

# Introduction to Perceptual Linearization of Video Display Systems for Medical Image Presentation

Bradley M. Hemminger, R. Eugene Johnston, Jannick P. Rolland, and Keith E. Muller

The perceptual linearization of video display systems should play a significant role in medical image presentation. It maximizes the faithfulness of information transfer to the human observer; it provides a method for standardizing the appearance of images across different display devices; and it allows for calculation of the inherent contrast resolution of different display devices. This paper provides insight into the process of perceptual linearization by decomposing it into the digital driving level-to-monitor luminance relationship, the monitor luminance-to-human brightness perception relationship, and the construction of a linearization function derived from these two relationships. A discussion of previous work in these areas is given. We then compare and contrast the results of previous work with recent experiments in our laboratory and related work in vision and computer science. We conclude that (1) sufficiently good visual models exist for agreeing on a standard method of calculating the perceptual linearization function; (2) improvements in the resolution and luminance distribution of the digital-to-analog circuitry in display systems are required for medical imaging; and (3), methods for calculating a linearization remapping from a perceptual linearization function currently have significant error and should be replaced with methods that minimize perceptual error.

Copyright © 1995 by W.B. Saunders Company

**KEYWORDS:** perceptual linearization, display standardization, gray-scale image presentation, video displays.

**P**ERCEPTUAL linearization of video display monitors should play a significant role in medical image presentation. First, it allows the maximum transfer of information to the human observer because each change in digital driving level of the display yields a perceptually equal step in perceived brightness by the human observer. Second, for an image to be perceived as similarly as possible when seen on different displays, the two displays must be standardized, which can be done when they have been perceptually linearized. Third, perceptual linearization allows us to calculate the perceived dynamic range of the display device, which allows comparing the maximum inherent contrast resolution of different devices.

Perceptual linearization was first suggested for medical image presentation by Pizer and Chan,<sup>1</sup> and in follow-up work<sup>2-5</sup> at the University of North Carolina at Chapel Hill (UNC). To

best visually present an image represented as digital data to the human observer, we would like to maximize the information transferred in mapping the digital driving levels (DDLs) to perceived brightness levels. Perceptually linearizing the mapping from the image data space to the human observer's visual sensory space most faithfully transmits changes in intensities in the image to the human observer.<sup>3,6,9,10</sup> This simply means that to the human observer, equal absolute changes in the input values to the display system should result in equal absolute changes in the perceived visual sensation.

Many advantages have been attributed to linearization. Most of these were first described by Pizer and Chan,<sup>1</sup> and have been further qualified more recently by others.<sup>11,12</sup> We will categorize them as error minimization, standardization, and characterization.

Minimization of error in the display system refers to minimizing distortions in the relationship between input data and perceived sensations, so that equal changes in DDLs are reflected as equal changes in perceived brightness. We have intentionally chosen not to refer to this as optimization, to carefully distinguish perceptual linearization from choosing an optimal gray-scale processing for an image. Perceptual linearization by itself is not intended to be the optimal gray-scale presentation of the original data. The important choice of best gray-scale presentation is dependent on the specific image content and visual task, and occurs before the linearization. For instance, some image-process-

---

*From the Departments of Radiology, Computer Science, and Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC.*

*Supported by National Institutes of Health Grants No. P01-CA47982, R01-CA60193, and R01-CA44060. An early version of this paper was presented at the SPIE Medical Imaging Conference in Newport Beach, February 13-14, 1994, and is used with the permission of SPIE.*

*Address reprint requests to Bradley Hemminger, Department of Radiology, 513 Old Infirmary Bldg, University of North Carolina, Chapel Hill 27599-7510.*

*Copyright © 1995 by W.B. Saunders Company  
0897-1889/95/10801-0003\$3.00/0*

ing technique (window and level, or adaptive histogram equalization) might be performed on the original image, resulting in the desired gray-scale processed image, whose data values are distributed in a linear uniform manner. Then the perceptual linearization is responsible for making sure the relationships in these data are properly conveyed to the human observer by having the display system reflect the equal changes in the input data as equal changes in the perceived sensations of the human observer.

Standardization is the attempt to make images presented on different display devices appear similar. As modes of radiologic acquisition become increasingly computerized, more and more of the display media are digitally based (computed tomography, nuclear medicine, magnetic resonance imaging, positron-emission tomography, computed radiography). This has increased the need for standardization as more images are viewed on monitors as well as on light boxes. Blume et al<sup>12</sup> provide a list of several advantages of standardization: predictable and reproducible gray-scale rendition, similarity between presentations of the same image on different display devices, and the ability to make comparisons between quantitative observer performance measurements over different display systems.

Perceptual linearization provides a quantitative characterization of the display system. First, the quantitative information generated from the linearization provides a better description of a display system than simply the luminance range of the monitor. This would help in comparing display systems. Second, the quantification provides specific information that helps the manufacturer of the display system make the best design choices.

With the increasing use of video monitors for the display of medical images in radiology departments, we are seeing an increased awareness of the problems of video display, especially in the areas of obtaining and maintaining high-quality reproduction of the images, and in the area of standardizing presentations of images across different display systems. To achieve the benefits of error minimization, standardization, and characterization of the display systems, the medical community needs to fully understand perceptual linearization and agree upon a meth-

odology for computing the linearization. Others have advocated this need and proposed display function standards.<sup>11,12</sup> Before a decision is made on such a standard it is necessary to understand all the issues involved. This paper will attempt to provide a basic understanding of the issues involved in perceptual linearization. A complete description of perception linearization as applied to medical imaging in radiology is given in an earlier paper.<sup>13</sup> In the second section of the paper, we present a paradigm for describing perceptual linearization in its three basic components: monitor luminance to perceived brightness, DDLs to monitor luminance, and calculation of a resulting linearization given the first two relationships. In the next three sections, we address each of these components in turn. In each of these sections we cover previous work in this and related fields, as well as new results from our laboratory and tie these results into a single framework for analysis. Finally, the last section summarizes what conclusions can be drawn from the earlier sections, and discusses what areas still require investigation.

#### PARADIGM FOR PERCEPTUAL LINEARIZATION

The process of displaying an image on a video display monitor to the human observer is depicted in Fig 1. This paradigm applies equally well to the display of images on film. Initially, an object such as the human body is scanned, and the resulting signal (for instance tissue density) is represented on the computer as a matrix of points, called pixels. This scanning samples the original source data (continuous analog function) into discrete data (set of digital values). Each pixel is represented by a scalar value, usually in the range of 0 to 4,096 for medical image data. These are the values referred to as "recorded intensities" in Fig 1. The optional second step is that some set of image-processing operations such as intensity windowing or contrast enhancement may be performed on the recorded intensities resulting in the displayable intensities. These values are then scaled into DDLs, which must be in the range accepted by the digital-to-analog converter (DAC) of the display system. This scaling is done by a table lookup operation, often referred to as a lookup table (LUT) or color map table. LUTs are often

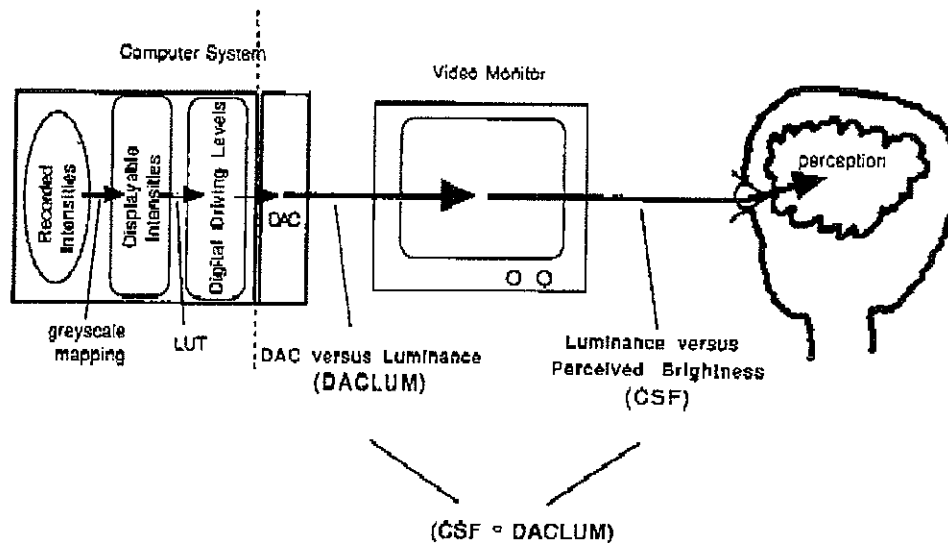


Fig 1. Diagram of components of perceptual linearization.

used to do intensity windowing dynamically, or to implement a linearization LUT (these are sometimes called gamma correction curves). The output of the LUT goes to the DAC, which takes the input DDL and converts it to an analog voltage level which is used to drive the monitor at different luminance levels. The luminance generated by the monitor is then recorded and processed by the eye-brain human visual system, resulting in the sensation of brightness by the human observer.

Two important topics relevant to the discussion of perceptual linearization are described elsewhere and are not discussed in this paper: imperfect display devices and imperfect source images. The standard video display device in use today is the cathode-ray tube (CRT) and there are known problems with the reproduction of luminance values on CRTs. Discussion of such problems including spatial and temporal nonlinearities, CRT noise, internal scatter, and distortion are well described.<sup>11,14,15</sup> The second area is noise in the source image. Noise in the source image has been discussed in some of the linearization work<sup>12</sup> and also incorporated into some of the recently proposed visual models<sup>16,17</sup> discussed below in the section on applying human visual models.

From the standpoint of linearization there are two important relationships in this process, that of the DDLs of the computer's DAC and the luminance of the monitor, and that of the

monitor luminance and the brightness perceived by the human observer. The first relationship of DAC to luminance will be referred to as DACLUM. The second relationship, that of luminance to perceived brightness, is best examined using a luminance contrast sensitivity function (CSF). CSFs measure the change in luminance ( $\Delta L$ ) required for a target, so that it may be detected from the surround luminance ( $L$ ) as a function of the surround luminance. More specifically, contrast thresholds are defined as  $\Delta L/L$ , whereas CSFs are defined as its reciprocal, ie,  $L/\Delta L$ . CSFs in this paper will refer to  $L/\Delta L$  versus  $L$ , whereas in vision literature, CSFs usually refer to  $L/\Delta L$  versus spatial frequency of the target.

If we think of the DACLUM and CSF curves as functions, and compose them on their common variable of luminance, we arrive at a CSF\*DACLUM function that defines the overall effect of the DACs, monitors, and human perception in the display system. The inverse of this function can be determined and used to remap the image values to perceptually linearize the relationship between the gray levels of the image in the computer and the sensation of brightness to the human observer. Thus, we have defined the three important components of perceptual linearization: the DACLUM curve, the CSF curve, and the linearization curve, which is the inverse of CSF\*DACLUM. We will first consider the CSF curve, which can be

considered fixed because of the wiring of the visual system. Second, we will consider the DACLUM curve, which is decided by the manufacturers and designers of the video display system, and should be based on the CSF curve. Finally, we will consider the calculation of the linearization, which depends on the results of both the CSF and DACLUM components.

#### RELATIONSHIP OF LUMINANCE TO PERCEIVED BRIGHTNESS

The CSF curve shows the relationship between the luminance displayed on a monitor and the brightness perceived by the human observer. The important attributes are the luminance range of the monitor and the distribution of distinguishable gray levels over that range that are perceivable by the human observer. The luminance range of display devices may vary between devices, as well as grow larger as higher brightness monitors are developed. By characterizing the CSF over the entire luminance range of the human visual system, we can define the CSF response at all possible monitor luminance levels.

Our goal is to quantitatively model the human observer's sensitivity to contrast differences, ie, an observer's ability to distinguish between different luminance levels. Quantitatively modeling the CSF allows us to calculate proportional changes in contrast sensitivity of the human observer that will correspond to equal proportional changes in DDLs. It also allows us to calculate the perceived dynamic range (PDR) of the human observer for a display system, where the PDR is the number of different gray levels that can be distinguished for that display system. The problem that arises in defining a CSF is that there is no overall function valid across the many variables affecting the presentation of an image (size of the image, luminance of the surround, ambient light, viewing distance, adaptation, etc). We can define CSFs for specific experimental tasks and measure them. However, these would only be accurate models for tasks that exactly match the experimental conditions. This suggests two different avenues of exploration. One method is to take experimental or theoretical models from vision research and parameterize them to fit the clinical presentation task as closely as possible.

The second avenue is to create experimental tasks that match generic clinical image presentation situations and empirically define a CSF for that task. These are discussed and compared in detail below.

#### Applying Models From Vision Research

Researchers have been studying the human observer's ability to make contrast distinctions for well over 100 years. Early experiments often consisted of determining relationships between two physical patches of possibly differing luminance. Experiments that measured this relationship were graphed as  $\Delta L/L$  versus  $L$ ; where  $\Delta L$  is the change in luminance required to detect a difference from the surrounding field of luminance,  $L$ . These curves are generally referred to as contrast threshold curves. An example study depicting this relationship is shown in Fig 2.<sup>18</sup> The contrast threshold, or the minimum difference between two luminances at which an observer can detect a difference between the luminances, is also referred to as a just noticeable difference.<sup>5</sup>

By the early twentieth century it was recognized that the overall luminance range of the human observer<sup>19</sup> is  $\sim 0$  to  $10^8$  candela/m<sup>2</sup> and contrast threshold over this range could be broken into three areas: scotopic, mesotopic, and photopic. In the low-luminance scotopic region (below  $\sim 5 \times 10^{-3}$  cd/m<sup>2</sup>), the lumi-

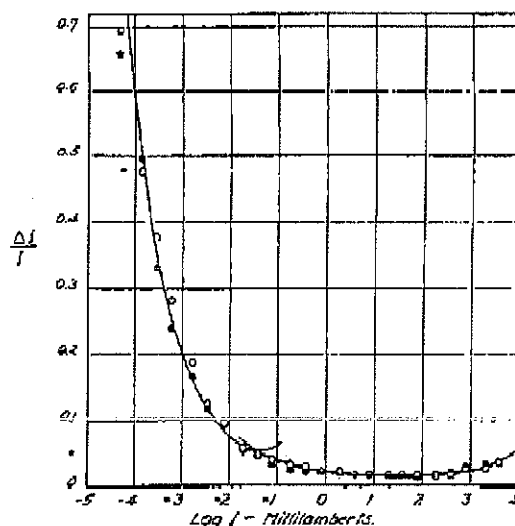


Fig 2. Representative experimental results showing contrast threshold relationship,  $\Delta L/L$  versus  $L$  (from Murchison<sup>18</sup>).

nance detection is mainly via rods, with significant contribution from the parafoveal area. In this region, referred to as the Rose-De Vries region, contrast detection is dependent on luminance and roughly follows a power ( $-1/2$ ) law. In the photopic region beginning around  $1 \text{ cd/m}^2$  to  $10 \text{ cd/m}^2$ , referred to as the Weber region, the contrast threshold is generally constant, and the response is mainly from the foveal area, which is tightly and homogeneously packed with cones.<sup>20</sup> The curve in this region is described by the Weber-Fechner law, which states that for a luminance,  $L$ , and a change in luminance,  $\Delta L$ , the ratio  $\Delta L/L$  is constant. The most sensitive (smallest) measured value of this constant is in the range 0.003<sup>21</sup> to 0.01.<sup>20</sup> Finally, the mesotopic region describes the middle area between these two regions where there is a combination of the effects of the scotopic and photopic regions.

In more recent times, especially since the advent of computers and video displays, work has concentrated on the presentation of gratings (square or sine wave, generally) centered on a large surround background luminance, where the observer's task is to detect the grating. This has led to the development of CSFs that define the contrast sensitivity of an observer, defined by  $L/\Delta L$ , versus the spatial frequency of the grating (ie, cycles/degree).<sup>22,23</sup> More recently some researchers have extended the stimulus targets to include different types of objects, eg, gaussian blobs.<sup>24,25</sup>

Many experimental or theoretical models have been proposed, including models based on empirical evidence and physiologic measurements on animals. Although several of these models, including both empirically oriented models such as log relationships (Weber-Fechner, etc), power law ( $1/2$  law, etc), and exponential density relationship and physiological based models (local cone, global cone, etc), offer good general descriptions of the relationship, they do not provide the necessary parameters (such as stimulus descriptions, image and visual noise, luminance of stimulus and surround, ambient light levels, etc) to model important attributes of the presentation that are present in the clinical environment.

However, several authors<sup>16,17</sup> have recently defined models based on mathematical descrip-

tions of the components of the human visual system that more completely and accurately represent the visual function. To test their models, they have taken empirical studies from the literature, encoded the parameters of the study into their visual model, and found that their models accurately predict the experimental results<sup>16,26</sup> (S. Daley, personal communication, January 1994). If we can parameterize more comprehensive models such as these to match clinical situations, we should be able to reasonably predict the CSF for a specific presentation situation. Furthermore, by studying the range of possible CSF values, we may be able to determine a single CSF that represents the majority of clinical conditions. Blume et al<sup>12</sup> have calculated estimates for several individual presentation conditions for each of the Barten, Daly, and Rogers-Carel models. Importantly, they found good agreement between the predictions of these models, and have suggested the adoption of either the Barten or Daly models for use as a display function standard.<sup>12</sup>

We have studied the three models<sup>16,17,27</sup> and implemented the Barten model based on the descriptions in Barten<sup>16,28,29</sup> and Blume.<sup>12</sup> We then performed preliminary investigations to see (1) whether the different models were complete enough (ie, did they take into account factors we found to be significant in our previous experimental work); and (2) whether by examination of the ranges of parameter values to the Barten model we could find a single set of parameter values that would be suitable for the purpose of specifying a representative CSF. After this, we examine how well our recent experimental results correlate with the representative CSF model predictions. In separate work we are investigating whether specific clinical viewing tasks can be accurately modeled, including the detection of mammographic features on digital mammograms.

#### *Are Vision Models Complete Enough?*

Some of the variables from our experiments that had significant effect on the measured CSF function were not represented by the visual models. The most significant difference was caused by multiple levels of surround. Most recent empirical results are based on the presentation of a grating of slight luminance difference

from a constant surround. The stimulus definitions of the models are similarly defined to have only a single surround. In our experimental work<sup>7,13</sup> and in general vision research,<sup>30-32</sup> experimenters have found that the luminance of the surround plays a significant role in the CSF. Because the models calculate contrast sensitivity for targets differing slightly from a single uniform surround luminance, they model an artificial condition where the human observer is most sensitive to distinguishing contrast differences. In most clinical situations, an image will have a certain expected mean luminance (surround) for the overall image, and the stimulus will be a different luminance, often located in smaller local surround, which has yet another mean luminance (and which the observer may concentrate most of their foveal gaze on). A model that allows specification of the more general condition of a large background surround mean luminance, a local target surround mean luminance, and the stimulus luminance should more accurately predict clinical presentation results.

A second area not modeled was the type of stimulus. Because such a large body of research has been done using gratings, the models were also based on gratings with variables to allow for size, number of cycles, and amplitude of the gratings. Objects we wish to detect in medical imaging are more varied, often including blobs or other structures not easily or accurately modeled as gratings. Thus, another enhancement would be to allow the specification of different types of basic visual stimuli, eg, gaussian blobs based on Bijl et al's<sup>24,25</sup> or others work. Human observers are most sensitive to line or bar-type objects in detection tasks, so gratings by themselves do serve as a good upper bound for our most sensitive responses.

The third concern was the length of time the observer viewed the presentation. Radiologists generally scan images in one of two modes: directed, eg, to rule out a mass in the upper left lobe, versus undirected, eg, looking at the lungs as an aside during a shoulder bone x ray. In a directed search, the clinician will likely spend more time carefully and exhaustively searching the area of interest. On an undirected search, a quicker, more cursory search is made. One way of modeling these two modes would be for the

visual models to have a parameter corresponding to the length of time the image is presented to the observer. Previous vision experimental work has shown that detection can depend on the length of presentation. For example, in Bijl et al<sup>24</sup>, they found that for presentation times of 0.13 and 0.25 seconds, the temporal properties of the stimulus play a role, whereas for longer presentation times of 0.50 and 1.0 seconds, detection is mainly determined by the spatial characteristics of the stimulus. In our experimental work, including our most recent experiment (see Experimental Results) we generally found that performance increases with longer presentation times (although our times usually varied between 1 and several seconds). Although presentation time seems to be an important factor to quantify, it is not currently a parameter of the visual models.

In addition to the above three effects, there are other variables that affect the CSF, notably image content and visual task.<sup>5,33</sup> Barrett suggests two classes of tasks: classification and estimation, with classification (including detection) being the usual task in radiology.<sup>34</sup> Burbeck and Pizer<sup>35</sup> suggest classifying the visual tasks as detection, object and structure extraction, and recognition. In modeling the CSF for contrast threshold detection, we are only considering the detection aspect, and effort should be made to study the effects of the higher level functions as well. Although it would be desirable to incorporate all of these effects, we are not currently aware of theoretical or experimental results that would allow the definition and incorporation of the other effects into the visual models.

#### *Can a Representative CSF Be Determined?*

Because the predictions made from visual models are expected to be accurate for the specific presentation conditions described by the parameters and not necessarily for the more general situation encountered in the display of medical image data, we face the problem of choosing a specific visual model, and more importantly, parameters of that model that are representative of a range of clinical conditions. To address this issue, we determined the range of values for each of the parameters of the Barten model,<sup>13</sup> and then calculated the result-

ing CSF values for all possible combinations of these parameter values.<sup>13</sup> Analysis of the resulting CSF values showed a concentration of values in one region. Furthermore, when the CSF values were plotted versus the luminance axis as needed for modeling the luminance to perceived brightness relationship, the result was a family of curves with similar shape with respect to the luminance axis, and slightly different heights with respect to the CSF axis.<sup>13</sup> From this result a single representative CSF curve, matching the most sensitive CSF values from the family of curves was chosen.<sup>13</sup> Although this single choice of parameters can not accurately model all possible clinical applications, it serves two important purposes. First, it describes the shape of the curve that is representative of the family of curves. This is important because perceptual linearizations are not affected by multiplicative changes (height of CSF curve), which simply scale the size of the threshold steps, but only the shape of the curve. Second, it provides us with a good upper bound target for designing our DACLUM curves because it represents the highest sensitivity achievable under expected clinical conditions. Importantly, the specific parameters of the CSF identified in our work closely matched those arrived at independently by Blume et al.<sup>12</sup>

#### *Experimental Results*

Little work has been done in this area. The only previous published work the authors are aware of was the initial work performed at UNC during the development of perceptual linearization.<sup>1,6,7</sup> Although these experiments were performed mainly in the interest of actually measuring and implementing perceptual linearization on display systems in our laboratory, effort was made to make the experimental tasks realistic for medical image presentation conditions.

We recently conducted two additional experiments<sup>13</sup> to address three specific questions: how would an experimentally measured clinical task compare with the predicted Barten CSF curve (as well as with our earlier experiments); what was the interobserver variability; and did the presentation time difference modeling the effect of directed versus undirected search (4 seconds v 1 second) cause a significant difference in CSF values? The CSF curves were

found to be similar in shape to those predicted by the visual models, but with some differences caused by insufficient contrast resolution of the DAC on our display system, and visual effects not accounted for by the model.<sup>13</sup> Measurement of the interobserver variability is important if perceptual linearization is used for display function standardization. For the five observers used in our previous experiments<sup>6</sup> we did not find a significant difference in different observers contrast threshold values. However, with the larger number of subjects in these experiments, we found a statistically significant difference in contrast threshold values for the different observers using a nonparametric randomized block analysis of variance calculation.<sup>13</sup> However, importantly, the shape of the curve was consistent for all observers, meaning that the same perceptual linearization can be used because linearization is insensitive to multiplicative changes as discussed earlier. However, this does imply that for perceptual linearization to work as a method of standardization, it must be based on the more sensitive subjects, ie, the smallest contrast threshold values. This also implies that the steps in luminance values between adjacent DDLs must be less than the observer's smallest contrast threshold values for those luminances. Shortening presentation times generally has the effect of decreasing performance as the task becomes more difficult. In these experiments, detecting the target was more difficult when the local surround was significantly different from that of the global surround. A significant interaction was found between presentation time and local surround luminance, with poorer performance (smaller CSF values) when local surrounds differed more from the global surround and when presentation time was short.<sup>13</sup>

#### *Discussion of CSF Choices*

Although there are differences between the predictions from models and our experimental results, we know that both the Daly and Barten models provide good predictions for empirical vision research results. Also, the models predict similarly shaped curves compared with our results, only shifted upwards reflecting higher (more sensitive) CSF values. Because the model can easily be used to provide predictions for the CSF for specific clinical situations, as well as

generalized or representative CSF curves to compare against, using the models provides an advantage over using results from specific experiments. More importantly, the adoption of a specific mathematical formula for calculating the CSF as part of the perceptual linearization would accrue the significant advantage of allowing it to be standardized and easily implemented on all display systems. Thus, as recently proposed by Blume et al,<sup>12</sup> we also recommend using a model such as Barten's or Daly's for the CSF function as a basis for calculating the resulting linearization display function from the inverse of CSF<sub>0</sub>DACLUM and the adoption of specific parameters to these models (based on either the previous Blume et al<sup>12</sup> or Hemminger et al<sup>13</sup> results). However, we would recommend that such models be extended to incorporate multiple levels of surround, different types of stimuli, and presentation times as parameters. Also, until the differences between the model's predictions and our empirical results are more completely explained, one should bear in mind that the model may overestimate the CSF values compared with experimental results for stimuli more similar to medical image presentations.

#### DAC TO LUMINANCE RELATIONSHIP

The DACLUM curve shows the digital driving levels and how they correspond to luminances generated on the monitor. An example for our Sun Sparc2 workstation (Sun Microsystems Inc, Mountain View, CA) is shown in Fig 3. There are three important attributes: the overall luminance range of the monitor; the number of digital driving levels of the DAC; and the distribution of output voltages of the DAC for these levels.

#### *Luminance Range of Monitors and Film*

The luminance range of a standard workstation monitor is ~0 to 100 cd/m<sup>2</sup>, with some medical image displays ranging up to ~200 cd/m<sup>2</sup>, and monitors capable of 600 cd/m<sup>2</sup> currently under development. For comparison, the maximum luminance of a standard light box is 2,056 cd/m<sup>2</sup>, a mammography lightbox is 3,426 cd/m<sup>2</sup>, and a hot lamp is 17,130 cd/m<sup>2</sup>. These values are from standard clinical equipment in our department, and are not through

film, so the maximum luminance when emitted through the lowest densities on film will be slightly less.

#### *Number of DAC DDLs*

Except for special purpose or prototype DACs, all the DACs made today for gray-scale monitors support 8 bits of contrast information resulting in 256 input levels. Most digital representations of medical image data are 4,096 levels (12 bits), although in some modalities sometimes slightly less than this number of levels contain significant information.<sup>19</sup> Thus, the 4,096 possible input image data values must be represented as one of only 256 output DAC values. How many output levels are clinically necessary depends on the image, the image processing, and the specific clinical task.

Clearly, a trade-off is being made in not presenting all the information in the image data in a single presentation. Thus, from an information transfer standpoint, we are compromising the data if we use anything less than a 12-bit DAC to achieve 4,096 levels equally spaced in perceptual space, although the spacing between DDLs would be less than one contrast threshold step in this case. However, from a clinical standpoint, we may be able to make a satisfactory clinical decision with fewer DDLs than the upper bound of 4,096 levels. To date, the authors are not aware of careful scientific studies that have evaluated the number of levels needed for specific clinical tasks. Some authors have attacked the lower bound of this problem by trying to answer the question, "how many DDLs is too few?" Most of these efforts have addressed the question of when the observer sees texture-contouring artifacts caused by the use of too few DDLs. Sezan et al<sup>36</sup> found that texture artifacts in radiographs could be avoided by using at least 9 to 10 bits with a visual model-based CSF (logarithmic, 1/2 power, and local cone), whereas using a default identity mapping still showed quantization artifacts at 12 bits. However, it is important to remember that these studies have addressed the question of whether a visual artifact is detected, not whether clinical performance changes.

The contrast threshold relationship between  $\Delta L/L$  and  $L$  at each DDL is shown for our Sun monitor with an 8-bit DAC in Fig 3. An identity



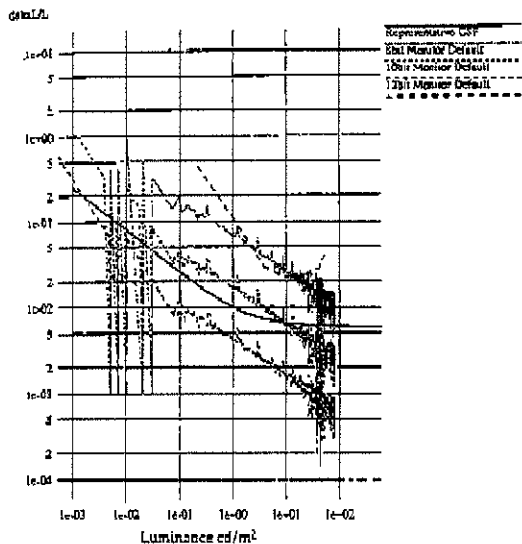


Fig 3. Monitor curves for the Sun Sparc station with 8-bit DAC, extrapolated 10- and 12-bit DAC curves for same system, a representative CSF curve, and recent UNC experimental data are plotted as contrast threshold curves. Vertical axis is  $\Delta L/L$ , and horizontal axis is luminance in candelas per square meter. Both axes are plotted in log scale. The large oscillations in the monitor curves at small luminance values (less than  $10^{-1}$ ) are caused by the bouncing back and forth between almost no change in luminance ( $\log$  of  $\Delta L/L$ , when  $\Delta L/L$  is almost zero) and very small changes in luminance divided by very small average luminance ( $\log$  of  $\Delta L/L$ , when  $\Delta L/L$  is almost equal to 1).

mapping (no perceptual linearization) is used, and the analog contrast and brightness levels on the monitor are set to maximize the luminance range of the monitor while not introducing visual artifacts such as blooming. To represent the monitor characteristics as a contrast threshold function, we calculated the contrast threshold,  $\Delta L/L$ , as

$$\left(\frac{\Delta L}{L}\right) = \frac{(L[DDL_{n+1}] - L[DDL_n])}{\left[\frac{L[DDL_n] + L[DDL_{n+1}]}{2}\right]} \quad (1)$$

where  $n$  ranges across all the DDL values (1 to 256), and  $L[DDL_n]$  is the luminance at DDL  $n$ . Overlaid on Fig 3 is our representative CSF curve and the most recent UNC experimental results, as described in the previous section. By interpolating between the values of an 8-bit DAC, we have estimated what the 10-bit and 12-bit DACs with similar distributions might look like (Fig 3). Ideally, to see as many distinct gray levels as possible, the contrast threshold step sizes for the monitor curves should be equal to or less than the CSF curve. However, in

Fig 3, the 8-bit DAC curve is entirely above the CSF curve, whereas the 10-bit DAC curve is close to the representative CSF curve, and the 12-bit DAC curve is completely below it. In Fig 3, the default monitor curves are based on identity mappings rather than ones that properly match the CSF curve, so we would expect from the work on quantization artifacts<sup>36</sup> that fewer levels will be required when the distribution of luminance levels of the DAC are better matched to a CSF. Figure 3 suggests that around 10 bits for a CSF matched function or 12 bits for a default identity mapping would be required to represent each contrast threshold step. This correlates well with the results of Sezan et al.<sup>36</sup>

#### Distribution of DAC Luminance Levels

As important as the number of levels supported by the DAC is the distribution of output voltage levels produced. DACs generally produce a uniform linear function distribution of voltage levels versus the input DDLs. These voltage levels are input to the monitor's CRT, which directs a beam of electrons onto the phosphorescent material coating the surface of the monitor. The luminous output of the phosphor is not directly proportional to the input driving voltage level, but instead, roughly follows a nonlinear power function. Ideally, the step sizes between the adjacent DDLs should be constant on a perceived brightness scale. This matching of the DACLUM and CSF curves maximizes the transfer of contrast information to the human observer.<sup>37</sup> However, in actual practice, there are often significant variations.

Although the monitor curves are similar in shape to the representative CSF curve in Fig 3 above, there are several important differences. First, the slope on the monitor curves is different from the representative CSF at low luminances ( $10^{-3}$  to  $10^{-1}$ ) and at high luminances (0 to  $5 \times 10^2$ ). This implies that the DACLUM curve does not match the CSF curve well, thus it does not optimize the transfer of contrast information. A second point is that there is significant variation in the contrast threshold step size along the monitor curves. Variations in  $\Delta L/L$  values appear as up and down movements (spikes) along the contrast threshold curve in Fig 3. At low luminances, there are extreme

variations because at some points  $\Delta L/L$  is nearly 0 between adjacent DDLs, whereas in other cases there is a very small luminance change at a very small average luminance values, resulting in  $\Delta L/L$  being nearly equal to 1. From 0 to 10 cd/m<sup>2</sup>, the  $\Delta L/L$  values are mostly the same, with only a few values significantly different (seen as small spikes); however, from 20 cd/m<sup>2</sup> to 80 cd/m<sup>2</sup>, there is greater variation, often with one interval having  $\Delta L/L$  values twice that of their adjacent neighboring intervals. The large variations in the low luminance and higher luminance ranges undermine the proportionality needed to achieve perceptual linearization.

#### *Discussion of Monitor Design Choices*

Several changes in the design of monitors would help improve the number and distribution of distinguishable luminance levels producible on the monitor. First, increasing the luminance range of a monitor would increase its potential PDR. However, it is apparent that, at least for our representative CSF curve and for the latest UNC experiments as shown in Fig 3, almost all of the 256 levels are at least one contrast threshold step apart, and thus increasing the luminance range would not result in a significantly greater PDR. To take advantage of a larger luminance range, more digital driving levels on the DAC are required.

Second, the number of bits on the DAC needs to be increased to increase the number of DDLs available. The exact number required is difficult to determine, and may well depend on the task. However, according to studies measuring quantization artifacts, it seems likely that 10 bits are required if the distribution matches the CSF; more are required if the distribution does not. This correlates well with analysis of the CSF representative curve which suggests that about 10 bits are required for an example workstation monitor. It may well be useful to increase the number of bits all the way to 12 bits to completely represent the input displayable intensities, as well as provide additional levels to help compensate for poor distributions of luminance levels. Although these arguments suggest the need for 10 or more bits in the DAC, work needs to be done to evaluate whether

clinical performance improves as the number of bits is increased above 8.

Third, whereas the DACLUM curve for existing display systems is somewhat similar to CSF curves, changing the distribution of the DAC luminance levels to more accurately match the curvature of the CSF curve and minimizing the fluctuations in the DACLUM curve would improve the proportionality of the final DAC-to-perceived-brightness relationship. The authors are not aware of previous work examining nonstandard DAC distributions based on matching CSF distributions.

#### CALCULATING THE LINEARIZATION FUNCTION

The characteristics of the human observer's visual system, ie, the range of perceived brightness and the contrast sensitivity over that range, are essentially fixed, although these vary somewhat depending on viewing conditions, image content, and visual task. The provider of the display system controls the three factors of the DACLUM relationship: the monitor luminance range, the number of discrete levels of the DAC, and the distribution of the resulting luminance levels. These are usually determined by economic factors. In the ideal situation, the manufacturer of the display system would support a perceptually uniform system through a DAC luminance distribution that matches the human visual CSF. If this is not practical, then the display system can be perceptually linearized by a postlinearization remapping step. Several methods for computing a linearization function from the DACLUM and CSF curves have been suggested. Essentially the task is to compose the DACLUM and CSF functions into a single function and then derive the inverse of this function. Applying this inverse function to the displayable intensities, eg, through a display lookup table (ie, LUT in Fig 1) will result in a proportional relationship between the displayable intensities and the sensation of perceived brightness.

#### *Calculation of Perceptual Linearization Function*

Pizer and Chan, in their initial description of perceptual linearization, gave both an intuitive and a formal analytical approach.<sup>1</sup> In the intuitive approach one calculates

$$L_i = L_{i-1} + (L_{i-1}) \times [1/(CSF(L_{i-1}))], \quad (2)$$

until  $L_i$  reaches or exceeds the luminance of the maximum DDL.  $L_i$  represents the luminance at the  $i$ th DDL value, and  $(1/[CSF(L_i)])$ , the contrast threshold at luminance  $L_i$ . Thus in the intuitive formulation one simply steps 1 contrast threshold in luminance at each step, starting at the minimum luminance, until the maximum luminance is reached. The analytical formulation is given by Pizer and Chan and the specifics of implementing the linearization by Cromartie et al.<sup>8</sup> Also, an approximation that further simplifies the analytical solution is given by Ji et al.<sup>37</sup> In similar work on color scales, other authors have developed methods that super sample in the perceptual scale, and then choose the closest digital driving scale of the monitor.<sup>38,39</sup>

In all of these approaches, the final step takes a calculated desired luminance level and then selects the DDL that most closely matches this luminance. Because there are limited discrete samples in the DDL range (256), and because they are often not distributed in a fashion matching the CSF function, errors may be introduced during each of these matching steps. Overall, this error may negate the intended effect of equal perceptual steps if the values are not chosen with regard to minimizing the perceptual error of the overall process. Examples of this can be seen in Fig 4, which shows a linearization function previously used in our laboratory versus the standard monitor curve and the representative CSF curve. As observed earlier, there is significant variation in the monitor step sizes at very low luminance levels, small spikes in the midrange, and larger variation (up to 200% changes in step sizes) in higher luminance levels. Surprisingly, though, the linearized curve is flawed as well. The variation in step sizes of luminance output levels of the linearization curve is consistently larger than that of the default monitor curve throughout most of the range. This is mainly caused by the distribution of DAC luminance levels not matching the CSF well and suggests that a larger number of DDLs may be necessary to compensate for display systems with suboptimal distributions of DAC luminance levels.

Another important issue is the number of DDLs in the resulting map. As noted in the DACLUM section, the luminance step size

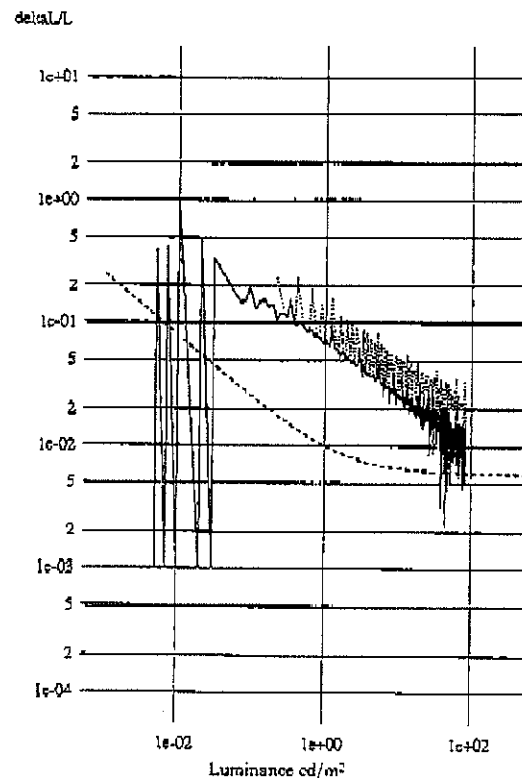


Fig 4. Monitor DACLUM curve for the Sun Sparc2 Station and resulting linearization. Linearization is based on CIELUV (½ power law) algorithm for CSF, measured monitor luminance levels of the display for the DACLUM, and a 128-level linearization remapping calculated using the standard approach of choosing the DDL with the luminance level nearest to that given by the linearization function. (—), 8-bit monitor default; (·), 8-bit monitor linearized; (---), representative curve.

between adjacent DDLs should ideally be less than the contrast threshold at that luminance. For a system with a limited number of DDLs with which to work, increasing the number of DDLs used in the linearization remapping usually means increasing the perceptual error in the mapping. Thus, one trades improved contrast resolution (size of contrast threshold step) for quality of the rendition (perceptual error in linearization). Ideally, one would have a way to quantify this tradeoff.

#### Optimally Calculating the Linearization Remapping

None of the above techniques attempt to minimize the error introduced during this matching of  $(CSF \cdot DACLUM)^{-1}$  desired luminances and actual available discrete luminance levels. General solutions exist for the similar signal

quantization problem of mapping a continuous variable into a discrete one.<sup>40,41</sup> This problem differs in that (1) we have fixed nonuniform spacing of the luminances resulting from the DDLs; (2) we can use any or all of the DDLs in the mapping; (3) we want to minimize the equalness of the steps, not simply the distance from the result sample points to the desired ones; and (4) we would like to maximize the number of DDL levels steps used (to avoid over quantizing the input data) but not at the cost of compromising the accuracy of the linearization. We have proposed the development of an optimal solution to this problem, one that minimizes the perceptual error in the resulting linearization and describes the actual resulting PDR or achievable PDR. In recent work, we have suggested a methodology for calculating the minimum perceptual error in the linearization based on the statistical variance of the contrast threshold of each step of the linearization remapping.<sup>42</sup>

#### *Discussion of Linearization*

To date, the methods developed have been aimed at simply implementing a reasonable  $(CSF \cdot DACLUM)^{-1}$  function. For many monitors, the limitation of 8-bit DACs and suboptimal luminance distributions (not matching the CSF) mean that calculated linearizations may not be very optimal, and in some cases like the one shown in Fig 4, turn out to be worse than not linearizing. Developers of DACs need to better match their distributions to CSF distributions, both to improve the inherent perceptual linearity of their system, and to better allow for after-market perceptual linearization corrections of the display system. Finally, work needs to be done to compare methods for calculating the remapping function matching the linearization function,  $(CSF \cdot DACLUM)^{-1}$ , for a given set of luminance values of a display system.

### CONCLUSIONS

#### *Use of Visual Model for CSF*

Fairly good models for the CSF (Barten and Daly) exist, but they need to be expanded to include multiple levels of surround, more general types of stimuli, and different viewing times. Although there is no single CSF that can

represent all different visual tasks and different clinical viewing conditions, we have derived a single CSF that is representative of most clinical conditions. This CSF function matches the CSF previously recommended by Blume et al, and also, the predictions from it are similar to our most recent experimental results, although the experimental results suggest less sensitivity overall, especially when the stimulus and its local surround are significantly different in luminance from the overall surround to which the eye is adapted. Still, we recommend using the Barten or Daly models for the CSF part of the linearization process. Work needs to be done in comparing visual model-based CSF linearizations with existing default monitor configurations for specific clinical tasks to evaluate whether improvements in clinical performance occur.

#### *Display System Characteristics (DACLUM)*

Current DACs are not sufficient, especially as we go to higher luminance monitors. More driving levels are required and, thus, more bits in the DAC. For an optimal DAC luminance distribution (ie, matching the CSF), ~10 bits should be sufficient to eliminate quantization artifacts and to allow step sizes of around one contrast threshold. The distribution of luminance levels of the DAC should more closely match the CSF of the human observer. Increasing the number of available DDLs can compensate somewhat for suboptimal DACLUM distributions by providing more choices for the linearization step. Work needs to be done evaluating whether the increases in bits in DAC and improved DAC luminance level distributions actually improve clinical performance in observer experiments. Such improvements would have to be traded off against the cost of manufacturing systems with larger numbers of DDLs and better DAC distributions.

#### *Calculation of Linearization Function*

Ideally, the linearization would occur in the choice of the DAC luminance distribution. If further correction of the luminance distribution is required, it can be accomplished by applying the inverse of the  $CSF \cdot DACLUM$  relationship via a post remapping step. Several methods for

this exist, but they do not attempt to minimize the error in matching the desired luminance values of the linearization function versus the actual ones available on the DAC. We propose an optimal solution to the linearization remapping calculation that minimizes perceptual error in the linearization and determines the achievable PDR of the display system. Lineariza-

tions need to be evaluated to test the benefit of different methods of calculating the linearization remapping.

#### ACKNOWLEDGMENT

We thank Steve Pizer for his insights and comments and Scott Daly for his comments on the visual models.

#### REFERENCES

1. Pizer SM, Chan FH: Evaluation of the number of discernible levels produced by a display in DiPaulo R, Kahn E (eds): *Information Processing in Medical Imaging*. Paris, France, Editions INSERM, 1980, pp 561-580
2. Pizer SM: Intensity mappings: Linearization, image-based, user-controlled. *Proc SPIE II* 271:21-27, 1981
3. Pizer SM: Intensity mapping to linearize display devices. *Comput Graphics Image Processing*, 17:262-268, 1981
4. Pizer SM, Johnston RE, Zimmerman JB, et al: Contrast perception with video displays. *Proc SPIE* 318:223-230, 1982
5. Pizer SM: Psychovisual Issues in the Display of Medical Images. *Proc NATO Adv Stud Inst Pictorial Information Systems Med*, 1985.
6. Johnston RE, Zimmerman JB, Rogers DC, et al: Perceptual Standardization. *Proc SPIE III* 536:44-49, 1985
7. Rogers DC, Johnston RE, Pizer SM: Effect of Ambient Light on Electronically Displayed Medical Images As Measured by Luminance Discrimination Thresholds. *J Opt Soc Am* 4:976-983, 1987
8. Cromartie RC, Johnston RE, Pizer SM, et al: Standardization of electronic display devices based on human perception. Chapel Hill, NC, University of North Carolina, 1987. (technical report 88-002)
9. Ji T, Roehrig H, Blume H, et al: Optimizing the display function of display devices. *Proc SPIE VI* 1653:126-139, 1992
10. Briggs SJ: Photometric technique for deriving a "best gamma" for displays. *Proc SPIE* 199:134-145, 1979
11. Blume H, Roehrig H, Browne M, et al: Comparison of the Physical Performance of High Resolution CRT Displays and Films Recorded by Laser Image Printers and Display on Light-Boxes and the Need for a Display Standard. *Proc SPIE* 1232:97-114, 1990
12. Blume H, Daly S, Muka E: Presentation of Medical Images on CRT displays: A Renewed Proposal for a Display standard. *Proc SPIE* 1897:215-231, 1993
13. Herpminger BM, Johnston RE, Rølland JP, et al: Perceptual linearization of video display monitors for medical image presentation. *Proc SPIE* 2164:222-240, 1994
14. Naiman AV, Makous W: Spatial nonlinearities of grayscale CRT pixels. *Proc SPIE III* 1666:41-56, 1992
15. Roehrig H, Blume H, Ji TL, et al: Noise of the CRT display system. *Proc SPIE* 1897:232-245, 1993
16. Barten FGJ: Physical Model for the Contrast Sensitivity of the Human Eye. *Proc SPIE III* 1666:57-72, 1992
17. Daly S: The visible differences predictor: an algorithm for the assessment of image fidelity. *Proc SPIE III* 1666:2-15, 1992
18. Murchison C (ed): *A Handbook of General Experimental Psychology*, Worcester, MA, Clark University, 1924, p 769
19. Hill CR, in Webb S (ed): *The Physics of Medical Imaging*, Philadelphia, PA, IOP 1988, pp 567-583
20. Wyszecki G, Stiles WS: *Color Science: Concepts and Methods, Quantitative Data and Formulae* (ed 2). New York, NY, Wiley, 1982, pp 567-570
21. Blackwell HR: Contrast threshold of the human eye. *Opt Soc Am* 36:624-643, 1946
22. Cornsweet TN: *Visual Perception*. New York, NY, Academic, 1970
23. Haber RN, Hershenson M: *The Psychology of Visual Perception*. New York, NY, Holt Rinehart, 1973
24. Bijl P, Koenderink JJ: Visibility of elliptical gaussian blobs. *Vision Res* 33:243-255, 1993
25. Bijl P, Koenderink JJ, Toet A: Visibility of blobs with a gaussian luminance profile. *Vision Res* 29:447-456, 1989
26. Daly S: Quantitative performance assessment of an algorithm for determination of image fidelity. *SID Digest of Technical Papers* 24:317-320, 1993
27. Rogers JG, Carel WL: Development of design criteria for sensor displays devices. Culver City, CA, Hughes Aircraft Company, 1973. (report HAC reference no. 66619)
28. Barten P: Spatio-temporal model for the contrast sensitivity of the human eye and its temporal aspects. *Proc SPIE IV* 1913:2-14, 1993
29. DeGroot SG, Gebhard JW: Pupil size determined by adapting luminance. *J Opt Soc Am* 42:492-295, 1952
30. Lie I: Visual detection and resolution as a function of adaptation and glare. *Vision Res* 21:1793-1797, 1981
31. Wildt GJ, Waarts RG: The influence of the surround on contrast sensitivity. *Second International Conference on Visual Psychophysics and Medical Imaging*, Brussels, Belgium, July 2-3, 1981
32. Hood DC, Finkelstein MA, in Boff KR, Kaufman L, Thomans JP (eds): *Sensitivity to Light: Handbook of Perception and Human Performance*, vol 1. New York, NY, Wiley, 1986, pp 5-28
33. Kundel HL: Visual clues in the interpretation of medial images. *J Clin Neurophysiol* 7:472-483, 1990
34. Barrett HH: Objective assessment of image quality: Effect of quantum noise and object variability. *J Opt Soc Am*, 7:1266-1278, 1990
35. Burbeck CA, Pizer SM: *Object Representation by Cores*. Chapel Hill, NC, University of North Carolina, 1994 (technical report 94-004)
36. Sezaa MI, Yip KL, Daly SJ: Uniform perceptual

quantization: Applications to digital radiography. *IEEE Transactions on Man, Machine, and Cybernetics SMC-17*: 622-634, 1987

37. Ji T, Rochrig H, Blume H, et al: Optimizing the display function of display devices. *Proc SPIE VI* 1653:126-139, 1992

38. Robertson PK: Color image display: Computational framework based on a uniform color space. Doctoral thesis. Australian National University, Melbourne, Australia, 1985 (CSIRO Technical Report No 27)

39. Levkowitz H, Herman GT: Color scales for image data. *IEEE Computer Graphics and Applications* 12:72-80, 1992

40. Jain AN: *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ, Prentice Hall, 1989, pp 99-123

41. Max J: Quantizing for minimum distortion. *IEEE Trans Inform Theory IT-6*: 7-12, 1960

42. Hemminger BM: Minimum perceptual error calculations for perceptual linearization. Chapel Hill, NC, University of North Carolina, 1994 (technical report TR94-032)