Perceptual Linearization of Video Display Monitors for Medical Image Presentation

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

Department of Radiology,
Department of Computer Science*, and
Department of Biostatistics†
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-7510

ABSTRACT

The perceptual linearization of video display monitors plays a significant role in medical image presentation. First, it allows the maximum transfer of information to the human observer since 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.

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. Based on these analyses we give recommendations for using existing methods when appropriate, and propose new methods or suggest additional work where the current methods fall short. Finally, we summarize the significant issues from all three component areas.

1. INTRODUCTION

Perceptual linearization was first suggested for medical image presentation by Pizer[1], and in follow-up work[2,3,4,5,6,7,8] 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 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.[3,6,9,10] 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[1], and have been further qualified more recently by others[11,12]. 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 digital driving levels are reflected as equal changes in perceived brightness. We have intentionally chosen to not refer to this as optimization, in order to carefully distinguish perceptual linearization from choosing an optimal greyscale processing for an image. Perceptual linearization by itself is not intended to be the optimal grey scale presentation of the original data. The important choice of best greyscale presentation is dependent on the specific image content and visual task, and occurs prior to the linearization. For instance, some image processing technique (window and level, or adaptive histogram equalization) might be performed on the original image resulting in the desired greyscale 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 (CT, nuclear medicine, MRI, PET, Computed Radiology). This has increased the need for standardization as more images are viewed on monitors as well as on lightboxes. Blume[12] provides a list of several advantages of standardization: predictable and reproducible grey scale

0-8194-1459-X/94/$6.00

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, we are seeing an increase in the sub-optimal display of images. This is especially true on workstations when used with their default identity display functions, which are often poor choices because of their overweighting of the low end luminance values. 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 methodology for computing the linearization. Others have advocated this need and proposed display function standards[11,12]. Before a decision is made on such a standard it is necessary to understand all the issues involved. This paper will attempt to carefully describe the issues involved in perceptual linearization. In section 2 of the paper we present a paradigm for describing perceptual linearization as its three basic components: monitor luminance to perceived brightness, digital driving levels to monitor luminance, and how to calculate a resulting linearization given the first two relationships. In sections 3, 4, and 5 we address each of these components in turn. In each of these sections we cover previous work in this field and related fields, as well as new results from our laboratory, and tie these results into a single framework for analysis. Finally, section 6 summarizes what conclusions can be drawn from the earlier sections, and discusses what areas still require investigation.

2. PARADIGM FOR PERCEPTUAL LINEARIZATION

The process of displaying an image on a video display monitor to the human observer is depicted in figure 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 4096 for medical image data. These are the values referred to as *Recorded Intensities* in Figure 1. The 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 *Digital Driving Levels* (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 colormap table. LUTs are often 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 non-linearities, CRT noise, internal scatter, and distortion are well described[11,13,14]. The second area is noise in the source image. Noise in the source image has been discussed in some of the linearization work[12] and also incorporated into some of the recently proposed visual models[15,16] discussed below in section 3.1.

From the standpoint of linearization there are two important relationships in this process, that of the DDLs of the computer's DAC versus the luminance of the monitor, and that of the monitor luminance versus 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*, abbreviated as CSF in this paper. CSFs measure the change in luminance (DL) 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 $\Delta L/L$, while CSFs are defined as its reciprocal, i.e. $L/\Delta L$. CSFs in this paper will refer to $L/\Delta L$ verus L, while in vision literature, CSFs usually refer to $L/\Delta L$ versus spatial frequency of the target.

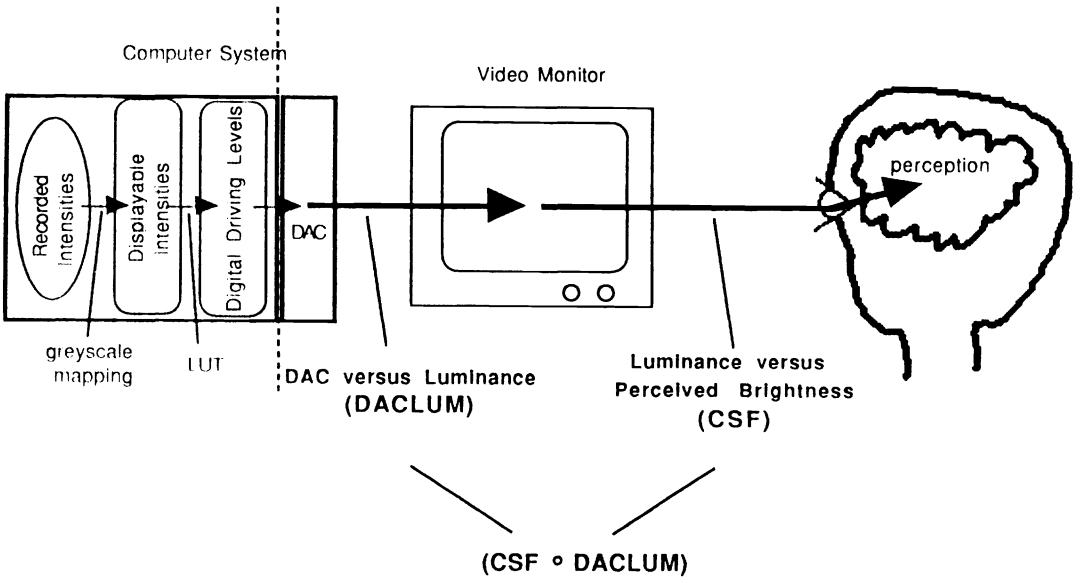

Figure 1. Diagram of components of Perceptual Linearization

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 grey 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, that is the inverse of CSF°DACLUM . We will first consider the CSF curve, which can be considered fixed due to 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.

3. LUMINANCE TO PERCEIVED BRIGHTNESS RELATIONSHIP

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 grey levels over that range which 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, that is, 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 digital driving levels. 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 grey 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.

3.1 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 ΔL/L versus L; where DL 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 (CT) curves. An example study depicting this relationship is shown in figure 2[17]. The minimum luminance difference required to see a contrast threshold is referred to as a Just Noticeable Difference (JND).

By the early twentieth century it was recognized that the overall luminance range of the human observer[18] is approximately 0 to 10^8 cd/m^2 and contrast threshold over this range could be broken into three areas: scotopic, mesotopic, and photopic. In the low luminance scotopic region (below approximately $5x10^{-3}$ cd/m^2),the luminance 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 cd/m^2 to10 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[19]. The curve in this region is described by the Weber-Fechner Law, which states that for a luminance L, and a change in luminance DL, the ratio ΔL/L is constant. This most sensitive (smallest) measured value of this constant is approximately 0.01[19]. Finally, the mesotopic region describes the middle area between these two regions where there is a combining of the effect of the scotopic and photopic regions.

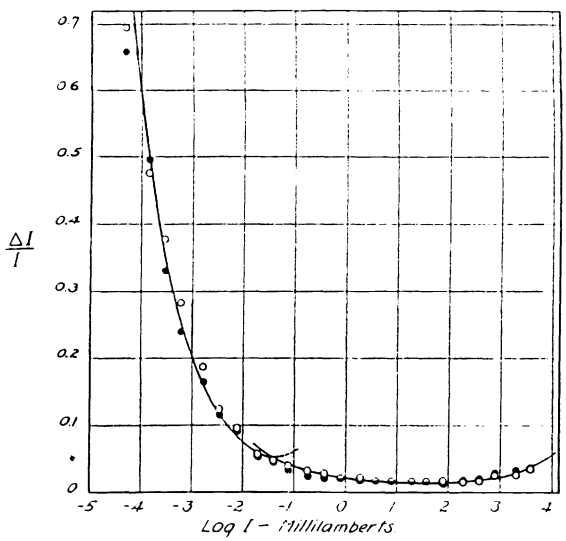

Figure 2. Representative experimental results showing contrast threshold relationship ΔL/L versus L (from Murchinson[17]).

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 lead to the development of CSFs which define the contrast sensitivity of an observer, defined by L/ΔL, versus the spatial frequency of the grating (i.e. cycles/degree)[20,21]. More recently some researchers have extended the stimulus targets to include different types of objects, for instance gaussian blobs[22,23].

Many experimental or theoretical models have been proposed, including models based on empirical evidence and physiological measurements on animals. While several of these models, including both empirically oriented models such as log relationships (Weber-Fechner, etc.), power law (1/3 law, etc.), 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.

Recently, however, several authors[15,16] have defined models based on mathematical descriptions of the components of the human visual system, which 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[15,24,40]. 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. And, perhaps, in the more general case, we could predict a resulting area of possible CSF values given ranges of input parameters to the CSF that match the range of possible clinical conditions. Blume, et al[12] 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[12].

We have studied the three models[15,16,25] and implemented the Barten model based on the descriptions in Barten[15,26] and Blume[12]. We then carried out preliminary investigations to see: (1) whether the different models took into account factors we found to be significant in our previous experimental work; (2) whether we could arrive at a general model for clinical presentations by specifying *ranges* of parameter values in the Barten model; (3) if there existed a single set of parameter values to the Barten model that would be suitable for the purpose of specifying a representative CSF. Section 3.3 examines how accurately the Barten model predicts the results of our latest experiments (described in section 3.2). In separate work we are investigating whether specific clinical viewing tasks can be accurately modeled, including the detection of mammographic features on digital mammograms.

3.1.1 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 due to **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[7] and in general vision research[27,28,29] 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 luminance (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, for instance gaussian blobs based on Bijl's[22,23] 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 views the presentation. Radiologists generally scan images in one of two modes: directed, for instance, to rule out a mass in the upper left lobe; versus, undirected, for instance, 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[23] they found that for presentation times of 0.13 and 0.25 seconds the temporal properties of the stimulus play a role, while 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 (section 3.2) we generally found that performance increases with longer presentation times (although our times usually varied between one and several seconds). While 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[5,30]. Barrett suggests two classes of tasks: classification and estimation, with classification (including detection) being the usual task in radiology[41]. Burbeck and Pizer suggest classifying the visual tasks as detection, object and structure extraction, and recognition[31]. 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. While 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.

3.1.2 General CSF model

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 general situation of medical image display, we must be careful about drawing conclusions for any individual result. One way of attempting to provide a more generalized medical image presentation CSF would be to consider ranges of values for each of the significant parameters. By calculating the possible CSF values for all the parameter combinations likely to occur in clinical situations we can determine a CSF multidimensional volume. Projecting this result on the contrast sensitivity versus luminance plane shows the distribution of possible CSF values for this generalized situation. Using Barten's model, as described in equation 20 by Barten[15], we have calculated such a distribution and plotted the result in the scatter graph in figure 3. The resulting points can be thought of as a family of curves, where each individual curve represents a fixed combination of all parameters except luminance, with the resulting luminance versus CSF relationship being plotted. The parameters considered significant for the generalized model and their chosen ranges are shown below.

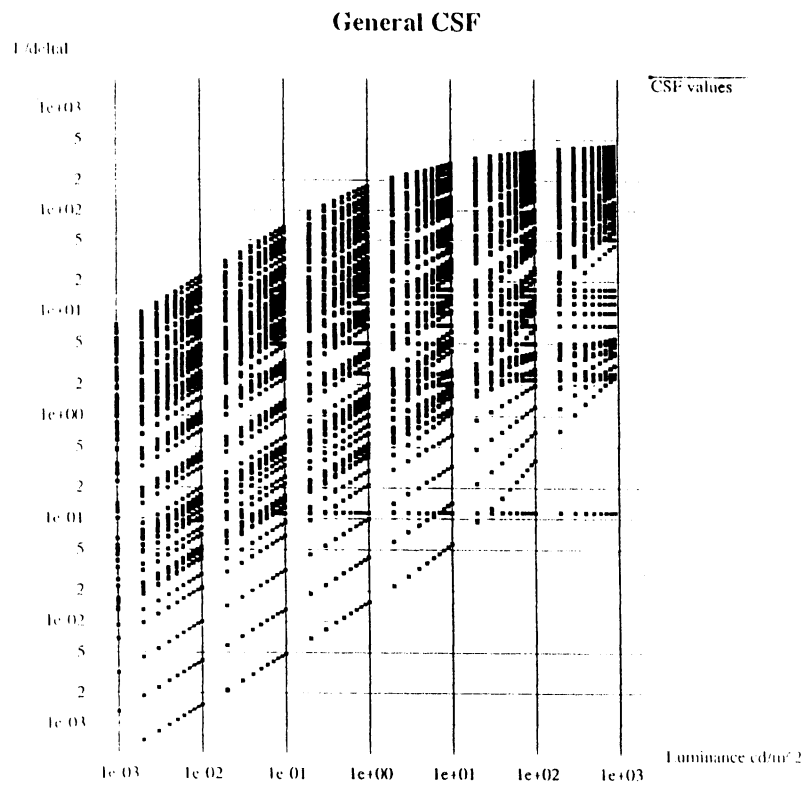

Figure 3. Shows *General CSF*, as scatter plot of contrast sensitivity versus luminance values for multidimensional calculation across ranges of parameter values for each of target size, cycles/degree of target, and luminance. Vertical axis is L/ΔL in log scale. Horizontal axis is luminance (cd/m^2) in log scale.

Luminance ranges from 0.001 to 1000 cd/m^2. At 0.001 cd/m^2 and below there are few monitor driving levels and few JND steps (assuming an average medical image for mean luminance of surround).

Cycles/Degree of Target Stimulus ranges from 0.10 cycles/degree to 20 cycles per degree. 20 cycles/degree was chosen because this accounts for most of the range of human vision (which tails off at 60 cycles per degree when all other parameters are set for maximum sensitivity) and because most medical targets would not be higher frequency than this. The low end range was not significantly constrained, i.e. 0.10 cycles/deg was chosen because this is the normal minimum value calculated or tested in vision models.

Target size ranges from 0.0235 degrees to 3.757 degrees of visual angle. 0.235 degrees corresponds to a 0.25mm target (width of two pixels on current high resolution greyscale monitors) at a nominal viewing distance of 610 mm. A 40mm target, or 3.757 degrees at 610 mm viewing distance, was chosen as the maximum size.

Figure 3 demonstrates the wide range of CSF values, $L/\Delta L$, that exist. For instance at 10 cd/m^2 the range is approximately four orders of magnitude (10^{-2} to 10^{2}). Because of the wide range it is difficult to model all different clinical situations with a single choice of parameters for Barten's model. However, the curves (except for a small portion of the curves at the bottom of the figure) have a uniform slope from 10^{-3} to about 10^{1}, and then have a different (smaller) uniform slope from 10^{1} to 10^{3}. Thus we might choose a *representative* curve from this family of similar curves.

In upcoming section 3.2.4 we describe the reasoning for choosing certain visual tasks for the two recent (Hemminger et al.) experiments. Figure 4 shows the resulting representative curve, *Barten-Hemminger*, which is computed from Barten's model with the parameters matched as similarly as possible to the Hemminger experiments: target size of 1.88 degrees, and angular spatial frequency of 2.0 cycles/degree. Three parameters that vary per experimental paradigm were set according to recommendations in Barten[15], i.e. constant K of 3.3, quantum efficiency of eye of 0.025, and photon conversion factor of 342*3600 (the 342 value more closely matches the phosphor light of monitors versus the standard 357 value for natural light). Also the value of the diameter of the pupil was calculated from the formula given in DeGroot[32].

While 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, shown in figure 3, especially those in the top portion of the graph. Perceptual linearizations are not affected by multiplicative changes, which simply scale the size of the threshold steps. Thus, figure 3 shows that the differences between the curves at the top of the graph is simply multiplicative (which appear as shifts on the logarithmic scale of the vertical axis) since they have the same shape, and are offset only in height from one another. 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, (i.e. there is only a small region between this curve and the top of the multidimensional CSF curve in figure 3, and this region is occupied by large width targets and high frequency gratings that do not generally occur in clinical conditions).

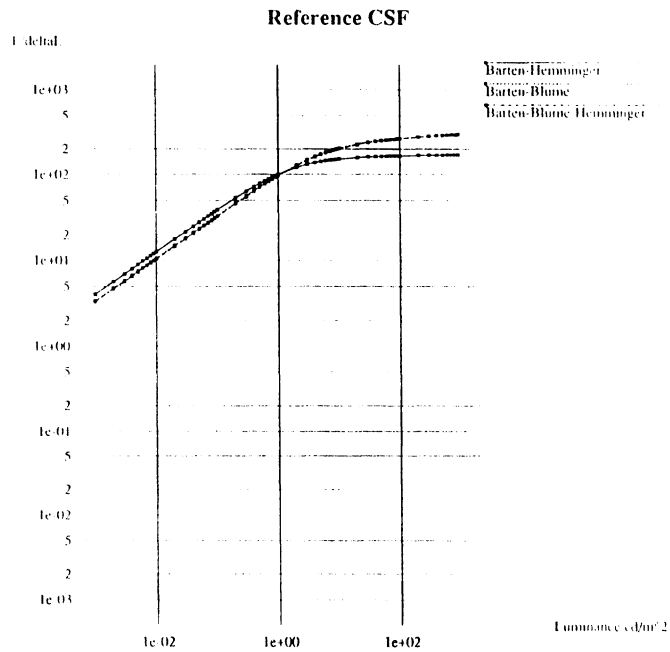

Figure 4. CSF reference curves. Shows Barten model predictions for experimental paradigm matching Hemminger experiments (*Barten-Hemminger*), as well as reference curve matching Blume paper (*Barten-Blume*), and the same curve with modification that p equal 342 (*Barten-Blume-Hemminger*). The *Barten-Blume* and *Barten-Blume-Hemminger* are nearly identical and appear as one curve in this graph.

The representative curve *Barten-Hemminger* used in this paper is very similar to the curve recommended by Blume[12] for the purpose of defining a display standard. The only differences are choice of the square targets being 2 degrees versus 1.88 degrees in our experiment; the choice of 4 cycles/degree versus the 2 cycles per degree in this experiment; and the choice of constant p (called h by Blume[12]) relating photon flux to illuminance to be 357 versus 342 in our calculations. For the purpose of choosing a reference standard, Blume's choice of angular frequency (4 cycles/degrees) may be more appropriate as this would better include the higher spatial frequencies. Also since the CSF is most sensitive around 4-5 cycles/degree, this would better represent the upper bound. For the choice of the constant p we would recommend the value of 342 instead of 357 used in Blume. Plotting both the Blume original recommendation (*Barten-Blume* in figure 4) and our suggested modification of setting p equal to 342 (*Barten-Blume-Hemminger*) shows the curves to be essentially identical. Thus either choice would function equally well as a model of the CSF.

3.2 Medical Image Task Experimental Results

Little work has been done in this area. The only previous published work the authors are aware of was the initial work carried out at UNC during the development of perceptual linearization[1,6,7]. While 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. To simplify referencing the experiments we have labeled them by the name of the first author on the paper describing the experiments.

3.2.1 Pizer et al. experiments

In the first experiments[1], each trial consisted of displaying two separate squares on a black background for 5 seconds. The bottom square was a reference luminance, while the top square varied between the reference luminance and the reference luminance plus a slight increase. The targets were intentionally separated because "real [clinical] discernments must be made where boundaries between target regions are very diffuse and because the Mach effect eases discernment for observers, the target regions used in the measurement should not have a sharp boundary with each other"[1]. An initial experiment was carried out on an analog TV display, and a second experiment used a Tektronix monitor and Ikonas framebuffer. The results of the second experiment have been converted to the standard CSF measure of $L/\Delta L$ versus L and plotted in log scale in figure 5.

3.2.2 Johnston et al. experiments

The first of the second series of experiments[6] used the same presentation technique. Two significant changes were added. First, the room lighting was maintained in the 3-9 lux range throughout the trial. Second, the surround was changed from black to a variable amount, which when combined with the individual trial stimulus squares always resulted in total scene luminance being equal to the average luminance of a CT scan when shown on a video monitor, which they calculated to be 8.6 cd/m^2 (2.5 footlamberts). In the follow-up experiment[6] the trial was modified to support 2AFC analysis which was found to be significantly less time consuming than the previous ROC analysis of the first experiment. Also, measurements at higher luminance values were recorded. In this portion two pairs of squares were presented, the first for 2 seconds, followed by 0.5 seconds of background, and then the second pair for 2 seconds. The task in this case was to determine whether the first or second pair of squares were not matched, i.e. of identical luminance. Results from this experiment have also been converted to the standard CSF scale and are shown in figure 5. Also, this experiment estimated the intraobserver variation to be less than 20% of the JND step sizes of that experiment.

3.2.3 Rogers et al. experiments

The third experiment[7] was an extension of the Johnston experiment above. All the conditions were the same, except that ambient light was more carefully controlled, measurements were made to higher luminance levels than previously, and the parameter of ambient light was added as a variable. An important result of this work was that the CSF did not vary in shape while the ambient light remained in the range of 4 lux to 40 lux. Above 40 lux, however, there were significant interactions, with the CSF curve shifting downwards as contrast thresholds increased. Results from the 4 lux experiment are shown in figure 5.

3.2.4 Hemminger et al. experiments

Several factors motivated us to carry out another experiment which we have just completed. The first was to make the detection task more realistic. One purpose of the CSF curve is to measure the most sensitive JND steps, thus we would like a task that is as sensitive as possible. This must be balanced against desiring the task to be clinically realistic.

First, we chose to have the targets be circles, which were inside of squares that acted as local surrounds, which were inside of a larger overall surround (see Figure 6). The circles were blurred into the squares to avoid mach band effects and to make them more clinically realistic (for instance the appearance of a typical lesion).

Second, this was a spatially separate 5AFC trial versus the temporally separate trials of the latter experiments above. This was because the task of detecting a blob within a local surround was more clinically realistic than comparing two pairs of temporally separate squares, which often involved a comparison from memory rather than a detection task.

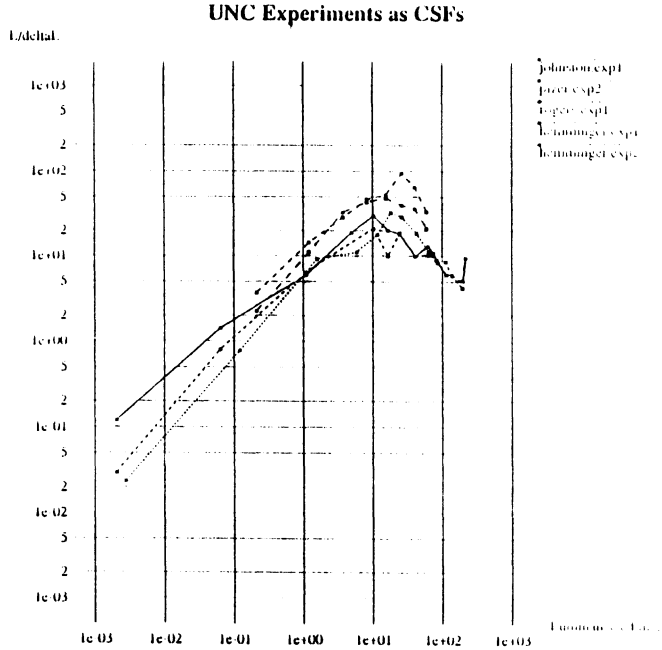

Figure 5. Shows results of most of the UNC experiments plotted as Contrast Sensitivity functions with Luminance on the horizontal axis and L/ΔL on the vertical axis. Both axes are plotted in log scale.

Third, we choose the overall background to remain at a constant luminance that corresponded more closely to the mean luminance of X-rays and mammography images. This resulted in a higher overall mean luminance than the previous experiments, which were based on chest CT images.

Fourth, we wanted to add the variable of presentation time to model directed and undirected searches as discussed in section 3.1.1. A 1 second presentation time was chosen to model undirected search, and a 4 second presentation time was chosen to model directed search.

The fifth reason was to try to help explain differences between the results of the earlier experiments (especially PDR range estimates) and the predictions from various visual models (which all have much higher PDR estimates[12]). Because the task in this experiment has the stimulus adjacent to the surround, the removal of the temporal comparison, and closer similarity to other vision research experiments, we expect observers will have a higher detection rate for this task versus the previous tasks, and probably be more similar to the results predicted by visual models.

Stimulus

Each trial consisted of presenting to the observer 4 squares vertically arrayed. The squares (local surround) were centered on the screen. The rest of the screen was set to a constant luminance of 11.58 cd/m^2. This was determined by measuring the average luminance of several X-rays and mammograms and taking the average photometer luminance reading. The digital driving level that produced the nearest matching photometer reading at the observer viewpoint was then selected (DDL 122). Choice of the size of the squares and location on the screen was determined by measurements of the screen which showed significant variations in luminance outside of the center area and pilot experiments which showed observer biases for other configurations (2 over 2 array and horizontal array). One or none of the squares had a circle (stimulus) in the center of the square. The circles were first combined into the square and then gaussian blurred with a frequency standard deviation of 0.5 to remove the sharp edge between the circle and the square. The circle stimulus was 30 pixels (10mm, 0.94 degrees of visual angle) in diameter. The square local surround was 60 pixels (20mm, 1.88 degrees of visual angle) in diameter.

Observers were seated at a distance of 610mm from the screen. While the observers were not fixed to this specific distance at all times, and were able to physically range between 305mm and 762mm, they generally remained at approximately the 610mm distance during viewing.

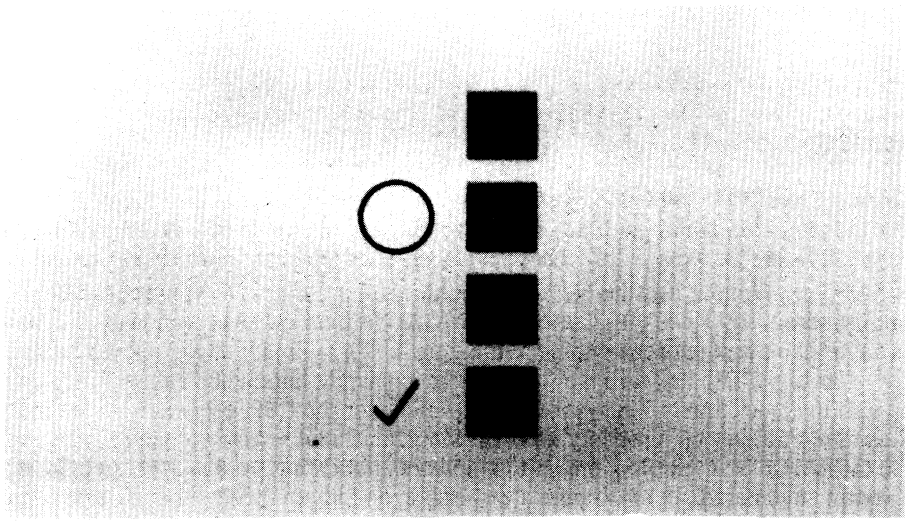

Figure 6. Shows an example screen from the Hemminger experiment. Four targets are presented in a vertical array. The next to top target has a circular gaussian stimulus in center. The screen shows a feedback training example, with the check mark indicating the observers (incorrect) answer and the circle mark indicating the correct response.

Presentation

The local surround and stimulus were presented for 4 seconds (first experiment) or 1 second (second experiment). The observer responded by pressing one of five keys on the keyboard corresponding to which of the four squares on the screen contained the stimulus, or to the condition that none of the squares contained the stimulus. Immediately upon responding the user was shown a check mark to the side of their selection as a feedback mechanism. Additionally, during training sessions, the observers were shown a circular mark to the side of the correct answer. In between presentations the screen would return to showing only the background luminance over the entire monitor for 0.5 seconds. Experiments were run in a special laboratory room controlled for light, noise, and other distractions. Room ambient light was set to 4.1 lux during all the experiments. Maximum and minimum driving levels of the monitor were checked at the start of each run, and the monitor adjusted as necessary to bring it in alignment with the standard readings for the study by adjusting the contrast and brightness knobs as described previously[8].

Trials

Eight different local surround values were used. These were chosen to be equally spaced on a log scale (assuming Weber-Fechner to be a reasonable rough approximation of CSF response). Five stimulus values were used for each local surround value. The stimulus values were also chosen to be equally spaced in perceptual space. Each stimulus value was depicted at each of the five locations (4 squares plus none of the squares condition). Since the "none of the above" condition matched the stimulus presentation condition where the stimulus was equal to the local surround, there were a total of 8 x 5 x 4 distinct trials. These distinct 160 trials were called a run. Each observer viewed 5 runs, for a total of 800 trials in all. The order of trials in each run was randomized before presentation to the observers. Training consisted of two sets of data. First the observers were presented with 60 trials of feedback training, during which they were shown the correct answer in addition to their response. Second, they were shown 30 trials without the correct answer. The second set, called realistic trials, was to simulate the real trials as best possible. The purpose of the training was to allow the subjects to understand the task, improve their performance to asymptote, and to reduce input errors. A special response key (undo) was added that allowed the observer to repeat the last trial in the case of pressing the wrong answer key. If the undo key was used more than several times during one trial, the observer would have to repeat the trial (to discourage their using this as a method to view a trial longer).

Observers

There were 12 subjects in the first experiment, 10 in the second. Age of subjects ranged from early 20s to 60s, with most being graduate students in the younger age range. Each subject completed the run in two separate sessions of approximately 30 minutes each. To avoid fatigue each session was separated by a minimum of 15 minutes (in most cases by approximately one day) and each individual trial by at least a 3 minute break. The subjects were dark adapted to the room for at least 10 minutes before beginning real trials. To encourage optimal performance over all the tasks, the subjects received performance based pay.

Results and analysis

Three specific questions were asked in the data analysis of this experiment: what are the CSF curves from the two experiments; what was the interobserver variability; and did the presentation time difference between experiment 1 (4 seconds) and experiment 2 (1 second) cause a significant difference in CSF values?

The **CSF values** for experiments 1 and 2 are shown in figure 5. The local surround luminance is referred to as L, and the luminance difference between the target luminance and L as DL. A probit analysis was performed for each subject and values of L, with the log of the ratio L to DL as the predictor. Each probit analysis determined mean (mu) and spread (sigma) parameter estimates. The mean in this 5AFC experiment was the 60% point on the probit curve. DL values (in cd/m^2) were calculated using the mu estimates. Contrast sensitivity values for each local surround luminance level were subsequently calculated as the log of ($L/\Delta L$). These values have been averaged across subject in Figure 5. The primary drawback we encountered was that JND step sizes were often less than one DDL, and thus we were unable to sample luminance values finely enough to accurately measure the actual JND step size in some cases. Thus, some of the values for the CSF, especially in the middle range where the target is near the global surround luminance, probably underestimate the actual sensitivity values. In general the CSF curves are similar in shape to the earlier UNC results, but shifted upwards indicating smaller JND step sizes. We believe this is due to the differences in experimental paradigm as discussed earlier (i.e. no temporal comparisons, and stimulus adjacent to surround). An important point is that the sensitivity falls off the farther the surround luminance is from the target luminance. While this is documented in vision research[28] and our recent experiments, the Barten and Daly models do not currently incorporate it, and existing perceptual linearizations do not take it into account.

Measurement of the **inter-observer variability** is important if perceptual linearization is used for display function standardization. Inter-subject variance was estimated by calculating the sample variance of the log CSF values across subject. This calculation was performed separately for each local surround luminance level. For the 5 observers used in the previous Johnston experiments 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 non-parametric randomized block ANOVA calculation. Importantly, however, 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. This does imply, however, that for perceptual linearization to work as a method of standardization, it must be based on the more sensitive subjects, i.e. the smallest contrast threshold values. This also implies that the steps in luminance values between adjacent DDLs must be less than the observers smallest contrast threshold values for those luminances. Another alternative would be to customize separate perceptual linearizations for each individual observer. However, this is impractical because of requiring measurement of the CSF for each user of the display system.

Shortening **presentation times** generally has the effect of decreasing performance as the task becomes more difficult. In this experiment, detecting the target was more difficult when the local surround was significantly different from that of the global surround. This is seen in the increased curvature (downwards turning) in the experiment 2 data relative to the experiment 1 data at the low and high luminances values (figure 5). To test the significance of the effect of presentation time in our experiments we used a Geisser-Greenhouse corrected univariate approach to repeated measures analysis[33]. The Geisser-Greenhouse test was used to correct for a possible lack of sphericity in the variance-covariance matrix. In order to consider both location and variability differences between probit analysis parameter estimates, the dependent variables for this test were taken as mu + 1*sigma. The test of a presentation time interacting with the local surround luminance level was significant at the 0.01 level ($F=52.58$, 4 numerator and 20 denominator df, $p=.0006$). This interaction corresponds to larger differences in CSF values for the more difficult tasks. The effect can be seen in Figure 5 where CSF values decrease (i.e. the contrast threshold increases) as the difference in local surround and global surround gets larger.

3.3 Discussion

Comparison of UNC experiments with Barten's model

Figure 5 shows that UNC experiments generally depict similar CSF curves. The datapoints of the earlier experiments are somewhat noisy due to fewer observers and smaller number of overall observations. The paradigm of the most recent experiment (Hemminger) results in more sensitive measurements of CSF values, resulting in the CSF curve being shifted upwards slightly from the other experiments. These values are likely a little higher, but we were unable to measure them

because of being limited to coarsely sampling the luminance axis due to the distribution and the small number of digital driving levels available (256). The Barten-Hemminger curve is also similar in curvature, but higher still, predicting significantly smaller contrast thresholds ($\Delta L/L$) across the range of luminance values. Some of this difference is attributable to the paradigm of the Barten model (single surround near stimulus luminance, stimulus grating target) measuring more sensitive contrast thresholds than our experiment (global surround differing from local surround, gaussian circle target). Notably, the experimental curve falls off at the low and high luminance ends when the surround (global surround) is significantly different from the local surround and target luminances. This is expected since the most sensitive responses will be when the entire surround is as close as possible to the target. Although the underestimation of thresholds in our experiment due to the DAC and the differences in paradigms likely explain most of the differences between the curves, more work needs to be done to accurately explain the discrepancies.

While 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 to our results, only shifted upwards reflecting higher CSF values. Because the model can easily be used to provide predictions for the CSF for specific clinical situations, as well as provide generalized or reference CSF curves to compare against, using the models has an advantage over using results from specific experiments. Thus, as recently suggested by Blume[12], 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°DACLUM. Additionally, we suggest the adoption of specific parameters to these models to describe a reference CSF curve (either the *Barten-Blume* or the equivalent *Barten-Blume-Hemminger* described earlier). We would, however, 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 to experimental results for stimuli more similar to medical image presentations.

4. 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 is shown in figure 7. 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.

4.1 Luminance Range of Monitors and Film

The luminance range of standard workstation monitor is approximately 0 to 100 cd/m^2, with some medical image displays ranging up to approximately 200 cd/m^2, and monitors capable of 600 cd/m^2 currently under development. For comparison, the maximum luminance of a standard lightbox is 2056 cd/m^2, a mammography lightbox is 3426 cd/m^2, and a hot lamp is 17130 cd/m^2. These values are from standard clinical equipment in our department, and are not through film, so the maximum luminance when emitted through low densities on film will be slightly less.

4.2 Number of DDLs

Except for special purpose or prototype DACs, all the DACs made today for greyscale monitors support 2^8, or 256 input levels. Most digital representations of medical image data are 4096 levels (12 bits), although in some modalities sometimes slightly less than this number of levels contain significant information[18]. Thus, the 4096 possible input image data values must be represented as one of only 256 output DAC values. To answer the question of how many output levels are clinically necessary depends on the image, the image processing, and the specific clinical task.

Clearly, a tradeoff 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 utilize anything less than a 12 bit DAC to achieve 4096 levels equally spaced in perceptual space, although the spacing between DDLs would be less than one JND 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 4096 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 does the observer see texture contouring artifacts caused by the use of too few DDLs. A comparison of the number of bits required to quantize a radiograph found that 9-10 bits were required to avoid texture artifacts when using logarithmic, 1/3 power and local cone models, while using a default identity mapping still showed quantization artifacts at 12 bits[34]. It is important to remember,

however, that these studies have addressed the question of whether a visual artifact is detected, not whether clinical performance decreases.

The contrast threshold relationship between ΔL/L and L at each DDL is shown for our Sun monitor with an 8 bit DAC in figure 7. To represent the monitor characteristics as a contrast threshold function, DL is calculated as the difference in luminance between adjacent DDLs, and L is the average luminance of those two adjacent DDLs. Overlaid on figure 7 is our reference Barten-Hemminger curve generated from section 3, and the results of the second Hemminger experiment. 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 (figure 7). Immediately obvious is that the luminance step sizes on the monitor when driven by an 8bit DAC are significantly larger than JND step sizes estimated from Barten's model. At a luminance value of 10 cd/m^2, the monitor contrast threshold value (ΔL/L) is approximately 25%, compared with 5% on the reference Barten curve. Even comparing with the 2nd Hemminger experiment at levels where the stimulus is similar to the surround, the contrast threshold value of the experiment is less than the 8 bit DAC. Ideally, the contrast threshold step sizes for the monitor curves should be equal to or less than either of these two values. In figure 7, the 10 bit DAC curve is much closer to the reference Barten curve, and the 12 bit DAC curve is completely below it. Since the default monitor curves are closer to *identity* mappings than ones that properly match the CSF curve, we expect from the work on quantization artifacts[34] that fewer levels will be required when the distribution of luminance levels of the DAC are better matched to a CSF. Figure 7 suggests that around 10 bits for a CSF matched function or 12 bits for a default linear mapping would be required to represent each JND step. This correlates well with the Sezan results[34].

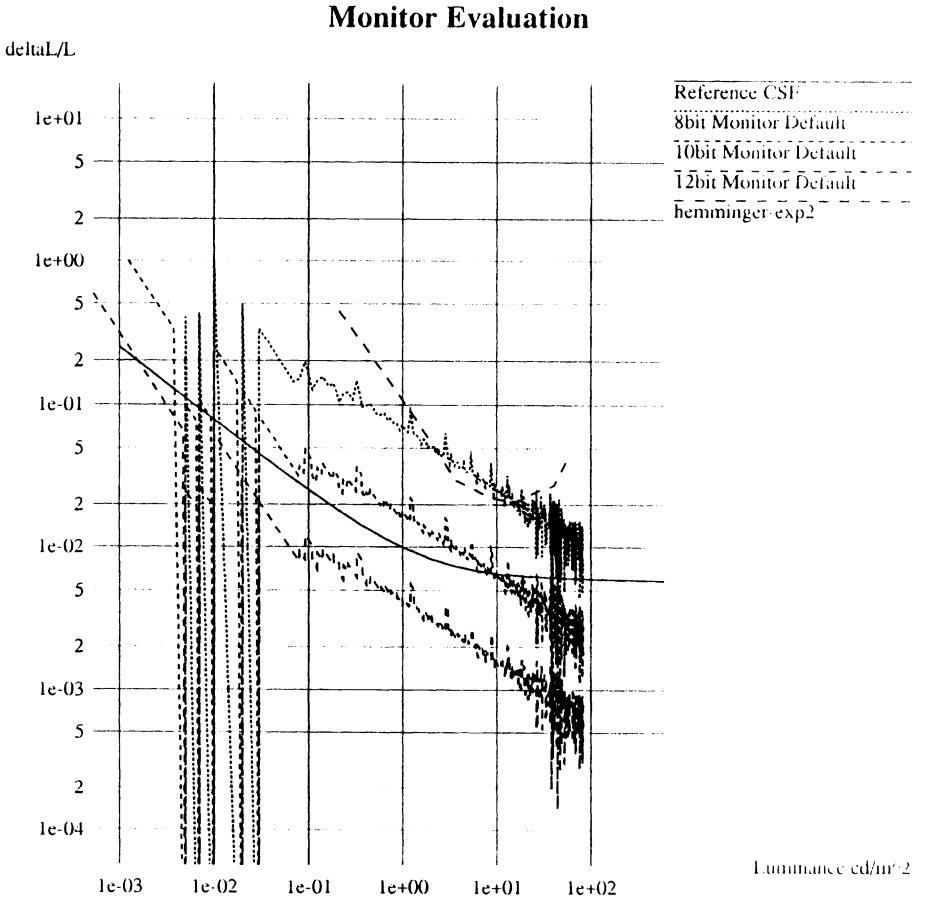

Figure 7. Monitor curves for Sun Sparc station with 8 bit DAC, extrapolated 10 and 12 bit DAC curves for same system, Barten-Hemminger reference curve, and Hemminger experiment 2 curve are plotted as contrast threshold curves. Vertical axis is ΔL/L, and horizontal axis is luminance in cd/m^2. Both axes are plotted in log scale.

4.3 Distribution

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 *cathode ray tube* (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 non-linear 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[35]. In actual practice, however, there are often significant variations.

While the monitor curves are similar in shape to the reference Barten curve in figure 7 above, there are several important differences. First, the slope on the monitor curves are different from the reference Barten at low luminances (10^{-3} to 10^{-1}) and at high luminances ($5*10^{0}$ to 10^{2}). This implies that the DACLUM curve does not match the CSF curve well, thus not optimizing the transfer of contrast information. A second point is that there is significant variation in the JND step size along the monitor curves. At low luminances, there are extreme variations in luminance step size between DDLs on the monitor curves due to minuscule changes in luminance in adjacent DDLs. From 0 to 10 cd/m^2 there are few values significantly different (seen as spikes); however, from 20 cd/m^2 to 80 cd/m^2 there is greater variation, often with one interval having $\Delta L/L$ values twice that of their adjacent neighboring intervals. These large variations undermine the proportionality needed to achieve perceptual linearization.

4.4 Summary

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 reference Barten-Hemminger curve, and from the latest UNC experiments as shown in figure 7, that almost all of the 256 levels are at least one JND 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 digital driving levels available. The exact number required is difficult to determine, and may well depend on the task. However, from studies measuring quantization artifacts, it seems likely that 10 bits are required if the distribution matches the CSF; more if the distribution does not. This correlates well with analysis of the CSF reference 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. While 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, while 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 non-standard DAC distributions based on matching CSF distributions.

5. CALCULATING THE LINEARIZATION FUNCTION

The characteristics of the human observer's visual system, that is 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. The final component, and one over which the end user has control, is the ability to remap the image grey values to revised ones based on a CSF°DACLUM curve. Several methods for computing a linearization 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 will result in a proportional relationship between the Displayable Intensities and the sensation of perceived brightness. Pizer, in his initial description of perceptual linearization gave both an intuitive and a formal analytical approach[1]. In the intuitive approach one calculates

$$L_i = L_{i-1} + (L_{i-1}) * (1 / (CSF(L_{i-1}))$$

until L_i reaches or exceeds the luminance of the maximum DDL. L_i represents the luminance at the ith DDL value, and (1 / (CSF(L_i)) the contrast threshold at luminance L_i. Thus in the intuitive formulation one simply steps 1 JND in luminance at each step, starting at the minimum luminance, until the maximum luminance is reached. The analytical formulation is given by Pizer[1] and the specifics of implementing the linearization by Cromartie[8]. Also an approximation that further simplifies the analytical solution is given by Ji[35]. In work on color scales other authors have developed methods that supersample in the perceptual scale, and then choose the closest digital driving scale of the monitor[36,37].

In all of these approaches, the final step takes a calculated desired luminance level and then selects the DDL that best matches this luminance. Because there are limited discrete samples in the DDL range (256), and since they are often not distributed in a fashion matching the CSF function, errors may be introduced during this matching step. Examples of this can be seen in Figure 8 which shows a linearization function previously used in our laboratory (based on a 1/3 power law model for the CSF, measurements of all the monitor luminance levels on our Sun Sparc2 for the DACLUM, and a 128 level linearization function calculated using the intuitive approach) versus the standard monitor curve and the reference Barten 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. While it avoids much of the very low luminance levels entirely, for the remaining range up to $5x10^1$ cd/m^2 the variation in step sizes is larger than that of the default monitor (up to 300% change in step sizes). This is mainly due to the distribution of DAC luminance levels not matching the CSF well, and suggests that a larger number of DDLs levels may be necessary to compensate for display systems with suboptimal distributions of DAC luminance levels.

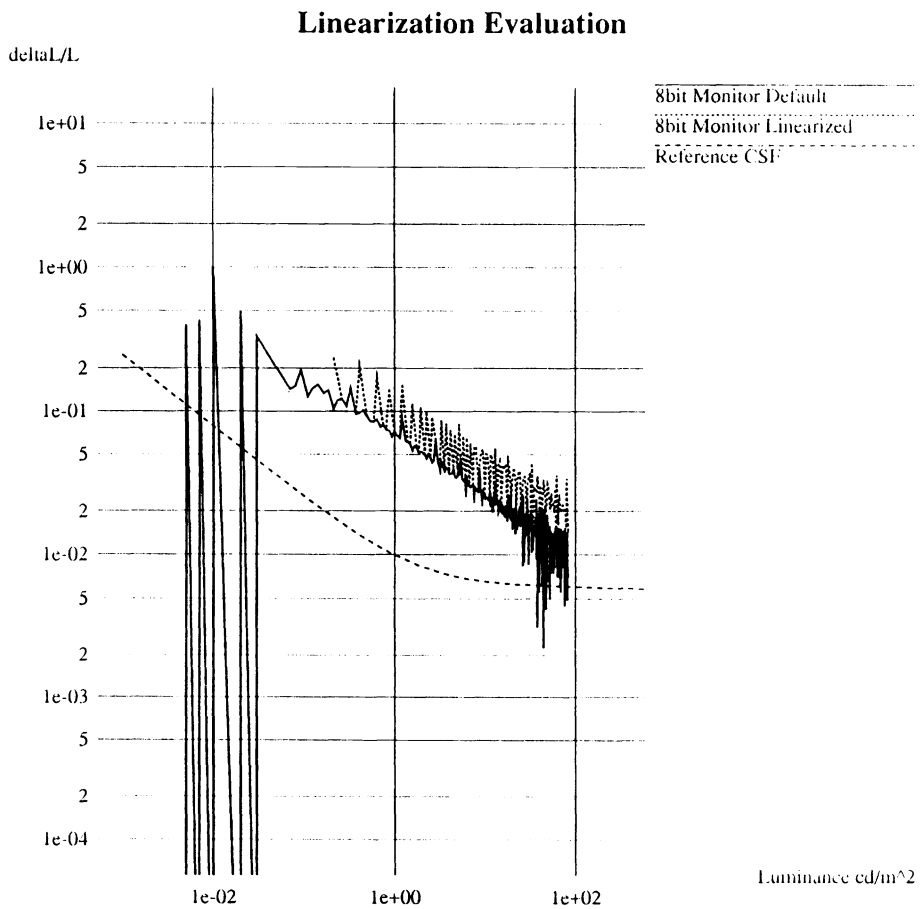

Figure 8. Monitor DACLUM curve for Sun Sparc Station and resulting linearization based CIELUV algorithm, DACLUM for monitor, and intuitive linearization.

Another important issue is the number of DDLs in the resulting map. Both of the methods suggested by Pizer attempt to create mappings with each DDL step being an equal fraction of a JND step apart, with the default implementation creating a table of steps being one JND apart. If the PDR range of the monitor is significantly less than the number of available DDLs on the DAC then we face the issue of whether to use more DDLs. Choosing not to do so means the contrast resolution must be downsampled to the smaller value of PDR rather than number of DDLs available. For instance, on a display system with a PDR of 80, we would have to downsample the input greyscale range of 4096 levels to just 80 levels. This coarse quantization of the input data may be undesirable. In order to use more DDLs one would have to resample the desired $(CSF \circ DACLUM)^{-1}$ curve, similar to the previously described supersampling methods.

5.2 Optimally Calculating actual LUT from $(CSF \circ DACLUM)^{-1}$

None of the above techniques attempt to minimize the error introduced during this matching of $(CSF \circ DACLUM)^{-1}$ 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[38,39]. This problems differs in that: (1) we have fixed non-uniform 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 suggest the development of an optimal solution to this problem, one that *minimizes the perceptual error* in the resulting linearization, and that describes the actual resulting PDR, or *Realizable PDR*. We plan to address solutions to this problem in a later paper.

5.3 DISCUSSION

To date, the methods developed have been aimed at simply implementing a reasonable $(CSF \circ DACLUM)^{-1}$ function. For many monitors the limitation of 8 bit DACs and sub-optimal luminance distributions (not matching the CSF) mean that calculated linearizations may not be very optimal, and in some cases like figure 8, 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 develop a method for determining the optimal solution for the problem of matching $(CSF \circ DACLUM)^{-1}$ to a given set of luminance values for a monitor.

6. SUMMARY

CSF

Fairly good models for the CSF (Barten and Daly) exist, but they need to be expanded to include better surround models, more general types of stimuli, and viewing time. UNC experimental work seems roughly similar to predictions from Barten's; however, the experimental results suggest less sensitivity overall, and especially when the stimulus and its local surround are significantly different in luminance from the overall surround the eye is adapted to. Still, we recommend using the Barten or Daly models for the CSF part of the linearization process.

While there is no single CSF curve that can represent different visual tasks and different clinical viewing conditions, we have attempted to provide a description of the range of possible values (general CSF) and a single choice of parameters for the CSF (reference CSF) based on the results of our experiments and others recent work.

Work needs to be done in comparing visual model based CSF linearizations versus existing default monitor configurations for specific clinical tasks, to evaluate whether improvements in clinical performance occur.

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 (i.e. matching the CSF) approximately 10 bits should be sufficient to eliminate quantization artifacts and to allow step sizes of around 1 JND. The distribution of luminance levels of the DAC should be changed to 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 digital driving levels and better DAC distributions.

Linearization

Methods for linearization exist, but they do not attempt to minimize the error in matching the desired luminance values versus the actual ones available on the DAC. We suggest the need for an optimal solution to the linearization calculation that minimizes perceptual error in the linearization and determines the realizable PDR of the display system. Results should be evaluated to test the benefit of such an optimal method versus the existing ones.

8. ACKNOWLEDGMENTS

This work was supported by NIH grants P01-CA47982, R01-CA60193, and R01-CA44060. I would also like to thank Steve Pizer for his insights and comments, and Scott Daly for his comments on the visual models.

9. REFERENCES

1. Pizer, SM, Chan, FH, "Evaluation of the number of discernible levels produced by a display", *Information Processing in Medical Imaging*, R DiPaola and E. Kahn, Eds., Editions INSERM, Paris, pp. 561-580, 1980.

2. Pizer SM, Intensity mappings: liearization, image-based, user-controlled", SPIE Display Technology II, Vol 271, pp21-27, 1981.

3. Pizer, SM, "Intensity mapping to linearize display devices", *Computer Graphics and Image Processing*, 17, pp 262-268, 1981.

4. Pizer SM, Johnston RE, Zimmerman JB, Chan FH, "Contrast perception with video displays", SPIE Vol 318, PACS for Medical Applications, pp223-230, 1982.

5. Pizer SM, "Psychovisual Issues in the Display of Medical Images", Proc NATO Advanced Studies Institute on Pictorial Information Systems in Medicine, 1985.

6. Johnston RE, Zimmerman JB, Rogers DC, Pizer SM, "Perceptual Standardization", Proceedings of Picture Archiving and Communcations Systems III", SPIE Vol 536, pp. 44-49, 1985.

7. Rogers DC, Johnston RE, Pizer SM, "Effect of Ambient Light o Electronically Displayed Medical Images As Measured by LUmikance Discrimination Thresholds", Journal Optical Society of America, 4, 5, May 1987.

8. Cromartie RC, Johnston RE, Pizer SM, Roger DC, "Standization of Electronic Display Devices Based on Human Perception", University of North Carolina at Chapel Hill, Technical Report 88-002, December 1987.

9. Ji, T, Roehrig, H, Blume, H, Guillen, J, "Optimizing the display function of display devices", *Proceedings Medical Imaging VI: Image Capture, Formatting*, and Display, Vol 1653, pp. 126-139, February 1992.

10. Briggs SJ, "Photometric technique for deriving a "best gamma" for displays", Advances in Display Technology, SPIE 199, pp 134-145, 1979.

11. Blume H, Roehrig H, Browne M, Ji TL, "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", Proceedings of Medical Imaging: Image Capture, Formatting and Display, SPIE volume 1232, pp97-114, 1990.

12. Blume H, Daly S, Muka E, "Presentation of Medical Images on CRT displays A Renewed Proposal for a Display standard", Proceedings of Medical Imaging: Image Capture, Formatting and Display, SPIE volume 1897, pp215-231, 1993.

13. Naiman AV, Makous W, "Spatial non-linearities of grayscale CRT pixels", Proceedings of Human Vision, Visual Processing, and Digital DIsplay III, SPIE volume 1666, pp41-56, 1992.

14. Roehrig H, Blume H, Ji TL, Sundaresham MK, "Noise of the CRT Display System", Proceedings of Medical Imaging: Image Capture, Formatting and Display, SPIE volume 1897, pp232-245, 1993.

15. Barten PGJ, "Physical Model for the Contrast Sensitivity of the Human Eye", Proceedings of Human Vision, Visual Processing, and Digital Display III, SPIE vol 1666, 1992.

16. Daly S, "The visible differences predictor: an algorithm for the assessment of image fidelity", Proceedings of Human Vision, Visual Processing, and Digital Display III, SPIE vol 1666, 1992.

17. Murchison C (Editor), A Handbook of General Experimental Pschology, p769, Clark University Press, Worcester MA, 1934.

18. Hill CR, The Physics of Medical Imaging, p567-583, Webb S (Ed), IOP Publishing Ltd, Philadephia PA, 1988.

19. Wyszecki, G, Stiles, WS, Color Science: Concepts and Methods, Quantitative Data and Formulae, pp. 567-570, 2nd Edition, John Wiley and Sons, New York, 1982.

20. Cornsweet TN, Visual Perception, Academic Press, New York, 1970.

21. Haber RN, Hershenson M, The Psychology of Visual Perception, Holt Rinehart and Winston, New York NY, 1973.

22. Bijl P, Koenderink JJ, "Visibility of Elliptical Gaussian Blobs", Vision Research, Vol 33, No. 2, pp243-255, 1993.

23. Bijl P, Koenderink JJ, Toet A, "Visibility of Blobs with a Gaussian Luminance Profile", Vision Research, Vol 29, No. 4, pp447-456, 1989

24. Daly, S, "Quantitative Performance Assessment of an Algorithm for Determination of Image Fideltiy", S.I.D. Digetst of Technical Papers, V. 24, pp317-320, 1993.

25. Rogers JG, Carel WL, "Development of design criteria for sensor displays devices", Report HAC Ref. No. C6619, Huges Aircraft Company, Culver City, CA, Dec 1973.

26. Barten P, "Spatio-Temporal Model for the Contrast Sensitivity of the Human Eye and its Temporal Aspects", Proceedings of Human Vision, Visual Processing, and Digital Display IV, SPIE Vol 1913, pp2-14, 1993.

27. Lie I, "Visual Detection and Resolution as a Function of Adaption and Glare", Vision Research Vol 21, pp1793-1797, 1981.

28. Wildt GJ, Waarts RG, "The Influence of the Surround on Contrast Sensitivity", 2nd Intl.Conf.on Visual Psychophysics and Medical Imaging, Brussles, Belgium, 2-3 July, 1981 (NY, USA:IEEE 1981).

29. Hood DC, Finkelstein MA, Sensitivity to light Handbook of Perception and Human Performance, Vol 1, eds Boff KR, Kaufman L, and Thomans JP, pp5-28, Wiley, 1986.

30. Kundel HL, "Visual Clues in the Interpretation of Medial Images", Journam of Clinical Neuropysiology, 7(4):472-483, 1990.

31. Burbeck CA, Pizer SM, "Object Representation by Cores", UNC-CS Technical Report 94-004, 1994.

32. DeGroot SG, Gebhard JW, "Pupil Size determined by adapting luminance", Journal of Optical Society America, 42, p492-295, 1952.

33. Kirk RE, Experimental Design: Procedures for Behavioral Sciences, pp261, Second Edition, Wadsworth, Belmont CA, 1982.

34. Sezan MI, Yip KL, Daly SJ, "Uniform Perceptual Quantization: Applications to Digital Radiolography", IEEE Transactions on Man, Machine, and Cybernetics, Vol SMC-17, No 4, pp622-634, 1987.

35. Ji T, Roehrig H, Blume H, Guilllen J, "Optimizing the Display Function of Display Devices", Proceedings of Medical Imaging VI: Image Capture, Formatting, and Display, eds Kim Y, SPIE vol 1653, 1992.

36. Robertson PK, "Color Image Display: Computational Framework Based on a Uniform Color Space", PhD thesis, Australian National University, April 1985, (CSIRO Tech Report No. 27)

37. Levkowitz, H, Herman GT, "Color Scales for Image Data", *IEEE Computer Graphics and Applications*, Vol 12, No 1, pp72-80, Jan 1992.

38. Jain AN, Fundamentals of Digital Image Processing, pp99-123, Prentice Hall, Englewood Cliffs NJ, 1989.

39. Max J, "Quantizing for Minimum Distortion", IRE Trans. on Info. Th., Vol IT-6, No. 1, pp7-12, March 1960.

40. Daly S, personal communication, Jan 1994.

41. Barrett HH, "Objective Assessment of Image Quality: Effect of Quantum Noise and Object Variability", J. Opt. Soc. Am., 7, pp1266-1278, 1990.