Factors Affecting the Text Quality of Low-Resolution Information Displays

Scott F. O'Dell; Eastman Kodak Company; Rochester, NY/USA

Abstract

Grayscale text was introduced to improve text quality on low-resolution displays. Gray levels perceptually smooth out "jaggies," improving quality for some type sizes. This paper explores grayscaling methods, display design parameters, and type size effects on text quality. Magnitude estimation and paired comparison methods were used to quantify text quality for several of these factors. Display resolution, type size, and bit depth were found to be significant factors for each anti-aliasing method. Predictive models enable the optimization and design of information displays having limited contrast, brightness, and resolution.

Introduction

If display resolution is unlimited, text with high contrast and sharp, distinct features is more readable, more legible, and less tiring than text with low contrast, soft edges, and/or muddled features, all else being equal. In rendering a 1-bit, high-resolution representation of a character onto a low-resolution display, aliasing can occur, whereby diagonal and curved lines exhibit a stair-stepped ("jaggies") appearance. Stem widths and spaces might also become uneven within characters, and the same character may be rendered differently within a line of text. These artifacts degrade text quality as, at some sizes, the character features become corrupted, whereas their high-resolution ancestor is still legible.

To mitigate these effects, particularly the uneven stem and intra-character space widths, hinting instructions are encoded into fonts. Hints align and quantize the widths of character elements to the output pixel grid, slightly deforming character shapes in order to maintain element integrity and consistency while preserving the look of the font. Hint application reduces the number of elements that are aliased when down-sampled. However, remaining diagonals, curves, and serifs are still susceptible to aliasing.

The remedy for aliasing is anti-aliasing, whereby levels of gray are introduced to fill in the jaggies. Anti-aliasing effectively allows intensity information to be substituted for spatial information as sharp-edged jaggies are replaced by smoother (albeit softer-edged) shapes.

Various readability studies have examined the quality effect of grayscale text, using CRT displays or printed material where contrast and brightness are not deficient [1–4]. These studies are limited to a small range of visual angles (9 to 14 point text viewed at typical CRT workstation distances). Anti-aliasing by proprietary rendering engines was often used without consideration of gray level distribution and interaction with the underlying font hinting.

The goal of this study is to examine and quantify the effect on quality of various anti-aliasing methods, using display parameters appropriate for contemporary electronic reflective displays. These displays are typified by lower luminance contrasts (~5:1) and white state luminance factors (<0.6). They typically have display resolutions lower than CRTs, are viewed at a wide range of distances, and have limited grayscale capability. Understanding the application space, and overall value of implementing grayscale for information displays, would facilitate intelligent engineering.

The concept of readability, and its requisite optimal text rendering parameters, does not apply to simple informational displays. Instead of reading and comprehending long tracts of running text, the informational display requires interpreting pieces of data that are not laid out into paragraphs, and may be comprised of both alphanumerics and symbols. Point-of-purchase signage has a high degree of numerals and symbols, while airport displays of arrivals and departures would be comprised of acronyms and abbreviations that do not contain conventional word contours that are typically leveraged in the process of reading.

Previous studies [3,4], whether by actual lack of signal, insufficient reading tasks, or limited range of stimuli parameters, often do not show differences in objective readability measures. However, they usually report differences in ratings of perceived attributes such as "legibility," "ease of reading," or "text sharpness." These subjective ratings might be more applicable for signage. Therefore, the subjective evaluation for this study was perceived quality, as defined by the test subject.

Test Methodology

Hypothesis

Anti-aliasing should not affect the quality of small characters, offer some improvement for slightly larger characters, and impart a degrading or no effect for larger characters. For small characters, jaggies are invisible; they (and those of the gray tones introduced) will have negligible impact. For slightly larger characters, the gray tones introduced by anti-aliasing will blend with the visible jaggies to give the desired quality effect. As the characters get larger at low resolutions, the gray tones introduced begin to exhibit their own visibly jagged structure. These additional jaggies further degrade at this combination of type size and resolution. Edge sharpness, contrast, and letter integrity changes might also contribute to perceived quality degradations.

Experimental Factors and Design

Four factors (Resolution, Type Size, Metric, and Bit Depth) were studied at three levels, as shown in Table 1. The design was comprised of a four-factor central composite design (CCD) with center point replicates. Axial points comprised of 1-bit unhinted rendering were added. These extra points resulted from a two-factor (Type Size and Resolution), three-level full factorial with replication. The CCD design of 26 levels was repeated for each of three anti-aliasing methods. All stimuli were composed using Arial font viewed at a constant 24 in., with a simulated white state luminance factor of 28% and a black state luminance factor of 5%. The luminance factors were calculated relative to the monitor white calibrated to the chromaticity of a D50 illuminant with a luminance of 190 cd/m^2. All stimuli were negative polarity text.

Table 1: Experimental Factors and Levels

Factor	Low	Med	High
Resolution (DPI)	34	68	102
Type Size (pix)	11	15	20
Metric	0.45 (sRGB)	0.33 (CIELab)	0.20 (DIN)
Bit Depth	2	3	4
Method	Unhinted	Hinted	Autohinter

Type size is the height of the typeface in pixels. It is defined as the cap height (height of the letter “H,” for example) plus the length of the descender (i.e., the part of the letter “g” that extends below the baseline). Type size in pixels, rather than points, was used in order to maintain character structure across all resolutions tested. At low resolution, several legible type sizes can result in a vertical stem width of one pixel. Pixel heights corresponding to aim stem widths of 1, 1.5, and 2 pixels were determined via magnification of the uppercase “L.” For Arial font, it was found that 11, 15, and 20 pixel height characters achieve these aims, while maintaining text that still resembles the font.

When the character pixel height (H) is combined with display resolution (R) and viewing distance (D), one can calculate the angular subtense of the typeface (TAS) via

$$\text{Type Angular Subtense (TAS)} = \text{ArcTan}(H/(D*R)) \qquad (1)$$

Type Angular Subtense has the units of degrees per character, describing the visual angle subtended by a character height. TAS ranged from 15.4 to 84.2 arcmin in this design.

Resolution is the addressable pixels (or dots) per inch for the display. This factor determines the smallest element or text feature that can be displayed. Along with viewing distance (D), one can calculate the angular subtense of a pixel (PAS) via

$$\text{Pixel Angular Subtense (PAS)} = \text{ArcTan}(1/(D*R)) \qquad (2)$$

Pixel Angular Subtense has an inverse relationship to display resolution. PAS describes the visual angle subtended by a pixel. Within observer acutance limits, a model based on Pixel and Type Angular Subtenses could be used to predict quality at viewing distances other than those used in the study. Unlike previous studies [1–4], the combinations of TAS and PAS in this design produced a large range of point sizes from 8 to 42.

Metric (M) pertains to the visual lightness scale (V) in which the gray levels were quantized such that V equals L^M, where L is the luminance factor of the tone. This quantization is done keeping the white and black state luminance factor of the stimulus at 28% and 5%, respectively. Metric levels were based on the exponent values proposed for various viewing conditions found in [5].

Bit Depth (BD) defines number of gray levels, including white and black. The number of gray levels is simply 2^{BD}.

Method describes the anti-aliasing algorithm that determines the intensity and spatial aspects of the rendered character. In general, anti-aliasing is achieved by application of a low-pass filter to a high-resolution image before subsampling. Entire areas of study are devoted to anti-aliasing for image and text processing, and are beyond the scope of this study. For this study, we desired implementations of anti-aliasing text that can be found in common software, with algorithms that are openly documented.

To this end, the FreeType 2 (http://www.freetype.org/) font-rendering engine in GIMP 2.2.7 (http://www.gimp.org/) was used. GIMP is a freely distributed, open-sourced software package that closely resembles commercial image editing software in appearance and functionality. The anti-aliasing methods used three types of hinting along with bicubic interpolation, which performs both the prefiltering and decimation in one step. The first method utilized the hints embedded in the font file. The second used the FreeType Autohinter, which applies additional algorithmic hinting instructions. The third version is simply unhinted, which maintains macro shape of the text, but results in irregular type elements.

Test Stimulus

The purpose of the stimulus, and the quality criteria to rate it, were determined by the test subject. Hence, the content of the stimulus should be relevant to informational displays while encompassing enough of the character set that is typically used. The ASCII character set ISO 8859-1 codes #32–126 plus #162 (the “cents” character) comprised the English characters under study.

Several stimulus content constructions were considered. A roughly square array of the entire set was devised. This stimulus is free of variable layout and letter or word spacing concerns, and relates to character legibility. However, it is not relevant to informational signage or reading tasks, and its lack of context makes it easier for subjects to focus on a few particular characters.

An alternative is a pangram (a passage that contains every letter of the alphabet) that offers letters in the context of words, and is closer to actual reading. However, content could be a distraction, and suboptimal character and word spacing could influence quality. Another relevant stimulus is a simple sign comprised of only numerals and currency symbols, similar to electronic shelf labels. While this offers key characters in the context of an actual application, a simple sign alone would not suffice for the purposes of this test because of the absence of letters. Finally, a running text passage could be used, having many characters in the context of an actual application (reading). Because comprehension is not assessed, ratings could easily be biased by a few particular characters. Furthermore, not all ASCII characters are present, as the passage was too short to adequately sample the characters based on frequency of occurrence in a variety of texts.

The ASCII Table array was chosen to baseline the quality effects of the entire character set under study. Twenty-two subjects evaluated this test stimulus. Based on subjects’ comments regarding this stimulus, a follow-up experiment was executed. Twelve subjects also evaluated both a simple sign and a running text passage for all experimental levels.

Psychophysical Task

For each stimulus, subjects completed two tasks in a single session to determine just-noticeable differences (JNDs) of quality. The first task was a forced-choice paired comparison of levels that were previously rated as close in quality. Five of the nine type size and resolution combinations were used as subsets for this task.

Within each subset, five levels were selected, resulting in (5*4)/2 = 10 comparisons, for 50 total comparisons for the task. JNDs of quality can be deduced from the unsaturated comparisons [6].

Because the entire design spanned a large range of quality, a second task was required to rate all of the experimental levels. A mathematical relationship between the JNDs from the first task and the ratings of the second task could be derived and applied to all the experimental levels. The second task was a magnitude estimation whereby a ratio-scale rating is assigned to each experimental level relative to a reference image. A linear relationship exists between JNDs and the logarithm of ratio-scale magnitude estimates [6]. The reference for each of the nine combinations of resolution and type size was the hinted 1-bit rendering, isolating the effect caused by grayscale as a function of resolution and type size.

Test Implementation

The stimuli were presented on a Viewsonic Model VP2290b LCD Monitor calibrated to a D50 chromaticity white point at 190 cd/m^2. The room lights in front of the display were off; however, the gray wall behind the monitor was illuminated. The luminance from both the back wall and the monitor desktop was 38 cd/m^2. The monitor resolution was 204 dpi, enabling simulation of devices with resolutions equal to integral factors of 204 dpi. Presentation of the stimuli was controlled by custom MATLAB® code that used the Psychophysics Toolbox, a free set of functions designed for vision research [7].

Analysis

Calculation of JNDs

All magnitude estimate ratings were transformed to JNDs of quality using the results of the regression between JNDs derived from the paired comparison levels and the logarithm of the corresponding magnitude estimates. This relationship was based on the 22 subjects that rated ASCII Table only, as the limited number of subjects for Running Text and Simple Sign produced too many saturated samples.

Effect of Stimulus Content and Method

Tukey-Kramer HSD analysis was performed comparing results between all three signs (ASCII Table, Simple Sign, and Running Text) for each rendering method. The mean JNDs for each sign were not significantly different for all anti-aliasing methods. The unhinted 1-bit rendering, however, had different responses for the ASCII Table relative to the other two signs.

Tukey-Kramer HSD analysis was performed again, comparing results between rendering methods for each sign. This transposition of the previous analysis shows that the mean JNDs for the methods that involved hinted anti-aliasing were not significantly different. The unhinted anti-aliased rendering for the ASCII Table and running text was significantly worse than the other anti-aliasing methods. The unhinted 1-bit rendering was significantly worse than all grayscale renderings for all signs.

Analysis of Variance

Empirical models were created relating mean JNDs over test subjects to experimental factors for all signs. The Tukey-Kramer HSD analysis of stimulus content generally supported collapsing data across all three signs, while the analysis of the anti-aliasing method suggests different models for each method. Hence, models were built relating mean JNDs over test subjects to experimental factors for each anti-aliasing method over all three signs.

One might be tempted to include the 1-bit hinted reference, with its explicit zero JND value (and its implicit standard error of zero), in the dataset for each combination of Type and Pixel Angular Subtense. Inspection of the data supports exclusion of these points, as they introduced a quality discontinuity in the Bit Depth series. Introduction of a few gray levels initially degrades quality, improves as more levels are added, and finally levels off at some point. Hence, the hinted 1-bit reference value was not included in data for the models.

The unhinted 1-bit levels, which were actually rated, are a special case of the unhinted anti-aliasing method, where metric is not applicable. Inspection of the Bit Depth series data justifies inclusion of these levels in the case of unhinted methods. Because the unhinted 1-bit structure is poorer than the 1-bit hinted reference, no discontinuity is introduced when these levels are included in the Bit Depth series. A general model that includes bit depths from 1 to 4 for all unhinted text renderings can then be created.

Table 2 shows the model results. Factors listed in parentheses have a negative regression coefficient. Each linear experimental factor is significant for the hinted method. Because Pixel Angular Subtense correlates to pixel size, the coefficient for PAS is negative, i.e., larger pixel sizes (lower display resolutions) degrade quality, all else being equal. This model also has a negative quadratic term for Type Angular Subtense, resulting in limits and optima for type size for a given display resolution and viewing distance. Figure 1 shows the model trends as a function of the experimental factors for the hinted anti-aliased text. Curvature and interactions between factors are easily interpreted from this plot.

Table 2: Analysis of Variance Across All Signs

Data Set	Significant Factors ($p < 0.05$)	R^2	RMSE (JND)
Hinted AA	(PAS), TAS, BD, M (TAS*TAS), (BD*BD) TAS*PAS, BD*PAS, (BD*TAS)	0.99	0.14
Autohint AA	(PAS), TAS, BD (BD*BD), BD*PAS, (BD*TAS)	0.97	0.18
Unhinted AA	(PAS), TAS, BD (PAS*PAS), (TAS*TAS), (BD*BD), TAS*PAS	0.92	0.39
All Methods	(PAS), TAS, BD (PAS*PAS), (TAS*TAS), (BD*BD), TAS*PAS, (BD*TAS)	0.96	0.25

Each linear experimental factor, except Metric, is significant for the FreeType Autohinter method. There are higher-order terms for TAS or PAS, but only slight interactions with Bit Depth.

For the 1- to 4-bit unhinted renderings, the linear experimental factors that deal with visibility of inherent character structure are significant, namely PAS and TAS. Bit Depth is also significant, as the quality degradation of the 1-bit unhinted levels is included. The lack of hinting severely limits the potential of anti-aliasing.

A model was created collapsing all anti-aliasing methods using the means over all test subjects and signs. The many terms are indicative of the mixture of the methods whose mean JNDs are significantly different, with distinctly different text structure. If a number of rendering methods are to be used in a display, a better approach would be as follows. JNDs for each rendering method are calculated, which contribute to a weighted average JND for the display, where the weights are the usage frequency of each method.

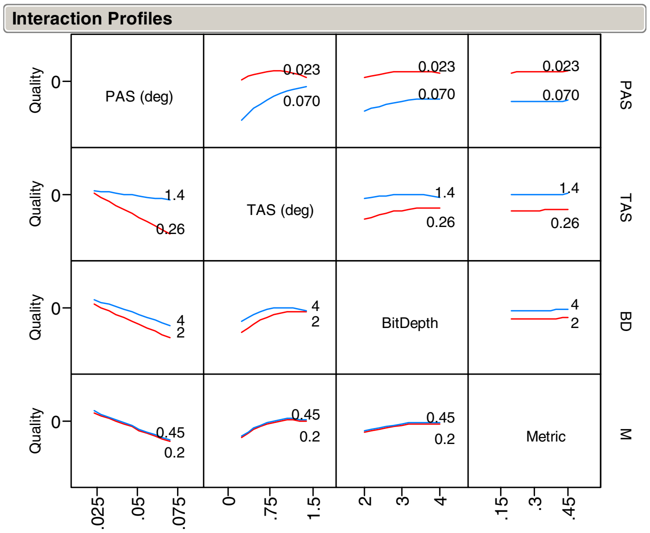

***Figure 1**. Model predictions for hinted anti-aliasing*

Discussion and Conclusions

Hinting improves inherent structure in 1-bit renderings of characters at lower resolutions. The higher mean quality ratings, and the significance of Bit Depth and Metric in the models for the hinted methods, indicate that anti-aliasing has the most effect when the underlying 1-bit structure is optimized. If characters are already spatially readjusted by hinting prior to down-sampling, then fewer character elements will alias. Conversely, the model for unhinted anti-aliased text must include factors that account for visibility of this degraded structure by introducing higher-order spatial terms.

Grayscale text can be used to improve text quality on low-resolution displays over a limited range of design parameters, type parameters, and viewing distances. Display resolution, type size, and bit depth were found to be significant factors for each anti-aliasing method. With display resolution, type size and viewing distances can be converted to angular subtense of the character and pixel, allowing application of these results to other viewing distances. Predictive models enable the optimization and design of information displays having limited contrast, brightness, and resolution. To illustrate, Table 3 shows the predictions for hinted grayscale Arial Text for a few applications.

These results show that grayscale text quality would degrade at high type and pixel angular subtense. Product designers should exercise caution with grayscale text for low-resolution displays intended to show large type sizes viewed closely. An outdoor electronic billboard would require large type in order to be seen at large distances, which if grayscaled, looks worse upon approach.

Table 3: Hinted Anti-aliasing Quality Model Predictions

System	DPI	Viewing Distance	Type Size (points)	Quality relative to 1-bit
E-Book	170	14 in.	8–10	Improvement
Point of Purchase Sign	80	12 in.	10–14	Degradation
	80	12 in.	16–18	No Difference
	80	36 in.	10–12	No Difference
	80	36 in.	14–18	Improvement
	50	12 in.	16–28	Degradation
	50	36 in.	16–28	No Difference

Application of font hinting significantly improves the quality of grayscale text, allowing smaller angular type subtenses to have equally perceived quality as slightly larger 1-bit and unhinted grayscale counterparts. It is highly recommended that hinting always be applied; however, the cost of sophisticated font rendering versus the resulting quality improvement would have to be considered for product design.

References

[1] E.M. Weisenmiller, "A Study of the Readability of On-Screen Text," Ph.D. Dissertation, Virginia Polytechnic Institute and State University, July, 1999.

[2] P. Oittinen, H. Saarelma, and K. Summanen, "The Quality of Text in Electronic Printing," Proc. SID, 28(4), pp. 459–464 (1987).

[3] D. Boyarski, C. Neuwirth, J. Forlizzi, and S. Regli, "A Study of Fonts Designed for Screen Display," Conf. Hum. Fact. Comput. Syst., pp. 87–94 (1998).

[4] M. Bernard, B. Chaparro, M. Mills, and C. Halcom, "Comparing the Effects of Text Size and Format on the Readability of Computer-Displayed Times New Roman and Arial Text," Intl. J. Hum. Comp. Stud, 59, pp. 823–835 (2003).

[5] G. Wyszecki, and W.S. Stiles, "Color Science: Concepts and Methods, Quantitative Data and Formulae, 2nd Edition" (John Wiley & Sons, New York, 1982) pg. 496.

[6] B.W. Keelan, "Handbook of Image Quality: Characterization and Prediction" (Marcel Dekker, New York, 2002) Ch. 5.

[7] D.H. Brainard, "The Psychophysics Toolbox," Spatial Vision, Vol. 10, pp. 433–436 (1997).

Author Biography

Scott O'Dell received a B.S. in Physics from Rensselaer Polytechnic Institute and is an alumnus of the Imaging Science & Technology Career Development (ISTCD) Program at Eastman Kodak Company. He is currently applying years of experience in imaging science, computer modeling and simulation, and image quality, as a Senior Research Scientist in the Display Science & Technology Group, Research & Development at Kodak.