4-15-1983

A decision theory approach to tone reproduction

Kevin E. Spaulding

Follow this and additional works at: <http://scholarworks.rit.edu/theses>

Recommended Citation

Spaulding, Kevin E., "A decision theory approach to tone reproduction" (1983). Thesis. Rochester Institute of Technology. Accessed from

This Thesis is brought to you for free and open access by the Thesis/Dissertation Collections at RIT Scholar Works. It has been accepted for inclusion in Theses by an authorized administrator of RIT Scholar Works. For more information, please contact ritscholarworks@rit.edu.

A DECISION THEORY APPROACH TO TONE REPRODUCTION

by

Kevin E. Spaulding

A thesis submitted in partial fulfillment
of the requirements for the degree of
Bachelor of Science in the School of
Photographic Arts and Sciences in the
College of Graphic Arts and Photography
of the Rochester Institute of Technology

Signature of the Author Keri E. Spaulding
Photographic Science and
Instrumentation

Certified by Name Illegible
Thesis Advisor

Accepted by Name Illegible
Coordinator, Undergraduate Research

ROCHESTER INSTITUTE OF TECHNOLOGY
COLLEGE OF GRAPHIC ARTS AND PHOTOGRAPHY

PERMISSION FORM

Title of Thesis: A Decision Theory Approach to Tone Reproduction

I, Keri E. Spaulding, hereby grant permission to
Wallace Memorial Library, of R.I.T., to reproduce my thesis in
whole or in part. Any reproduction will not be for commercial use
or profit.

Date: 6/15/83

A DECISION THEORY APPROACH TO TONE REPRODUCTION

by

Kevin E. Spaulding

Submitted to the Photographic Science and
Instrumentation Division in partial
fulfillment of the requirements for the
Bachelor of Science degree at the Rochester
Institute of Technology

ABSTRACT

The concepts of statistical decision theory were applied to the photographic tone reproduction system to develop an objective measurement of tone reproduction accuracy called the "expected loss" of the system. Reproductions were made of four scenes using systems with a wide range of tone reproduction characteristics. The expected loss for each system/scene combination was calculated, and the perceived tone reproduction accuracy was determined using the psychophysical scaling method of magnitude estimation. A statistically significant correlation was found between the subjectively determined accuracy, and system expected loss, but the degree of correlation was not high enough for the merit function to be of any great value in its present form.

ACKNOWLEDGEMENTS

The author wishes to thank Mr. Peter Engeldrum for his very valuable advice, and assistance throughout this project in his role as thesis advisor. Also, a word of thanks goes to Mr. Robert Chung for allowing me the use of the negatives which he generated for his earlier tone reproduction study, to Dr. Alan Frank for his valuable advice on the subjective testing procedures, and to Mr. Joseph Biegel for his aid with the statistical analysis. Thanks are also due to the Xerox Corporation for the use of their scanning densitometer. Gratitude is to be expressed to Dr. Ron Francis for his patience and wisdom over the last four years he has spent with our class.

TABLE OF CONTENTS

List of Tables	vi
List of Figures	vii
I. Introduction	1
II. Experimental	11
A. Preparation of Test Prints	11
B. Measurement of Expected Loss	19
C. Subjective Evaluation of Reproductions	24
III. Results	30
A. Scene Brightness Distributions	30
B. Brightness Reproduction Characteristics	30
C. Measured Expected Loss Values	32
D. Subjective Evaluation of Test Prints	33
E. Correlation of results	36
IV. Discussion	41
V. Conclusions	46
References	48
Appendix A: Sample Output From Expected Loss Calculation Program	50
Appendix B: Sample Subjective Testing Record Sheet ..	51
Appendix C: Sample Output from Data Filer Program ...	52
Appendix D: Measured Brightness Reproduction	53
Characteristics	
Appendix E: Calculated Expected Loss Values	57
Vita	58

LIST OF TABLES

No.	DESCRIPTION	PAGE
1.	Means and standard deviations of subjective quality values for scaled and unscaled data.	34
2.	Correlation coefficients for scene regressions using third order loss function.	40

LIST OF FIGURES

No.	DESCRIPTION	PAGE
1.	Perceived brightness as a function of log relative luminance for complex fields viewed with dark and light surrounds. (from ref. 5)	3
2.	Scenes which were used in testing. (a) man and test target, (b) rooftop, (c) dancing girl, (d) night scene.	13
3.	Variations used in tone reproduction systems.	17
4.	Layout of print and scale used for preparation of scenes.	18
5.	Schematic diagram of method used for the determination of tone reproduction from density distributions.	22
6.	Photograph (a), and schematic (b) of viewing booth for subjective testing.	27
7.	Measured Brightness Distributions for (a) man and test target, (b) rooftop, (c) dancing girl, (d) night scene.	31
8.	Scaled Vs. Unscaled Subjective Data.	36
9.	Correlation of data using quadratic loss function. (a) normal plot, (b) log-log plot.	37

I. Introduction

One of the first attempts to quantify the photographic process was made by Hurter and Driffield in 1890¹. The result of this investigation was the description of the photosensitive material in terms of the negative logarithm of opacity (i.e. density) as a function of the logarithm of the exposure. Today, this is better known as the "D-logH" curve. In 1920, Jones applied this concept to characterize the "tone reproduction" of a photographic system.² By cascading the transfer characteristics of each element of the photographic system through a multi-quadrant graph, the relative luminances of the original scene could be compared with those of the reproduction. He hypothesized the optimum photographic system to be that which would result in the exact 1:1 reproduction of the relative luminances. Essentially what this implied was a linear relationship between input log luminance and output density for the "optimum" system. As noted by Bartleson,³ this method served as the basis for most tone reproduction studies for many years.

More recently, experimentation which has been done with the visual perception of brightness has shown that Jones' approach is actually oversimplified. Stevens' work with brightness perception in a simple field led him to conclude that brightness perceived by an observer was not simply proportional to the logarithm of

luminance, but was more appropriately described by a power function of luminance,⁴ of the form:

$$B = a(L - L_0)^n$$

where: B = perceived brightness, L = luminance, L_0 = threshold luminance, and a and n constants. Bartleson and Breneman extended this concept and found that if a complex field, such as a photograph, were considered, perceived brightness was more closely a power function with exponential decay.⁵ They proposed the following function to describe this relationship:

$$B = 10^{a\beta} L / \text{antilog}(\gamma \exp(\delta \log L))$$

where B = perceived brightness, L = luminance (in mL), and a , β , γ , and δ are parametric constants which are functions of illuminance level, and surround luminance. Their results indicated that the rate of increase of relative brightness with log luminance was essentially constant with changes in absolute illuminance, but varied significantly with changes in the luminance of the picture's surround. This meant that perceived relative brightness levels of a picture were essentially independent of the actual viewing illuminance for identical viewing conditions. But comparatively, an image viewed with a dark surround (such as when viewing slides) exhibited a significant loss of the image contrast relative to when it was viewed with a

light surround. This is consistent with the work of Clark who found that "the rate of change of perceived brightness as a function of log luminance is less with dark surrounds than with illuminated surrounds."⁶ Graphically, Bartleson and Breneman displayed this relationship by plotting log relative luminance vs. relative perceived brightness for the different viewing conditions (see fig 1.). The difference in the shapes of these curves was similar to that reported by Breneman,⁷ and to the effects of visual induction as reported by Jameson and Hurvich.⁸

Figure 1: Perceived relative brightness as a function of log relative luminance for complex fields viewed with dark and light surrounds. (from ref. 5)

Later in the same year, Bartelson and Breneman published a paper which utilized their data on perceived brightnesses to analyze tone reproduction in the photographic process.⁹ They determined, using their brightness perception function, that optimum quality would result for the reproduction process and viewing condition combination, when the brightnesses relative to a reference white were reproduced in a one to one relationship. They showed that their results were consistent with the results of several subjective tone reproduction experiments which had been previously performed. The implications of this relationship were that the physical gradient of the system's tone reproduction should be dependant on the viewing conditions, so as to produce the desired 1:1 reproduction of brightnesses relative to a reference white. In general, Bartleson and Breneman determined that the gradient for a reflection print viewed under normal conditions should be approximately 1.0, and the gradient for a transparency viewed with dark surround should be about 1.5.

In 1968, Bartleson published the results of an experiment he had done to determine whether differences from this optimum tone reproduction would be a valid measurement of the quality of the tone reproduction system.³ By taking the square root of the sum of the squares of the differences between the relative brightness of the scene, and the relative brightness of the image for a variety of tone reproduction systems, he was able to

hypothesize a measure of the deviation from the optimum. Described mathematically:

$$\text{deviation from optimum} = \sqrt{\sum (B_o - B_i)^2}$$

where B_i = scene brightness, and B_o = reproduction brightness. Comparing his results to Clark's findings on the measurement of tone reproduction quality,⁶ Bartleson found that there was a significant relationship between quality and any decrements from the 1:1 relative brightness reproduction criterion. But to produce the degree of correlation which he desired, Bartleson found that it was necessary to use an eigenvector weighting function to scale the data; this removed some inconsistencies which he attributed to the difference in the relative importance of densities in the reproduction. He hypothesized, for example, that high density areas may not be as important as low density regions in many scenes. This seems to indicate that the importance of accurate reproduction of the relative brightnesses may be, in some way, a function of the actual scene brightness distribution.

Experimentation with the photographic system seemed to verify the validity of the brightness function which Bartleson and Breneman developed, but the complexity of the equation made it difficult to work with; this later led Bartleson to develop a simpler equation which would represent the relationship equally well.¹⁰ This new quantity he labeled "lightness" so as to avoid

confusion with earlier terminology:

$$L^{**} = A(100R/R_0 + B)^a - 16$$

Where L^{**} = Lightness, R = reflectance, R_0 = reflectance of the reference white, and A , B , and a are constants whose values depend on the surround luminance. The values he reports for these constants are:

	A	B	a
light surround	11.5	1.0	0.50
dim surround	17.5	0.6	0.41
dark surround	25.4	0.1	0.33

In 1981, Dom¹¹ applied Bartleson's lightness functions to the work done by Yule¹² on optimum tone reproduction, and was able to show that this function did yield the expected linear relationship between relative brightness of the scene and relative brightness of the original for optimum reproduction. An interesting result which he also found was that for some given tone reproduction system with a minimum brightness value which is different from zero, (as is with most real systems), the optimum brightness reproduction for the system is the line connecting the minimum brightness value (D-max) to the minimum brightness value.

Domalso warns in this paper that the psychological scaling

experiments of Yule, upon which his conclusions are based, were done with "more-or-less typical originals", and that the significance of the results could not be extended beyond such images. This statement is also true for most of the other experiments upon which the published results have been based.

But not all scenes are what could be called "typical." In the graphic arts field, photographs are often labeled as "high key", "normal key", or "low key" depending on the overall tonal content of the image. A high key image, for example, would be one which was dominated by primarily high brightnesses. Jorgensen¹³ has noted that the same system may not reproduce all these types of scenes equally well, i.e. a tone reproduction system that works well for low key scenes may not be the best for high or normal key scenes.

It is the hypothesis of this research, that a decision theory approach to the analysis of tone reproduction data would take into account the characteristics of each individual scene, and would accordingly result in a measure of how well a given scene can be reproduced by some specific system. In the terms of decision theory, decrease in the quality of the tone reproduction could be considered equivalent to an increase in the quantity called "risk." Formally defined, risk is the expected value of the loss function:¹⁴

$$R(\theta, d(Z)) = E \left[L(\theta, d(Z)) \right]$$

where $R()$ is the risk, $L()$ is the loss function, θ is the "state of nature", and $d(Z)$ is the "decision function." Applying this to tone reproduction, it would seem reasonable to associate the "state of nature" with the actual reproduced relative brightness which shall be called B_o , and to associate the "decision" with a 1:1 reproduction of brightnesses relative to white. Therefore if the relative brightness of the scene is called B_i , our decision function becomes:

$$d(B_i) = B_i$$

This is because the well established results have "decided" that the relative brightness out of the system should equal the relative brightness into the system at each brightness level.

The loss function is a measurement of the variation from the optimum conditions. It would seem logical that the loss function should equal zero when $B_i = B_o$, and that the loss function should be increasing with the absolute difference in the relative brightnesses: $|B_o - B_i|$. A typical loss function which has these properties is the so-called "quadratic loss function."¹⁴

$$L(\theta, d(Z)) = L(B_o, B_i) = (B_o - B_i)^2$$

Bartleson's approach to quantifying reproduction quality, which was discussed earlier, used a function which was similar to

this one³, and therefore lends some additional merit to the choice of this particular loss function.

At this point, it is necessary to make the assumption that B_i is a random variable, i.e. the distribution of B_i depends on the particular image, I . This distribution can be characterized by a conditional probability density function, $p(B_i|I)$. Using the well known mathematical relationship for the expected value of a function:

$$E[f(x)] = \int f(x) p(x) dx$$

the statistical risk, which has been hypothesized to be a measure of the tone reproduction quality for the particular scene-system combination, can now be expressed by the following relationship:

$$R(B_o, B_i | I) = \int_0^{100} (B_o - B_i)^2 p(B_i | I) dB_i$$

It is easy to see that the effect of this calculation is to measure the differences between the optimum, and the actual brightness reproduction curves at each scene brightness, and to weight them according to the fractional area of the scene which occurs at that brightness level. This would seem to be a reasonable method to take into account the effects, which were observed by Bartleson and others, of the scene on the observed tone reproduction quality. For example, this algorithm would indicate that a low key scene would reproduce better on a system with a high

D-max than one with a lower D-max, which seems to be a reasonable result. In effect, this measurement determines the average accuracy (actually lack of accuracy) over the entire image.

This research investigated the feasibility of using this merit function as a direct measurement of "perceived tone reproduction accuracy" which would provide a useful tool for the evaluation and optimization of a tone reproduction system.

II. Experimental

A. Preparation of Test Prints

In order to test the ability of the calculated "expected loss" of a tone reproduction system to predict psychophysically perceived tone reproduction accuracy, it was necessary to generate a series of test prints which could be examined using both the expected loss algorithm, and subjective scaling methods. First, it was necessary to find several scenes which had a variety of brightness distributions, which could be conveniently measured. It was decided that the easiest, and most convenient way to accomplish this was to select several high quality photographs with the desired characteristics, and to use them as "scenes" which could then be reproduced by subsequent photographic systems.

R. chung has done some research on the subject of image classification for the purposes of tone reproduction determination.¹⁵ Through the measurement of scene brightness distributions (which he called TDCs, "tonal distribution curves"), he found that he was able to accurately classify a series of scenes. Since the scenes that Chung used in his experimentation had been shown to exhibit a variety of brightness distribution characteristics, it was decided that they would be appropriate for the purposes of this research also. Four photographs were chosen

out of his original six which covered the entire range of scene types ranging from "low-key" to "high key." These four photographs are reproduced in figure 2. Scene (b) would be classified as a high-key scene, scene (d) as a low-key scene and the other two as somewhat "normal," intermediate scenes. Mr. Chung was very gracious to provide the negatives which were used to produce his prints, and these negatives were then printed carefully to make the "best possible" prints, which could then be used as scenes for this research.

For each of the four scenes discussed above, it was necessary to make reproductions using a variety of tone reproduction systems. Simonds reported that the number of ways that a tone reproduction can vary is actually very small.¹⁶ He showed that most differences between systems can be explained simply by differences in the contrast, and in what he calls "fog adjustment," which is essentially just the overall darkness of the print.¹⁷ For each of the four scenes, a series of nine reproductions was made which consisted of three variations in contrast, and three variations in density (see figure 3). The reproductions were made using 4x5 negatives exposed on a copy stand, which were then printed using a diffusion enlarger. The three contrast variations were controlled primarily in the negative preparation through the variation of the development time, whereas the variation in density levels were controlled in the printing stage. To facilitate the easy measurement of the

(a)

Figure 2: Scenes which were used in testing. (a) man and test target, (b) rooftop, (c) dancing girl, (d) night scene.

Figure 2 (b)

Figure 2 (c)

Figure 2 (d)

High Contrast Low Density	High Contrast	High Contrast High Density
Low Density	"Normal"	High Density
Low Contrast Low Density	Low Contrast	Low Contrast High Density

Figure 3: Variations used in tone reproduction systems.

system tone reproduction characteristics without the need for cascading D-Log H curves, a 20-step gray scale was photographed adjacent to the original as shown in figure 4. The reproduction characteristics could therefore be measured directly. (During the subjective scaling, the gray scales were removed from the prints.) To minimize differences in image tone, surface texture, and image structure all reproductions were made using a single type of photographic paper, and a fixed camera aperture. These conditions are similar to those used by Simonds in a tone reproduction study which he conducted,¹⁶ and should reduce any effects other than those due to the tone reproduction itself.

Figure 4: Layout of print and scale used for reproduction of scenes.

B. Measurement of Expected Loss

As derived in the introduction section, the value of the expected loss for a tone reproduction system can be represented by the integral:

$$R(B_o, B_i | I) = \int_0^{100} (B_o - B_i)^2 p(B_i | I) dB_i$$

Where, for the purposes of this evaluation, Bartleson's "lightness" value, L^{**} , will be used as a measure of the relative brightnesses. With the appropriate constants substituted for prints viewed with a "light surround," L^{**} is of the form:

$$L^{**} = 11.5(100R/R_o + 1.0)^{0.50} - 16$$

Where R is the reflectance of the print, and R_o is the reflectance of the reference white.

To calculate the value of the expected loss for each of the reproductions, it was necessary to determine both the brightness distribution of the scenes, and the tone reproduction characteristics of the photographic processes. It was determined by Chung,¹⁵ that to make an accurate measurement of the tonal distribution of a photograph it was necessary to sample approximately 20% of the image. Measurements using larger

fractions did not produce any significant gains in accuracy. The four originals were measured with a scanning densitometer at the Xerox Corporation using a 0.2" diameter aperture, sampling at a rate necessary to examine approximately 25% of the image. The instrument was interfaced with a Wang computer, and produced output in the form of a reflectance histogram (in 0.01 increments). These reflectance histograms were then properly transformed and scaled to produce brightness probability density curves.

To verify Chung's conclusion that 20% sampling was adequate to determine the scene brightness distribution, the night scene was selected and scanned using a 0.1"x0.1" aperture and 100% sampling. (This print was thought to be the most likely scene to produce measurement error because of the rapid changes in density.) Visual examination revealed very little difference in the measured brightness probability density curves, and a Chi-square test for lack of fit at a significance level of 0.95 failed to reject the hypothesis that the two distributions were identical.

As mentioned previously, a gray scale was included adjacent to the scenes when each of the reproductions was made. From these, the tone reproduction characteristics of each print could be directly determined, and hence the "loss function". Each of the reproduced gray scales (as well as the original) were read in density with a MacBeth reflection densitometer. These densities were then converted to L^{**} values to produce the brightness

reproduction curves, and the loss functions. Upon close examination of the reproduced gray scales it became apparent that there was a considerable amount of glare which could be observed on several of the scales. Apparently this was due to the lighting used when making the reproductions, and the curl of the original scale. Subsequently, two methods were examined to verify the accuracy of the tone reproduction measurements. By measuring the original and reproduced density distributions of an image, it is possible to determine the transformation that must have been necessary to cause the observed change.¹⁵ Schematically, this method is diagramed in figure 5. If the distributions are plotted in a cumulative form it is only necessary to determine corresponding input-output density pairs which correspond to the same cumulative frequency. These pairs can be found at all values along the curve, and plotted to recreate the density reproduction of the system.

It is also possible to measure the density reproduction characteristics of a print by direct measurement of corresponding regions on the original and reproduced prints. If enough points of varying densities are measured, it is possible to reconstruct the entire curve. Although the first of these methods is the more elegant of the two, it was also found to be somewhat impractical for this application for two reasons: by nature, the accuracy of this method is not very good in regions of low probability density, and secondly, the scanning densitometer which

Figure 5: Schematic diagram of method for the determination of tone reproduction from density distributions.

was being used was limited by a maximum value of 2.0 in the density mode, which was not sufficient to accurately characterize some of the prints. It was found that the brute force method, despite its lack of glamour, worked quite well in all cases. (In the mid-density regions where the first method performed well, it was found that there was good agreement between the two methods.) In many of the prints, it was determined that indeed there were errors with the direct gray scale measurements, and that corrections were necessary.

Before the value of the expected loss could be found for each of the reproductions it was necessary to convert the formula into one which could be used with data that is by necessity discrete rather than continuous. If we consider the terms in the expected loss integral: " $p(B_i) dB_i$ ", it is possible to interpret this product as the brightness probability for the differential brightness interval dB_i . In a discrete distribution, the analogous term is $p_j(B_i)$, the brightness probability of the j^{th} interval. The expected loss integral can therefore be represented by the following summation:

$$\text{Expected Loss} = \sum_{j=1}^N (L^{**}_{o,j} - L^{**}_{i,j})^2 p_j(L^{**}_{i,j})$$

where L^{**} has been used as a measure of the relative brightness.

This calculation is easily possible with the data which was collected as described above.

The amount of data, and calculations involved is obviously quite large, and when 36 images are considered the numbers rapidly grow. To make these calculations feasible, a program was written in Apple Pascal to store all the data, and to perform all of the necessary calculations for each data set. A sample set of output for one of the reproductions can be found in appendix A. (As will be described later, several alternate "loss functions" were examined throughout the course of the research; for those cases, this program was easily modified to obtain the necessary results.)

C. Subjective Evaluation of Reproductions

A variation of the method of "magnitude estimation" used by Clark in his tone reproduction quality tests⁶ was used to scale the perceived accuracy of the tone reproduction for each of the prints. This technique involves setting what is called, by statisticians, an "anchor" against which the magnitude of some response can be compared. For example, a judge might be told that some particular standard had a value of +40, and be asked to estimate the magnitude of some test stimuli with respect to the anchor's value. In a normal magnitude estimation test, the test stimuli are presented one at a time to the judge, and he then assigns each stimuli a value individually. On the recommendation

of a sociologist with much experience in this type of scaling experiment, it was decided that a better approach for this test would be to present all nine of the reproductions of a given scene at the same time to let the observers compare the reproductions to each other as well as to the anchor. His experience indicated that this would result in each judge having a much more consistent internal scale.

Since the model is predicting the accuracy of some tone reproduced image relative to the original scene from which it was reproduced, it was decided that the most appropriate "anchor" for this test would be the original scene itself. Since, by definition, the optimum tone reproduction system should reproduce the scene exactly, brightness for brightness, it could never be possible to have a reproduction which is "more accurate" than the original. The originals were therefore assigned a value of one hundred and the judges were told that they could give the reproductions magnitudes between zero and one hundred. In a test like this, it is common to let the "zero point" correspond to an "indifference level,"¹⁸ that is, the worst imaginable reproduction. The judges were told in this case that a quality value of zero would correspond to a reproduction which was totally uniform in density; this should have given all judges a common reference, and hopefully should have forced each judges scaling to be consistent with all of the others.

A "viewing booth" was constructed to provide for even,

repeatable illumination, and a convenient location to conduct the testing. Illumination was provided by two standard 48 inch fluorescent tubes approximately 35 inches above the viewing surface. When the observers were making their judgements this was the only source of illumination in the room and provided approximately 1300 lux at the viewing surface. To alleviate reflection problems, and make viewing more comfortable, an inclined viewing surface was provided at an angle of approximately 50° . A physical scale was provided along which the observers could lay out the prints to aid them in estimating a scale. The values of the scale ranged from zero to one hundred and covered a total of 4 feet of the viewing surface. Figure 6 displays a diagram, and an actual photograph of the viewing area.

The results of a study done by Yule¹² demonstrate the importance of how the test question is presented to the judges during a subjective test. In his experiment, he found that if three slightly different questions were asked, at least two distinctly different answers were obtained. Care was therefore taken to phrase the question in such a way as to make the criterion clear to the judges. It was stressed that the values which were assigned should be representative of the "accuracy" of the reproduction, regardless of aesthetic quality. So that the instructions to each judge would be consistent, a statement was prepared which was read to each observer. This statement appears below:

(a)

(b)

Figure 6: Photograph (a), and schematic (b) of viewing booth for subjective testing.

One at a time, you will be given four sets of nine prints, and the reference print from which they have been reproduced. By definition, the reference print has "perfect" quality, and a corresponding quality rating of 100. Examine each group of prints, and place them on the quality line which has been provided at a value which you feel is representative of the accuracy of the reproduction. For example, if the densities of a copied print were exactly the same as those of the original print it would be assigned a quality rating of 100. A quality of zero would correspond to a totally uniform density image. Please consider the entire image when evaluating the prints. Judgement should be based wholly on the overall accuracy of the tone reproduction. When rating the prints, please ignore any defocus, color differences, pinholes, scratches, or other physical defects. When you have placed all of the prints on the scale, you will be asked to report the corresponding quality value for each print so that it can be recorded. You will be given up to ten minutes to examine each of the four sets.

The ratings for each judge were recorded on a record sheet, an example of which is reproduced in appendix B. Thirty judges were used most of which had at least moderate experience in some aspect of photography.

To easily manipulate the large volume of data which was collected (30x36 values), a data-handling program was written to perform all the necessary grouping, plotting, and statistics for the subjective values. The program calculated the average and standard deviation for each print, and histogrammed the data by print, and by observer to get a visual picture of the response. An option was also included in the program to find the average and standard deviation of each observer's response and then scale

the data for that observer to a "z-score" where:

$$z = \frac{(\text{actual rating for print} - \text{average rating for observer})}{\text{standard deviation of ratings for observer}}$$

The effect of this was to force each observer's scale to be "normal." This was necessary because, despite the careful instructions, it was apparent that observers were using a variety of different scales. For example one observer might rate all of the prints between 40 and 60 while another, would use the entire range of 0 to 100. After each observer's data was scaled individually, the program would again output the statistics, and histograms for each print using the scaled print values. For sample output from this program see appendix C.

III. Results

A. Scene Brightness Distributions

As expected, measurement of the four scenes demonstrated very different brightness distributions similar to those found by Chung. The range represented in the prints should cover the typical range encountered in a normal circumstances. Figure 7 displays the brightness histograms which were transformed and scaled from reflectance data as described earlier. Replication of measurements indicated that the brightness probability density should be accurate to approximately 0.01 (brightness units)⁻¹.

B. Brightness Reproduction Characteristics.

The calculated brightness reproduction characteristics for each of the reproductions for the four scenes can be found in appendix D. Since each set of reproductions for the four scenes was reproduced independently, there is not an exact relationship between the four sets of curves. For example, the low contrast, low density image of scene A does not necessarily have the same brightness reproduction characteristics as that of the corresponding print for scene B. Since each system was characterized individually, this should not be an important factor.

L** PROB. DEN. REFD-A

(a)

L** PROB. DEN. REFD-B

(b)

L** PROB. DEN. REFD-C

(c)

L** PROB. DEN. REFD-D

(d)

Figure 7: Measured Brightness distributions for
 (a) man and test target, (b) rooftop, (c) dancing girl,
 and (d) rooftop.

The plots which are displayed in appendix D were generated from the measured density reproduction data using the Apple Pascal program which was described in the experimental section.

C. Expected Loss Values

Using the numerical integration described in the experimental section, the expected loss value for each of the scene/system combinations was calculated. Initially this was done using the "quadratic loss" function which had been chosen somewhat arbitrarily in the original derivation of the expected loss function. As will be described later it was found necessary to explore other loss functions. Expected loss values for each print were calculated using the general form of the loss function to be:

$$\text{Loss}(B_o, B_i) = |B_o - B_i|^N$$

where the order of the loss function, N , was varied from 1 to 4 in integer increments. The results of the expected loss calculations for the various order loss functions can be found in appendix E.

An error analysis was performed on the expected loss calculation using the method described by Y. Beers¹⁹. Making some assumptions as to the accuracy of the various measurements which were made, expected errors were calculated for each of the

reproductions. It was estimated that the values were accurate to approximately $\pm 20\%$ on the average. At the lower end of the expected loss scale the percent error was typically somewhat higher, but the actual magnitude was much smaller.

D. Subjective Evaluation of Test Prints

Data from thirty observers was gathered, and subsequently processed in the manner described earlier in the experimental section. Statistics for the unscaled, and "z-scored" data is presented in Table 1. Particularly on the unscaled data, it can be observed that there is a large variance from observer to observer. The large standard deviations represent data that often covered almost the entire scale from 0 to 100 on a single print. Typically, the prints that were either very good, or very poor exhibited less variance, but there was very little agreement among judges on the prints that were of intermediate quality. It was apparent that some of the variance was due to different perception of the scale from one judge to the next. As mentioned earlier, the scaling of each observer's values into z-scores should to a large extent remove this effect due to the difference of observer scales. Indeed, it can be seen from the standard deviations of the scaled data that there was a significant decrease in the variability for many of the prints. This effect becomes even more apparent when histograms of the scaled, and unscaled data are examined. In appendix C, the histograms of one

Table 1: Means and standard deviations of subjective quality values for scaled and unscaled data.

prints 1- 9: man and test target
 10-18: rooftop
 19-27: dancing girl
 28-36: night scene

unscaled data			scaled data		
DATAFILE = TDATA2			DATAFILE = TDATA22		
PRINT	MEAN	S.D.	PRINT	MEAN	S.D.
1	27.73	18.15	1	-1.26	0.58
2	63.83	16.41	2	0.39	0.74
3	76.18	15.82	3	1.86	0.68
4	43.67	19.35	4	-0.58	0.72
5	55.47	18.15	5	0.81	0.61
6	65.28	18.18	6	0.49	0.64
7	44.83	21.67	7	-0.55	0.85
8	58.93	22.91	8	-0.28	0.82
9	42.48	22.12	9	-0.61	0.98
10	55.63	17.68	10	0.88	0.66
11	77.97	13.95	11	1.89	0.68
12	46.88	19.11	12	-0.48	0.72
13	29.43	17.15	13	-1.21	0.43
14	76.37	13.16	14	1.85	0.56
15	46.83	22.21	15	-0.46	0.93
16	48.23	17.97	16	-0.66	0.58
17	74.43	13.39	17	0.94	0.61
18	64.58	14.51	18	0.44	0.71
19	32.87	18.94	19	-1.85	0.49
20	41.63	18.22	20	-0.63	0.52
21	46.68	18.56	21	-0.38	0.68
22	54.28	16.99	22	-0.87	0.72
23	72.17	16.75	23	0.84	0.67
24	75.28	18.26	24	0.98	0.74
25	55.98	20.48	25	0.86	0.82
26	72.83	20.13	26	0.85	0.74
27	59.97	25.81	27	0.28	0.94
28	37.78	24.86	28	-0.77	0.84
29	57.97	19.77	29	0.18	0.85
30	64.83	16.38	30	0.44	0.64
31	54.58	21.92	31	-0.81	0.87
32	68.55	17.18	32	0.68	0.64
33	55.48	17.66	33	0.85	0.69
34	69.38	28.17	34	0.71	0.84
35	36.37	21.88	35	-0.86	0.76
36	35.87	24.84	36	-0.92	0.93

print using both sets of data are contrasted. As expected, the main effect of the transformation was to move several of the extreme values into better agreement with the majority of the data. But, despite the significant decrease in the variance, it was found that there was no significant effect on the mean of the values. The scaled values were regressed against the unscaled values (see figure 8) and it was found that an extremely good relationship existed. The effect of the transformation could be explained by the equation:

$$\text{scaled SQ} = (0.04)\text{SQ} - 2.62$$

with an r^2 correlation coefficient of 0.998. Thus we must conclude that although this transformation helps to remove some of the variability of the data, it has no significant effect on the central tendency for each print. This implies that, on the average, there are just as many observers whose scale was on the low side, as there were whose scale was high. Since this transformation did not alter the relative position of the various reproductions, it was decided to use the unscaled data throughout the rest of the calculations.

Figure 8: Scaled vs. unscaled subjective data.

E. Correlation of Results

As a first examination of the data, a plot was made of subjective quality vs. expected loss using the quadratic loss function (see figure 9). From this graph it was easily apparent that any relationship between the subjectively determined tone reproduction accuracy and the expected loss of the system was not linear. A standard procedure used when there may be a power relationship between data is to plot the data in log-log form. This was done, and as can be seen in figure 11, there appeared to be a much more linear relationship between the log of the

(a)

(b)

Figure 9: Correlation of data using quadratic loss function.
(a) normal plot, (b) log-log plot.

subjective quality, and the log of the expected loss value.

The degree of correlation was not as good as had been hoped for having an r^2 value of only 0.46. Upon scrutinizing the various parameters involved in an attempt to find the source of the lack of correlation, it appeared that the expected loss algorithm may be failing to weight deviations from the optimum tone reproduction heavily enough. To examine this possibility, the order of the loss function was varied, as has been described previously. Regressions were performed on the data generated with each loss function which indicated that increasing the exponent of the loss function increased the correlation coefficient, and lowered standard error until the order of the function was increased to above 3.0. Using the SPSS statistics package on the VAX computer system, a stepwise multiple regression was run with the log subjective quality as the dependant variable as a function of the scene number, and the log expected loss values using the first, second, third and fourth order loss functions. It was found that the most significant relationship was with the third order expected loss value. Functionally:

$$\text{Log SQ} = (-0.14)\log \text{EL}_{(3)} + 2.15$$

with an r^2 value of 0.53. Where $\text{EL}_{(3)}$ is the expected loss calculated with the third order loss function. An ANOVA

showed the regression to be significant with an F regression of 37.9, as compared to an F-critical of 4.17 at a 0.05 significance level. The multiple regression also found the first order expected loss term to be significant increasing the r^2 value to 0.60 with the relationship:

$$\text{Log SQ} = (-0.33)\log \text{EL}_{(3)} + (0.47)\log \text{EL}_{(1)} + 2.27$$

where $\text{EL}_{(1)}$, and $\text{EL}_{(3)}$ are the first and third order loss functions respectively. No other terms in the regression were found to be significant. This indicates that there is not a significant effect on the regression due to the scene used in the system. However, individual regressions of the four scenes separately does indicate that there is a difference in the correlation between scenes. Table 2 indicates the various correlation coefficients associated with each scene for the third order loss function regression. In particular, scene D, the night scene, appears to have a significantly lower level of correlation, although the actual regression coefficients which were calculated were not found to be significantly different.

Table 2: Correlation coefficients for scene regressions using third order loss function.

<u>Scene Description</u>	<u>Correlation Coefficient</u>
Man and Test target	0.704
Rooftop	0.671
Dancing Girl	0.637
Night Scene	0.533

IV. Discussion

In its present form, the expected loss value as a measurement of the tone reproduction accuracy for a photographic system has apparently only a limited utility as is evidenced by the relatively low correlation coefficient which was obtained. It appears that the model is failing to take into account some aspect of the subjective process which an observer uses to evaluate the accuracy of the reproduction of a scene by a photographic process. There are several factors which seem to be reasonable possibilities, and it may be that all, or some are responsible for the lack of correlation observed in this research.

One effect may be tied to the observer's perception of the aesthetic quality of the various reproductions. When being given testing instructions, it was stressed that the judge should not consider any personal preferences he may have for one image when giving a reproduction its weighting, but that he should only consider what he perceived as the level of accuracy with which the reproduction had been made. But, despite any amount of instructions, it is impossible for an observer to completely ignore his personal reaction to an image. For example, if he likes the picture, the observer may tend to give it a higher rating regardless of the accuracy with which it was reproduced, and conversely, if the observer does not like the scene he may

unconsciously rate that reproduction lower than he would have with another scene of identical characteristics. This is an element that no mathematical formula can ever hope to describe, and its effect is primarily to add noise to the data. However, experimentation conducted by Corey et al.²¹ resulted in the conclusion that image content did not have a large influence on tone reproduction quality, and therefore it is not expected that this effect is very large. A second factor which may have contributed significantly to the lack of correlation, is one which is a little more tangible. The derivation of the expected loss value was based on the assumption that all parts of a scene are of equal importance in the reproduction of a scene. The effect of the expected loss calculation is to weight the ability of a system to reproduce a brightness by the fractional area of the image which contains that brightness. Thus if two equal areas of uniform brightness in a photograph are considered, the merit function will assume that they are equally important, whereas to an observer, it is very possible that they are not. The low correlation in the night scene may be an example of this effect. Much of the image was made up of a uniform density background of very high density, with the main object occupying a smaller fraction of the image. The expected loss algorithm tended to weight the background as being very important because of its relatively high probability, so that if the

density of the reproduced image were close to that of the original in this density range, it would get a low expected loss value. In contrast, it appeared that this region was not as important to the observers. In examining trends in the subjective data, it is apparent that the observer was more concerned about what the reproduction was like in the brightness interval occupied by the building. Thus it was typically preferred by an observer to have a reasonable reproduction in the building and lose density in the black region, than to reproduce the high densities well at the sacrifice of the less frequently occurring brighter areas. For images in which there are important details in all regions of the photograph, this concern would not be as great; this may be the reason why the other scenes had a greater degree of correlation. Contrary to the first factor which was discussed, the effect of this would not be to add random noise, but should produce some repeatable relationships in the results.

Although it may be impractical to consider quantifying this type of response in some fully automated process, it does seem conceivable that a model could be developed which could include some sort of "importance factor" for various regions of a scene, and then perform some form of a expected loss calculation weighted by this factor. Instead of summing the loss function multiplied by the probability density function over all brightness intervals, which in effect sums the loss for each differential area element of the scene, an alternate merit function might sum the loss multiplied by an importance factor for

each differential area element:

$$\text{Loss} = \int L(a) W(a) da$$

where $L(a)$ is the loss of the tone reproduction system as a function of the image position, $W(a)$ is the importance factor as a function of the image position, and da is a differential area element of the image. Although it may be possible to develop such a model, and test it successfully, it seems that much of the utility of the merit function would be lost in the process. Any such algorithm would have to somehow determine the relative importance of various scene elements which would not seem to be feasible without some sort of subjective input, which is the very factor the model was attempting to eliminate.

One method for consideration which could to some extent automated would be to eliminate, or weight less strongly, areas of uniform density when constructing the brightness probability density function of the scene. This approach makes the assumption that the regions of interest in a scene would correspond to those areas which contained information. This approach has been used for a slightly different type of merit function by Ovchinnikov et al.²⁰ As noted by Chung,¹⁵ the problem with this method is the sophisticated image processing equipment which is necessary to sort the image properly. The approach described by Ovchinikov et al. offers an interesting alternative in itself; their method

was to reproduce the scene so that the histogram of the regions containing tonal gradation (information content) was reproduced as close as possible to a gaussian distribution. Essentially this is just a different tone reproduction objective: rather than attempting to reproduce brightnesses in a one to one relationship, their goal was to reproduce the image in such a way as to produce a gaussian distribution. This approach could also be analyzed using a decision theory derived model. The expected loss value derived for this research used Bartleson and Breneman's tone reproduction criterion as the "decision function," but an analogous expression could be derived using the gaussian distribution criterion as a "decision" for the tone reproduction system. the loss function in this case would be related to deviations from the gaussian distribution criterion rather than the one-to-one brightness reproduction criterion. It is possible that an approach using a method similar to one of these may yield results which would be more consistent with subjective evaluation, because they would attempt to determine the relative importance of various scene elements which apparently was the biggest failure of the model tested in this research.

V. Conclusions

From the data collected in the experimental work done for this thesis, it can be concluded that there is a significant relationship between the expected loss value for a tone reproduction system as derived using statistical decision theory, and subjectively evaluated tone reproduction accuracy. However, the degree of correlation is not high enough for this merit function to be of any great value for the optimization, or classification of atone reproduction system. It was determined that a loss function of the form:

$$\text{Loss}(B_i) = B_o - B_i^N$$

with an exponent of $N = 3$ produced the best fit of the various functions tested; it was found that the significance of the regression could be increased slightly by the inclusion of a term which represented the expected loss with a first order coefficient of the loss function as a variable in the relationship, but even then an r^2 value of only 0.61 could be obtained.

It is hypothesized that the major failure of the model was in its inability to take into consideration the relative importance

of various areas of the scene. Apparently an observer will judge some regions of a reproduction more critically than others depending on its image content, rather than considering all regions of the print equally as was assumed in the derivation of this model. Several alternate approaches have been suggested which may better simulate the subjective process which an observer uses when evaluating a tone reproduction system.

REFERENCES

1. F. Hurter and V. C. Driffield, J. Soc. Chem. Ind., 9, 455 (1890).
2. L. A. Jones, J. Franklin Inst., 190, 39 (1920).
3. C. J. Bartleson, "Criterion for Tone Reproduction," J. Opt. Soc. Amer., 58, 992 (1968).
4. S. S. Stevens, Psychol. Rev., 61, 153 (1957).
5. C. J. Bartleson and E. J. Breneman, "Brightness Perception in Complex Fields," J. Opt. Soc. Amer., 57, 953 (1967).
6. L. D. Clark, "Mathematical Prediction of Photographic Picture Quality from Tone-Reproduction Data," Photogr. Sci. Eng., 11, 306 (1967).
7. Edwin J. Breneman, "The Effect of Level of Illuminance and Relative Surround Luminance on the Appearance of Black-and-White Photographs," Photogr. Sci. Eng., 6, 172 (1962).
8. D. Jameson and L. M. Hurvich, Vision Res., 4, 135 (1964).
9. C. J. Bartleson and E. J. Breneman, "Brightness Reproduction in the Photographic Process," Photogr. Sci. Eng., 11, 254 (1967).
10. C. J. Bartleson, "Optimal Image Tone Reproduction," J. SMPTE, 84, 613 (1975).
11. B. Dom, "Gray Scale Compression in Reprography," Photogr. Sci. Eng., 25, 216 (1981).
12. J. A. Yule, in Advances in Printing Science and Technology Vol. 3 - Halftone Printing, W. H. Banks (Ed.), Permagon Press, New York, 1964, pp. 17-42.
13. G. W. Jorgensen, "Preferred Tone Reproduction for Black and White Halftones," Paper presented at the 13th IARIGIA Conf., Wildhaus, Switzerland, 1975.
14. Bernard W. Lindgren, Statistical Theory, Macmillan Publishing Co., New York, 1976.

15. Robert Chung, "A Statistical Method for Image Classification and Tone Reproduction Determination," J. Appl. Photogr. Eng., 3, 74 (1977).
16. J. L. Simonds, "A Quantitative Study of the Influence of Tone-Reproduction factors on Picture Quality," Photogr. Sci. Eng., 5, 270 (1961).
17. J. L. Simonds, "Analysis of the Variability Among Density-Log Exposure Curves of Black-and-White Negative Films by the Method of Principal Components," Photogr. Sci. Eng., 2, 205 (1958).
18. J. Bartleson and W. W. Woodbury, "Psychophysical Methods for Evaluating the Quality of Color Transparencies: III. Effect of Number of Categories, Anchors, and Types of Instructions on Quality ratings," Photogr. Sci. Eng., 9, 323 (1965).
19. Y. Beers, Introduction to the Theory of Errors, Addison-Wesley Publishing Company, Cambridge, N.Y., 1953.
20. Yu. Ovchinnikov, et al., "A New Approach to Programming in Photomechanical Reproduction," The 12th IARIGIA Conference Proc., Versailles, France, Ed. W. Banks, IPC Science and Technology Press, Guilford, England, 1974, pp. 160-163.
21. C. P. Corey, M. J. Clayton, and K. N. Cuprey, "Scene Dependence of Image Quality," Photogr. Sci. Eng., 27, 11 (1983).

Appendix A: Sample output from expected loss calculation program.

FILE = BLCND

L**IN	L**OUT	LOSS F.
1.537	11.463	98.527
2.923	11.726	77.488
4.409	12.261	61.652
6.327	13.092	45.771
7.718	14.253	42.709
10.203	15.793	31.253
12.261	17.099	23.406
15.477	19.189	13.780
18.475	21.445	8.824
21.838	24.302	6.070
26.050	27.419	1.874
30.313	31.326	1.026
36.202	37.360	1.340
42.927	44.921	3.973
47.692	51.346	13.354
56.839	64.603	60.272
64.603	73.232	74.465
77.369	86.247	78.808
86.247	93.470	52.174
103.920	98.573	28.585

EXPECTED LOSS FOR FILE: BLCND IS EQUAL TO --> 61.482

CALCULATED WITH BRIGHTNESS DISTRIBUTION: REFD-B

Appendix B: Sample subjective testing record sheet.

Judge # _____ Name: _____

Level of Expertise (1-none, 5-extensive): _____

Print A (Man and Target)

number 1 _____

2 _____

3 _____

4 _____

5 _____

6 _____

7 _____

8 _____

9 _____

Print B (Rooftop)

number 1 _____

2 _____

3 _____

4 _____

5 _____

6 _____

7 _____

8 _____

9 _____

Print C (Dancing Girl)

number 1 _____

2 _____

3 _____

4 _____

5 _____

6 _____

7 _____

8 _____

9 _____

Print D (Night Scene)

number 1 _____

2 _____

3 _____

4 _____

5 _____

6 _____

7 _____

8 _____

9 _____

Appendix C: Sample output from data filer program.

Appendix D: Measured Brightness Reproduction Characteristics.

Man and Test Target

Appendix D (cont.)

Rooftop

Appendix D (cont.)

Dancing Girl

Appendix D (cont.)

Night Scene

Appendix E: Calculated Expected Loss Values

Print	order of loss function			
	1	2	3	4
ALCLD	26.65	735.6	20669	591644
ALCND	9.22	87.9	867	8850
ALCHD	3.83	19.4	112	692
ANCLD	14.97	250.5	4759	102062
ANCND	2.74	12.8	77	524
ANCHD	5.07	38.3	347	3492
AHCLD	7.27	76.6	1073	17862
AHCND	4.75	30.8	233	1912
AHCHD	9.64	132.7	2135	37405
BLCLD	17.28	311.7	5808	111036
BLCND	7.54	61.5	513	4346
BLCHD	13.15	180.9	2555	36893
BNCLD	23.73	648.7	19816	663127
BNCND	5.15	32.3	228	1764
BNCHD	15.51	250.9	4141	69416
BHCLD	16.39	288.7	5341	102341
BHCND	11.35	142.5	1887	25943
BHCHD	4.95	27.9	169	1097
CLCLD	13.17	205.0	3533	64555
CLCND	6.96	87.5	1235	18284
CLCHD	9.64	103.7	1172	13706
CNCLD	11.97	151.6	2013	27865
CNCND	4.66	28.2	188	1324
CNCHD	6.33	60.1	690	8732
CHCLD	10.62	124.3	1610	22822
CHCND	4.31	25.5	175	1314
CHCHD	5.61	46.7	482	5688
DLCLD	14.68	221.0	3409	53934
DLCND	5.38	34.8	242	1745
DLCHD	5.30	38.6	349	3653
DNCLD	8.06	72.4	751	9105
DNCND	3.13	14.4	83	597
DNCHD	4.72	46.2	635	9581
DHCLD	3.13	23.6	283	4250
DHCND	4.43	39.6	449	5611
DHCHD	6.18	90.7	1729	36645

print code: letter #1 : A=man and target, B=rooftop,
C=dancing girl, D=night scene
letter #2-3 : contrast level (Low, Normal, High)
letter #4-5 : density level (Low, Normal, High)

VITA

Kevin Edward Spaulding was born and raised in the Hebron, Connecticut area. After graduation from RHAM High School in 1979 as Valedictorian of his class, Kevin attended Rochester Institute of Technology to pursue a Bachelor of Science degree in Photographic Science and Instrumentation.

He spent two summers working in the photographic industry: the first at the DuPont Photo Products Division in Rochester, NY as a Photo-Scientist in the Test Methods department where he was responsible for the development and control of film testing procedures, and the second working for the Eastman Kodak Company, also in Rochester, N.Y., in the Commercial Publications department where he served as an editor/writer for technical publications, and designed experiments to collect related data. During his last year at R.I.T., he served as a technical assistant to Dr. John Schott for the Remote Sensing program. In 1982, Kevin was honored by the Society of Photographic Scientists and Engineers with the prestigious Raymond Davis Scholarship, and in 1983 he was inducted into the Sigma Pi Sigma physics honor society in the Society of Physics Students.

This Thesis is in partial fulfillment of the requirements for the Bachelor of Science degree in Photographic Science and Instrumentation. Upon graduation, Kevin will return to Kodak to work in the optical engineering area of Kodak Apparatus Division's Research and Engineering department.