The Effect of Opponent Noise on Image Quality

Garrett M. Johnson* and Mark D. Fairchild
Munsell Color Science Laboratory, Rochester Institute of Technology
Rochester, NY 14623

ABSTRACT

A psychophysical experiment was performed examining the effect of luminance and chromatic noise on perceived image quality. The noise was generated in a recently developed isoluminant opponent space. 5 spatial frequency octave bands centered at 2, 4, 8, 16, and 32 cycles-per-degree (cpd) of visual angle were generated for each of the luminance, red-green, and blue-yellow channels. Two levels of contrast at each band were examined. Overall there were 30 images and 1 "original" image. Four different image scenes were used in a paired-comparison experiment. Observers were asked to select the image that appears to be of higher quality.

The paired comparison data were used to generate interval scales of image quality using Thurstone's Law of Comparative Judgments. These interval scales provide insight into the effect of noise on perceived image quality. Averaged across the scenes, the original noise-free image was determined to be of highest quality. While this result is not surprising on its own, examining several of the individual scenes shows that adding low-contrast blue-yellow isoluminant noise does not statistically decrease image quality and can result in a slight increase in quality.

Keywords: Image quality, Chromatic Noise, Contrast Sensitivity, Image Difference Modeling

1. INTRODUCTION

The study of image quality (IQ) has traditionally focused on two distinct approaches; a systematic modeling of individual perceptions combined into a computation image quality metric, and through the use of human vision based models. While these techniques are not mutually exclusive, they do represent tackling the same fundamental problem from two different sides. The Image Quality circle presented by Engledrum provides an admirable description of these IQ modeling techniques.[1]

The systems-based approach to IQ modeling typically takes several measurable attributes, such as pixel-addressability, luminance range, system MTF, streaking, noise, and grain and scales those directly against perceived image quality. These scales can then be combined with some sort of weighting scheme to generate a metric of overall image quality. The human vision approach typically feeds images into a computation model with no knowledge of where the images came from, and attempts to predict the perceived differences between the images. This difference can then be related to a measure of change in image quality. The modular framework presented by Johnson and Fairchild[2] and S-CIELAB[3] represent two of these types of visual models.

Many researchers have studied the effect of noise on image quality, usually through measuring variation on solid patches. In a seminal paper by Dooley and Shaw[4] they proposed a metric that integrated the Weiner noise power spectrum with properties of the human visual system, specifically the contrast sensitivity function. A similar approach that measured the variance of filtered noise patches was used successfully by Johnson *et al* to predict overall image quality.[5] Recently Kuang *et al* proposed a new noise metric based upon optimized weightings of the CIELAB color difference equations.[6]

All of these approaches have shown a degree of success in regards to predicting the effect of noise on image quality, for uniform patches. The effect of noise perception in existing complex images cannot be addressed by these systematic approaches, as they all require the measurement of patch data. In order to understand the relationship these systematic metrics have with the perception of overall image quality, we need to have a better understanding of noise in complex images. This study attempts to address this problem through psychophysical experimentation. The goal is to use the data on chromatic noise perception in complex images as a means of developing and testing future computation models of image quality and image differences.

---

* garrett@cis.rit.edu, http://www.cis.rit.edu/people/faculty/johnson/

1.1. Opponent color spaces

Most models of the human visual system rely on spatial filtering in an opponent color space. Oftentimes image encoding itself is done in a similar opponent type space, typically black-white, red-green, and yellow-blue. The choice of opponent space can also have an effect on overall image quality, especially if there is image compression or if noise is introduced into channels. This experiment was performed using a recently introduced opponent space that was designed to be orthogonal and isoluminant. Additional information on color space encoding can be found in Poyton[7] and Song *et al.*[8]

2. EXPERIMENTAL DESIGN

A paired comparison experiment was performed scaling the effect of additive opponent noise on perceived image quality. Noise was generated in opponent color space representing luminance, red-green, and yellow-blue channels. The noise was filtered using 5 octave band filters centered at 2, 4, 8, 16, and 32 cycles-per-degree of visual angle. Three levels of "contrast" were examined for each channel, and four scenes were used.

2.1 Stimulus generation

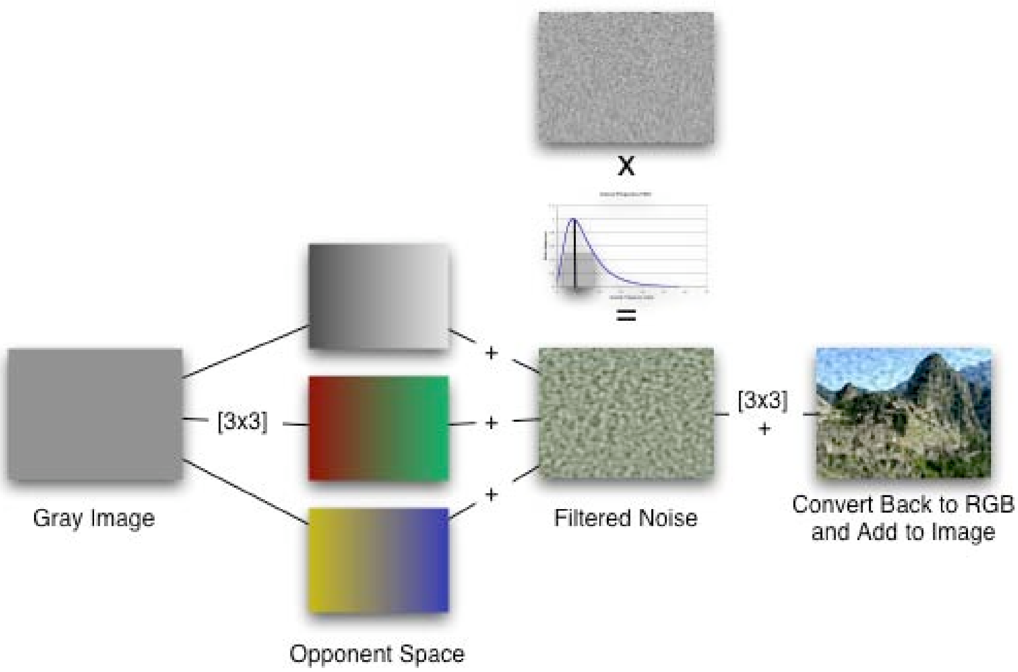

Figure 1: Procedure followed to generate image stimuli

The images used in the experiment were generated as described with the flowchart in Figure 1. The first stage was to generate noise patterns. A uniform gray patch (D65) was first transformed into the opponent color space using Equation 1. This color space was specifically designed to be orthogonal and isoluminant, meaning that no luminance information is contained in the red-green or yellow blue channels.[8] Additionally the yellow-blue axis falls on unique yellow, while the red-green axis is orthogonal to that.

$$\begin{bmatrix} Y' \\ r-g \\ y-b \end{bmatrix} = \begin{bmatrix} 0.0556 & 0.9981 & -0.0254 \\ 0.9510 & -0.9038 & 0 \\ 0.0386 & 1.0822 & -1.0276 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{D65} \qquad (1)$$

Random noise was then added to one of the opponent channels. This noise was generated by creating a uniform "white" noise pattern and filtering with an octave filter in the frequency domain. An octave filter is essentially a Gaussian filter in a log space, designed such that the half-width is at half and twice the center frequency. Figure 2 shows an example octave filter, centered at 8 cycles-per-degree (cpd) of visual angle, with a half-width at 4 and 16 cpd respectively. Five such filters were used, centered at 2, 4, 8, 16, and 32 cycles-per-degree. It is important to stress that these frequencies are only valid for a specific display device and viewing condition. Figure 3 shows red-green opponent noise at the five spatial frequency bands.

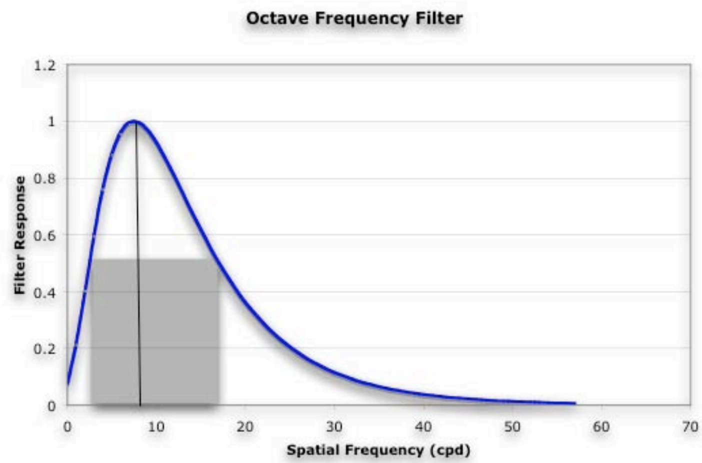

Figure 2: Example octave filter centered at 8 cycles-per-degree of visual angle

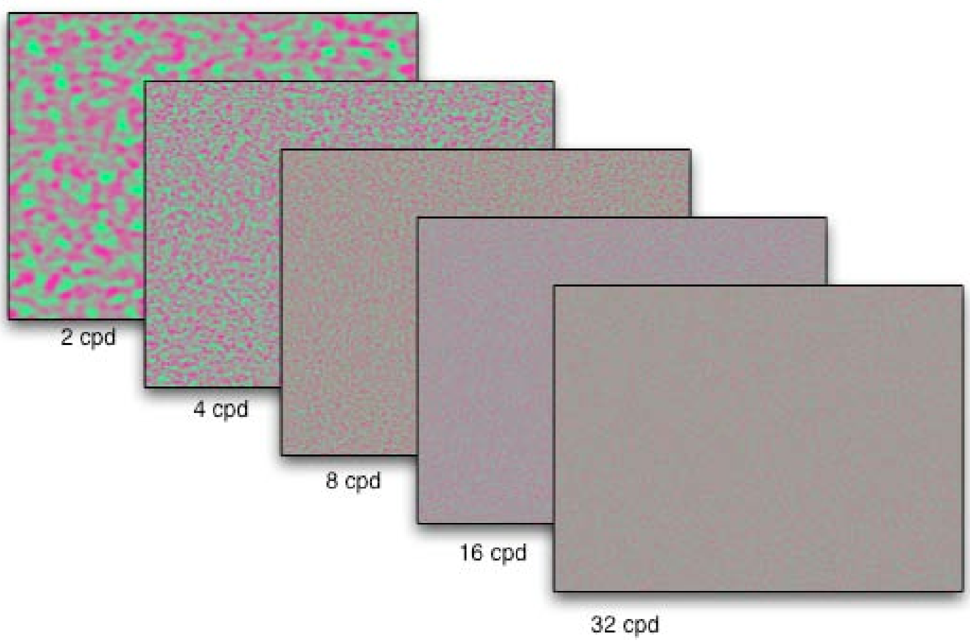

Figure 3: Red-green opponent noise at 5 frequency bands

The opponent noise was then transformed back to XYZ tristimulus values and then converted to display RGB using the device characterization. The RGB noise remained in floating point precision, and ranged from approximately -0.5 to 0.5. The mean of the noise was zero, so that when added to an image it would not change the overall mean of the image. The RGB noise was then scaled to "contrast" ratios representing 5%, 10% and 15% 8-bit digital counts for the luminance noise, and 10%, 20%, and 30% digital counts for the red-green and yellow blue noise. The noise was then quantized to 8-bits, and added to the RGB images. Negative values and values greater than 255 were clipped when necessary. An example of luminance, red-green, and yellow-blue noise is shown in Figure 4. The noise was then added to the four scenes shown in Figure 5.

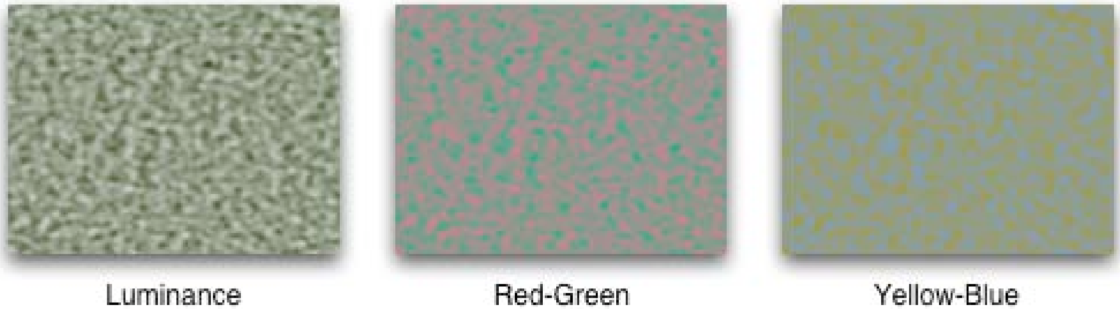

Figure 4: Example noise in the three opponent color channels. Note, this noise is designed to be isoluminant for a specific display characterized to D65

Figure 5: Scenes used in experiment.

2.1. Experimental procedure

A pilot experiment was run using all four scenes, with noise frequencies at 4, 8, and 16 cpd, and three contrast levels. There were 27 manipulations, and 1 original image. A paired comparison experiment was performed on an Apple 23-inch Cinema HD display, viewed at 24 inches. The maximum spatial frequency of the stimulus was approximately 40 cycles-per-degree of visual angle (80 pixels) and each image spanned 10 degrees of visual angle. The 28 images, and four scenes combined for 1512 pairs. Ten observers completed the pilot experiment and were asked to choose the image they thought to be of "higher image quality." Thurstone's law of comparative judgments was then used to generate an interval scale of perceived image quality, using Gulliksen's method of solving for unanimous decisions.[9]

From the pilot experiment it was determined that the extreme contrast levels (15% for luminance, 30% for chromatic channels) were too visible, and deemed too difficult to judge. A second experiment was then designed that removed the higher contrast images, and added higher and lower spatial frequency noise, centered at 2 and 32 cpd. The tiger was judged to be similar to the puffin image in the pilot experiment, and was removed to reduce the number of trials. For the second experiment there were 31 images (5 spatial frequencies, 2 contrast levels, 3 opponent channels) and 3 scenes, resulting in 1395 trials. Observers were again asked to select the image they thought was of higher quality, and were not given further specific details. On average the experiment took 45 minutes to complete. A total of 19 observers took part in the second experiment, drawing from a pool of experienced graduate students and scientists.

3. EXPERIMENTAL ANALYSIS

Thurstone's law was again applied to the second experiment, resulting in interval scales of quality. These scales are plotted for the average of all three scenes in Figure 6. Error bars were calculated using the method described by Montag.[10]

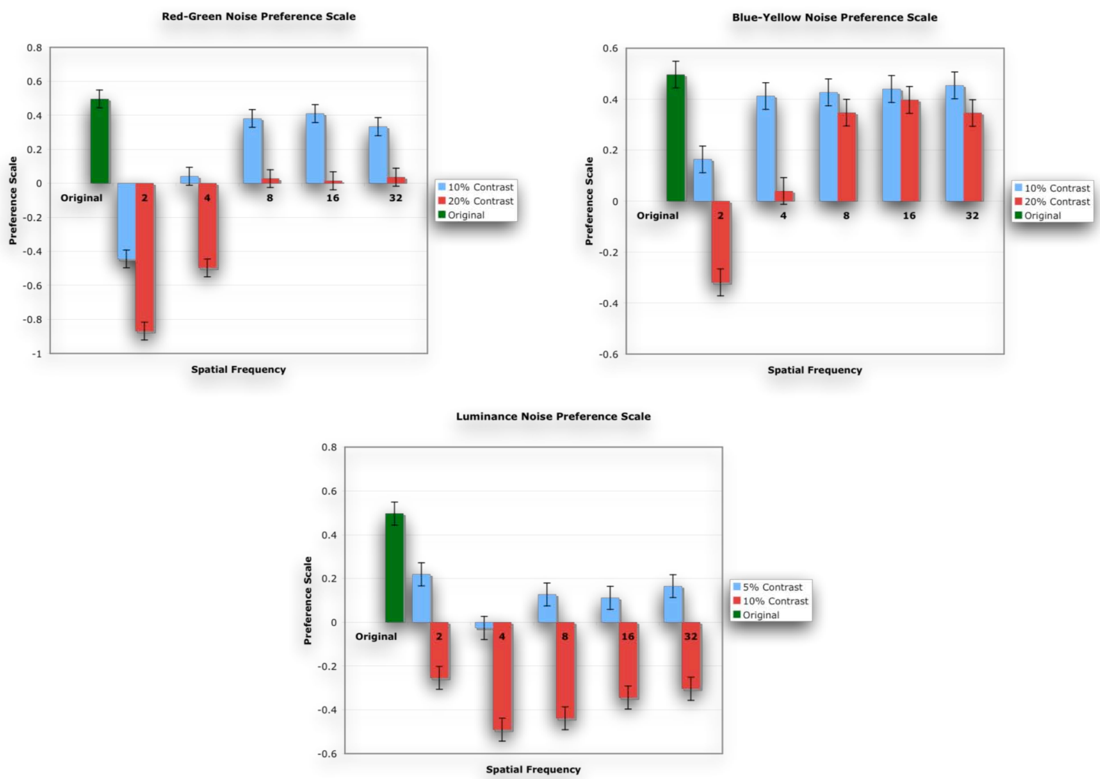

Figure 6: Interval scales of quality for red-green (top left), yellow-blue (top right) and luminance noise.

Examining these scales reveals some interesting trends. On average the low spatial frequencies bands, centered at 2 and 4 cycles-per-degree, are the most objectionable in all color channels and at all contrasts. The only exception to this trend would be the 4 cpd low-contrast yellow-blue noise, which is not statistically different than the original images. An interesting note was that for the two chromatic channels the most offensive image was judged to be lowest frequency, while for the luminance noise the most offensive image was the 4 cpd frequency band. This behavior follows what might be predicted from the human CSF, and is discussed in more detail below.

Overall, the most objectionable noise for all frequencies and opponent channels was found to be the high-contrast low-frequency red-green noise, followed by all of the high-contrast luminance noise. For both of the chromatic channels the higher frequency noise (above 8 cpd) was considerably less objectionable at all contrast levels. This is especially evident in the yellow-blue noise at low contrast levels as the only frequency band judged statistically different from the original image was the lowest (2 cpd) frequency band. At the higher contrast levels only the 2 and 4 cpd noise in the yellow-blue level was judged to be different in quality than the original, indicating that the high-frequency yellow-blue noise was perhaps not very visible. A similar trend holds for the low contrast red-green noise as well. The high-contrast noise, however, was significantly more detrimental to overall image quality at all spatial frequencies..

3.1 Scene dependencies

Figure 6 shows the interval scales (z-scores) averaged across all three scenes. Perceived image quality is often shown to be very image dependent. It is useful to determine whether this is true for noise perception as well. If the effect of noise on image quality were indeed scene dependent, then it would be impossible to get an accurate noise quality measurement using simple uniform patches. The individual interval scales for each scene are plotted against the average interval scale, in Figure 7, and the correlations between the individual scales and the average are shown in Table 1.

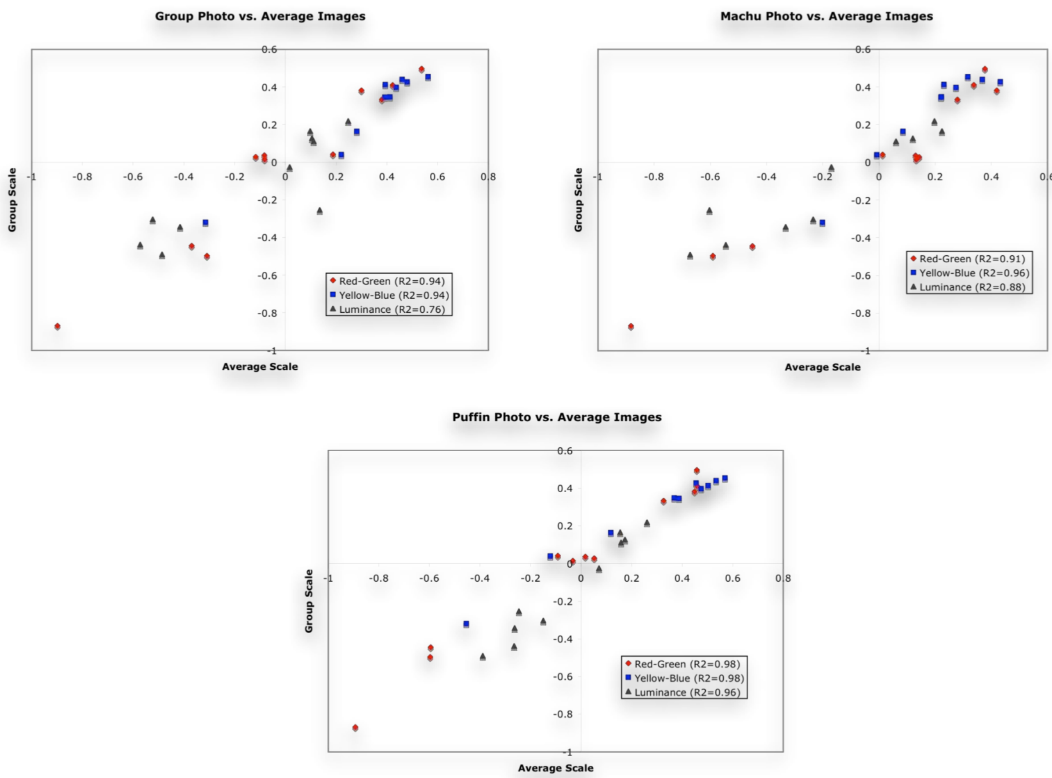

Figure 7: Examining scene dependencies. Each individual scene quality scale is plotted against the average scale

Table I: Correlation coefficients between individual scenes and average scale

Scene	Red-Green R^2	Yellow-Blue R^2	Luminance R^2
Group	0.94	0.94	0.76
Machu	0.91	0.96	0.88
Puffin	0.98	0.98	0.96

The plots in Figure 7 show that in general there is an acceptable correlation between any individual scene and the average of all the scenes. This is encouraging, as it would suggest that an image independent algorithm might be capable of predicting the effect of noise on image quality. The correlations coefficients are above 0.91 for all scenes for the chromatic channels, and generally higher. This indicates that the chromatic noise perception is similar across all the scenes. The luminance channel, however, reveals a different story. The correlation between the Group image and the average scene is only 0.76. The Group image, as seen in the top left of Figure 5, is a portrait containing several people with varying skin-tones. These experimental results suggest that the judgment of luminance noise is different when there are human faces involved. This is also true, though to a lesser degree, for the Machu image, which contains a large uniform blue-sky region. As shown in Figure 6 on average the luminance noise was judged to have the most detrimental effect on perceived image quality, and it also has the most variation between the different scenes. This suggests that an image independent noise metric might not be adequate for predicting overall image quality in complex images.

4. DISCUSSION

The results from both the experiments performed provide some insight into the perception of opponent noise, and its effect on image quality. In general the noise was most offensive, providing the largest decrease in image quality, at lower spatial frequencies. For the chromatic channels the noise was most offensive in the band centered at 2 cycles-per-degree of visual angle, while for the luminance noise it was most offensive when centered at 4 cycles-per-degree. It is important to note that because of the octave filtering these frequency bands have some overlap.

This behavior follows what we might assume from the contrast sensitivity function (CSF) of the human visual system, and provides some encouragement for their use in vision-based models of image quality. We know the human CSF typically behaves as a band-pass function for luminance channels, and as low-pass functions for the chromatic channels. It may be that the low frequency noise was deemed most unacceptable because it was the most perceptible. By inverting the interval scales this behavior becomes more clear. The inverted scales can be thought of as a scale of perceptibility. This perceptibility scale is plotted in Figure 8 for the high contrast luminance noise. A band-pass type curve representing an iconic CSF is overlaid on top of the perceptibility scale which has been normalized so that the peak value (representing the most negative judgment of image quality).

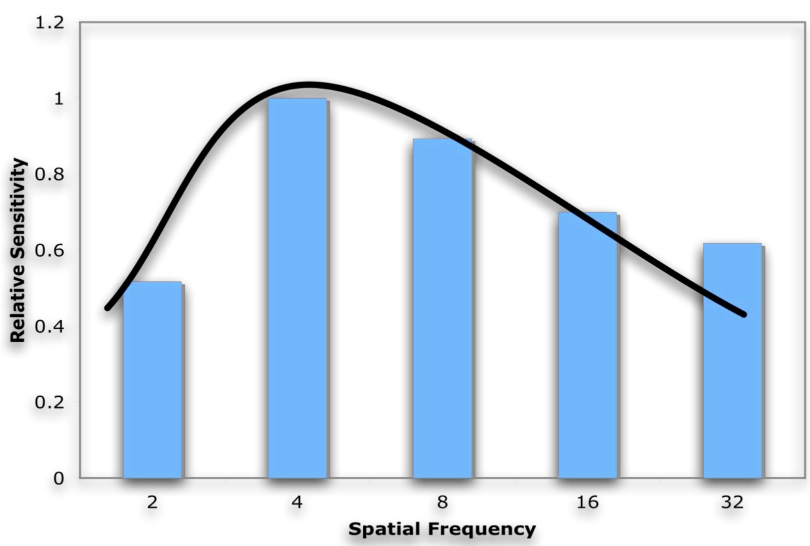

Figure 8: Normalized perceptibility scale for high-contrast luminance noise, with iconic CSF

The chromatic channels can be analyzed in a similar fashion. Only the two lowest frequencies have a large impact on image quality, most dramatically for the red-green channel. The higher frequencies are not statistically different from the original image, indicating a low degree of perceptibility for those frequencies, as would be predicted by a low-pass contrast sensitivity function.

These types of data can provide a useful starting point for designing and analyzing image difference and quality models, through the optimization of both the opponent color spaces and the contrast sensitivity functions used. The general approach that image difference models use is to accept two images as input, convert those images into an opponent color space and then apply spatial filters based upon the human contrast sensitivity functions. A per-pixel color difference calculation is then applied between the two filtered images. The general thought is that the spatial filtering removes information that the visual system cannot see, and attenuates the color differences such that the perception is the same at all spatial frequencies. By optimizing the choice of both the opponent color space as well as the spatial filters it is hoped that we can build an effective computational model of image quality.

5. CONCLUSIONS

A paired-comparison psychophysical study was conducted testing the effects of opponent noise on image quality. Noise patterns were generated in an isoluminant color-space, representing a luminance channel, a red-green channel, and a yellow-blue channel. Five different spatial frequency bands were tested, along with three contrast levels. Four scenes were examined to determine any image dependencies. An interval scale of image quality was generated from the experimental data.

The interval scale tends to follow the behavior predicted from the human contrast sensitivity function, which provides some hope for using vision-based models for predicting image quality for images with noise. The lowest frequency red-green noise, centered at 2 cycles-per-degree of visual angle, was determined to have the most detrimental effect on perceived quality. Overall the luminance noise was most perceptible, resulting in the biggest decrease in quality. The high-frequency chromatic noise was judged to have the small effect on quality, suggesting it was almost imperceptible. This may effect may be enhanced by the fact that the noise was generated in a specifically designed isoluminant space ensuring that there was no luminance information in the chromatic channels.

The experimental results showed a high correlation between the quality scales of the chromatic noise between the individual scenes. This suggest that it may be possible to predict the perception of chromatic noise using an image independent metric (such as measurements off a uniform patch.) The luminance scales were not as well correlated, with the largest difference coming from the human portrait image.

6. REFERENCES

1. P.G. Engledrum, Extending Image Quality Models, *Proc IS&T PICS Conference*, 65-69 (2002).
2. G.M. Johnson and M.D. Fairchild, Measuring images: Differences, Quality, and Appearance, *Proc. of SPIE/IS&T Electronic Imaging Conference, Santa Clara*, **5007** 51-60 (2003).
3. X.M. Zhang and B.A Wandell, A spatial extension to CIELAB for digital color image reproduction, *Proceedings of the SID Symposiums* (1996).
4. R.P. Dooley and R. Shaw, Noise perception in Electrophotography, *Journal of Appl. Photogr. Eng.,* **5** 190-196 (1979).
5. G.M. Johnson, E.D. Montag, R.A. Patil, M.D. Fairchild, Image quality scaling of electrophotographic prints, *SPIE/IS&T Electronic Imaging Conference, San Jose*, **5294** 165-175 (2004).
6. J. Kuang, X. Jiang, S. Quan, A. Chiu, Perceptual color noise formulation, *SPIE/IS&T Electronic Imaging Conference, San Jose*, (2005).
7. C. Poynton, Digital Video and HDTV, *Morgan Kaufmann, San Francisco CA*, (2003).
8. X. Song, G.M. Johnson and M.D. Fairchild, Minimizing the Perception of Chromatic Noise in Digital Images, *IS&T/SID 12th Color Imaging Conference, Scottsdale* (2004).
9. P.G. Engledrum, Psychometric Scaling, *Imcotek Press, Winchester MA* (2000).
10. E.D. Montag, Louis Leon Thurstone in Monte Carlo: creating error bars for the method of paired comparison, *SPIE/IS&T Electronic Imaging Conference, San Jose*, **5294** 222-230 (2004).