for binary images such as letters as the sum of the inside and outside perimeters of the foreground p, squared, divided by the foreground area a, divided by 4π, or $c = p^2/4\pi a$. Figure 2 shows examples of the perimetric complexity of geometric shapes and some letters typeset in different fonts. Figure 3 shows a subset of the candidate glyphs, with glyphs highlighted in black being glyphs selected for inclusion in the PseudoSloan font.

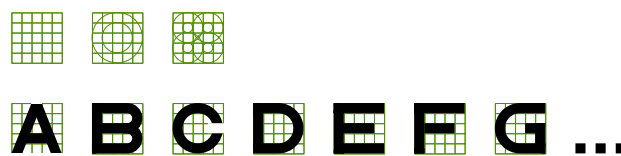Figure 4 shows the ink area and perimetric complexity of the preliminary corpus of 264 glyphs

Figure 1. Design language of the Sloan font. The Sloan font is based on a 5×5 grid. The thickness of the lines equals the thickness of the white spaces and the thickness of the gap in the letter "C." The height and width of the optotypes is five times the thickness of the line. In addition to the 5×5 grid, there are two sizes of inscribed circles that are used to compose letters with curves.
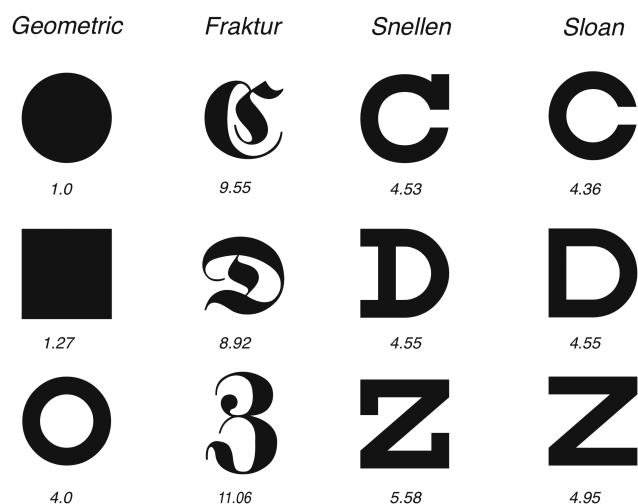
Figure 2. Examples of perimetric complexity of different geometric shapes and glyphs. Note that perimetric complexity for glyphs decreases substantially between Fraktur glyphs and Snellen optotypes with a smaller decrease for the Sloan C and Z optotypes. The D optotype has the same complexity in the Snellen and Sloan fonts.

Figure 3. Examples of PseudoSloan glyphs of varying levels of complexity. Examples in dark ink are included in the PseudoSloan font given their close match to a corresponding Latin letter.

constructed according to the Sloan rubric along with the 26 Latin letters. The ink area values are expressed in units of "typographic points," squared, and correspond to the glyphs drawn on a high-resolution grid of 500 × 500 points. The inset shows the neighborhood of the Latin letter V. From this, it can be seen that there are several glyphs that are similar in ink area, perimetric complexity or both. We chose two matches for each Latin letter, one that was the closest in complexity on the higher complexity side and the other on the lower complexity side, with most matches having complexity difference from corresponding Latin letters less than 1% (mean = 0.5%, max = 3.69%).
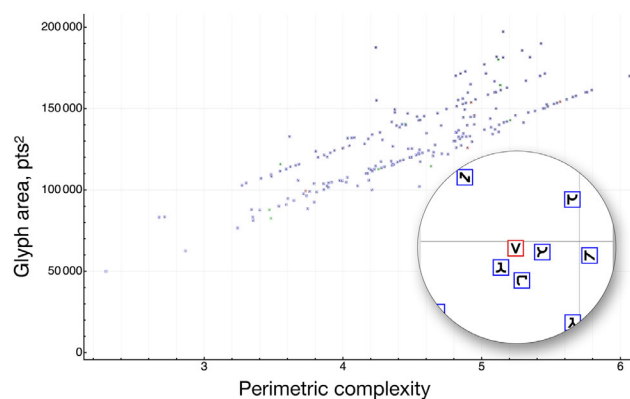

Figure 4. Perimetric complexity and ink area of the combined set of 290 glyphs (26 Latin, 264 artificial) from which matches to the 26 Latin letters were selected. The ink area values correspond to glyphs of 500 point size.

Figure 5. Character table of PseudoSloan font.

Note that there are lines or stripes of glyphs within this space. These come from the number of closed curve elements in the glyph. The glyphs scattered along the lower line have no closed loops, while the glyphs forming the higher stripe have one loop, and there are two loops in the glyphs along the top line.

OTF font creation and use

The hand-drawn Latin letters and matching glyphs were converted into PseudoSloan computer font in TrueType and OpenType formats (https://osf.io/qhj2b/). The font includes 26 capital Latin letters, two sets of letter-like glyphs, and the Landolt C and tumbling E symbols (Figure 5.)

All font symbols have bounding rectangles of the same height, which corresponds to the vertical size of the 5 × 5 design grid, and for each symbol, the width of the rectangle equals to the horizontal extent of the symbol, that is, in typography terms, the font provides neither horizontal nor vertical bearings.

The font table is organized as follows. The Latin letters occupy their standard upper-case slots with the decimal ASCII codes of 65 to 90. The codes 97 to 122, used in the standard fonts for the lower-case letters, correspond to the first set of artificial symbols that match the Latin alphabet by complexity and area on a letter-by-letter basis. The similarly matched second set of artificial symbols has ASCII codes in the 193 to 218

range. The Landolt C and the tumbling E symbols have ASCII codes 161 to 164 and 165 to 168, respectively. The font contains two white space characters. One, with ASCII code 32, is five grid steps wide and corresponds to the space key on computer keyboards. There is also a narrow space character, which is one grid step wide, and it is mapped on the ' (grave accent) key on U.S. keyboards.

The table layout provides for a convenient way of creating text displays by either using keyboard input or manipulating numeric ASCII codes in software. After installing and selecting the font in a text editor, the Latin letters can be typed by using the keyboard's upper case. Entering the same keys in lower-case replaces the text with the symbols from the first set of artificial glyphs. To type the symbols from the second set, one needs to enable the keyboard input of ASCII codes. For example, on Mac computers, add Unicode Hex Input in the Keyboard panel of the System Preferences and select it as current input method using system's toolbar menu. After that, holding the <option> key and typing four-digit hexadecimal ASCII codes in the range 00C1-00DA (decimal 193–218) inserts corresponding symbols from the font table. Alternatively, when using custom software programs for display control, in the program code add decimal numbers 32 or 128 to the ASCII codes of uppercase Latin letters to replace them with matching letter-like symbols.

Results and discussion

The evolution of optotype design was motivated by decreasing their complexity, in the colloquial sense. To show how perimetric complexity depends on the base font, in Figure 6 we plot perimetric complexity of the Courier New font (the basis of the BACS2 font, blue) and compare it to the Sloan letters (red). In each font, there is a range of complexities, with the letter I being the least complex. The most complex letter in the Sloan font is S, but in Courier New, it is M, so the rank order of complexity is similar, but not identical across fonts. Complexity increases smoothly from least to most complex glyphs in the Sloan font, but the difference in complexity for a given glyph between New Courier and Sloan is quite variable, and New Courier letters are all more complex than any Sloan letter.

By design, Sloan and PseudoSloan glyphs are precisely matched in terms of complexity. This can be seen by comparing the red and orange dots in Figure 6. The differences in complexity for Courier New glyphs (blue) and their matching glyph in the BACS2 pseudofont (green) for the upper-case set are larger and more variable. Although BACS2 and Courier New are matched on the number of strokes, junctions, terminators and symmetry, this design rubric leads to considerable differences between the complexity
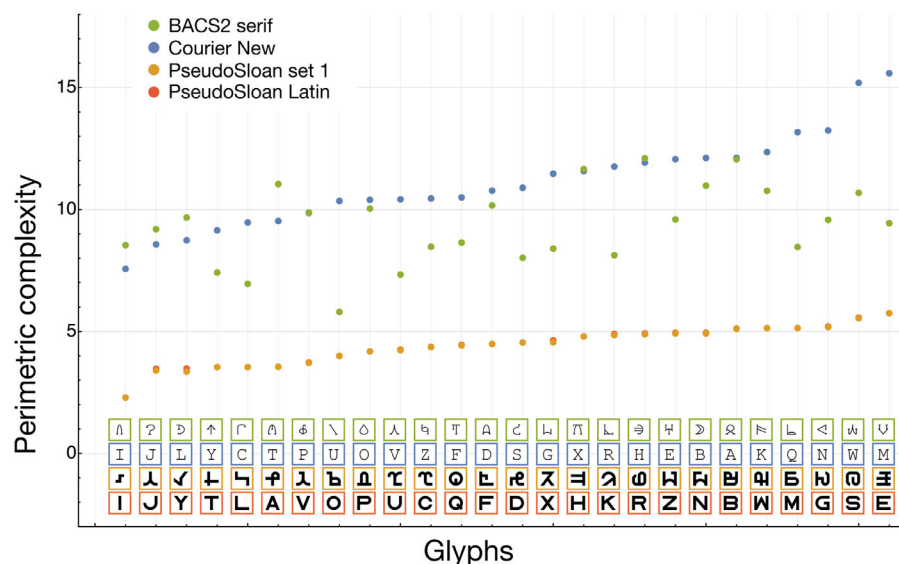

Figure 6. Comparison of uppercase Sloan, Courier New, BACS2, and PseudoSloan font complexities. Sloan and PseudoSloan character order was determined by the rank order of Sloan letter complexity, whereas BACS2 and Courier New character order was based on the rank order of Courier New. Sloan letters are less complex than are Courier New letters, as well as being less variable in complexity (red vs blue). BACS2 glyphs are overall less complex than those of Courier New (21 of 26 cases; blue vs yellow, respectively). Sloan and PseudoSloan are closely matched in terms of complexity (red vs orange). BACS2 characters are more complex and variable in terms of complexity than either Sloan font (green vs red and orange).

of the font and pseudofont glyphs. On average, the BACS2 glyphs are less complex than their New Courier counterparts for 21of 26 cases. So, when one would make an experimental contrast between words and pseudofonts by replacing each Courier New letter with its BACS2 counterpart, the contrast would be matched for number of strokes, junctions, terminators, and symmetry, but not likely on complexity. There is a subset of BACS2 glyphs that closely match their New Courier counterparts (P, O, D, and A, and to a lesser extent I, J, L, and H). PseudoSloan glyphs, in addition to being less complex overall, are less variable, as a set, than are BACS2 glyphs (compare orange with green).

All the files can be downloaded from https://osf.io/qhj2b/. The set represents a wider range of complexity and ink area values, and its SVG format supports vector-graphics display techniques. These glyphs could be used to develop alternative matching principles, say matching the BACS2 feature list within certain limits of complexity and ink area matches to the Sloan letters. The full set of glyphs we provide is by no means exhaustive and could be extended to suit the needs and constraints of experiments using artificial orthographies.

Keywords: text, letters, reading, acuity, false font, pseudo font

Acknowledgments

References

Arnoult, M. D., & Attneave, F. (1956). The quantitative study of shape and pattern perception. *Psychological Bulletin, 53*(6), 452–471.

Bigelow, C. (2019). Typeface features and legibility research. *Vision Research, 165*, 162–172.

Changizi, M. A., & Shimojo, S. (2005). Character complexity and redundancy in writing systems over human history. *Proceedings. Biological Sciences, 272*(1560), 267–275.

Colenbrander, G. (2008). The historical evolution of visual acuity measurement. *Visual Impairment Research, 10*, 57–66.

Indefrey, P., Kleinschmidt, A., Merboldt, K. D., Kruger, G., Brown, C., Hagoort, P., . . . Frahm, J. (1997). Equivalent responses to lexical and nonlexical visual stimuli in occipital cortex: a functional magnetic resonance imaging study. *Neuroimage, 5*(1), 78–81.

Loomis, J. M. (1990). A model of character recognition and legibility. *Journal of Experimental Psychology. Human Perception and Performance, 16*(1), 106–120.

Miton, H., & Morin, O. (2021). Graphic complexity in writing systems. *Cognition, 214*, 104771.

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

Petersen, S. E., Fox, P. T., Snyder, A. X., & Raichle, M. E. (1990). Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli. *Science, 249*(4972), 1041–1044.

Sloan, L. L. (1951). Measurement of visual acuity: a critical review. *AMA Archives of Ophthalmology, 45*(6), 704–725.

Sloan, L. L. (1959). New test charts for the measurement of visual acuity at far and near distances. *American Journal of Ophthalmology, 48*, 807–813.

Vidal, C., Content, A., & Chetail, F. (2017). BACS: The Brussels Artificial Character Sets for studies in cognitive psychology and neuroscience. *Behavior Research Methods, 49*(6), 2093–2112.

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J. P., Sigman, M., & Cohen, L. (2007). Hierarchical coding of letter strings in the ventral stream: Dissecting the inner organization of the visual word-form system. *Neuron, 55*(1), 143–156.

Watson, A. B. (2012). Perimetric complexity of binary digital images (No. ARC-E-DAA-TN3185).