MEASUREMENT OF DIGITAL PHOTOGRAPHIC IMAGE QUALITY: SURVEY OF PSYCHOPHYSICS JUST NOTICEABLE THRESHOLD DIFFERENCE METHOD

Pierre Lindeque[1], Andre Nel[2], Philip Robinson[3]
[1]Department of Electrical and Computer Engineering Technology
[2,3]Department of Mechanical Engineering Science
University of Johannesburg, South Africa
pierrel@uj.ac.za, andren@uj.ac.za, philipr@uj.ac.za

ABSTRACT
The modeling and quantification of digital photographic image quality has, from a psychophysics perspective, traditionally followed two paths, one of which is the discriminable small or just noticeable difference (local psychophysics) as detected in an image pair; further extended to cover a wide range of attribute artefactual quality variation. This method has its roots in the mathematical and psychological modeling of psychophysics and boasts a long history starting with the work of researchers such as Bernoulli, Weber and Fechner (18^{th}, 19^{th} century). The method models human perception of difference as a full scale logarithmic law and will be surveyed for its value in the determination of the quantitative quality of digital images.

KEY WORDS
Image quality, local psychophysics, JND, Weber's law, Fechner, psychophysical function.

1. Introduction

The first ever printed photographs of Niépce in 1826 and Daguerre in 1837 introduced the era of photographic image capture via chemical means; in time followed by the introduction of digital imagery with the landmark release in 1999 of the Nikon D1 as the first ever fully *digital* single lens reflex camera. Since that time, photographic image assessment and assessment quantification has grown in importance to photographers, equipment manufacturers and people in general. The public has shown a lively interest therein albeit indirectly, as the viewers (or consumers) of displayed photographic artifacts, whether directly involved in the capture process (as subjects) or not. An increasingly important criterion for the prospective photographic equipment purchaser is that of the quality of captured digital images that leads to the overall attainment of improved-quality photographs [1 et al.].

2. Measurement, Physical and Perceptual

Humans do not perceive "the entire earthly reality" but only a subset thereof. We are visually aware of only a small part of the total range of energies associated with the electromagnetic spectrum, different for instance from the range perceived by other members of the animal kingdom, each governed by its own limitations. The concept of "light" is defined only in human terms, referring to that small range of the electromagnetic spectrum that our human eyes are responsive to as developed according to evolutionary needs. In our quest for a better understanding, a need to measure emerged.

Physical measurement has a long development history with a, by now, well-established modus operandi, using instruments to objectively measure physical phenomena. Although the initial development of this science was challenging it has emerged as a well-defined process with well-defined goals. Physical measurement is mostly unaffected by external influences, unlike human perceptions that is affected by both external context and internal psychological activity and emotion. This led to the belief that the quality of an image should be measured by objective means alone. The mere thought of trusting the human intellect to "measure" just about anything was met with suspicion [2].

Human beings are indeed susceptible to external influence and internal disposition, possessing the ability to perceive a range of different types of stimuli, all challenging for attention from the brain, which in its own right is capable of coping rather well albeit not ideally suited to the act of measurement. Several pairs of cranial nerves make possible the interconnection between the cortex and the sensing and motor control systems of the human body, with large areas of the cerebral cortex and brain stem dedicated to the servicing of such stimuli and corresponding motor action. As human beings we therefore find ourselves in the middle of two places; on the one hand we are quite capable of playing the role of "the rational man or woman"; but on the other hand we are emotional beings that are not capable of invariance. A system of measurement was however needed [1 et al.] and [2].

The means to achieve such a measurement originated from the field of psychophysics, using physical measurement as a well-defined domain to study the psychological human functional (or psychophysical) co-domain value. From a physical point of view it was of importance to measure in terms of *objective measures* providing well established and repeatable methods of measuring with the least degree of ambiguity, the measurand. At the same time it was essential that the human subjectively informed measurement be obtained and integrated with the physical variable values, as ultimately only real people can judge any sensory input in qualitative terms – and this may be especially true of visual material input [4].

3. Objective Measures

Psychophysics took to this task by, in general, employing a system of sensation response capture and mathematical modeling that is based on the appropriate artefactual physical measurement, resulting in a useful perceptual measurement. The sensation response and physical measurement are then functionally aligned and referred to as the psychophysical function. An example of a physical measure is the modulation transfer function (MTF), an objective measurement of image sharpness. The MTF scale becomes the domain for an investigation of the psychophysical co-domain, a representation of the perceptual attribute of "unsharpness". Two main paths have been pursued namely local or small difference discrimination on the one hand and global or magnitude estimation on the other. This paper will focus on the former, namely that of local psychophysics. The psychophysical method laid a foundation that is adaptable to numerous modalities of scientific interest of which photographic attribute image evaluation is one.

From a photographic point of view the Image Quality Circle described by [5] was an attempt to bring together some of the major concepts in image quality assessment. In this paper the view taken on image quality assessment is that of comparison of an image degraded in some way being compared to an image that shows no such degradation *perceptually*. Furthermore a single, rather than multiple, attribute assessment method will be considered in a manner that views an image in its entirety. This distinction requires careful attention to the instructions given to a viewer that is to assess an image. It will thus be expected form an observer to focus on only one attribute such as unsharpness, but viewing the image in a holistic manner. That implies that this paper focuses on images of which the image quality attribute nature is that of being artefactual rather than aesthetic or personal. Aesthetic or personal image quality will not be considered as this is regarded as a different question altogether. Examples of artefactual typed attributes are colour accuracy, noise and sharpness (or unsharpness) [3] and [5].

The main motivation for this research lies in the development of a mathematical model to characterize human perception in terms of photographic image quality. The logarithmic relation was the outcome of the work of Weber, Fechner and others that lay an important part of the foundation for image quality assessment research [1 et al.], [2], [3], [5], and [6].

4. Small Discriminable Difference (or Local Psychophysics)

The method of small discriminable difference originates from the fundamentally important work of Bernoulli, Weber, Fechner and Thurstone. Synonymous terms for this class of image quality assessment are: minimalist constraint, local psychophysics, threshold detection and threshold analysis. This paradigm puts the focus on small differences between objects, whether they are objects of weight or photographic objects of which an attribute has the focus of the research. The objective is to find a way of modeling the human mind when faced with the task of detecting some kind of value appreciation or depreciation from an arbitrary starting point. Of the earliest research came from the field of economy and in particular the work of Bernoulli. His research identified the required variables and relations that were needed to make a formal start [1 et al.], and [7].

5. The Utility Function of Bernoulli as Viewed by Masin

As far back as the 18th century, Bernoulli devised a relation between personal monetary asset value and the utility value thereof to an individual. Masin [7] reflected thereon, crediting Bernoulli for what must be some of the very first work with regard to modeling. What follows is a summary of the observations of Masin with regard to Bernoulli's analysis [7].

For any particular person a distinction is made between the (objective) asset value x owned by the person at a point in time and the subjective affluence or wealth value y. The variable x and y were considered adequate for the task; however variation variables were needed too, fulfilled by Δx and Δy. The domain is the positive reals, including 0, and was assumed continuous. The co-domain represented human perceptions of another's wealth also in a numerical continuous range. An increase in x, denoted as Δx results in an increase in y, denoted as Δy. The implication is a direct proportionality between Δx and Δy:

$$\Delta y \propto \Delta x \qquad (1)$$

Bernoulli argued that Δy is less for the rich than for the poor. In other words, an equivalent increase in asset value is not appreciated by the rich as much as by the poor. That is to say that as the level of asset value increases from a particular level, the corresponding feeling of increased wealth Δy progressively diminishes and therefore decreases in absolute terms. Therefore:

$$\Delta y \propto \frac{1}{x} \tag{2}$$

A basic relation was constructed from equations 1 and 2, yielding $\Delta y = b\left(\frac{\Delta x}{x}\right)$, with b a constant coefficient derived from empirical conditions and data. By assumption of infinitesimal scale delta values and manipulation we have:

$$\frac{dy}{dx} = b\left(\frac{1}{x}\right) \tag{3}$$

As our interest lies with the summation of y as a function of x, equation 3 is integrated to yield the following relation, with k providing the perceptual threshold:

$$y = b\ln(x) + k \tag{4}$$

Intuitively the relation makes sense, showing a trend of flattening out as the asset value increases, resulting in a non-linear negatively accelerated perception of well-being to the individual [7].

6. Weber's Law

Weber was an anatomist and physiologist with an interest in human perception and the sense of touch which led to his study of the sensation of weight. Weber devised the relation: $\frac{\Delta x}{x} = c$, also referred to as the Weber fraction, which after manipulation becomes:

$$\Delta x = cx \tag{5}$$

Weber argued that a person that lifts a weight x will only detect a difference Δx should it be heavier at least by a constant coefficient c. This means that for a c value of 0.1, a weight x of 10 N would require to be increased by 1 N to a total weight of 11 N to be detected as being heavier. So a weight of 100 N would have to be increased by at least 10 N before it is felt to be different. More specifically, the average person will only detect weight gain for a Weber fraction of 0.05. It must be noted that Weber used a probabilistic method in his work to model the degree of human variance that emerged from the data, thereby modeling the human tendency to fluctuate in choice despite virtually identical experimental conditions. He determined that the value of an increment (in weight) to be such that 75% of respondents recognized the difference. This was generally applicable to other sensory modalities as well, such as sound etc. Weber termed this the "just noticeable difference" (JND). This is significant as it has gained popularity as unit of measurement of small discriminable or detectable differences. Weber's law therefore is not applicable to large differences, in principle. The question of large difference was a separate question that Fechner and others considered [1 et al.], [4], [6], and [8].

Figure 1 displays a simple graph showing the fundamental characteristic of Weber's law.

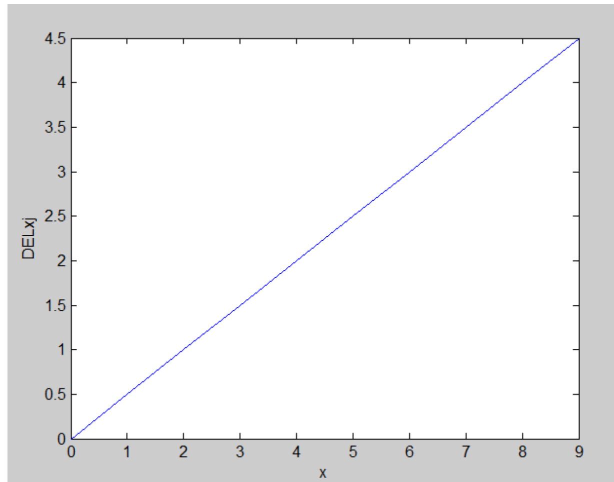

Figure 1: The increment function modeling $DELx_j$ as a function of the objectively measured variable x.

The graph shows that with the actual weight on the abscissa the relation represented on the ordinate is linear with respect to x, denoted as $DELx_j(x)$ or simply $DELx_j$. Weber's function therefore may be expressed as: $DELx_j(x) = cx + d$, but if it assumed that the perceptual threshold is zero , or close thereto, we may state that:

$$DELx_j(x) = cx \tag{6}$$

The unit is N per JND, dependant on the weight x in N. What subsequently is of importance is that the $DELx_j(x)$ function value resembles what might be referred to as an *increment function*. This function thus has the task of yielding the value of the equivalent weight corresponding to a single JND increment. In other words, the function yields the number of units of the abscissa that corresponds with the perceptual unit of one JND. Weber confirmed the law through experiments that covered a range of stimuli values. The law has been found to hold quite well over the mid-range stimuli values, but holds less well for incremental changes at the extremes [4], and [9].

It is clear that careful empirical work would be required to evaluate this function. What is required is to obtain, for a specific x value as a constant parameter, the needed delta or incremental value of x, such that 75% of respondents become aware of the increased mass. A simplified regression analysis applied to the measurements from a large group would yield the 75% increment value for that particular x weight. This is then the desired $DELx_j(x)$ value.

At this stage we therefore are in possession of a series of $DELx_j(x)$ increment values that are needed to characterize, over the total required weight range, as accurately as possible the increment. It would appear that in many modalities this characteristic was found to be close to linear in the middle sensory range, thereby confirming Weber's law for such range at least.

The empirical data obtained from experiments of the type Weber conducted involve relatively large absolute values as well as relatively large increments. It simply is not possible to accurately characterise human perceptual performance in infinitesimal terms. What is of utility however is to make reasonable assumptions that might bridge the gap between human oriented "delta" values and scientific infinitesimal values. If the increment function were a constant function, we could readily use it, but our increment function has a linear (sloped) characteristic which makes matters more challenging. This point has been elaborated by [6] and [10], arguing that the situation calls for unique functions. However, if care is taken with the increment function, we will have a fairly accurate figure of rate of change of change-of-weight versus change-of-perceptual-difference.

Considering the increment function once again, it may be stated that the $DELx_j(x)$ function represents the $\frac{\Delta x}{\Delta j}$ ratio, meaning the ratio of a change in weight to a change of one single JND. This means that the expression represents the relation of an incremental amount in x versus an incremental and corresponding amount in perceptual JND's, thus implying the JND increment relation:

$$DELx_j(x) = \frac{\Delta x}{\Delta j} \tag{7}$$

Fechner pointed out the importance of this fact; namely that Weber created the much needed link between the human perception and a quantitative value that could represent it in physical terms [1 et al.]. With the previous assumption in mind, Δx may be viewed as approximately equal to dx and Δj approximately equals dj. Fechner too assumed that the empirical dimension could be considered to be the same as for the infinitesimal dimension, referred to by him as the "auxiliary principle". [7]

We thus have:

$$DELx_j(x) = \frac{dx}{dj} \tag{8}$$

7. The Logarithmic Relation

Fechner further investigated the question of human perception, based on the work of Weber. Having started off in medicine, his focus moved to mathematics, physics and eventually the relation between matter (or energy) on the one hand and mind (or sensation) on the other attempting to formally describe the relation [1 et al.].

As discussed in the previous section, Fechner saw in Weber's law the utility of the increment function, from which a relation may be obtained as a linear function with a strong empirical base. Fechner essentially started off with $DELx_j(x) = cx$, and proceeded to $DELx_j(x) = \frac{dx}{dj}$. The next step is a reciprocal function of equation 8, yielding:

$$\frac{1}{DELx_j(x)} = \frac{dj}{dx} \tag{9}$$

As we are interested in the accrual or summation of infinitesimal pieces of $j's$, it is essential to obtain the reciprocal function of the increment function. We will denote this function as $RECx_j(x)$ and express it as:

$$RECx_j(x) = \frac{dj}{dx} \tag{10}$$

The function has the basic form of figure 2 below. The $RECx_j(x)$ function implies:

$$\frac{dj}{dx} = \frac{1}{cx} \tag{11}$$

The units are JND's per gram and $1/c$ is a constant. Figure 2 shows the reciprocal function still based on the original x domain.

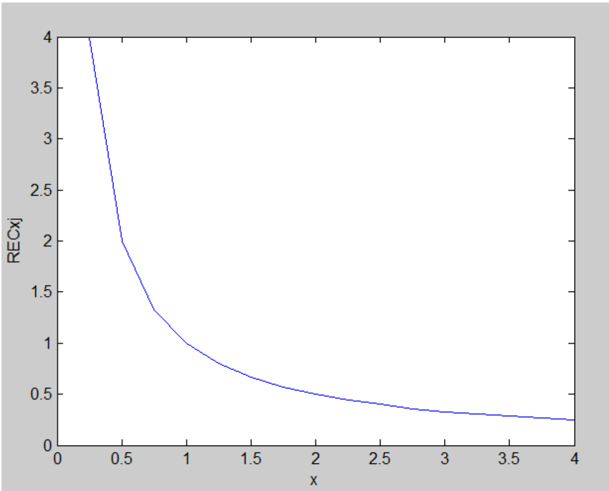

Figure 2: The reciprocal function modeling $RECx_j(x)$ as a function of the objectively measured variable x.

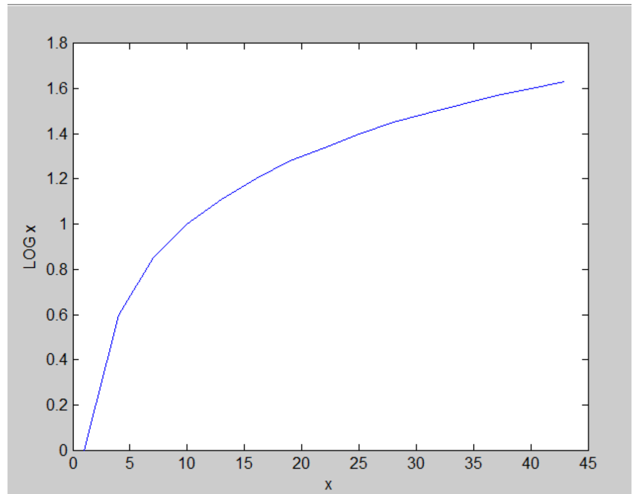

Figure 3: The natural log function modeling the perceptual j as a function of the objectively measured variable x.

The function is the rational function $\left(\frac{1}{c}\right)\left(\frac{1}{x}\right)$ which is asymptotic at the origin along both axes. This implies that the reciprocal function will receive as input argument a real valued set that will be a subset (due to the exclusion of 0) that is yielded by the increment function. The final value from the reciprocal function therefore is a functional composition of the increment function and the reciprocal function.

From this point the accumulation of $j's$ (or JND's) is possible through the integral: $dj = \int\left(\frac{1}{cx}\right)dx$ which yields:

$$j = \frac{1}{c}\ln x + k \qquad (11)$$

The natural log (or ln) function has the form shown in figure 3 below.

Fechner concluded that the psychophysical function is a logarithmic relation [7]. Fechner's method of cumulating JND's was criticized by Luce ([6]) arguing that the log law does not always hold, comparing the results of magnitude estimation or fractionation with that of the log based results. The discrepancies were also evident across different modalities, such as sound and smell and so too in weight. Fechner was also criticized for inconsistent definitions, arguing that he wrongly defined the magnitude of each and every JND on the perceptual or sensational scale to be equal perceptually. The evidence from magnitude estimation is once again cited as the reason for this disagreement. Luce further cites the results of tone experiments where 20 JND's does not sound half as loud as 40 JND's, when compared to the results of magnitude estimation [6]. Stevens also criticized Fechner with regard to the difference in results between the cumulative JND method versus that of magnitude estimation [10]. Stevens introduced magnitude estimation to psychophysics [6], and [10].

8. The Law of Comparative Judgement

Thurstone developed a probabilistic model in support of the just noticeable difference by considering a dual normal distribution for both objects under investigation through the concept of a difference limen or threshold. This establishes a model for psychophysical judgement by obtaining a difference (assumed normal) distribution as a measure of human pairwise judgement of the relevant modality, such as two weights or two photographs. For [11] the JND was of primary importance and was considered as having, by definition, a constant value in the eyes of the observer. The value thereof lay in the characterization of the relation between the pairwise physical stimuli on a continuum as domain and the perceptual separation or difference between the two objects, on the co-domain and well so as equal appearing JND's. This is of importance as it further supports the conjecture that JND's are primarily fit for small discriminable difference typed assessment rather than for large differences.

Thurstone [11] achieves this by "presenting" a fictitious series of stimuli on a continuum alongside a series of postulated *discriminable processes* that are the observer's correlate "response" to any specific stimulus from either of the two objects. This response is not deterministic but instead varies about a mean as the observer demonstrates a certain amount of variance in assessing any one particular object. As much as there is reference to an observer, the sketched process is not practical as the judgement would be too difficult and inaccurate. The second stimulus is "presented" as well, also along the earlier physical continuum, also represented as a likewise normally distributed response. Once again, this is fictitious and only serves to provide a theoretical base for establishment of the *difference distribution*. At this stage the only known fact is the physical (measured) difference between the two stimuli, such as weight in N.

The choice made in a practical psychophysics type of assessment is done by an observer in such a way that the judgement yields a useful figure of difference between the two objects. However the choice presents itself as a derived distribution, also assumed normal. For the difference distribution a z calculation is performed in a reversed process; by working back from the probability figure of the experimental judgement from observers. The reversal takes on the following form. Through experimental judgement of a response pair (photographic or otherwise), a percentage of correct responses become a base for the calculation of a difference distribution deviance z value. This value takes on a range of real typed values that may then be correlated with the known physically measured difference between the two objects. In this way, an assessment in the perceptual space is transformed into a quantitative value, which represents the perceptual difference of the average observer. This process is essentially the same as Weber's elucidation of the 75% JND figure [11].

Photographic quality assessment has for some time based its work on that which was established by authors of the field of psychophysics. Keelan and others made extensive use of these models [2].

9. Photographic Image Quality Currently

Currently, in terms of JND small discriminable threshold modeling, many utilize the models developed by Weber, Fechner, and Thurstone etc. The increment function and the logarithmic law form the base that relates perceptual degradation in image quality to physical *objective metric* values. By the term objective metric is meant an objective measurement that attempts to model human perception. If this were totally achieved the increment function would be a constant function but this is difficult to achieve.

Keelan [3] provides an integrated hyperbolic increment function that performs the earlier described reciprocal function. The final model presents a zero value in the case of no perceptual degradation and drops off in a negative numeric direction as the amount of attribute objective metric value increases. Thurstone's difference distribution is utilized, converting it to an "angular cdf" that models perception, yielding a quantitative z value which, along with the percentage figure from the psychometric experiment, is used to construct a Weber-like increment function.

A range of images are laid out alongside one another, each with the JND degradation figure and corresponding value of objective metric stipulated along with the image, to form an image quality ruler. Figures 4 and 5 below give examples of photographs that respectively show: no degradation in terms of sharpness (in this case), and a certain measure of degradation [3]. Such images may be utilized in the construction of a quality ruler.

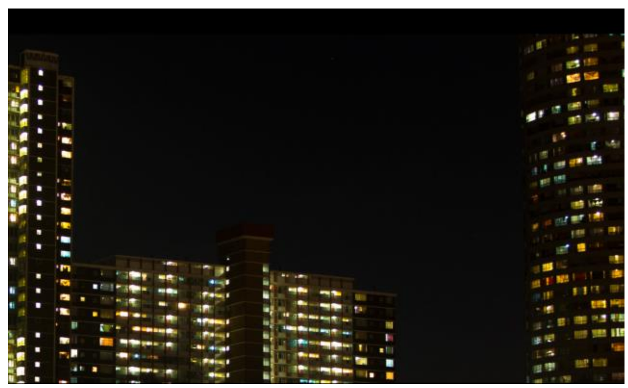

Figure 4: A reference image with no (sharpness) degradation

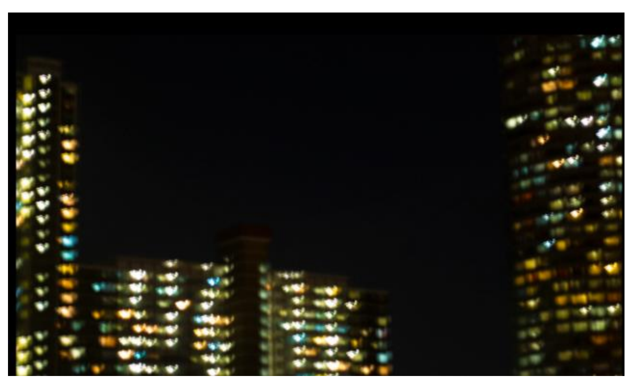

Figure 5: An image that shows degradation in terms of sharpness, in this case due to camera shake.

10. The Focus of this Research

The aim of this work is an empirically constructed mathematical small discriminable threshold model using a 35mm digital single lens reflex Canon camera and 85mm diffraction limited high quality lens for the perceptual measurement of the quality attribute of unsharpness. It needs to be added that this writing represents the one part of a two part initiative of which the second will investigate magnitude estimate based image quality assessment. The stages that will be followed will be in accordance to what was described in this writing: the MTF objective measure, the increment function, the reciprocal function, and the final logarithmic function.

The process will start with the taking of one single excellent (best-possible) black and white image that is to suffer no visible degradation. The scene will comprise a range of suitable indoor artifacts that exhibit the type of sharp edges required for the measurement of MTF unsharpness as objective measure, but will also be reasonably interesting to the average person. A series of experimental parameters will be established. They are camera settings of: base ISO of 100, a lens aperture of f1.8, a fast shutter speed of 1/500 of a second, and lighting that is adequate to ensure working at the base ISO of the camera. The camera will be secured so as to inhibit any camera shake. The camera will be utilized in a mode that will result in the least amount of in-camera (digital) signal processing.

From the original reference image a series of images will be created spanning an adequate range of unsharpness in terms of MTF figure. This will be achieved by DSP means, introducing the required amount of unsharpness through filtering. The images will be presented in image pairs for assessment of a large number of observers. This data will allow the increment function according to Weber's law to be constructed in a detailed manner. Although this relation is expected to be near linear, there is no guarantee for this, and will be investigated with care. Once this is in place the next step is the construction of the reciprocal followed by the logarithmic function.

With the log function in place a series of images will be selected so as to build an image quality ruler, which may be used in a relatively short and simple psychometric experiment whenever an image suffering from unsharpness is to be judged.

11. Conclusion

What has been of particular interest in this incomplete study is the mathematical modeling whereby image degradation may be understood. The increment function as introduced by Weber is of pivotal importance, requiring careful empirical work and has been investigated. The probabilistic modeling of pairwise comparative assessment is equally pivotal in terms of empirical importance as it needs to collaborate with the establishment of the increment function.

The search for the increment function forms the heart of the system, following which the next steps are relatively simple: the reciprocal and log functions, in a functional composition. The integrated overall approach taken in this research study has been the focus of this work as the overall perspective provides insight into the method.

An image ruler is a convenient way to easily assess images in terms of JND's and will be established. The outcome of this author's research will compare this method, the small discriminable threshold with the magnitude estimate method for the ultimate attainment of better image quality quantification.

12. References

[1 et al.] J.M. Wolfe et al, *Interval analysis sensation & perception second edition*. Sinauer Associates, Sunderland Massachusetts, USA, 2009.

[2] F.S. Roberts, *Measurement theory with applications in decision making utility and the social sciences*. Cambridge University Press, New Brunswick, New Jersey, 1985.

[3] B. Keelan, *Handbook of image quality characterization and prediction*. CRC Press, Boca Raton, USA, 2002.

[4] E. Hunt, *The mathematics of behavior*. Cambridge University Press, New York, USA, 2007.

[5] P.G. Engeldrum, Image Quality Modeling: Where Are We?, *IS&T's 1999 PICS Conference*, Savannah, Georgia, Society for Imaging Science and Technology, Springfield, Virginia, 1999, 251-255.

[6] R.D. Luce, The Derivation of Subjective Scales From Just Noticeable differences. *Psychological Review, 65*(4), 222-237, 1958.

[7] S.C. Masin, V. Zudini, and M. Antonelli, Early Alternative Derivations Of Fechner's Law. *Journal of the History of the Behavioral Sciences, 45*(1)*, 56-65, 2009.*

[8] S. Suthaharan, S-W. Kim, and K.R. Rao, A New Quality Metric Based on Just-noticeable Difference, perceptual Regions, Edge Extraction and Human Vision. *Canadian Journal of Electrical and Computer Engineering, 30*(2), 81-88, 2005.

[9] H. Altner, *Fundamentals of Sensory Physiology*. Springer-Verlag, Heidelberg, 1978.

[10] S.S. Stevens, The Psychological Review. *American Psychological Association, 64(3), 153-181, 1957.*

[11] L. L. Thurstone, *The measurement of value*. University of Chicago Press, Chicago, 2007.