6. References

[1] Debra Anne Adams. A dialogue of forms: Letters and digital font design. Master's thesis, Massachusetts Institute of Technology, September 1986.

[2] Aries Arditi, Kenneth Knoblauch, and Ilana Grunwald. Reading with fixed and variable character pitch. *J. Optical Society of America A*, 7(10):2011–2015, 1990.

[3] Charles Bigelow and Kris Holmes. The design of Lucida: an integrated family of types for electronic literacy. In J.C. van Vliet, editor, *Text Processing and Document Manipulation: Proceedings of the International Conference*, New York, April 1986. Cambridge University Press.

[4] James Craig. *Designing with type.* Watson-Guptill Publications, 1515 Broadway, New York, NY 10036, paperbound edition, 1978.

[5] J.E. Farrell and M. Desmarais. Equating character-identification performance across the visual field. *Jour. Opt. Soc. Amer.*, 7(1):152–159, 1990.

[6] Lewis O. Harvey, Jonathan O. Roberts, and Martin J. Gervais. The spatial frequency basis of internal representations. In *Modern Issues in Perception.* North Holland, 1983.

[7] Donald E. Knuth. *The METAFONTbook*, volume C of *Computers and Typesetting.* Addison-Wesley, Reading, MA, 1986.

[8] Donald E. Knuth. *Computer Modern Typefaces*, volume E of *Computers and Typesetting.* Addison-Wesley, Reading, MA, 1986.

[9] Henry Kuĉera and W. Nelson Francis. *Computational Analysis of Present-Day American English.* Brown University Press, Providence, 1967.

[10] Gordon E. Legge, Gary S. Rubin, and Andrew Luebker. Psychophysics of reading—V. The role of contrast in normal vision. *Vision Res.*, 27(7):1165–1177, 1987.

[11] Ruari McLean. *The Thames and Hudson Manual of Typography.* Thames and Hudson, Inc., 500 Fifth Avenue, New York, NY 10010, 1980.

[12] Sandra Ernst Moriarty and Edward C. Scheiner. A study of close-set text type. *J. of Applied Psychology*, 69:700–703, 1984.

[13] Robert E. Morrison and Keith Rayner. Saccade size in reading depends upon character spaces and not visual angle. *Perception & Psychophysics*, 30(4):395–396, 1981.

[14] Richard Rubinstein. *Digital Typography—An Introduction to Type and Composition for Computer System Design.* Addison-Wesley, Reading, MA, 1988.

[15] Richard Southall. First principles of typographic design for document production. *TUGBoat*, 5(2):79, November 1984.

[16] Richard Southall. Designing new typefaces with Metafont. Technical Report STAN-CS-85-1074, Stanford University, September 1985.

[17] Miles A. Tinker. *Legibility of Print.* Iowa State University, Ames, 1963.

At small sizes, the typographically scaled fonts, however, are somewhat bolder than are the linearly scaled ones. At this writing, we have no realistic way to isolate the effect of type weight, because the bold variant of a typeface is generally wider than its roman counterpart. If we simply thickened strokes without changing character widths, we might make the fonts harder to read because we would have changed the character shapes (e.g., reduced the counterforms). On the other hand, data we took on Computer Modern Bold vs. Computer Modern Bold Extended (a wider version of Computer Modern Bold) gave inconclusive results, as did data we took with the Macintosh Times 10 and its bold variant, which the Macintosh software makes by emboldening the bitmaps.

4. Applicability of the results

Applying our results, as well as many of the results in the previous literature, to questions of actual type design and typography presents several problems.

First, reading from screens is not the same task as is reading from paper[14].

Second, few of the effects we measured—even where statistically significant—are very large. Arditi et al. report that pseudo fixed-width fonts are more than twice as easy to read than are variable-width ones. We found that "well rasterized" fonts are about twice as easy to read as are comparable "naively rasterized" ones (see Figure 3.4 for the reading rates and Figure 3.2 for an illustration of the fonts). Most of our other effects (and most of Arditi et al.'s at reading sizes or larger), however, are generally 20% or less. This agrees with the typographic literature summarized by Southall[15], who concluded that interword and interline spacing are the most important factors in readability.

Third, the flash card method of measuring reading rate might not be a reasonable test of typographic variation. But a study[12] which used timed readings of text on paper also reported effects of only around 20% in concluding that tightly spaced fonts are easier to read than are normally spaced ones.

Finally, readability is not the sole criterion for font choice. In fact, economic factors probably dictate most choices: for instance, variable-width fonts can squeeze the same text into a smaller area than fixed-width fonts would use; linear scaling is cheap in terms of design and storage; and so on. Unquantifiable design criteria, i.e., aesthetics, can also determine font choice. Both our results and those in the literature suggest that fonts chosen according to such criteria, if reasonably spaced and shaped, will not be hard to read.

5. Summary

Most typographic variation has little or no effect on reading rates for characters at normal reading sizes. Fonts corresponding to 6-point and perhaps smaller type require adequate letterspacing. However, the means by which this is obtained, e.g., using fixed-width fonts, typographic scaling, or simply increasing the sidebearings, does not matter. Crowding effects due to inadequate spacing are significantly worse for badly rasterized fonts than for well rasterized ones.

Figure 3.4: Reading rates for Times 11-point vs. Times 12-point. The 11-point fonts are made by bitmap scaling. The well-formed 12-point font is up to three times more readable independent of whether or not letterspacing is tight (NSB) or loose (PFW). Graphs depict words per minute vs. character size in degrees of visual angle.

3.3. Linear vs. typographic scaling, and other manipulations

Most contemporary digital fonts are produced by rasterizing a scaled version of a single master outline for each character. But typographic tradition contends that as you make smaller fonts for a given typeface, you should, among other things, increase both their character and stroke widths relative to their x-height. Logically, this "typographic scaling" should either weaken or eliminate the crowding effect described above.

To test this, we generated two variants of the Computer Modern Roman typeface[8] using METAFONT[7]. We produced one by rendering at 150 dpi an outline designed for a 12-point size. We produced the other by rendering an outline designed for a 6-point size at 300 dpi. Both fonts had a cap-height of 17 pixels. Figure 3.5 compares a linearly to a typographically scaled font (magnified fourfold, for convenience of illustration).

RQENbaegnov
RQENbaegnov

Figure 3.5: Linearly vs. typographically scaled Computer Modern Roman.

For .16 deg high characters, all subjects showed a slight preference, ranging from 10–30%, for the typographically scaled font.

Figure 3.3: Reading rates for the poorly formed Times 11-point fonts. For very small characters the naively but loosely spaced PFW shows an advantage over the tightly spaced NSB. Graphs depict words per minute vs. character size in degrees of visual angle.

to the normal 12-point Times—Lucida fonts have shapes which survive the abuse of rendering them with only a few pixels. There is no advantage to fixed-width fonts in either the well-designed Lucida fonts, or even the rather tightly set 12-point fonts we used. Hence, we conclude that incorrect spacing and perhaps poor shape, rather than crowding per se, is the likely explanation of the results of Arditi et al. which showed an advantage for fixed-width fonts at small sizes.

3.2. Resolution

As mentioned above, we conducted the Times and Lucida experiments with both the nominal 12-point fonts (8 and 9 pixel cap-heights, respectively) and the nominal 24-point (16 and 17 pixel cap-heights). No subjects showed any preference for either of the two resolutions. On the other hand, the bitmap-scaled Times NSB 11-point font was 2–3 times harder to read than its Times 12 progenitor (see Figure 3.4). We took no data with fonts made by bitmap scaling the 24-point Times to, say, nominal 23-point because such scaling does not distort characters nearly as much as does bitmap scaling 12-point to 11.

From this we conclude that, for character heights down to .16 deg and resolutions down to 8 pixel cap-heights, letter shape quality probably influences readability more than does resolution. (But note that the Arditi et al. data were taken with low resolution characters, and resolution might be a factor in sizes smaller than .16 deg.)

cases were the effects very large. Arditi et al., on the other hand, report as much as two-fold faster rates for the *fixed*-width fonts at .16 deg.

Having shown that tight letterspacing does not cause crowding effects, we then asked ourselves whether or not other factors could explain Arditi et al.'s findings about pseudo fixed-width fonts. Because we were unsure of the Amiga fonts' quality, we decided to produce a bad font. Fortunately, the Macintosh makes this easy: when a program asks for a font for which there is no available bitmap, the Macintosh produces one by scaling the *bitmaps* of the closest sized one. We used this notoriously inappropriate method to produce an 11-point Times NSB from our 12-point Times NSB. Figure 3.2 illustrates these two fonts.

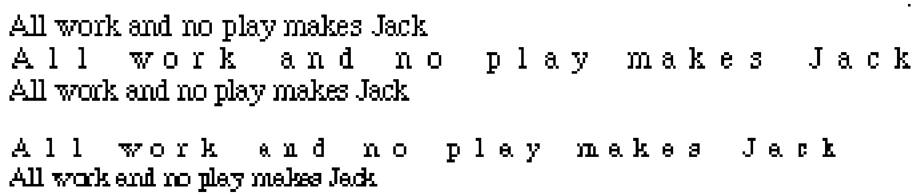

Figure 3.2: The second and third fonts, Times12PFW, and Times12NSB represent artifically spaced variants of the Macintosh supplied Times12, shown at the top. The lower two fonts are formed by bitmap scaling Times12PFW and NSB to 11-point. The result is so ill-formed that the fonts are substantially harder to read than their 12-point progenitors. See Figure 3.4.

With the 11-point fonts, we get the crowding effect at small sizes. However, the difference, in reading rates between the 11-point Times NSB and PFW is still less than a factor of two; see Figure 3.3.

We thus concluded that loose letterspacing alone does not make fonts easier to read; rather, shape or other font-quality factors do. Furthermore, we concluded that reasonable vs. poor letterspacing, not fixed-width vs. variable-width fonts, reduces crowding effects. We argue this further below.

3.1.2. Designed fixed-width All the pseudo fixed-width fonts described above suffer from a major typographic defect: uneven letterspacing. To make all the characters of a variable-width font take up the same space, the amounts of space we must add to their sidebearings varies widely from character to character. This makes pseudo fixed-width fonts' letterspacing highly irregular. To avoid this, a designer of a fixed-width font generally expands the narrower characters. In a well-designed typeface family, the fixed-width variants have the same overall typographic contrast ("color") as do the corresponding variable-width variants, so comparison between them may be more meaningful.

To further investigate questions of fixed vs. variable-width, we gathered data using 9 and 17 pixel cap-height (nominal 12 and 24-point, as before) Lucida Sans and Lucida Sans Typewriter fonts. The designers Bigelow and Holmes made Lucida Sans Typewriter by adapting Lucida Sans' shapes[3]. Typographers would predict that the variable-width variant would be easier to read at or below normal reading sizes (.32 deg, e.g., nominal 12-point type read at 18 in). To our surprise, our subjects favored neither variant, for sizes down to .16 deg.

Since Bigelow and Holmes purposely designed the Lucida typeface family for digital representation, we expect it to have better shapes at low resolution than those of the Times family, which wasn't necessarily designed to digitize well. Compared to the ill-conceived 11-point Times NSB—and even

3.1.1. Pseudo fixed-width We first attempted to duplicate the conditions of Arditi et al. The most striking thing typographically about their fonts is the unusual way in which they constructed their fixed-width ones. To make these “pseudo fixed-width” fonts (as they call them), they modified the Times font supplied with their Amiga computer by adding equal white space to both sides of each character until the sum of that character’s width and its surrounding white space was equal to the width of the font’s uppercase ‘W’ (the widest character in the font). They varied character height by varying viewing distances from the monitor. They used a 9 pixel cap-height font for experiments with sizes of less than .165 deg, and an 18 pixel cap-height font for all other sizes.

We similarly modified standard Macintosh Times fonts, using ones that had 8 and 16 pixel cap-heights (nominal 12 and 24-point, respectively). We called this “Times PFW,” short for “Times Pseudo Fixed Width.” To help determine whether or not crowding alone caused Arditi et al.’s results, we also created a particularly crowded variant by removing *all* side-bearings from the Times fonts. We called this “Times NSB,” short for “Times No Side Bearings.” Figure 3.1 shows our reading rate results for the 16 pixel cap-height fonts.

For most subjects, there is little or no significant difference between the crowded NSB font and the artificially spaced PFW font, suggesting that letterspacing alone can not account for reading rate differences.

Figure 3.1: Reading rates in words per minute as a function of character size in degrees of visual angle for 16 pixel cap-height Times NSB and PFW.

Except for KB, our subjects read our PFW and NSB Times at very similar rates. When they favored fonts, it was mostly the variable-width ones—despite those fonts’ lack of sidebearings. This agrees with typographic wisdom. Generally, our reading-rate differences for PSW vs. NSB were less than one standard error (the standard error varied from 5%–35% of the mean reading rates). In no other reading studies—we simply measured reading rate. We computed this as the product of the percentage of correctly read words and the rate of text presentation in words per minute.

We measured readability using a "flash-card" method. In each presentation, the subject first fixated on the center of the screen of a Macintosh II computer. She or he then clicked the mouse button, causing the computer to display for a particular duration the initial part of a sentence, divided into two lines. The subject read the text aloud. A trial consisted of a series of presentations.

If the presentation's string included a proper noun or a non-English word, we ignored it. If not, and if the subject correctly read more than 50% but less than 100% of its words, the trial ended and we recorded the reading rate. Otherwise, we modified the duration for the next presentation in the trial: if the subject read all the words correctly, we decreased the duration by 33 ms; if the subject read fewer than 50% of the words correctly, we increased the duration by 33 ms.

For a particular distance, we gathered data for each subject at least six trials. We considered this a "run." In most cases, we made runs for up to four distances. For convenience, we usually interleaved runs with different fonts, presenting the fonts in random order for each trial.

For most experiments, we used a standard Macintosh II grayscale 12 in monitor. We conducted the experiments about typographic scaling on a RasterOps 19 in grayscale monitor. In all cases, the Macintosh system was set to display 256 levels of grayscale, but we displayed the characters as nominal black on white. We left the room lights on, controlling neither for ambient illumination nor for total stimuli contrast (a factor which we—and many typographers—believe should be investigated).

We broke the sentences into from 24–29 characters per line. We chose to keep constant the character count—as opposed to the string width—because (as mentioned above) previous experiments[13] showed that, in normal reading, saccades cover a constant number of characters.

We used text from one million English words in the Brown Corpus of English[9]. It contains a variety of American English documents—including fiction, and newspaper and magazine articles.

Three of the authors served as principal subjects. All were typographically sophisticated; in addition, subjects KH and KB have letterform design and typography training. A fourth subject (DG) was naive both typographically and about the experiment's purpose. We did not gather data for DG on all experiments, but his results did not appreciably differ from those of the other subjects.

3. Results

3.1. Fixed vs. variable-width

Arditi et al. found that fixed-width fonts were easier to read than were variable-width fonts at small character sizes, but that the reverse was true at large sizes. They speculated that this was due to "crowding effects" present only at small sizes. This is contrary to conventional typographic wisdom[11,14] that says variable-width fonts are *always* more readable than are fixed-width fonts, regardless of character size.

We hypothesized that poorly-spaced fonts and/or poorly-shaped characters, rather than variable-width fonts per se, caused crowding effects. To investigate this, we replicated the Arditi et al. experiments with several pairs of fonts, as described below.

1.1. Character size, reading distance, and resolution

There are many ways to measure type size. Some ways make it hard to compare one typeface to another. Traditionally, typographers measure characters in printer's points, where 1 in = 72.27 pt. But to speak of a "ten-point" font does not mean that, say, one of its capital letters is 10 points high. With hot metal type, typographers used "body size," i.e., the depth of the metal block upon which the character's raised printing surface sits, which is larger than the character itself.[4] For digital type, we will speak of a font's "cap-height" (the height of its capital letters, even though the lowercase letters with ascenders may be taller).

Such absolute linear measures do not consider the size of the image on a reader's retina. This size depends on the distance of the text from the reader. However, when we express size in degrees of visual angle, it is *independent* of viewing distance. For example, the cap-height of nominal 12-point type (which for the fonts we examined had a cap-height of 8–9 pixels) at a typical reading distance (18 in) subtends approximately .32 deg of visual angle. Since .16 deg-high type (nominal 6-point) is rather small from a typographic point of view, we did not examine type smaller than that.

However it is measured, type size over a large range doesn't seem to affect reading. Over viewing distances of 14–28 inches, Morrison and Rayner found[13] that the length of a saccade—the movement the eyes make between fixations—was generally constant when measured in *characters*. Put another way, they found that a reader's gaze always shifts by a constant number of characters (about seven, for normal adult readers), regardless of the type size. Correspondingly, we and other experimenters[2] have found that reading rates measured in words per minute are relatively constant over a wide range of type sizes.

Morrison and Rayner used a monospaced font whose width was about .35 deg of visual angle at 28 in (about .70 deg at 14 in). Because our main findings are about letterspacing of fixed and variable-width fonts, we use cap-height instead of width as a measure of size. Usually we express it in degrees of visual angle—the measure also used by Arditi et al.[2], some of whose findings we seek to explain. However, when discussing the quality of particular fonts we sometimes find it more meaningful to express cap-height simply as an absolute number of pixels, since many people believe that low-resolution characters are harder to distinguish from one another than are high-resolution ones.

Our display devices have fixed resolution (75 dots/inch), so we can change the character size in two ways: we can change either the viewing distance or the size of the characters. We can do the latter either by changing the nominal type size or by producing the letterforms at a different resolution.

For example, we produced fonts at a resolution of 300 dpi for some experiments (described below). When viewed from the same distance, characters in such a font appear four times larger when displayed on a 75 dpi screen than they do when rendered on a 300 dpi device. However, when viewed from four times the distance as their low-resolution counterparts, they will subtend the same visual angle as those counterparts, allowing us to see how resolution affects reading rates (cf. below). (We were somewhat surprised to find that resolution had little effect—at least for fonts with cap-heights of nine pixels or more.)

2. Methodology

We know of no quantitative models either for type quality or for text readability, so—as with most

How typeface variation and typographic scaling affect readability at small sizes

Robert A. Morris
University of Massachusetts at Boston
Harbor Campus
Boston, Massachusetts, USA
ram@cs.umb.edu

Karl Berry
UMass/Boston
karl@cs.umb.edu
Also at the Free Software Foundation

Kathryn A. Hargreaves
UMass/Boston
letters@cs.umb.edu
Also at the Free Software Foundation

Dimitrios Liarokapis
UMass/Boston
dimitris@cs.umb.edu

1. Introduction

Many experimenters have drawn conclusions about how typographic variables[17,2] or psychophysical variables[10] related to typographic variables affect reading. Typographic variables include typeface choice, line length, boldness, and letterspacing. Psychophysical variables include contrast, polarity, and spatial frequency. However, we noticed that most of the cited studies controlled either haphazardly or not at all for "font quality," and wondered if some of the results were confounded by this.

For example, several experimenters change the size of the stimulus—in our case, typeforms—by placing the subject at various viewing distances. This is "linear scaling" of type, which for centuries typographers have considered poor practice; instead, punch cutters created different letterform masters for different type sizes. However, ever since the advent of phototypesetting in the 1950's linear scaling has increasingly become the norm, and remains so with many digital technologies.

Although some psychophysical experiments use individual characters as stimuli[6,5], we know of only one series of reading rate studies[10] of people with normal vision in which the experimenters attempted to control for psychophysical variables other than the size of the stimuli. (They controlled for contrast and spatial frequency content.) Yet some typeface design parameters directly impact corresponding visual parameters—namely, contrast, spatial frequency content, and noise.

Here we try to explain how font quality—a font being an instance of a typeface[16,1]—might influence conclusions about typographic variables and readability. Specifically, we examine a number of font variations for which visual "crowding effects"[2] appear to affect reading performance. We believe that some of the previous conclusions drawn about this performance may be irrelevant to fonts that are properly designed.

We will address several issues that will interest anyone using type—especially digital type. These include the relative merit of: (a) fixed-width vs. variable-width fonts; (b) typographic vs. linear scaling; (c) tight vs. normal vs. loose letterspacing.