A Report on a Subjective Print Quality Survey Conducted at NIP16

Gretchen Gast and Ming-Kai Tse
Quality Engineering Associates (QEA), Inc.
Burlington, Massachusetts, USA

Abstract

We conducted a subjective survey on image quality using a set of photographs printed on commercially available inkjet printers. The purpose of the survey was to gain first hand insight into customer preferences in digital photographic printing.

The observers in this survey were participants at the IS&T NIP16 Conference and the Diamond Research Corporation's Digital Imaging Conference. Both conferences took place in October 2000. In this survey, an image of a newlywed couple was used and six attributes were studied including: image defects such as blur (unsharpness), noise (graininess), and banding (effect of a clogged nozzle); personal preferences such as color rendition and tone reproduction; and finally, printer type. The observers were asked to rank order four images of different quality levels for each of the six attributes.

The results show that: a) the human visual system is very acute at detecting blurriness in an image, b) the presence of image noise in luminance is much more detectable than in the color channels, and c) banding due to missing cyan ink and yellow ink appears to be more readily detectable than banding due to missing magenta. In terms of color rendition and tone reproduction, a greenish cast is objectionable to most survey participants and darker images are preferable to lighter ones. In terms of printer type, the results suggest that the participants have a consistent preference for images from certain brands of printers over others. In this paper, the design of the experiment and the subjective analysis results will be discussed in detail.

The significance of this subjective survey is in the unusually large number of participants (close to 130) and also the worldwide representation of the participants (12 countries). We believe that the methodology used in this study and the survey results should be of interest to most in the business of digital imaging.

Introduction

QEA has noticed for some time through quick informal experiments that observers have a consistent preference for certain images. Recently a systematic subjective print quality survey was conducted to investigate which qualities most affect personal preference and to gain some idea as to the preferences of the majority. We have also investigated potential factors influencing personal preference, such as cultural differences. The survey can be considered as a popularity contest and the report shows which prints won the popular vote and by how much.

Survey Design

An image of a newlywed couple was used and a total of six attributes were studied. The five perceptual attributes are:

- Blur
- Noise
- Banding
- Color rendition
- Tone reproduction

The final attribute studied was:

- Printer type

Each print (stimulus) is assigned a letter (A-D) for identification. A set of four prints was used where one print was unadjusted (control) and the other three were altered to varying degrees for each perceptual attribute.

The observers in this survey were participants at the IS&T NIP 16 conference held in Vancouver BC and the Diamond Research Corporation's Digital Imaging Conference held in Santa Barbara, California. A total of 130 observers participated in the study. Data from four observers were excluded due to incompleteness in the response.

To create the first five attributes, the original image was manipulated in Adobe Photoshop: blur and noise were added using the appropriate filters; banding was created by "subtracting" cyan, magenta, or yellow lines periodically to simulate a clogged or misdirected jet in a printer; and color and tone were adjusted using the curves function in Photoshop. The original image, left unaltered, was the control for each print set. The level of adjustment was intended to be enough for the average observer to detect differences, while not so much so that the control image would always be preferred.

The printer type set was created using four top-of-the-line printers for the consumer market at that time, including the one used to print the other five attributes. For all the prints, the best photo-glossy paper recommended by the printer manufacturer for each of the four printers was used.

In terms of print quality control, we used the "best mode" printer driver settings without any additional image processing or color management.

The prints were placed in plastic sleeves and presented to the user in a binder. Each observer was asked to rank order the prints in order of personal preference according to each given attribute. Observers were asked to fill in 1 – 4 for the six print sets on a form provided, 1 being the best and 4 being the worst.

Results and Analysis

The results are shown in the following tables and graphs organized according to the attributes studied. For each print a subjective score was calculated using Equation 1.

$$\text{Subjective Score} = \frac{4a + 3b + 2c + d}{n} \tag{1}$$

Where

n = the number of observers (in this case 126)
a = number of observers who ranked each print as 1^{st}
b = number of observers who ranked each print as 2^{nd}
c = number of observers who ranked each print as 3^{rd}
d = number of observers who ranked each print as 4^{th}

This is similar to the computation for a GPA. Therefore, a score of 4.0 would mean that every observer ranked that specific print as the best or 1st, and a score of 1.0 would mean that every observer ranked that print as the worst or 4th.

The distribution of the data can be seen in the graphs of the values a, b, c, and d from Equation 1.

Further, we analyze the data to explore the presence of any correlation with cultural differences, sorting the data into the categories of American, Asian, and European.

Blurriness

Starting with the control image and adding the blur filter in Adobe Photoshop created the prints for blurriness.

A = Control (No blurring)
B = Slight (Blur filter applied once)
C = Medium (Blur filter applied twice)
D = Severe (Blur filter applied three times)

As shown in Figure 1 the control print was the most preferred. The severely blurred print was the least preferred. There is a clear majority preference as demonstrated by the narrow distributions. This suggests that the sensitivity of the human visual system to blurriness is very high.

Table 1. Subjective Score for Blurriness.

Print ID	A	B	C	D
Blur Magnitude	Control	Slight	Medium	Severe
Subjective Score	3.6	2.9	2.3	1.3

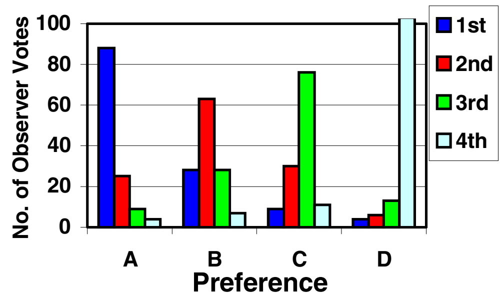

Figure 1. Blurriness Results Plot

Blurriness: Cultural Analysis

There was no clear cultural difference for blurriness as shown in Table 2.

Table 2. Cultural Subjective Score Breakdown for Blurriness.

Print ID	A	B	C	D
Blur Magnitude	Control	Slight	Medium	Severe
American	3.6	2.9	2.3	1.3
European	3.5	3.1	2.2	1.3
Asian	3.4	2.7	2.6	1.4

Image Noise (Graininess)

The noise prints were created by adding noise using the Photoshop filter at 10% to each of the three channels, a*, b*, and L*.

A = 10% L* (10% noise added to the L* channel)
B = 10% b* (10% noise added to the b* channel)
C = 10% a* (10% noise added to the a* channel)
D = Control (No noise)

As shown in Figure 2, print A is clearly the majority's choice for the 4^{th} place ranking. In other words, L* noise is highly objectionable. The data shown here reflects the higher spatial resolution of our visual achromatic channel.[1] There is less perceived difference between the votes for prints B, C, and D.

Table 3. Subjective Score for the Graininess Image Defect.

Print ID	D	B	C	A
Noise Magnitude	Control	10% b*	10% a*	10% L*
Subjective Score	3.1	2.9	2.6	1.4

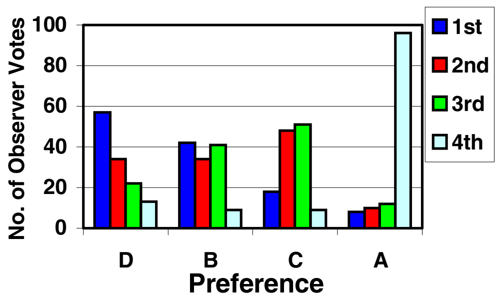

Figure 2. Image Noise Results Plot

Image Noise: Cultural Analysis

There was no clear cultural difference for image noise as shown in Table 4.

Table 4. Cultural Subjective Score Breakdown for Image Noise.

Print ID	D	B	C	A
Noise Magnitude	Control	10% b*	10% a*	10% L*
American	3.0	2.8	2.7	1.5
European	3.2	2.9	2.5	1.5
Asian	3.3	3.0	2.5	1.3

Table 5. Subjective Score for Banding.

Print ID	C	A	D	B
Banding Type	Control	Missing M	Missing C	Missing Y
Subjective Score	3.6	3.0	1.8	1.6

Banding (Simulated Clogged Inkjet Nozzle)

"Subtracting" the specified color from the image in lines every 3 mm created the banding images.

A = Missing Magenta (Simulated missing magenta ink)
B = Missing Yellow (Simulated missing yellow ink)
C = Control (No banding)
D = Missing Cyan (Simulated missing cyan ink)

The data indicates that banding due to missing magenta is less objectionable than missing cyan or yellow, as seen in Table 5 and Figure 3. Prints C and A have the clear majority of votes for 1st and 2nd ranks respectively. Prints D and B are more evenly distributed between ranks 3rd and 4th.

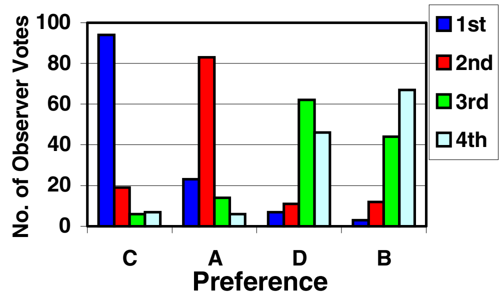

Figure 3. Banding Results Plot

Banding: Cultural Analysis

There was no clear cultural difference for banding as shown in Table 6.

Table 6. Cultural Subjective Score Breakdown for Banding.

Print ID	C	A	D	B
Banding Type	Control	Missing M	Missing C	Missing Y
American	3.5	3.0	1.8	1.8
European	3.6	2.9	2.1	1.4
Asian	3.8	3.0	1.7	1.8

Color Rendition

The color adjustment made is described by how many digital counts are added to the red, green, or blue curve in Photoshop at the midpoint of the curve to manipulate the color of the image. The endpoints of the curve remained anchored. For example 0, +8, 0 (R, G, B) means that 8 digital counts were added to the image in the green curve at a digital count of 128, giving this image a slightly greenish cast.

A = 0,+8,0 (Green cast)
B = Control (No adjustment)
C = 0,+8,+8 (Cyan cast)
D = 0,-8,0 (Magenta cast)

The results show that print B, the control, was the most preferred. This is seen in Figure 4. Print A was the least preferred overall. Print D was ranked as 3rd overall. Figure 4 illustrates that the color rendition data has a more even distribution. The control print is the only one that has a clear majority vote. Prints D and A, which are opposites in hue shift, are very similar in distribution. This can be seen in the same figure.

Table 7. Subjective Score for Color Rendition.

Print ID	B	C	D	A
Color Adjustment	Control	0,+8,+8	0,-8,0	0,+8,0
Subjective Score	3.4	2.5	2.2	2.0

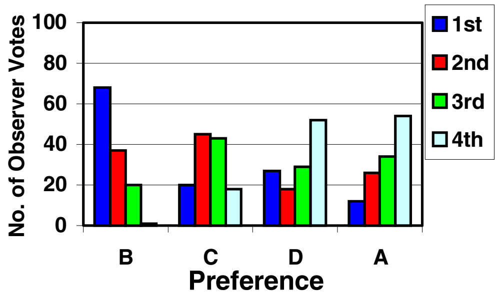

Figure 4. Color Rendition Results Plot

Color Rendition: Cultural Analysis

Cultural differences in the preference of color rendition are not surprising.[2] As seen in Table 8, the European observers appear to object strongly to the magenta cast in print D, while the Asian observers preferred print D and voted the greenish cast of print A as the least preferred.

Table 8. Cultural Subjective Score Breakdown for Color Rendition.

Print ID	B	C	D	A
Color Adjustment	Control	0,+8,+8	0,-8,0	0,+8,0
American	3.4	2.5	2.2	1.9
European	3.3	2.7	1.6	2.4
Asian	3.4	2.5	2.8	1.5

Tone Reproduction

Adding, or subtracting, digital counts at the midpoint of all three curves (R,G,B) in Photoshop created the tone reproduction prints. The endpoints of the curves were anchored. For example, the lightest image was created by adding 8 digital counts to the RGB curves at a digital count of 128, lightening the overall appearance without applying any hue shift.

A = Darkest (-16)
B = Lightest (+8)
C = Medium Dark (-8)
D = Control (No adjustment)

Table 9 shows that print C was preferred overall, although looking at Figure 5, the votes are closely split between prints A and C for 1st place. The somewhat bimodal distribution of print A brought the subjective score down. Print B (lightest) was the least preferred overall. The preference for a darker print may be due to the print giving the impression of having a higher contrast and therefore appearing sharper. The distributions, shown in Figure 5, indicate less perceived difference overall.

Table 9. Subjective Score for Tone Reproduction.

Print ID	C	A	D	B
Tone Adjustment	Med. Dark	Darkest	Control	Lightest
Subjective Score	3.0	2.7	2.4	1.9

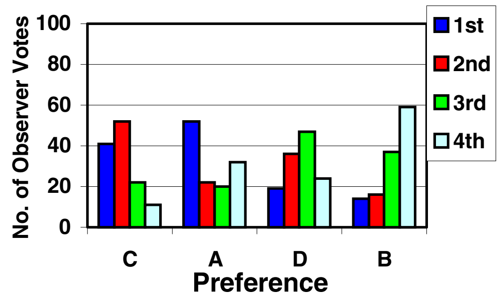

Figure 5. Tone Reproduction Results Plot

Tone Reproduction: Cultural Analysis

As seen in Table 10, the largest difference appears in the subjective scores of print A, where the American and European observers show more of a preference than the Asian observers.

Table 10. Cultural Subjective Score Breakdown for Tone Reproduction.

Print ID	C	A	D	B
Tone Adjustment	Med. Dark	Darkest	Control	Lightest
American	3.0	2.8	2.4	1.8
European	2.8	2.9	2.4	1.9
Asian	3.1	2.3	2.5	2.0

Printer Type

Printing the control image on four commercially available printers, which were all within a similar price range, created the prints for the printer type. The best quality photo glossy paper recommended by the printer manufacturer was used. The "best mode" or "highest quality" printer driver setting was used.

Although the prints were from similarly priced printers and on the recommended best quality paper, Table 11 shows that there is a clear majority preference for printer A. Printer C was the least preferred. Printer B has a bimodal distribution, seen in Figure 6. Printer D has more or less a Gaussian distribution.

Table 11. Subjective Score for Printer Type.

Printer ID	A	B	D	C
Subjective Score	3.3	2.7	2.4	1.7

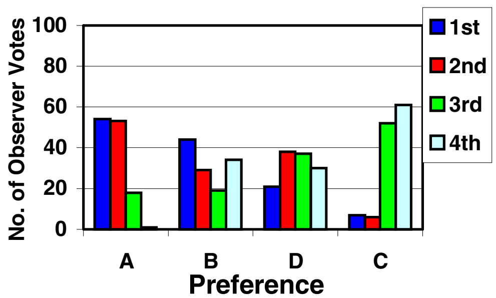

Figure 6. Printer Type Results Plot

Printer C was very light compared with the other prints. As a result, the color was unsaturated and the overall contrast was low.

The slight greenish cast of printer D may have contributed to printer D's lower ranking.

A possible explanation for printer B's bimodal distribution is the presence of graininess. Although the tonal and color quality of printer B was good, some observers might have been responding negatively to the graininess, while others either did not perceive the graininess, or did not strongly object to it.

Printer: Cultural Analysis

The largest difference appears in the distinction between printers B and D. The American observers had a stronger preference for B and much less of a preference for C, where the European and Asian observers were more evenly distributed.

Table 12. Subjective Score Cultural Breakdown for Printer Type.

Printer ID	A	B	D	C
American	3.3	2.9	2.3	1.6
European	3.1	2.5	2.6	1.8
Asian	3.4	2.4	2.6	1.7

Summary

The following observations can be made from this subjective print quality survey:

- The human visual system is very acute at detecting blurriness in an image and there appears to be no correlation with cultural difference.
- The presence of noise in the luminance channel is the most detectable, whereas noise in the a* and b* is less so. Once again, there appears to be very little cultural difference.
- Most observers picked out missing cyan and yellow more readily than missing magenta. No cultural difference was observed in the banding data set.
- Most observers, out of the prints presented, dislike a greenish cast in the image, and magenta is next in the rank order. Cultural difference is very strong in color preference.
- Most observers appear to prefer a darker image than a lighter one. This may be the result of a darker image giving the impression of having a higher contrast and more sharpness. There is some degree of cultural preference in the tone of an image.
- Observers consistently show a preference for the images from certain brands of printers over others.
- While most of the observations listed above may have been reported by various researchers, the significance of this subjective survey is in the large number of participants (close to 130) and also the worldwide representation of participants (12 countries). We believe that the methodology used in this study and the survey results should be of interest to most in the business of digital imaging.

Acknowledgments

We are grateful to all our colleagues and friends who participated in the survey. We would especially like to thank John McCann for his advice and assistance in the survey.

References

1. Roy S. Berns, Principles of Color Technology, John Wiley & Sons, Inc., 2000.
2. Mark D. Fairchild, Color Appearance Models, Addison Wesley Longman, Inc., 1998.

Biography

Gretchen Gast is an Applications Engineer at Quality Engineering Associates (QEA), Inc. Ms. Gast received her B.S. degree in Imaging and Photographic Technology from the Rochester Institute of Technology in May 1999. She joined QEA in September of 2000.