A Predictive Model for Text Quality Analysis: Case Study

Ming-Kai Tse, Quality Engineering Associates (QEA), Inc., Burlington, MA, USA

Abstract

Text quality is a key aspect of overall print quality. Assessing text quality objectively and quantitatively has remained a challenge, despite our longstanding desire to reach this goal. The range of quality attributes is still seen by many as too broad and the definitions too vague and subjective. In this study, we aim to help overcome these obstacles by exploring whether key attributes exist that can be easily quantified and dependably correlated with subjective perceptions of print quality. If such attributes can be found, we believe a simple predictive model can be developed. For insight into which perceived attributes are critical and to help us select and design objective measurement algorithms, we started by conducting a subjective survey. Guided by the results, we performed quantitative stroke quality measurements and found good correlations between basic stroke properties (e.g., blurriness, stroke width and contrast) and the subjective survey results. We also found that text defects introduced complicating factors into the predictive model. This study provides the foundation for a more comprehensive future study.

Introduction

It has long been a goal of the imaging community to quantify image quality objectively and quantitatively. In this study, we address issues specific to text quality. Despite many previous studies [1-2], a definitive approach to objective text quality analysis remains a challenge due to the perceived complexity and multidimensional nature of the problem. A group of imaging professionals is actively working to develop an international standard as part of a larger program to develop a perception-based image quality standard [3]. Taking advantage of recent advances in image analysis equipment [4] and our long experience in objective image analysis, we aim to push the effort further with an instrumented approach and determine whether there are objective metrics that can predict subjective text quality preferences.

Subjective Survey

The author of this paper is an active member of the ISO Text Quality Working Group. Print samples prepared by the Working Group for its research were used independently in a survey conducted at QEA to investigate subjective text quality preferences. The results presented here are based on QEA's internal data and analysis and do not reflect opinions or conclusions of the ISO Text Quality Working Group.

The survey was conducted using 10 samples representing three different printing technologies (imagesetter, electrophotographic and inkjet) and a range of print quality. Each sample used a different substrate, a factor that may have affected perceived quality. Ten observers participated in the QEA survey and each performed a complete combinatorial pair-wise comparison of all 10 samples. During this process all the samples were supported on white cardboard, and the comparisons were conducted in a well-lit room (illuminated by daylight or fluorescent light) under normal viewing conditions and without the use of viewing aids other than the corrective lenses worn by some of the staff. Prior to each round of comparisons, each observer was asked to read a brief introduction to text quality analysis as well as some very high-level general guidelines for how it can be done (e.g., considering sample text quality at three different levels – characters, words, and paragraphs) [5].

In this paper, we use the ISO Working Group's original sample designations to simplify future reference. The samples are designated B3, B4, B8, F1, F8, F9, X1, X2, X6, and X9. At the end of each round of comparisons, observers were asked to comment on why they preferred one sample over another, and their responses provided insight into the subjective evaluation process. At the completion of the observer evaluations, the data were tallied, analyzed, and reported on a scale of 0 (least preferred) to 10 (most preferred).

The subjective survey results are shown in Table 1 and Figure 1. Within the ten-point scale, scores range from 0.78 to 8.17. Three of the samples (B3, B8 and X9) have significantly lower scores than the others, whose scores range from 5.5 to 8.17.

Table 1. Subjective Survey Results

Sample	Score (0 to 10)
X9	0.78
B8	1.06
B3	1.72
B4	5.50
X6	5.67
F9	5.83
F1	6.22
F8	7.44
X2	7.61
X1	8.17

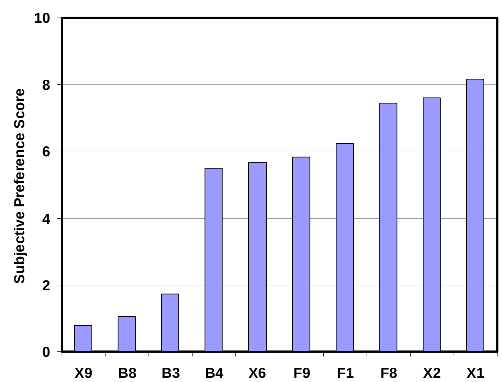


***Figure 1**. Subjective preference scores and rankings.*

Observer Comments and Sample Images

The observers' comments on the evaluation criteria they used were a useful indicator of which text quality attributes were seen as most important. Comments were received from eight of the ten participants, as follows:

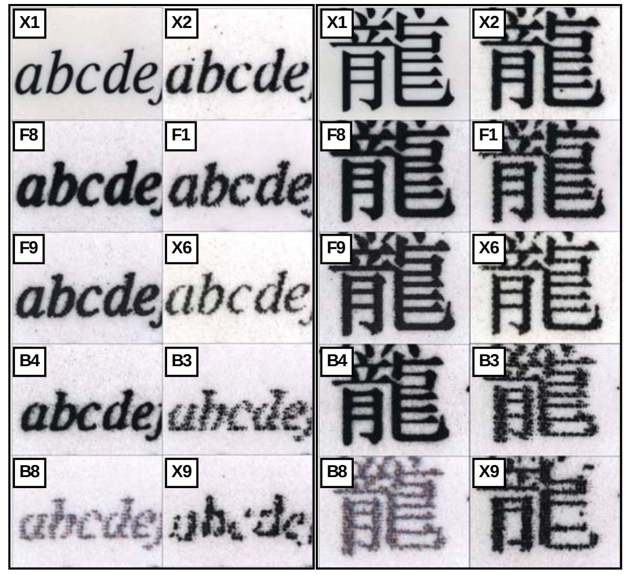
- Contrast is important to me; the clarity of the lines leaves a good impression of the printing.
- I like text to be dark and sharp. I think I was biased by the quality of paper (i.e., the roughness and tint).
- I looked at the sharpness of the 4 pt. fonts, the overall darkness and smoothness of slanted edges, and the gloss uniformity of the ink.
- Individual strokes should be sharp, distinct and continuous, without voids.
- I looked first at the large letters and then read the 'Quick Brown Fox' text to judge which sample was easier to read.
- Sample #2 (B8) is much too light. Samples #9 (X6) and #10 (B3) seem grainier than the others.
- I looked at these attributes in the same order with each comparison: sharpness, contrast, density, and stroke width.
- I decided based on the darkness of the print, the clarity of the fonts (particularly the small fonts), and whether thin lines showed clearly.

While there are differences among these comments, there are similarities as well. Indeed, it is apparent that they fall into three main categories of consideration:

1. Clarity, sharpness, distinctness (goodness measures)
2. Contrast, darkness, density (goodness measures)
3. Discontinuities, voids, graininess (defect measures)

What do observers really see and experience as they make preference decisions? Obviously, beyond a point it is impossible to know; but a qualitative impression is given by the images in Figures 2a and 2b. These are high-magnification images that highlight the appearance of individual characters.

The images in Figure 2 are arranged in descending order of observer preference, and it is safe to say that the most preferred (X1) and least preferred (X9) samples are consistent with what we see in the figure. Indeed, the assignment of B3, B8 and X9 to the last three positions also appears very reasonable. We see that despite a) the large difference in scale between the subjective-survey samples and the magnified images in Figure 2 and b) the difference in complexity between the letters and the Chinese characters, the rankings of sample quality come out the same. Therefore, it seems reasonable to assume that there are a set of critical image attributes that trigger consistent sensations and quality judgments in our minds. The question is, which ones are they? Our observers' comments may offer clues – contrast, darkness, density, clarity, sharpness, distinctness, etc. – that can provide a basis for designing objective analysis experiments. To test this, the logical first step is to measure stroke properties, including stroke width, edge quality (e.g., blurriness and raggedness), darkness, and contrast, attributes consistent with our observers' criteria for their subjective quality judgments.



(a) (b)

Figure 2. *(a) 4 pt Times Roman fonts and (b) 9 pt Chinese characters. High-resolution images at 4680dpi. Note that the images in (a) and (b) are arranged in descending order of observer preference (left to right, top to bottom).*

Objective Stroke Quality Analysis

The correspondence between the subjective rankings and quality judgments based on the magnified images in Figure 2 suggests that indeed there are critical attributes that determine our perception of text quality. If such critical attributes exist, wouldn't it be a boon to the industry if they could be measured quickly, easily and quantitatively? We begin our quest with an instrumented study of stroke properties.

Using a handheld image analysis system [6] with 5.5μm/pixel resolution, properties of vertical strokes in the characters "I", "L" and "T" in the 12 pt. Arial font set were measured. The stroke properties measured included stroke width, blurriness, raggedness, density, contrast, and fill (voids). The algorithms for determining these properties are based on an international standard for image quality measurements (ISO13660) [7].

The results are tabulated in Table 2, where the quality scores obtained in the subjective survey are included to facilitate the investigation of correlations between the subjective scores and objective measurements. The reported value of each attribute is the average of the measurements obtained for each character.

Table 2. Stroke Properties of "I", "L" and "T" in 12 pt Arial font

Sample ID	Subjective Score	Width (μm)	Blurriness (μm)	Raggedness (μm)	Density	Contrast
X1	8.17	373.7	74.2	0.5	0.76	0.80
X2	7.61	389.7	105.3	3.5	0.70	0.79
F8	7.44	430.7	110.5	3.5	0.69	0.78
F1	6.22	423.3	117.2	8.8	0.66	0.77
F9	5.83	419.3	115.2	6.0	0.56	0.69
X6	5.67	384.7	115.2	2.7	0.67	0.79
B4	5.50	369.7	122.0	3.8	0.68	0.77
B3	1.72	380.0	185.3	21.7	0.55	0.70
B8	1.06	310.7	202.7	18.7	0.32	0.47
X9	0.78	362.3	146.2	10.7	0.58	0.71

A total of 5 objective attributes are reported for each sample. It should be noted that while in principle these are all independent variables, empirically we found in this sample set a strong correlation between blurriness and raggedness and between density and contrast. Therefore, for the purposes of this analysis, we selected three of the five variables that are clearly independent: stroke width, blurriness and contrast. The questions to be answered are whether stroke width and blurriness relate to perceptions of clarity, sharpness and distinctness and whether stroke contrast relates to contrast, darkness and density – the first two categories of observer criteria in our survey. Figure 3 shows the empirical correlation between subjective scores and the three objective attributes.

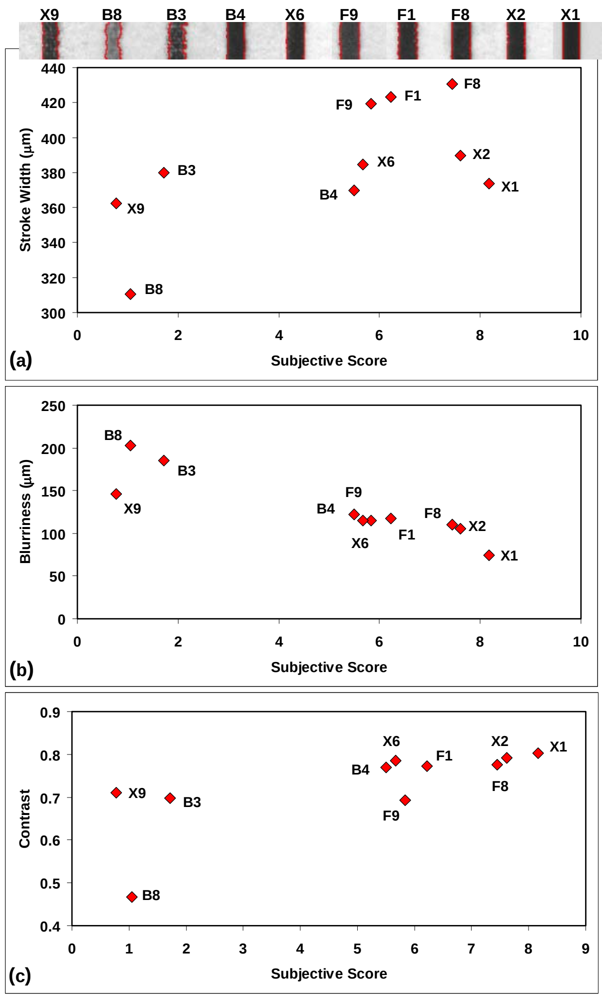


__Figure 3.__ Correlations between stroke properties - a) stroke width, b) blurriness, and c) contrast – and subjective scoring.

Qualitatively, in Figure 3 we can see a clear negative correlation between blurriness and subjective scores, and a weaker positive correlation between stroke width or contrast and subjective scores. The strong correlation between blurriness and subjective scores is particularly noteworthy (R^2=0.82). In other words, the measured blurriness of a text stroke may in fact be a good predictor of subjective perceptions of sharpness or lack of sharpness. In fact, focusing on Figure 3b, if we exclude sample X9 (for reasons to be elaborated later), the R^2 improves further to 0.95! Therefore, without excluding the possibility that other factors are also involved in our judgment of text quality, we must conclude that edge blurriness is an important factor in perceived text quality. The other two variables, stroke width and contrast, may play a somewhat secondary role.

A Predictive Model for Perceived Text Quality

For the purposes of developing a rudimentary predictive model based on the selected objective text quality attributes, we performed a linear regression (least-squares method) between the subjective scores and the three objective measures. The resulting model is:

$$\text{Score} = -37.7\,\text{mm}^{-1}\times \boldsymbol{B} + 5.38\times \boldsymbol{C} + 16.4\,\text{mm}^{-1}\times \boldsymbol{W} \quad (1)$$

where

B = edge blurriness, mm
C = stroke contrast
W = stroke width, mm

Note that in the above model, sample X9 is again excluded from the regression.

To illustrate the application of the empirical model, the "predicted" scores using the original objective attribute values are shown in comparison with the original subjective scores in Figure 4. The intent of this figure is not to prove the validity of the model, but to check the reasonableness of the linear regression and the methodology in utilizing the objective measurements for text quality prediction purposes.

Limitations of the Stroke Property-Based Model

Again, we have seen that there are three main categories of criteria that our observers used in the subjective text quality survey. The first (clarity, sharpness and distinctness) and second (contrast, darkness and density) have been discussed. The third category (discontinuities, voids, graininess) suggests the need for some sort of defect measurement. Just what are text defects? Clearly, there is no simple answer, but some common defective conditions are: poor formation, missing serifs, jitter, voids, distortion, unattractive character spacing, and the like. We believe such defects are key factors in the perception of text quality, but devising objective analysis algorithms has been a stumbling block. As an illustration, let us take a closer look at our outlier, sample X9. Several factors distinguish X9:

1. X9 has appreciably larger stroke width, lower blurriness and higher contrast than B3 and B8, yet its subjective score is lower than that of the other two samples.
2. We obtained the results shown in figure 5 with the handheld image analysis system and using the "slant edge" analysis technique [7] to measure the MTF (modulation transfer function) of the samples. The MTF ranking of the samples is consistent with the subjective ranking, with the exception of X9. X9 has better MTF than B3 and B8, but once again it ranked lower.
3. In Figure 2, a close examination of the 4 pt letters and 9 pt Chinese characters in X9 clearly suggests that X9 has a character formation problem, i.e., the reasonable stroke properties and MTF notwithstanding, the printer has problems forming characters and text. Consequently, observers ranked X9 lower than all other samples.

4. Figure 6 is another illustration of the poor formation, or text defect problem, in sample X9. In this figure, the outer contours of a 10 pt Arial “B” are shown. Sample X1 was rated the best and sample X9 was rated the worst: X9 shows conspicuous edge roughness and distortion.

The two categories of goodness measures (clarity, sharpness, distinctness; and contrast, darkness density) enhance our perception of text quality. In contrast, defects such as distortion, voids, graininess, poor formation, erosion of serifs and corners, and the like detract from our perception of text quality.

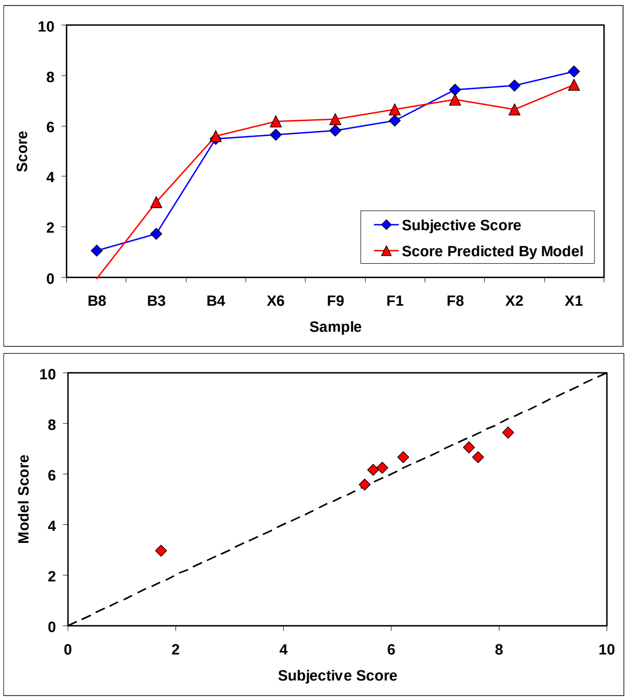


Figure 4. Developing a predictive model of text quality based on objective stroke quality measurements.

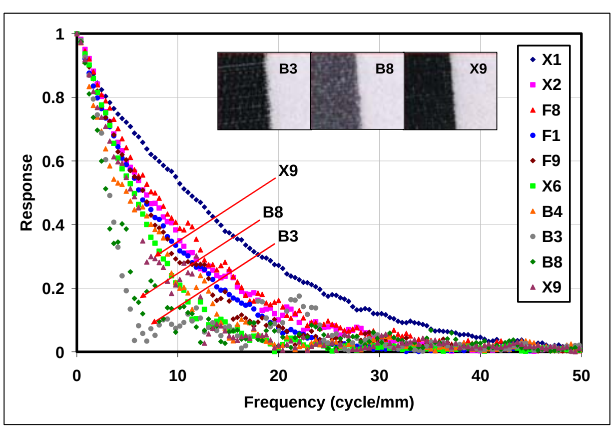


Figure 5. MTF of all samples measured using the "slant edge" method.

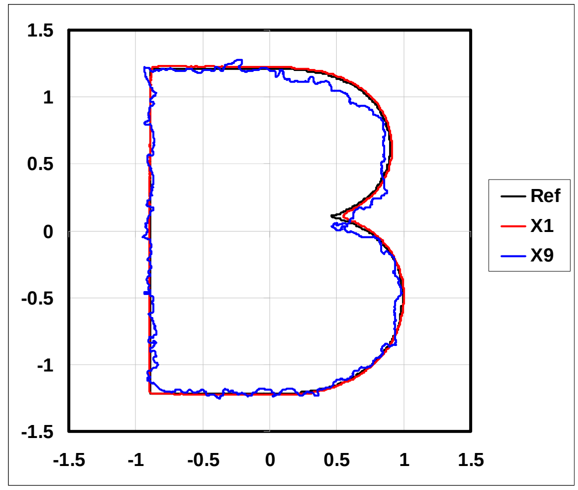


Figure 6. Outer contours of 10pt Arial "B" - comparing X9 with X1 and the original electronic reference.

Summary

1. This study aims at discovering a simple method for objectively evaluating text quality and predicting subjective text-quality preference.
2. A subjective survey was conducted using 10 samples of varying print quality and produced with several different printing technologies. Subjective scoring served as a reference for objective analysis results.
3. Comments by survey participants provided valuable insights into the factors influencing participants’ preferences, e.g., clarity, sharpness, distinctness; contrast, darkness, density; and defects such as voids and graininess.
4. Guided by the results of the subjective survey, the same test samples were measured quantitatively using a portable, high resolution (5.5µm/pixel) image analysis system. Stroke properties measured included width, blurriness, raggedness, density and contrast.
5. Quantified edge blurriness was found to exhibit the highest correlation with subjective scores. Similar but weaker correlations exist for stroke width and contrast.
6. A linear regression model was obtained to demonstrate a simple predictive model of subjective text quality preference. The model is based on straightforward stroke property measurements.
7. The identification of key attributes influencing subjective perception of text quality and the simplicity of the predictive model developed are very encouraging. However, much more work is needed to expand the effort to encompass other controlling factors such as character and text defects.
8. Text defects such as lack of fidelity, distortion, poor formation, and erosion of serifs, etc. are critical factors that detract significantly from our perception of text quality. Developing simple, practical quantitative measures for such defects would be an important step toward the goal of a robust predictive model.
9. This study is our first step toward building a more comprehensive text quality prediction model.

Acknowledgment
The author would like to thank the INCITS W1.1 Line and Text Quality Ad Hoc Group, which is working towards the ISO19751 Perceptual Image Quality Standard, for providing a stimulating environment for our many valuable discussions and insights. The Ad Hoc Group Chairman is Dr. Edul Dalal of Xerox Corporation. The author appreciates the opportunity to use the sample set provided by the working group in an independent internal study which produced the subjective evaluation results reported in this paper.

References
[1] T.Itoh, S. Hirota, Evaluation of Legibility, PICS, p.213. (1999).
[2] M.Sato, Analysis of Quantitative Indicators of Kanji Text, PICS, p.218. (1999)
[3] INCITS W1.1 Line and Text Quality Working Group for the ISO19751 standard.
[4] Image analysis systems by QEA. www.qea.com or info@qea.com
[5] Instructions for subjective survey provided by the W1.1 Line and Text Quality Working Group (2007)
[6] PIAS™-II A Second Generation Portable Image Analysis System, QEA. www.qea.com or info@qea.com
[7] ISO/IEC13660:2001 Information Technology - Office equipment- Measurement of image quality attributes for hardcopy output-Binary monochrome text and graphic images
[8] ISO 12233:2000 Photography-Electronic still-picture cameras-resolution requirements.

Author Biography
Dr. Ming-Kai Tse is the founder and president of Quality Engineering Associates (QEA), Inc. in Burlington, MA, USA. Prior to founding QEA, Dr. Tse was an Associate Professor of Mechanical Engineering at the Massachusetts Institute of Technology (MIT). His research interest is in measurement technology, with special application interest in digital printing and image quality analysis. He earned his BS at Cornell University and his SM and PhD at MIT. His is a Senior Member of IS&T. He can be reached at info@qea.com or +1-781-221-0080.