VISUAL PERFORMANCE IN MEDICAL IMAGING USING LIQUID CRYSTAL DISPLAYS

by

Philip Marcel Tchou

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Nuclear Engineering and Radiological Sciences)
in The University of Michigan
2007

Doctoral Committee:

Adjunct Professor Michael J. Flynn, Co-Chairperson
Professor Edward W. Larsen, Co-Chairperson
Professor Ronald F. Fleming
Professor Mitchell M. Goodsitt

To my parents,
who have always been there for me
with compliments, criticisms, support, and love.

And to Gregory, Christopher, Michael, and Jeremy,
whose lives have paralleled, crossed, and contrasted with mine.
Our camaraderie and our varying paths have been a source of spirit and insight,
perhaps more than you realize.

ACKNOWLEDGMENTS

I am deeply grateful to my advisor, Michael J. Flynn, whose patient guidance has made this work possible. The many in-depth discussions and the countless hours of collaboration and critical review have been indispensable in ensuring the quality of this work. I wish to thank him for sharing his knowledge, experience, and enthusiasm with me over the course of this project. Most importantly, I would like to thank him for introducing me to the field of medical physics.

I wish to express my gratitude to Edward Larsen, for accepting the role of co-chair and for his comments and review of this work. His advice has been influential and much appreciated. I would also like to thank the staff and students from the University of Michigan, Wayne State University, and Henry Ford Health System that I had the pleasure of working with during my time as a graduate student. I would especially like to thank Mary Brake, Wendy Derby, Pam Derry, Charles Dodge, Peggy Gramer, Helen Lum, Yue Ying Lau, Donald Peck, Phil Rauch, Allen Seifert, and Cari Lehner (now Cari Seifert). Each of them played a significant role in the many experiences and decisions which have led me to where I am now.

On a personal note, I would like to thank the many family members, near and far, that have provided support and encouragement throughout my college years. And finally, an acknowledgement to the many friends that I made in Ann Arbor who have helped me keep my sanity despite appearances to the contrary. You know who you are.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table

LIST OF ACRONYMS

Acronym	
2AFC	two-alternative forced choice.....6
AAPM OR-03	American Association of Physicists in Medicine On-line Report No. 03.....27
AAPM TG-18	American Association of Physicists in Medicine Task Group 18.....27
ACR	American College of Radiology.....24
AFC	alternative forced choice.....56
CCD	charge-coupled device.....13
cLR	calibrated luminance response.....34
CR	computed radiography.....101
CRT	cathode ray tube.....5
CV	coefficient of variance.....120
DAC	digital to analog converter.....36
DL	display level.....116
DDL	digital driving level.....22
DICOM	Digital Imaging and Communications in Medicine.....5
DQE	detective quantum efficiency.....12
DVI	digital video interface.....119
EDID	extended display identification data.....43
FFT	fast Fourier transform.....101

GDI	Microsoft Windows Graphic Device Interface	43
GSDF	grayscale standard display function	6
GUI	graphical user interface	33
H & D	Hurter and Driffield	2
HFHS	Henry Ford Health System	33
IPS	in-plane switching	4
JND	just noticeable difference	8
LCD	liquid crystal display	3
LUT	lookup table	22
MLE	maximum likelihood estimation	7
MTF	modulation transfer function	12
NEMA	National Electrical Manufacturers Association	24
NEQ	noise equivalent quanta	101
NPS	noise power spectrum	111
NV	noise visibility	114
NVI	noise visibility index	115
OD	optical density	1
PACS	picture archiving and communication system	2
RMSE	root mean squared error	39
RPA	resting point of accommodation	24
RPV	resting point of vergence	24
SDK	Microsoft Windows Software Development Kit	33
TIFF	Tagged Image File Format	97

ABSTRACT

This thesis examined the contrast performance of liquid crystal display (LCD) devices for use in medical imaging. Novel experimental methods were used to measure the ability of medical LCD devices to produce just noticeable contrast. It was demonstrated that medical LCD devices are capable of high performance in medical imaging and are suitable for conducting psychovisual research experiments.

Novel methods for measuring and controlling the luminance response of an LCD were presented in Chapter 3 and used to develop a software tools to apply DICOM GSDF calibrations. Several medical LCD systems were calibrated, demonstrating that the methods can be used to reliably measure luminance and manipulate fine contrast.

Chapter 4 reports on a novel method to generate low contrast bi-level bar patterns by using the full palette of available gray values. The method was used in a two alternative forced choice (2AFC) psychovisual experiment to measure the contrast threshold of human observers. Using a z-score analysis method, the results were found to be consistent with the Barten model of contrast sensitivity.

Chapter 5 examined error distortion associated with using z-scores. A maximum likelihood estimation (MLE) method was presented as an alternative and was used to re-evaluate the results from Chapter 4. The new results were consistent with the Barten model. Simulations were conducted to evaluate the statistical precision of the MLE method in relation to the number and distribution of trials.

In Chapter 6, 2AFC tests were conducted examining contrast thresholds for complex sinusoid, white noise, and filtered noise patterns. The sinusoid test results were consistent with the Barten model while the noise patterns required more contrast for visibility. The effects of adaptation were also demonstrated. A noise visibility index (NVI) was introduced to describe noise power weighted by contrast sensitivity. Just noticeable white and filtered noise patterns exhibited similar NVI values.

The results indicated that medical LCD devices are capable of high performance for psychovisual testing and presenting complex patterns. However, 256 gray levels are insufficient for high fidelity. 10-bit grayscales can achieve high fidelity, but it is uncertain whether such changes would affect diagnostic performance.

CHAPTER 1

VISUAL PRESENTATION USING MEDICAL DISPLAYS

1.1 Radiographic Image Presentation

In November of 1895, Wilhelm Conrad Roentgen discovered x-rays and soon after took the first radiographic images. Most memorable among these were images of human anatomy, made using photographic materials. Less than three months later, in January of 1896, the first radiograph made for clinical diagnosis was taken by two Birmingham (UK) doctors, Ratcliffe and Hall-Edwards, to locate a needle in a woman's hand. This was soon followed by the first x-ray guided surgery, performed by J. H. Clayton who reviewed the image of the needle in the woman's hand from a bromide print[1]. Thus began the practice of radiography and the study of diagnostic medical imaging. Thus also began radiography's reliance on photographic plates and films.

For over a century, film has been the primary medium for radiographic image presentation. It has defined both the training and the technology that go into the practice of radiology. Most of the improvements in medical imaging over the past century have been designed with film in mind for the final presentation. After a century of research and development, the characterization of film systems is generally considered to be well-understood. The intensity of light transmitted by a backlit film is inversely proportional to the inverse-log of the optical density (OD), $I=I_o/10^{OD}$. The OD of a film varies with x-ray exposure, and is described by the Hurter and Driffield (H & D) curve. X-ray film systems are designed to operate in the "linear region" of this curve, where the OD increases linearly with log-exposure, $OD = \gamma\log(E)$. As a result, the transmitted light intensity becomes a function of the inverse exposure, $I= I_o/E^{\gamma}$. Film also has extremely low granularity and high resolution, owing to its molecular structure. The combination of all of these characteristics gives film exceptional fidelity as a display medium.

Despite film's excellent image quality, the use of film has declined in recent years. Motivated by cost savings and electronic access provided by picture archiving and communication systems (PACS), many radiology departments have been replacing their film systems with electronic display systems[2]. Liquid crystal displays (LCDs) are becoming the preferred device for review of medical images. Even for the demanding task of diagnostic interpretation by radiologists, specialized medical LCD devices have been introduced. In the last 5 years, many major medical centers have completely eliminated the use of film.

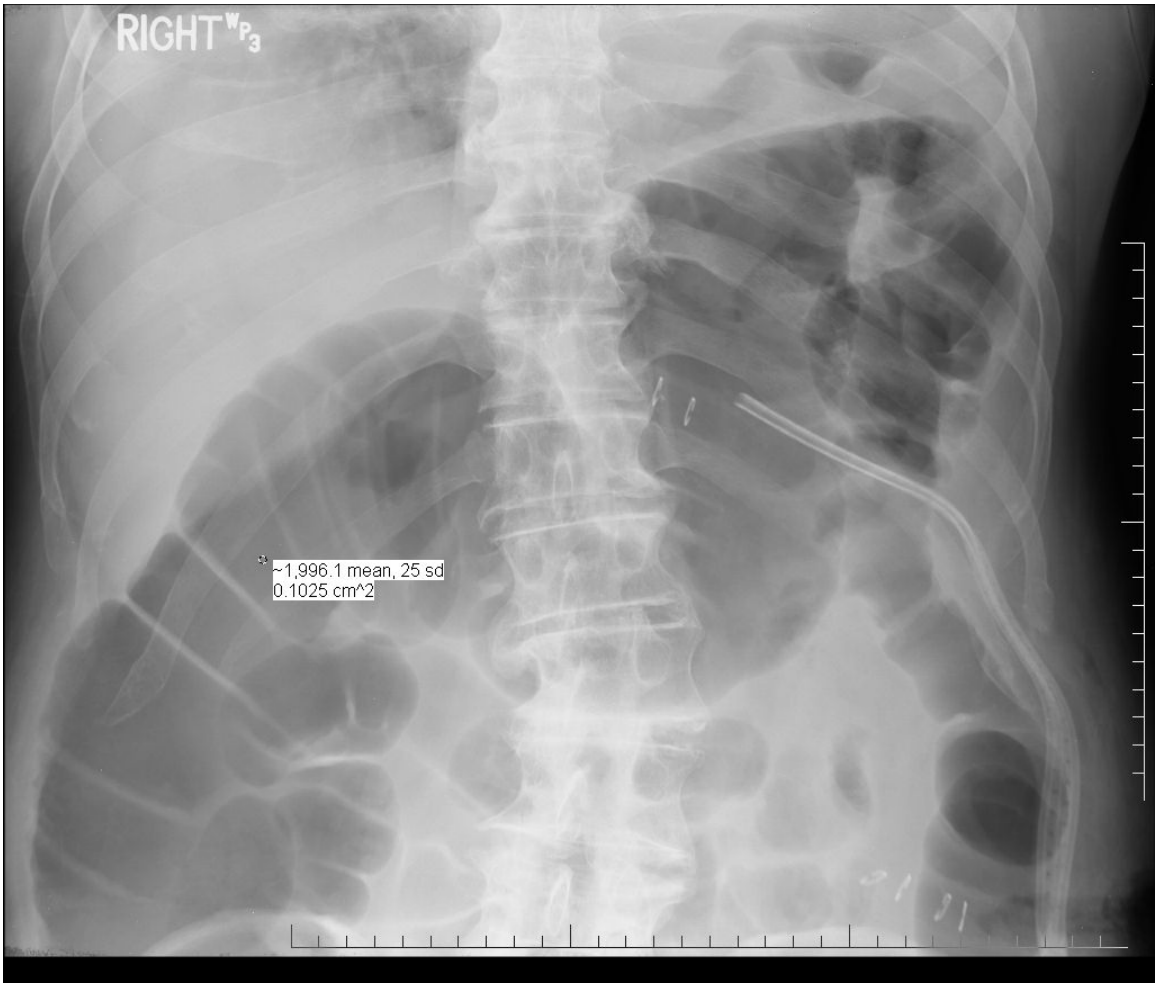

Figure 1.1: AP Abdomen study taken by a Kodak DR 7100 direct radiography system with a pixel size of 139 microns. The radiation exposure used was equal to the target value for this radiographic view and is similar to that of a 400 speed screen film system.

Figure 1.1 shows a typical radiographic image from an abdomen study taken by a Kodak DR 7100 digital radiography system. In this format, the image may appear to have relatively little noise. The background seems very uniform, whereas visible noise would be expected to look grainy. However, on a high-fidelity medically calibrated display, the noise is just visible and can impact the detection of disease. This image illustrates the demanding requirements for the display of medical images. In order to improve diagnostic performance in medical imaging tasks, it is important to understand how image features are perceived by human observers.

Visual performance with LCD devices is not nearly as well understood as it is with film. With such a significant technological change in medical imaging, it is important to examine critically any visual performance deficits that are associated with the use of LCD monitors. Diagnostic imaging tasks require high-fidelity display systems that maximize the performance of the human visual system. Psychovisual experiments can be used to determine the requirements for high-fidelity displays. LCDs are already capable of achieving spatial resolution that is optimal for the human observer[3]. However, contrast resolution for medical imaging is still an unknown, both in terms of what LCDs are capable of and what observers need for high-performance diagnostic imaging. Current medical imaging display systems are restricted to presenting images using 8-bit grayscales unless specialized hardware or software is used. It is not certain whether this is an actual limitation for diagnostic imaging tasks or if current standards are sufficient for optimum observer performance. This dissertation begins to address these issues by investigating demanding visual performance tasks using carefully calibrated medical LCD devices.

The following chapters will discuss the development of tools and methods for controlling the contrast of an LCD. Experimental procedures are then designed to validate these methods while studying aspects of human visual performance with LCDs. Experimental studies are conducted to measure psychovisual thresholds for simple contrast patterns, more complex noise patterns, and contrast thresholds under varying states of adaptation. The contrast threshold results are found to be consistent with previous research, confirming that our experimental methods are sound. The noise threshold and adaptation threshold results are presented as new research and are analyzed

in relation to current medical imaging standards. The results of these studies demonstrate that LCDs can be used for high performance medical imaging, while also identifying limitations in current medical LCD systems.

This chapter presents a summary of the issues facing LCD systems used for diagnostic medical tasks, methods for overcoming some of these issues, techniques for measuring observer performance using LCDs, and human perception studies conducted using these techniques. Later chapters, described below, go into further detail on each of these topics.

1.2 Liquid Crystal Displays

Early LCDs suffered from several key deficiencies. They were relatively limited in size, brightness, and contrast ratio (max/min brightness), and had poor off-angle viewing performance[4, 5]. However, recent improvements in LCD design and manufacturing have significantly ameliorated these problems, making them more suitable for the display of medical images[6].

LCDs can be described as a multi-layered optical "stack" composed of a backlight mounted behind a thin-film transistor liquid crystal matrix, color filters (in the case of color LCDs), light diffusers, and light polarizing films[7]. The optical characteristics of the liquid crystal change when an electric field is applied. The transmission of light is then associated with its polarization as it passes through he polarizer films and liquid crystal layer[3]. Light transmission is generally poor, with only 3-5% of the light generated by the backlight exiting the face of the display in color LCDs and 8-15% in monochrome LCDs[6]. As a result, high-efficiency backlights are needed to produce adequate levels of brightness over the lifetime of the display.

The light output from an LCD is also a strong function of the viewing angle. As the electric field is modulated, the liquid crystal molecules are oriented in an oblique direction relative to the display. Light traveling off the normal axis encounters differing optical thickness. As a result, the transmission drops off and the color can shift dramatically (or the grey level for monochrome displays). Several technologies have been introduced to improve off-angle performance, the most common being multi-domain, in-plane switching (IPS), and vertical alignment (VA). Multi-domain display

pixels are divided into sub-pixel regions that are each oriented differently so that the angular response is more even. IPS technology aligns the electric field such that changes in the liquid crystal molecules are limited to the plane parallel to the display, minimizing the asymmetry due to modulation. VA technology runs somewhat counter to IPS technology by actually improving on the perpendicular alignment of the molecules relative to the display. The purpose is to provide near perfect blockage of light, improving the dark states and the contrast ratio. Combined multi-domain and VA technologies have shown excellent viewing angle performance[6].

1.3 Response Characteristics and Calibration

After a century of use and research, the characterization and optimization of film systems for the viewing of medical radiographs is generally considered to be well-understood. One of the most important aspects of film is its linear region, over which its response is consistent in relation to input exposure, a very desirable quality in any analog recording device[8]. Film radiography systems have taken full advantage of this behavior, with technologies and techniques designed to maintain exposures within the linear region. As cathode ray tube (CRT) displays came into the clinical environment, their characteristics were studied for medical imaging. Similar to film, the CRT display is an analog device with a consistent output in response to input voltage. Medical industry standards have since been developed for the calibration of electronic displays based on these properties. This is known as the DICOM 3.14 standard, which is designed to produce a grayscale display function (GSDF) with even changes in perception at all display levels[9]. An effectively calibrated system will produce luminance at each level consistent with the DICOM GSDF. Further details regarding the DICOM GSDF are given in Chapter 2.

Liquid crystal displays (LCDs), which are now quickly replacing both film and CRT presentation systems, have significant differences from film. Most notable is their lack of a “linear region” of behavior, with transitions between digital image levels producing erratic changes in contrast[7]. As a result, the medical calibration standards developed for CRT displays are not viable for LCDs. Methods are presented in Chapter 3 of this dissertation to measure and calibrate the grayscale of LCD monitors by

systematically selecting gray values from large gray palettes. It is shown that these calibration methods produce gray level transitions that closely follow the DICOM 3.14 standard.

It is important to note that this work on LCD calibration has led to the development of an open source software package, pacsDisplay, for the calibration of medical LCDs. This software has been used by two major medical centers, Henry Ford Health System and the Mayo Clinic. In addition, this research has directly influenced calibration of commercial products by medical display manufacturers.[1]

1.4 Contrast Threshold Testing

In human visual perception, contrast threshold refers to the contrast of a small visual pattern that is just detectable by a human observer. The measurement of the contrast threshold has been performed many times over the last century using patterns presented with high performance optical projection display devices. Many of these studies were collected and used to form a model of contrast perception known as the Barten Model[10]. This model was in turn used to form the DICOM grayscale standard display function (GSDF) for the luminance response of medical monitors. Chapter 4 reports on an experiment designed to measure contrast threshold with an LCD using bi-level bar patterns with extremely small contrast. This experiment establishes that LCD devices have the ability to present very subtle contrast without interference from device noise or pixel non-uniformity.

The test format used for the experiment was a two-alternative forced choice (2AFC) test, in which a target is presented randomly in one of two positions and the observers must choose which position it is in. By pairing specific display levels, grating pattern targets with a variety of contrast levels were formed. Observers were tested to determine how well they could detect these targets at different contrast levels and were required to make a decision, even when they were uncertain where the target was. By definition, the contrast threshold is the contrast level where the observer has a 50% chance of detecting the target. For the 2AFC paradigm, this translates to a 75% chance of

[1] Direct collaboration with National Display Systems (NDS) led to their inclusion of these techniques into their MDCAL calibration software for their medical displays. Other medical display manufacturers have since adjusted their own calibration software to include similar features.

being correct. Since the results reported by the test were binary, they were converted into a percent-correct value using a detectability index, also known as the z-score method[11]. Multiple observers were tested, and it was found that the average performance was within one standard deviation of the expected threshold, based on previous literature. This result established that the measurement and calibration techniques presented could be used to precisely control the contrast at very low levels, and that the experimental methods could be used to accurately measure observer performance in perception tasks. In addition, several cases were recorded of observers who were able to detect contrast well below the expected threshold, thus providing evidence of significant variation among individual observers. Further details are provided in Chapter 4.

1.5 Maximum Likelihood Analysis

Use of the z-score method during the contrast threshold tests revealed limitations that affected the design of the 2AFC experiment and the analysis of the test results. It was determined that the z-score method distorted the original contrast information primarily in two ways. First, it involved binning the data and taking the average value. Since the contrast levels that were tested were not evenly spread, due to the uneven luminance response found in LCDs, binning would have caused uneven shifts in the original contrast values. Second, the z-score then converted the binned values into a linear model. Contrast detection is commonly accepted to follow a psychometric model, which is non-linear in behavior. Fitting the data to a linear model would have distorted the standard deviations at different contrast levels.

Other methods for analyzing binary data were investigated in order to improve the accuracy of the experimental method. Eventually, a maximum likelihood estimation (MLE) method was chosen. This process calculates the combined probability of all binary results and tests combinations of parameters to find the model that is most likely to have produced those outcomes. The method was customized for a 2AFC psychometric experiment. The data from the contrast threshold tests were then reevaluated, resulting in a slight increase in the observers' threshold values, although not so much as to significantly alter the conclusions. In addition, a number of simulations were set up using the MLE algorithm in order to examine how the standard error would be affected by the

range of test values chosen. The results of these tests, as well as derivations for the MLE method, are reviewed in Chapter 5. No previous 2AFC medical imaging studies have used the MLE method, and most new studies still use the z-score method.

1.6 Variable Pattern Threshold Testing

In Chapters 3-5, advanced statistical methods are developed to analyze the contrast threshold of bi-level bar patterns presented on calibrated LCD monitors in 2AFC experiments. Chapter 6 extends these concepts to perception measurements of visual patterns with more complex variations, requiring a range of low-contrast gray levels with perceptually uniform luminance intervals. For a typical medical display system with a GSDF calibrated 8-bit grayscale, the change in luminance between successive gray levels is perceptually large, approximately twice that of the human visual contrast threshold. By using a non-standard calibration with a very narrow luminance range, gray level changes can be presented that have a relative luminance change of much less than a just noticeable difference (JND).

Two test patterns were chosen, a sinusoid and a white noise pattern. The sinusoidal pattern is used to establish whether the contrast threshold measured using varied gray levels is consistent with the Barten model. It also provides a measure of the variability amongst each observer's contrast threshold. Responses to the sinusoid patterns were also tested under varying image background conditions to examine the effect of visual adaptation on contrast detection. Although this effect has been acknowledged by medical imaging researchers[6, 12], there have been no previous studies to measure it in medical LCDs.

The white noise test patterns were intended to mimic the quantum mottle noise commonly seen in direct radiography systems. They were modeled by a Gaussian noise distribution and were varied in their standard deviations relative to the image backgrounds. These patterns were used in a 2AFC test, in a manner similar to the sinusoid patterns, in order to measure the noise perception threshold for human observers. A second set of noise patterns were generated with a low-pass filter function applied in order to better simulate the more structured quantum mottle noise seen in indirect

radiography systems. All of the noise test results were measured against the contrast per gray level of the DICOM standard.

1.7 Summary

The goal of this dissertation is to determine if medical LCD systems are capable of presenting medical images with subtle contrast that is at the threshold for human visual detection. Future research might then examine whether presentation with improved contrast resolution would result in increased observer performance for diagnostic medical imaging tasks. The following summarizes the methods and results reported in the remaining chapters of this dissertation.

In Chapter 2, important concepts on human visual performance and image presentation are presented to help the reader understand this area of research. At the end of this chapter, current industry standards and professional guidelines are summarized.

Chapter 3 presents a method to measure the luminance of the available gray levels in an LCD system. This is referred to as the gray palette. A technique is then described to transform the luminance response of an LCD system to a set of gray levels with precisely controlled contrast between each successive level. The method is shown to produce calibrated responses that meet current medical imaging standards. The techniques described in this chapter are used to create controlled image characteristics for the observer performance experiments of Chapters 4 and 6. Software developed in conjunction with this chapter has otherwise been made available as a open source package for medical monitor calibration.

Chapter 4 uses the measured gray palette to generate bi-level bar patterns. Using a 2AFC experimental method, an observer performance study is conducted to measure the contrast threshold using a medical LCD and compare the results to previous research. No such studies had previously been done using medical LCDs. The results of the study matched the expected results from previous studies, validating the experimental methods and the techniques for producing low contrast.

Chapter 5 introduces a new MLE method for the analysis of 2AFC test results. Reasons are provided for using this method over the commonly used z-score method in order to reduce data distortion. Equations for the new MLE method are derived and the

results from Chapter 4 are re-evaluated. Some differences in the results are seen but the overall conclusions remain the same. The MLE method has an important advantage in that observer performance may be measured using a set of images with patterns of widely varying contrast.

In Chapter 6, the calibration methods of Chapter 4 are used to establish display conditions with 256 gray levels having very small contrast between successive levels. New observer performance methods are described to conduct 2AFC contrast threshold experiments using complex image patterns. Contrast threshold is measured for a sinusoid pattern under varying states of adaptation by varying the background. The results confirmed behavior that has been previously hypothesized. Contrast threshold is also measured for white and filtered white noise mimicking the quantum mottle seen in radiographs. The results suggest that the contrast of typical quantum mottle is just above the contrast threshold. A noise visibility index is proposed that is calculated from the noise power spectrum weighted by the expected human contrast sensitivity. The noise visibility index provides similar results for the just noticeable white and filtered noise patterns, suggesting that it is effective at describing the just noticeable condition.

The experimental results presented in this dissertation demonstrate that medical LCD devices are capable of increased contrast resolution for high-performance psychovisual testing, but that 8-bit grayscales common in LCD systems prevent them from achieving high-fidelity for the presentation of easily seen patterns such as a region with sinusoidal luminance modulation. However, tests using two dimensional patterns that are more typical of the noise in medical radiographs suggest that optimal observer performance may not require more than 8-bit grayscales. Further research is needed to determine the contrast resolution required for specific medical imaging tasks. This dissertation provides a set of validated experimental tools and methods that can be used for such future psychovisual research in medical imaging.

1.8 References

1. Webb, S., *The Physics of medical imaging*. 1988, Bristol ; Philadelphia: Hilger. xv, 633 p.
2. Nitrosi, A., et al., *A Filmless Radiology Department in a Full Digital Regional Hospital: Quantitative Evaluation of the Increased Quality and Efficiency.* Journal of Digital Imaging, 2007. **20**(2): p. 140-148.
3. Flynn, M.J., et al., *High-fidelity electronic display of digital radiographs.* Radiographics, 1999. **19**(6): p. 1653-69.
4. Badano, A. and D.H. Fifadara, *Goniometric and conoscopic measurements of angular display contrast for one-, three-, five-, and nine-million-pixel medical liquid crystal displays*. 2004, AAPM. p. 3452-3460.
5. Samei, E. and S.L. Wright. *Effect of viewing angle response on DICOM compliance of liquid crystal displays*. in *Medical Imaging 2004: PACS and Imaging Informatics*. 2004. San Diego, CA, USA: SPIE.
6. Badano, A., M. Flynn, and J. Kanicki, *High-Fidelity Medical Imaging Displays*. 2004, Bellingham: SPIE - The International Society for Optical Engineering.
7. Fetterly, K., et al., *Introduction to Grayscale Calibration and Related Aspects of Medical Imaging Grade Liquid Crystal Displays.* Journal of Digital Imaging, 2007.
8. Bushberg, J.T., *The essential physics of medical imaging*. 2nd ed. 2002, Philadelphia: Lippincott Williams & Wilkins. xvi, 933 p.
9. NEMA, *Digital Imaging & Communications In Medecine (DICOM), Part 14: Grayscale Standard Display Function*. 1998, National Electrical Manufacturers Association.
10. Barten, P.G.J., *Contrast sensitivity of the human eye and its effects on image quality*. 1999, Bellingham, Wash.: SPIE Optical Engineering Press. xix, 208.
11. Burgess, A.E., *Comparison of receiver operating characteristic and forced choice observer performance measurement methods.* Medical Physics, 1995. **22**(5): p. 643-55.
12. Samei, E., et al., *Assessment of display performance for medical imaging systems: executive summary of AAPM TG18 report.* Med Phys, 2005. **32**(4): p. 1205-25.

CHAPTER 2

HUMAN VISUAL PERCEPTION AND THE PRESENTATION OF DIGITAL RADIOGRAPHS

2.1 Introduction

This chapter provides background information and definitions to help the reader in understanding the concepts presented throughout the dissertation. This primarily includes information on common metrics for image presentation and how they relate to LCDs. Current standards for medical imaging display systems are also discussed.

The process of medical imaging can be grouped into several generalized steps, as depicted in Figure 2.1, starting with the production of radiation (1), then attenuation by interaction with various tissues (2), followed by energy-dependent detection of the radiation (3), transformation of the recorded signals into values intended for display (4), translation of those values into display brightness (5), and finally interpretation by an observer (6). To date, the majority of research on the engineering of imaging systems has focused on the first four steps – optimization of the radiation spectrum, development of improved detector systems, and image processing methods to improve contrast and resolution. Less research has been done to examine the display and interpretation. Traditional physical measures for imaging, such as the modulation transfer function (MTF) and the detective quantum efficiency (DQE), are effective metrics for describing detector system performance but are not necessarily good predictors of an observer's performance for a specific interpretation task. This requires metrics that are more appropriate for image displays and consideration of the human visual system and how an observer will respond to various factors influencing the imaging process.

This chapter presents a summary of the concepts and issues involved with the characteristics of human visual perception and the presentation of digital radiographs. Section 2.2 discusses the basic characteristics of images and how they interact with the

human visual system. Section 2.3 considers the properties of electronic displays and how they affect image presentation. Section 2.4 provides an overview of the current industry standards regarding digital image presentation in medicine.

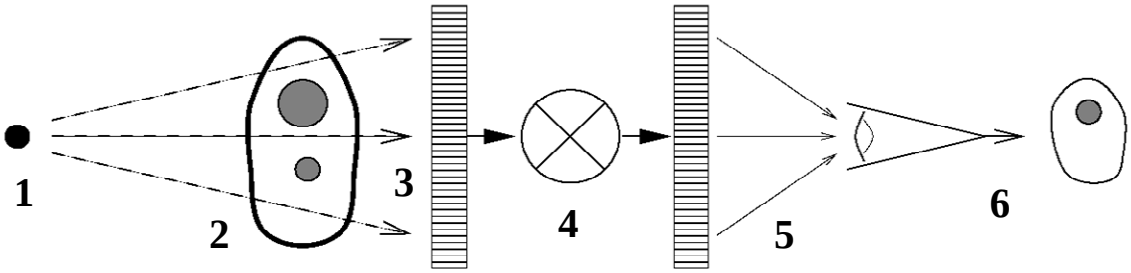

Figure 2.1: The radiographic imaging process in 6 steps.

2.2 The Human Visual System and Image Characteristics

In general, medical image presentation systems are designed with the goal of optimizing the response from the human visual system. The limits of the human visual system have been measured in various psychometric studies, providing us with an understanding of how an observer perceives information transferred from an image system[1]. In this section, a brief overview of the human visual system is given along with related standards of measurement for image characteristics. These concepts will be important when discussing the characteristics of electronic displays and the medical standards developed for them.

2.2.1 Basic human perception

Figure 2.2 shows a simplified anatomy of the human eye. The human eye is similar to a conventional photon detector system with the ability to detect 1.70-3.35eV photons (370-730nm). Incoming photons are focused by the eye onto the retina, in a manner similar to a camera and a charge-coupled device (CCD) sensor. The central point of image focus in the retina is called the fovea, which is responsible for detailed visual recognition of patterns in scenes with high brightness. When interpreting medical images, the observer will typically search the image for detailed findings using the fovea region[2]. The foveal response to image characteristics is the primary concern of this research.

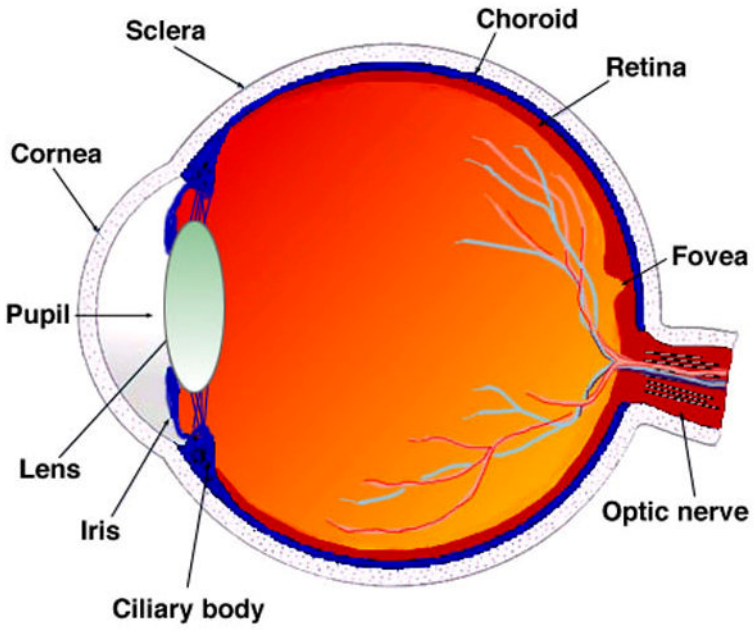

Figure 2.2: "Sagittal section of the adult human eye" from Webvision (webvision.med.utah.edu).

The fovea contains numerous rod and cone cells, shown in Figure 2.3, that act as photo-receptors, converting absorbed photons into neural activity. The cone cells are primarily found in the central region of the fovea, with a small region in the center that is packed exclusively with cone cells at a very high density. This region corresponds to approximately 2° of visual angle[2]. Given a typical viewing distance of 60cm, this translates to only a 2cm wide region within an image. Cone cells are divided into three different types, each of which is sensitive to a different range of visible light, as shown in Figure 2.4. These ranges approximate the red, green, and blue wavelengths, and allow the human visual system to distinguish differences in color. The rod cells are dominant in the peripheral regions of the fovea. They are limited to black-and-white perception, but are also highly sensitive to changes in luminance, improving motion-detection. Both types of cells have a columnar structure with an absorptive layer at the backside, providing significant directional sensitivity[3, 4]. This produces some anti-scatter filtering of incoming and backscattered photons, giving the eye better performance than most photographic recording systems.

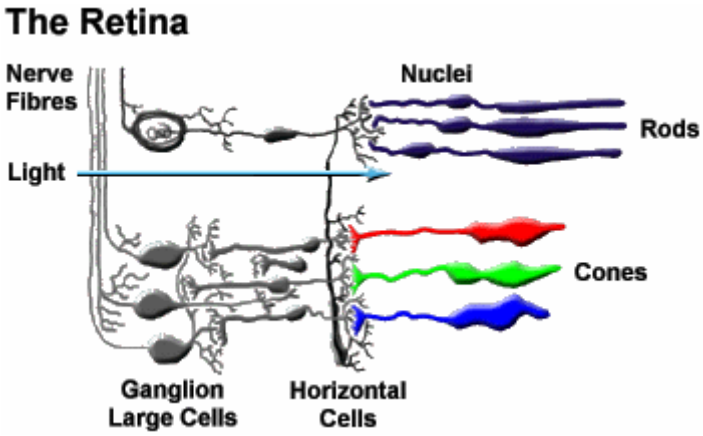

Figure 2.3: Illustrated anatomy of the retina from Photo.net ("Spectral Selectivity", Ed Scott).

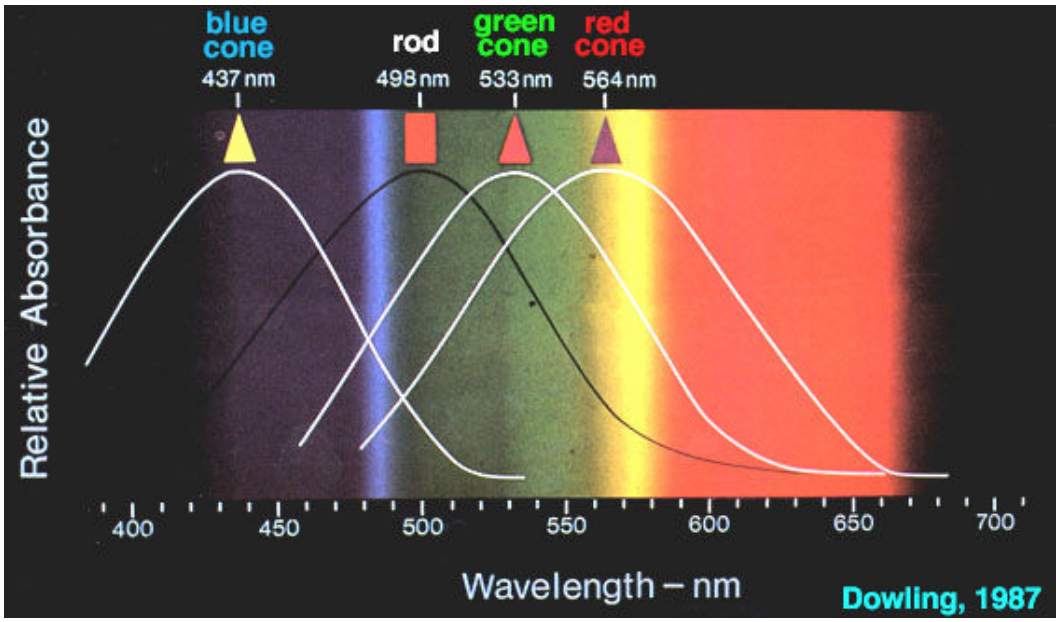

Figure 2.4: Relative spectral response of the rods and cones within the retina.

2.2.2 Luminance and contrast

The basic unit of measurement used when characterizing a displayed image is "luminance". Luminance (*L*) is defined as the brightness of a small region on a display surface, and has units of candela per square meter (cd/m^2). The candela is the SI unit for luminous intensity. Specifically the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency 540 x 1012 hertz and that has a radiant intensity in that direction of 1/683 watts per steradian. The SI unit for brightness is the nit, which is equal to 1 cd/m^2, or 0.2919 foot-lamberts (fL).

Differences in luminance within an image produce "contrast" (*C*). Modeled using a sinusoidal luminance pattern, *C* is defined as the difference between the minimum and maximum luminance divided by the average luminance of the scene,

$$C = \frac{L_{\max} - L_{\min}}{L_{avg}} = \frac{\Delta L}{L}. \tag{2.1}$$

This is also referred to as the "signal strength" or the "modulation". An alternate definition often used is the "Michelson contrast",

$$C_M = \frac{L_{\max} - L_{avg}}{L_{avg}} = \frac{\Delta L}{2L}, \tag{2.2}$$

which uses the amplitude of the sinusoidal variation rather than the peak-to-peak value. This research study uses the first definition unless stated otherwise.

Detection of contrast by the human visual system varies with luminance, spatial frequency, and several other parameters of secondary influence[5]. In terms of luminance, the probability of detection increases as the signal modulation strength increases. This behavior is non-linear and can be described by the "psychometric function". Mathematically, the shape of the psychometric function is often approximated by the cumulative Gaussian probability function, as shown in Eq. (2.3)[1].

$$p(s) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{s} e^{-\frac{(x-s_0)^2}{2\sigma^2}} dx \tag{2.3}$$

p = detection probability

s = signal strength

s_0 = signal strength for a detection probability of 50%

σ = standard deviation of the Gaussian distribution

The level of modulation at which the probability of detection is 50% is defined as the "contrast threshold", C_T. In psychophysics, the inverse of the contrast threshold is referred to as the "contrast sensitivity", C_s. Figure 2.5 shows the contrast threshold at a constant spatial frequency of 0.5 cycles/mm. C_T is nearly constant at the high luminance end, but rises below 10 cd/m^2[6]. It should be noted that C_T can be referenced in terms of contrast per just-noticeable-difference ($\Delta L/L$ per JND). In this work, all values of C_T are presented in this form, including those shown in charts and graphs. The term "JND" and how it is calculated are explained further in Section 2.4.

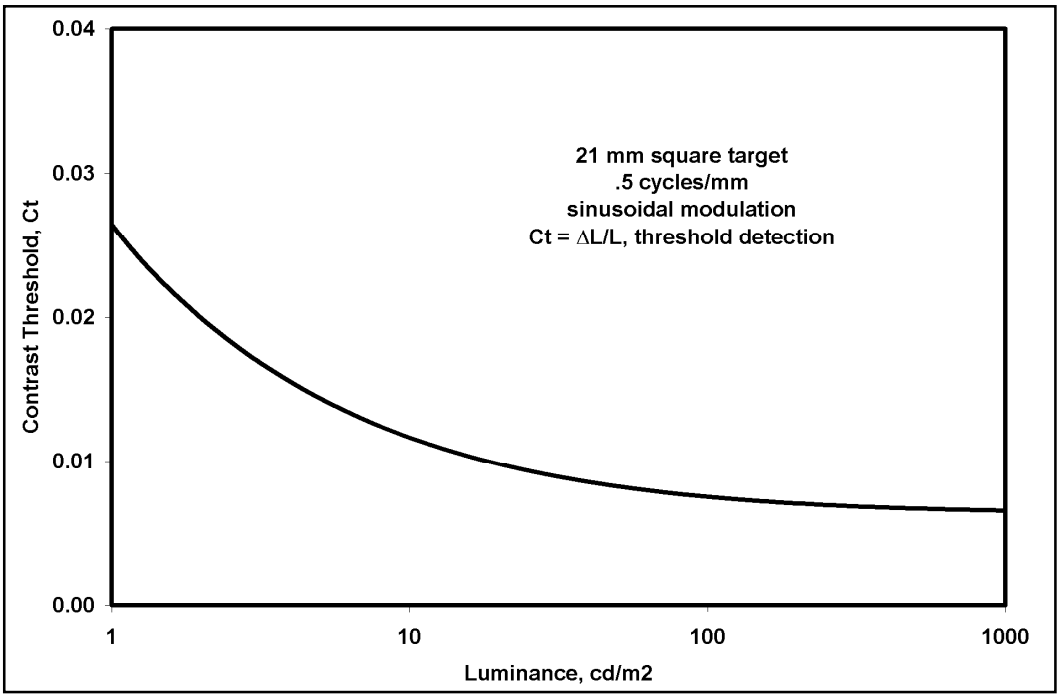

Figure 2.5: Human contrast threshold for a grating pattern as a function of luminance.

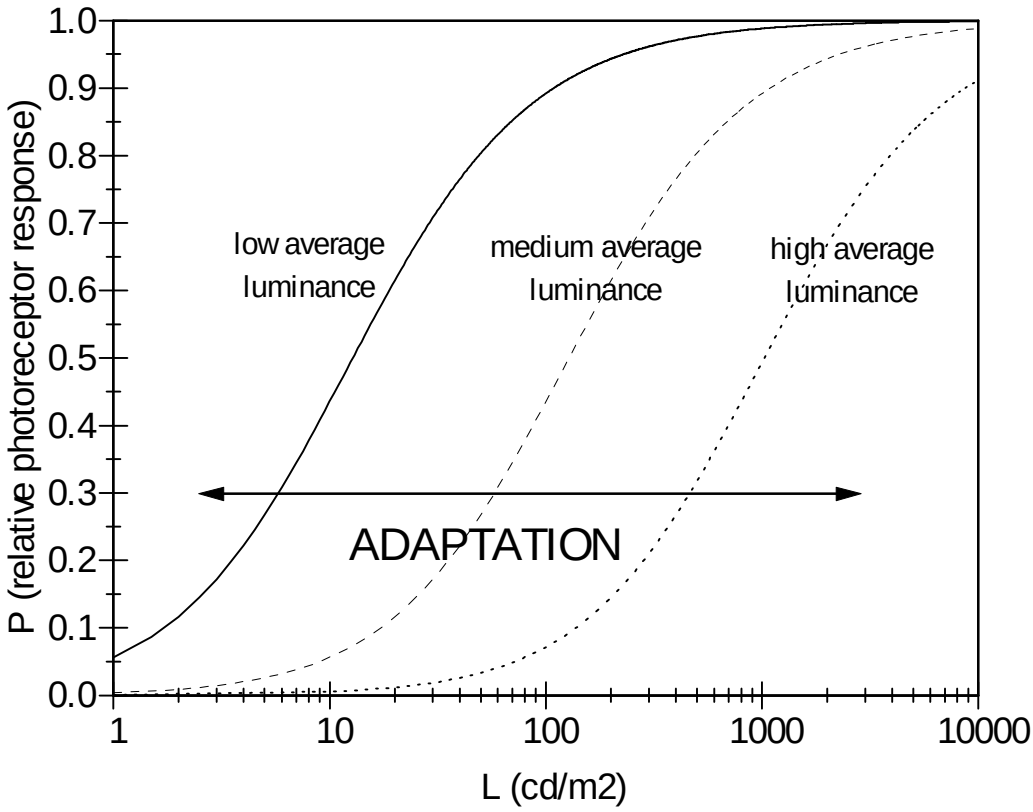

Figure 2.6: Relative photo-receptor response of the eye to luminance for three different average luminance levels. Similar incremental changes in stimulus cause a different response according to the adaptation state.

Contrast perception also varies with the average luminance, rather than simply the contrast. When viewing an image, the human visual system adapts to the average quantity of light falling on the retina. This is referred to as "fixed adaptation"[7]. The adaptation process results in a shift in the neuronal response to input luminance[8]. Baxter[9] has reported on a relationship for photo-receptor sensitivity involving the detection of low-contrast radiological features in non-uniform backgrounds. The photo-receptor response (R_{photo}) for the visual system can be approximated by the function

$$R_{photo} = \frac{I_{retina}}{I_{retina} + S_{adapt}}. \tag{2.4}$$

I_{retina} = retinal intensity

S_{adapt} = state of adaptation constant

This relationship is demonstrated in Figure 2.6, showing how the photoreceptors of the eye shift their response to luminance after adapting to different luminance levels[10]. This function was presented by Hecht as part of a photochemical theory of photo-receptor response[11], and has been confirmed in experiments by Baylor[12]. However, no experimental studies have yet been published applying this concept to human observers or medical imaging.

Figure 2.7 shows how adaptation affects the contrast threshold[2]. Line "B" shows the contrast threshold as a function of luminance, given that the eye has had time to adapt to each luminance level. However, for a typical radiograph with a fixed average luminance but varying degrees of luminance throughout the image, line "A" shows a more realistic response. After adaptation, the human visual system's perception of contrast drops off in brighter and darker regions, constraining the useful luminance range. In order to prevent information from being lost to the observer, the maximum and minimum luminance of a radiology display must be similarly constrained. This requirement is often described using the luminance ratio, L_{max}/L_{min}.

Spatial frequency is yet another major factor that affects contrast perception. As shown in Figure 2.8, contrast sensitivity drops off at low frequencies as well as at high frequencies[2]. The curve shape follows the form

$$C_S \sim f^2 e^{-f}, \tag{2.5}$$

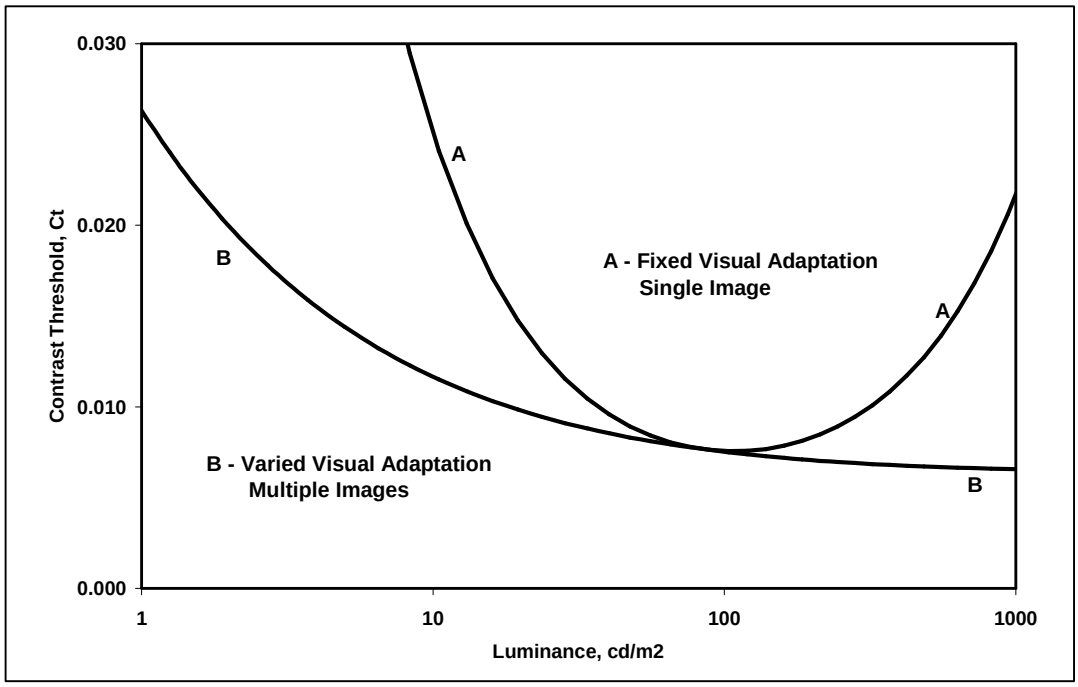

Figure 2.7: Contrast threshold predicted for conditions where the observer is globally adapted at 100 cd/m2 to a single image (A). This response is compared to the contrast threshold for an observer who is variably adapted to a series of images with different luminances (B).

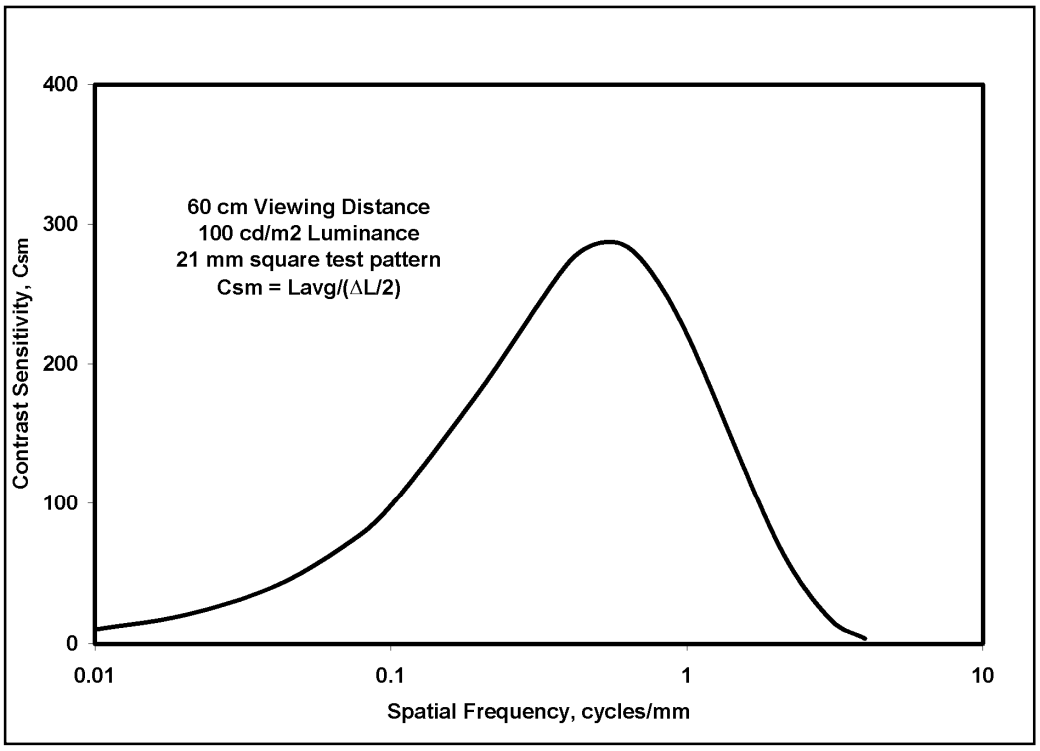

Figure 2.8: Contrast sensitivity based on the Michelson contrast, shown as a function of spatial frequency for a 21mm grating test pattern image viewed at a 60cm distance.

where *f* is the spatial frequency. The shape of the relationship suggests that the human visual system acts as a spatial bandpass filter[5]. For typical viewing distances, the response is best at a frequency of about 0.5-1.0 cycles/mm, dropping off rapidly beyond the maximum[13].

2.2.3 The Barten Model

As electronic displays became more common in medical imaging, the need arose for the development of a display calibration standard. Several semi-empirical models for visual contrast perception were proposed[14], including work by Rogers and Carel[15], Barten[6, 16], and Daly[17]. Of these, the "Barten Model" became the most widely accepted and forms the basis for the current medical imaging display calibration standards[18]. This section provides a brief overview of the Barten model and the major contributing factors involved. While this section consists of a large number of terms and variables, for the purposes of this study most will be set constant to match industry standards (see section 2.4.1) and will not be individually referenced in later chapters. However, the Barten model as a whole is integral to this work and an understanding of how it is formulated may be beneficial to the reader.

Barten's model considers several factors that affect vision, such as noise sources, filtering and processing by the eye, and orientation. Noise sources include internal or neural noise, photon noise, and external or image noise. The human visual system's processing is modeled by the limited integration capability, optical modulation transfer function, lateral inhibition, and temporal filtering[18].

"Lateral inhibition" refers to the attenuation of the luminance signal and added photon noise at low spatial frequencies. This phenomenon occurs in the neural system and has been experimentally verified in previous studies by Enroth-Cudgell & Robson (1966)[1]. Temporal filtering is developed in a similar manner and its use has some resemblances to models by Burbeck & Kelly (1980) and Fleet et al. (1985)[1]. The temporal and spatial frequency components of Barten's model can be treated separately, allowing for the exclusion of one or the other in a given implementation. For the purpose of this research, temporal filtering effects are not included.

The resulting equation for contrast sensitivity as a function of spatial frequency is shown in Eq. (2.6).[18]

$$S(u) = \frac{1}{k}\sqrt{\frac{T}{2}}\frac{M_{opt}(u)}{\sqrt{\left(\frac{1}{\eta h I_L} + \frac{\Phi_0}{(1-F(u))^2} + \Phi_{ext}(u)\right)\cdot\left(\frac{1}{X_0^2} + \frac{1}{X_E^2} + \left(\frac{u}{N_E}\right)^2\right)}} \qquad (2.6)$$

The optical modulation transfer function, $M_{opt}(u)$, is given by Eq (2.7), with u being the spatial frequency in cycles/degree of viewing angle. σ_0 is the optical MTF at small pupil sizes. C_{sph} is the spherical aberration, which is dependent on the pupil diameter d.

$$M_{opt}(u) = e^{-\pi^2\cdot\sigma^2\cdot u^2} \qquad \sigma = \sqrt{\sigma_0^2 + \left(C_{sph}\cdot d^3\right)^2} \qquad (2.7)$$

The pupil diameter in millimeters is given by Eq. (2.8)[18]. L is the average luminance of the target in cd/m^2.

$$d = 4.6 - 2.8\cdot\tanh\left(0.4\cdot\log_{10}(0.625\cdot L)\right) \qquad (2.8)$$

The illuminance of the eye, I_L, is given in trollands [td] by Eq. (2.9).

$$I_L = \frac{\pi}{4}d^2 L \qquad (2.9)$$

"Photon noise" is defined by the fluctuations of the photon flux h (photons/td·deg·deg^2), the pupil diameter d (mm), and quantum detection efficiency η of the eye. The temporal and spatial integration limits of the eye are represented by a time constant T (sec), a maximum angular size X_E (deg) for a square target, and the maximum number of cycles N_E.

$$\left(1-F(u)\right)^2 = 1 - e^{-u^2/u_0^2} \qquad (2.10)$$

Eq. (2.10) gives the low frequency attenuation of neuron noise due to lateral inhibition[18].

The Barten model has been compared to a large number of published studies measuring the contrast sensitivity of the eye and has been found to agree well with them. These studies used either CRTs or projection devices. There have been only two contrast perception studies since that have utilized LCD devices[19, 20]. Full details of the Barten Model and how it was developed are given in "*Contrast Sensitivity of the Human Eye and its Effects on Image Quality*", by Peter G. J. Barten[1].

2.3 Display Characteristics

2.3.1 Grayscale display values

Radiographs are typically displayed electronically as black and white images, with exposure values being represented by gray levels. Computer display systems are commonly understood to have an 8-bit grayscale, which provides 256 gray levels. For a radiographic image, image data is translated to luminance by mapping each of the image pixel values to one of the possible 256 gray levels (0-255). How the image values are translated can be modified by assigning window and level parameters. The "window" value is the total number of image values that will be displayed. The "level" is the image value that will be used as the median display value, and thus will have a number of image values above and below it each equal to half of the window parameter. Those image values within the range of the window will be evenly mapped to the 256 gray levels. Any image values above the window will be set to the maximum level (255), while those below the window will be set to the minimum level (0).

Once gray levels have been assigned, the display system maps each of them to a Digital Driving Level (DDL), which is used as input by the graphic controller to drive the display and produce luminance. For most display systems, the DDL values are described by a set of three sub-pixel color values, red (R), green (G), and blue (B), each of which can be driven at 256 different luminance levels. Typically, the gray levels are assigned to the 256 "true gray" DDLs for which the sub-pixels are driven at equal levels (R=G=B). This mapping of gray levels can be modified using a lookup table (LUT) to transform the DDL values sent to the graphics controller. This will be discussed further in Chapter 3.

2.3.2 Luminance response and ambient luminance

Once an image is translated into display values, the resulting output is still dependent on the luminance characteristics of the display device. The "luminance response" of a display describes the amount of luminance produced at each individual driving level. For CRTs, the luminance response follows a smooth curve as the driving voltage to the tube is increased and is simple to characterize. For LCDs, the luminance response is not smooth, with the rate of luminance change rising and falling erratically while stepping through the driving levels[21]. In addition, the LCD luminance response drops off sharply with viewing angle[22, 23], although this is being improved upon by

manufacturers. Because of these factors, more effort is required to accurately characterize an LCD. Methods for measuring the luminance response of an LCD are discussed in depth in Chapter 3.

The maximum luminance and minimum luminance of the display are particularly important in determining what contrast can be produced and how the final displayed image will be perceived. Typical consumer color LCD displays have a maximum brightness around 250 cd/m^2. Newer models are increasing in brightness, to 300 cd/m^2 and higher. Medical LCD displays, designed to produce high brightness grayscale images, can easily reach as high as 500 cd/m^2, if not higher. These displays typically retain the same basic pixel structures as their consumer counterparts, i.e., sets of RGB triplets, but with the color filters removed to improve brightness and contrast. Monochrome displays are often preferred for medical viewing because they can achieve much higher brightness than similar color displays. However, as will be demonstrated in Chapter 3, color displays still provide certain advantages even when dealing only with grayscale images.

Most displays are capable of producing very low minimum luminance. The minimum luminance of LCD displays can be limited by the luminance produced by the backlight, but even these can typically go as low as 0.2-0.5 cd/m^2. However, environmental sources of luminance, such as room lighting, will illuminate the display and add to the minimum luminance. The environmental light striking the surface of the display is referred to as "illuminance", and the added brightness that is reflected back towards the observer is the "ambient luminance". Ambient luminance can range from 0.1 cd/m^2 for a very dark environment such as a radiology reading room, to 1 cd/m^2 or more for a well lit area such as an office. Display measurements should take into account ambient luminance, which adds to the display luminance at all levels. The affect of ambient luminance is most noticeable at low luminance levels, where it is a much more significant portion of the total luminance.

2.3.3 Viewing distance and pixel size

The viewing of medical images is generally done at distances of 30-60cm. A viewing distance of around 60cm is recommended in order to reduce eye strain. This is based on the concepts of "convergence" and "accommodation". Convergence (or

vergence) refers to the rotating of the eyes inward in order to focus an object at the same place on each retina. The closer the object, the more the ocular muscles converge the eyes inward towards the nose. Accommodation refers to the changing of the shape of the lens capsule in the eye in order to focus on a close object. Both convergence and accommodation have resting points, or default viewing distances that they focus at when there is nothing to focus on. Viewing objects at distances closer than the resting point increases eye strain, while viewing from farther away does not, although visual resolution is of course reduced. The resting point of vergence (RPV) averages about 114 cm looking straight ahead, and 89cm at a 30° downward viewing angle[24]. The resting point of accommodation (RPA) has been found to be about 58-67cm[25, 26]. Research has shown that these resting points will tend to shift closer after working at near distances, and that the stress of convergence contributes more to eye strain than the stress of accommodation[27].

As explained in Section 2.2.2, the contrast sensitivity of the human visual system is dependent on spatial frequency. The distance at which the image is viewed affects the number of cycles per degree perceived by the eye, and different spatial frequencies are more visible at different viewing distances. In addition, the spatial frequencies in a displayed image are limited by the size of the pixels. In order to minimize noise in an image due to the pixel structure in a display, it is recommended that pixels sizes be small enough that the Nyquist frequency of the pixel pitch is above the resolution limit of the human visual system. At a viewing distance of 60cm, this value is about 0.250mm[2].

2.4 Medical Display Standards

2.4.1 DICOM PS 3.14

The Digital Imaging and Communications in Medicine (DICOM) standard was developed by the American College of Radiology (ACR) and the National Electrical Manufacturers Association (NEMA)[18]. While much of the standard deals with how digital image data can be transferred between systems, Part 3.14 provides specifications for how image pixel values should be interpreted and displayed in order to provide visual consistency in how that image appears on different display devices. The focus of this document is the Gray Scale Display Function (GSDF), which is the common standard

currently used for calibrating the grayscale of a medical monitor. This standard is based on the contrast sensitivity of the human visual system when viewing a Standard Target. A Standard Target is defined by DICOM PS 3.14 as follows:

> "A 2-deg x 2-deg square filled with a horizontal or vertical grating with sinusoidal modulation of 4 cycles per degree. The square is placed in a uniform background of a Luminance equal to the mean Luminance of the Target."

Figure 2.9: Sample target similar to a Standard Target as defined by DICOM.

The GSDF provides a set of calibrated gray levels referred to as "P-values", and are designed to produce changes in luminance that the average human observer can just perceive. In order to facilitate this, DICOM PS 3.14 introduces the "Just Noticeable-Difference" (JND) index. Each JND index value has a luminance associated with it, and each step in the index represents a change in luminance equal to C_T, with the luminance of the previous index value used as the average scene luminance. Intermediate values are determined by a mathematical interpolation based on 1023 JNDs.

Generating the GSDF involves first measuring the luminance produced by the available grayscale DDLs. The desired luminance range for the display is chosen and mapped to an equivalent range in the JND index. For an 8-bit grayscale, the available JNDs are divided evenly into 256 values. DDLs are selected and arranged to produce

luminance values that most closely match the 256 JND values. These DDLs are then assigned to the 256 P-values that make up the GSDF. The GSDF uses the Barten definition of contrast sensitivity with the following constants:

C_{sph} = 0.0001 deg/mm^3	(pupil spherical aberration)
σ_0 = 0.133 deg	(small pupil optical MTF constant)
k = 3.3	(signal to noise ratio required for detection)
η = 0.025	(quantum efficiency of the eye)
h = 357·3600 photons/(td sec deg^2)	(photon conversion factor)
Φ_0 = 3·10^{-8} sec deg^2	(contrast variance from neuron noise)
Φ_{ext} = 0 sec deg^2	(contrast variance from external noise)
T = 0.1 sec	(temporal integration)
X_E = 12 deg	(maximum angular integration size)
N_E = 15 cycles	(maximum cycle integration)
u_0 = 8 cycles/deg	(lateral inhibition frequency constant)

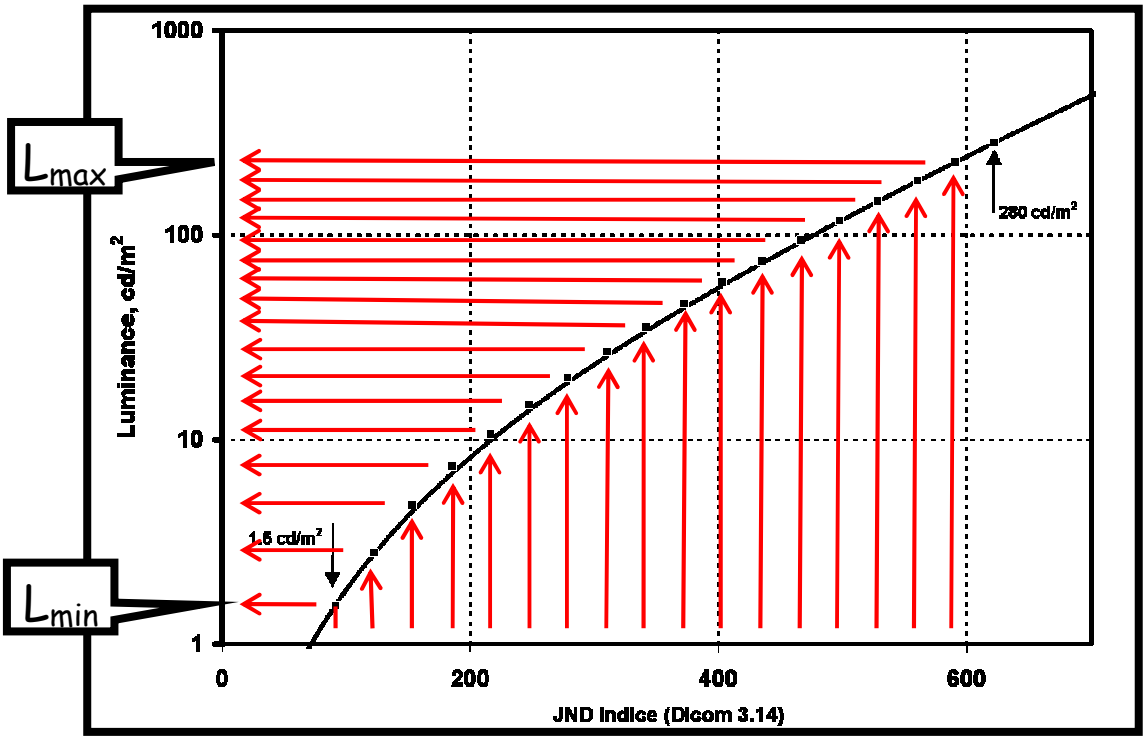

Figure 2.10: Example of the GSDF shown as the luminance corresponding to the JND index.

While the DICOM GSDF has become a common standard in medical imaging, it has certain limitations that should be noted. First, the DICOM GSDF relies on variable adaptation states, rather than fixed adaptation. The DICOM luminance response is derived from contrast threshold experiments where the background luminance is changed to equal the luminance of the target pattern, and the observer fully adapts to the new background. As was shown in Figure 2.7, the contrast response for fixed adaptation is much different than for variable adaptation. Thus, while the DICOM GSDF is a good predictor of the minimum contrast changes needed near the average luminance of an image, it does not accurately reflect the contrast threshold response for a mixture of bright and dark regions that vary significantly from the average luminance. Secondly, the GSDF reflects visual performance under highly specific conditions defined by the Standard Target and the threshold contrast. For medical images, the features of interest will vary in size and spatial frequency, will often have suprathreshold contrast, and will be viewed with differing levels and types of background noise[7].

For these reasons, the DICOM GSDF does not represent an optimal calibration for observing the features of any particular image. Rather, the GSDF provides a specific grayscale transformation that improves image presentations on display systems and produces similar appearance on all display systems that are GSDF-calibrated.

2.4.2 AAPM OR-03

The American Association of Physicists in Medicine On-line Report No. 03 (AAPM OR-03), also known as AAPM Task Group 18 (AAPM TG-18), provides guidelines and acceptance criteria for acceptance testing and quality control of medical display devices[7]. It includes visual, quantitative, and advanced testing methodologies for various display characteristics such as reflection, geometric distortion, luminance, resolution, noise, glare, and display artifacts. These methods have been tested in clinical environments using LCD devices[28]. Many of the tests and measurements done for this research study are based on the setup and assessment guidelines of the AAPM OR-03 report. Of particular interest is Appendix II: Equivalent Appearance in Monochrome Image Display. This section discusses methods for maintaining perceived information across different displays with different luminance characteristics. A key point is the

maintaining of a constant luminance ratio for all displays. This issue will be discussed further in Chapter 3.

The AAPM OR-03 document also provides an overview and discussion of the DICOM GSDF similar to what was presented in section 2.4.1. It also notes that the DICOM GSDF is not the only standard for display calibration in use. Some European imaging centers use the CIELAB function suggested by the International Illumination Commission. However, the AAPM OR-03 recommends the use of the DICOM GSDF for consistency among imaging centers.

2.5 References

1. Barten, P.G.J., *Contrast sensitivity of the human eye and its effects on image quality*. 1999, Bellingham, Wash.: SPIE Optical Engineering Press. xix, 208.
2. Flynn, M., *Visual requirements for High-Fidelity Display.* Advances in Digital Radiography: RSNA Categorical Course in Diagnostic Radiology Physics 2003, 2003: p. 103-107.
3. Stiles, W.S. and B.H. Crawford, *The Luminous Efficiency of Rays Entering the Eye Pupil at Different Points.* Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character, 1933. **112**(778): p. 428-450.
4. Strauss, O., *The Retinal Pigment Epithelium in Visual Function.* Physiol. Rev., 2005. **85**(3): p. 845-881.
5. Kelly, D.H., *Visual Contrast Sensitivity.* Optica Acta, 1977. **24**(2): p. 107-129.
6. Barten, P.G.J. *Physical model for the contrast sensitivity of the human eye*. 1992: SPIE.
7. Samei, E., et al., *Assessment of display performance for medical imaging systems: executive summary of AAPM TG18 report.* Med Phys, 2005. **32**(4): p. 1205-25.
8. Normann, R.A. and I. Perlman, *The effects of background illumination on the photoresponses of red and green cones.* J Physiol, 1979. **286**: p. 491-507.
9. Baxter, B., H. Ravindra, and R.A. Normann, *Changes in lesion detectability caused by light adaptation in retinal photoreceptors.* Invest Radiol, 1982. **17**(4): p. 394-401.
10. Flynn, M.J., et al., *High-fidelity electronic display of digital radiographs.* Radiographics, 1999. **19**(6): p. 1653-69.
11. Hecht, S., *The relation between visual acuity and illumination.* J. Gen. Physiol., 1928. **11**(3): p. 255-281.
12. Baylor, D.A. and M.G. Fuortes, *Electrical responses of single cones in the retina of the turtle.* J Physiol, 1970. **207**(1): p. 77-92.
13. Owsley, C., R. Sekuler, and D. Siemsen, *Contrast sensitivity throughout adulthood.* Vision Res, 1983. **23**(7): p. 689-99.
14. Reimann, D.A., M.J. Flynn, and J.J. Ciarelli. *System to maintain perceptually linear networked display devices*. in *Medical Imaging 1995: Image Display*. 1995. San Diego, CA, USA: SPIE.
15. Blume, H.R., S.J. Daly, and E. Muka. *Presentation of medical images on CRT displays: a renewed proposal for a display function standard*. in *Medical Imaging 1993: Image Capture, Formatting, and Display*. 1993. Newport Beach, CA, USA: SPIE.
16. Barten, P.G.J. *Spatiotemporal model for the contrast sensitivity of the human eye and its temporal aspects*. in *Human Vision, Visual Processing, and Digital Display IV*. 1993. San Jose, CA, USA: SPIE.
17. Daly, S.J. *Visible differences predictor: an algorithm for the assessment of image fidelity*. in *Human Vision, Visual Processing, and Digital Display III*. 1992. San Jose, CA, USA: SPIE.
18. NEMA, *Digital Imaging & Communications In Medecine (DICOM), Part 14: Grayscale Standard Display Function*. 1998, National Electrical Manufacturers Association.

19. Sund, P., et al. *A comparison between 8-bit and 10-bit luminance resolution when generating low-contrast sinusoidal test pattern on an LCD*. in *Medical Imaging 2007: Image Perception, Observer Performance, and Technology Assessment*. 2007. San Diego, CA, USA: SPIE.
20. Tchou, P., M. Flynn, and E. Peterson, *2AFC assessment of contrast threshold for a standardized target using a monochrome LCD monitor*. Proceedings of SPIE, 2004. **5372**(Medical Imaging 2004: Image Perception, Observer Performance, and Technology Assessment): p. 344-352.
21. Fetterly, K., et al., *Introduction to Grayscale Calibration and Related Aspects of Medical Imaging Grade Liquid Crystal Displays.* Journal of Digital Imaging, 2007.
22. Badano, A. and D.H. Fifadara, *Goniometric and conoscopic measurements of angular display contrast for one-, three-, five-, and nine-million-pixel medical liquid crystal displays*. 2004, AAPM. p. 3452-3460.
23. Samei, E. and S.L. Wright. *Effect of viewing angle response on DICOM compliance of liquid crystal displays*. in *Medical Imaging 2004: PACS and Imaging Informatics*. 2004. San Diego, CA, USA: SPIE.
24. Harisinghani, M.G., et al., *Importance and Effects of Altered Workplace Ergonomics in Modern Radiology Suites.* Radiographics, 2004. **24**(2): p. 615-627.
25. Leibowitz, H.W. and D.A. Owens, *Anomalous Myopias and the Intermediate Dark Focus of Accommodation.* Science, 1975. **189**(4203): p. 646-648.
26. Leibowitz, H.W. and D.A. Owens, *New evidence for the intermediate position of relaxed accommodation.* Documenta Ophthalmologica, 1978. **46**(1): p. 133-147.
27. Owens, D.A. and K. Wolf-Kelly, *Near work, visual fatigue, and variations of oculomotor tonus.* Invest. Ophthalmol. Vis. Sci., 1987. **28**(4): p. 743-749.
28. Samei, E., et al. *Clinical verification of TG18 methodology for display quality evaluation*. in *Medical Imaging 2003: Visualization, Image-Guided Procedures, and Display*. 2003. San Diego, CA, USA: SPIE.

CHAPTER 3

A METHOD FOR THE CHARACTERIZATION AND CALIBRATION OF MEDICAL LIQUID CRYSTAL DISPLAYS

3.1 Introduction

LCD display devices for medical imaging come in a variety of sizes and configurations. The maximum resolution, color depth, and luminance response may differ amongst displays, and thus the appearance of an image will differ if the displays are not properly calibrated. Most medical images are shown using a monochrome presentation with 256 gray levels. For consistent appearance, the display grayscale (i.e., the luminance versus gray level relationship) and the luminance ratio (i.e., the ratio of the maximum luminance to the minimum luminance) need to be the same. Chapter 2 described a standardized grayscale, the DICOM Grayscale Standard Display Function (GSDF), which is now widely used in medicine. Display devices whose grayscale follows this function produce smoothly varying contrast for incremental gray level steps, as shown in Figure 3.1.

In general, LCD computer monitors have a grayscale that is unsuitable for presenting medical images. Typically, the dark levels are too black and the mid-gray levels are too bright. Additionally, the luminance response is often not a smoothly varying function. For medical devices, grayscale calibration involves transformations done by the graphic adapter or, for specialized medical devices, within the LCD device. This chapter presents a method to calibrate medical monitors in order to achieve equivalent grayscale response and luminance range.

The calibration method presented involves the measurement of large gray palettes (i.e., the set of gray toned luminance states that can be set for a display), the generation of a calibrated look-up table (LUT), and verification of the calibrated luminance response. Small perturbations of the red, green, or blue sub-pixel values in the display are used to

achieve precise luminance changes between successive gray levels[1, 2]. The maximum and minimum luminance of the gray scale, L_{max} and L_{min}, are specified to achieve a desired luminance ratio, L_{max}/L_{min}, in the presence of ambient illumination. While commercial medical calibration methods have used similar sub-pixel perturbation methods, prior methods have not constrained the luminance ratio.

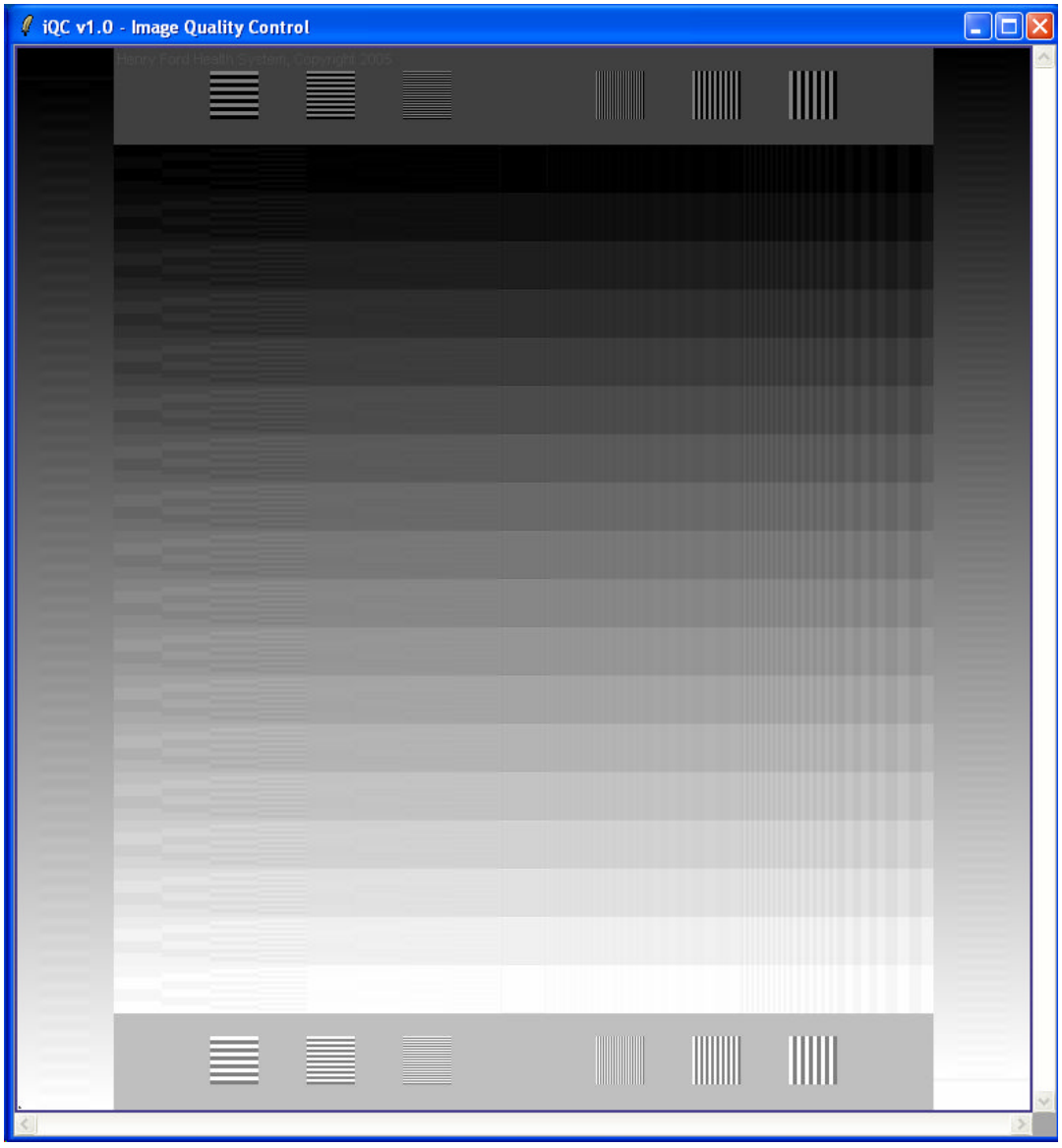

Figure 3.1: Quality control test image used to visually evaluate the grayscale contrast response of a medical imaging display device. Bar patterns are displayed with increasing brightness. Within each step are low contrast bar patterns with varying spatial frequencies. At the right and left are gray level ramps with sinusoidal modulation. For a properly calibrated display, bar patterns should be perceived as having similar contrast at all levels of brightness.

To date, most grayscale calibrations have been done using specialized medical software and equipment with proprietary medical graphic cards. In order to calibrate general purpose computer monitors, the methods presented in this chapter have been incorporated in a set of software tools referred to collectively as "pacsDisplay". The pacsDisplay software is intended to provide a simple interface between an IL1700 luminance meter, the DICOM GSDF, and a Windows operating system. Its computation functions are minimal, primarily involving data storage and table-lookups in accordance with formulas provided in the DICOM PS 3.14 document[3]. The majority of pacsDisplay's development focused on designing an efficient graphical user interface (GUI) for picture archiving and communication system (PACS) administrators in radiology departments. It includes tools to implement each step of the DICOM calibration process, utilities to build specialized or generic LUTs to suit a variety of needs, and an extensive library of LUTs for various makes and models of LCD monitors. The pacsDisplay software package was built using TCL/TK and Visual C++. It relies on standard function calls from the Windows Software Development Kit (SDK), thus making it applicable for use with any Windows-based system with a Windows-compliant graphics card. The current instruction manual for pacsDisplay is included in the Appendix at the end of the dissertation. Those interested in the software can request a copy by sending an email to pacsDisplay@rad.hfh.edu.

Initial testing of the software's functions was done by calibrating a number of different LCD devices according to the DICOM standard. Later versions were field-tested on LCD systems deployed at Henry Ford Hospital in Detroit, MI and the Mayo Clinic in Rochester, MN. These tests are detailed later in this chapter. The pacsDisplay software has been made publicly available and is currently in use by Henry Ford Health System (HFHS) and the Mayo Clinic to provide calibrated grayscales on monitors in Radiology, referring physician offices, and surgery.

Although the pacsDisplay software is important in terms of PACS development, for purposes of this dissertation, the primary function of pacsDisplay is to provide a means for measuring and controlling luminance produced by a medical LCD monitor. While relatively simple, these procedures are essential for the experimental research conducted in this dissertation. The pacsDisplay software package provides a tested and

streamlined means of implementing these procedures. This chapter explains and validates these methods in the context of the DICOM GSDF. Later chapters will use pacsDisplay's functions for non-standard measurements and calibrations in order to conduct human psychovisual research.

This chapter reports on the luminance characteristics of 94 LCDs of 10 different models. These monitors come from HFHS and the Mayo Clinic. The pacsDisplay software is used to calibrate these devices and the results are evaluated against the DICOM GSDF and criteria set down in the AAPM OR-03 report. In addition, this chapter examines luminance variability between several different display types and among those of the same make and model. The results demonstrate that the pacsDisplay software can accurately measure luminance response data for LCDs and effectively implement a DICOM GSDF calibration in accordance with AAPM OR-03 standards. In addition, newer display models are shown to have similar luminance responses, allowing for the use of common LUTs built from averaged luminance response data. Much of this chapter has also been previously reported in a conference proceeding article[4].

3.2 Methods

3.2.1 Measuring the luminance response

Grayscale medical monitors are commonly calibrated by transforming the image values sent to a graphic controller using a "lookup table" (LUT). The LUT references each of the original display values to ones that are adjusted for proper display calibration. The calibration LUT is deduced from the "uncalibrated luminance response" (uLR) of the display system. First the uLR is measured, providing a list of luminance values that the display can produce. Next, the LUT is built by matching these values to the expected DICOM GSDF calibration curve[3] (see Chapter 2 for details on the DICOM GSDF). Finally, the "calibrated luminance response" (cLR) is measured to verify the calibration.

The uLR for a CRT has a well-behaved functional shape for which the slope changes gradually with the controller output value. In relation to the driving voltage (V), the luminance response follows a power-law relationship of the form

$$L = L_{\min} + \left(L_{\max} - L_{\min}\right)V^{\gamma}. \qquad (3.1)$$

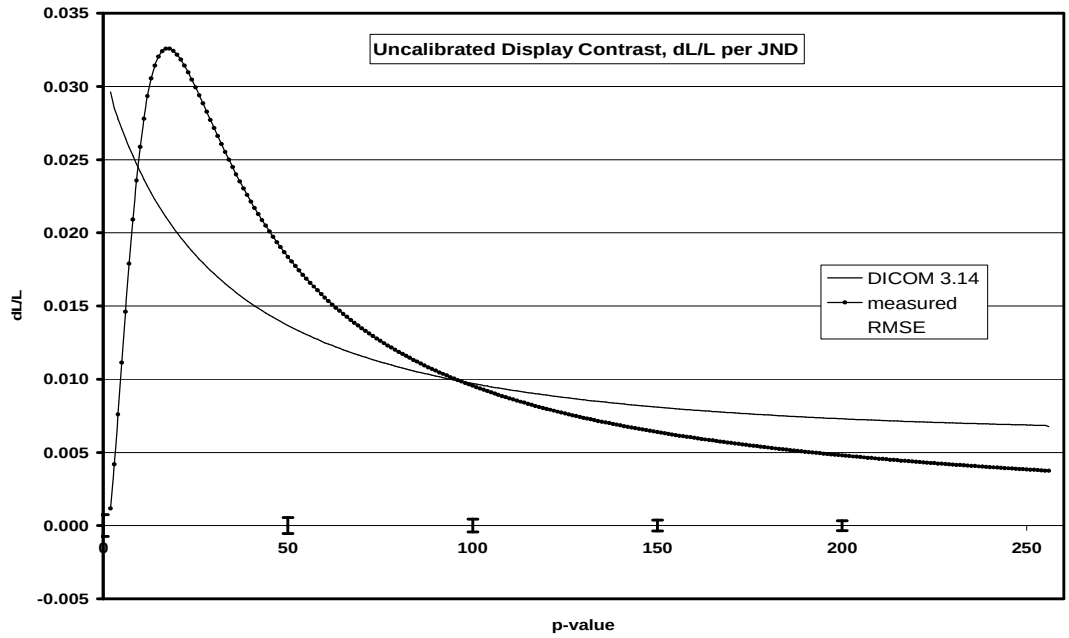

Figure 3.2: Simulation of the contrast from a CRT display (gamma = 2.2) in relation to the GSDF.

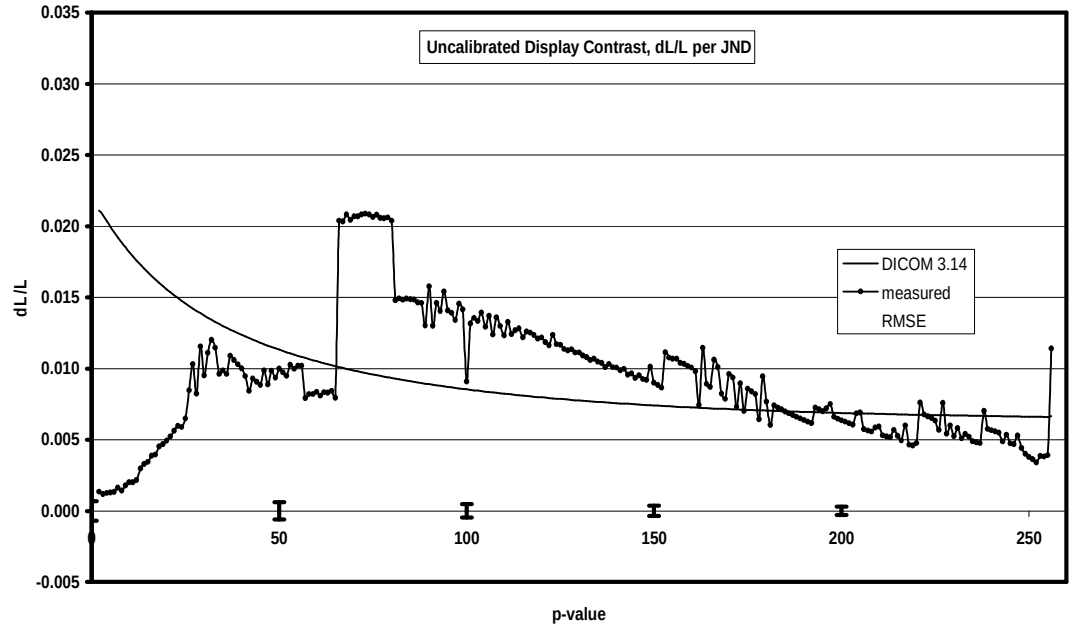

Figure 3.3: Uncalibrated contrast response from an NDS Nova 3MP LCD display in relation to the GSDF.

The value of gamma (γ) is determined by the design of the CRT electron gun and may vary from 1.4 to 2.4[5, 6]. The uLR for a CRT monitor can be easily estimated by fitting

such a curve to a small number of measured points. The calibration LUT is then obtained from the fitted curve and can be applied as a gamma-correction by the graphic controller. Because of the analog nature of the device, the resulting contrast response (i.e., $\Delta L/L$ vs. gray level) is a smooth function. The Digital to Analogue Converter (DAC) in a CRT controller determines the ability to precisely define the cLR. Conventional controllers with 8-bit DACs have limited control over the slope ($\Delta L/L$). Specialized controllers intended for medical applications will often use 10 or 12 bits to more precisely produce the intended luminance at each display level[6].

Unlike CRT systems, the uLR of an LCD panel typically has an irregular shape. Significant discontinuities occur in the relative luminance change associated with each change in the controller output value[7]. Figures 3.2 and 3.3 present very different contrast curves for a CRT power-law response and an uncalibrated LCD, respectively. An accurate grayscale calibration for LCD devices thus requires a full measurement of the luminance for the entire palette of possible output values.

LumResponse has three modes for measuring the luminance response: 256, 766, and 1786. In 256 mode, LumResponse sequentially sets each of the standard gray levels in increasing order. In an uncalibrated system, these would correspond to the "true gray" DDLs for which the red (R), green (G), and blue (B) pixels are set equal ($R=G=B$). This uncalibrated state is also referred to as a "linear calibration" since the RGB values advance linearly from 0 to 255 without perturbation. The 256 mode is not intended to measure the uLR since a calibration LUT derived from a palette of only 256 values will typically have redundant entries and an unacceptable calibrated response. Instead, this mode is used to obtain the cLR at the end of the calibration process and check the results. The two other modes, 766 and 1786, are used to obtain a much larger palette for use in building a calibration LUT.

For monochrome displays, a luminance response following the DICOM standard can be obtained by selecting 256 gray levels from a palette of 766 states. To measure a palette of 766 gray values, each true gray value from the primary sequence is modified by a sub-sequence of values from a perturbation table having 3 entries. These are three-digit numbers from the set (000, 100, 010, 001, 110, 101, 011). The gray value for each sub-sequence entry is computed by adding the perturbation value to the current true gray

value. For example, at a true gray sequence value of 200, the sub-sequence entry 010 would indicate that R and B are set to 200, while G is set to 201. The sub-sequence values used for this work were ordered to produce increasing luminance. For presently available monochrome LCD monitors, each pixel has three sub-pixels with the RGB color filters removed. In such devices, perturbations of 001, 010, and 100 produce an identical luminance change. Similarly, perturbations of 110, 101, and 011 are equivalent. This was experimentally tested for the monitors evaluated as a part of this work. Since the luminance response tends to be locally linear, perturbations involving changes of 2 or more are redundant with the above set. For example, a perturbation of 020 causes the same luminance change as 011. Thus, there are three unequal perturbations for each of the 256 true gray sequence values for a total of 766 different gray values. This perturbation series is shown in Table 3.1.

For a color LCD monitor, contributions to luminance from the R, G, and B sub-pixels are not equal. This property allows for calibration using a significantly larger palette[1, 2]. The three colored sub-pixels influence luminance in the amount of approximately .24 for Red, .65 for Green, and .11 for Blue as measured for a Dell 1900 FP LCD monitor. Therefore, each perturbation value from the set of seven defined above will produce a different luminance change. The values for other monitors will vary depending on the design of the color filters and backlight. In addition, measurements have shown that the total luminance produced by a particular RGB combination is sometimes greater than the sum of the R, G, and B luminance values when driven separately. This may be due to the way that the sub-pixel structures are designed, possibly resulting in some amount of voltage or light leakage between them. Therefore, one must still measure each of the proposed gray levels individually to determine their exact luminance value.

Table 3.1 shows a sub-sequence of seven perturbations that produces increasing luminance for color monitors. The palette of different luminance values for a color monitor has 256 primary values with a sub-sequence set of 7 for a total of 1786 different luminance values. The slight chrominance perturbation associated with the sub-sequence is thought to be insignificant.

Table 3.1: RGB sub-sequence steps for 766 and 1786 gray value measurements

Monochrome - 766 Gray Values

Step	R	G	B
1	0	0	0
2	1	0	0
3	0	1	1

Color - 1786 Gray Values

Step	R	G	B
1	0	0	0
2	0	0	1
3	1	0	0
4	1	0	1
5	0	1	0
6	0	1	1
7	1	1	0

3.2.2 Automated luminance measurements

For this work, luminance measurements were made with an International Light IL1700 research radiometer having a serial line output that reports luminance in cd/m^2 every 0.5 seconds. A high gain silicon photodiode detector (IL SHD033), photopic filter (ILY filter), and cylindrical collimator (IL H hood and K9 aperture) provided low noise measurements (see Figure 3.4). All measurements were made with room lights turned off and a dark cloth covering the probe and monitor (see Figure 3.5). Unlike typical flat luminance detectors (often referred to as "pucks" because of their flat, round shape), this meter has a narrow response solid angle. Since the luminance response of an LCD display changes with viewing angle, a narrow solid angle probe was used to specifically measure the luminance at a normal angle from the surface.

As LumResponse sequentially changes the test pattern, luminance is automatically measured for each palette state. A time delay is specified to allow the radiometer to stabilize before taking the next measurement. The software also includes the option to average several measurements in order to improve noise. Occasionally, luminance measurements that were anomalously high or low compared to previous measurements were observed. These were associated with certain luminance values changes where the radiometer adjusts its sensitivity. A provision was included to detect such outliers, log the occurrence, and repeat the measurement. If the outlier persists, LumResponse generates an error. Luminance values are stored in memory during data acquisition and then written to a text file at the end of the sequence. The minimum measurement time was approximately 4 minutes for 256 gray values, 12 minutes for a 766 palette, and 30 minutes for a 1786 palette.

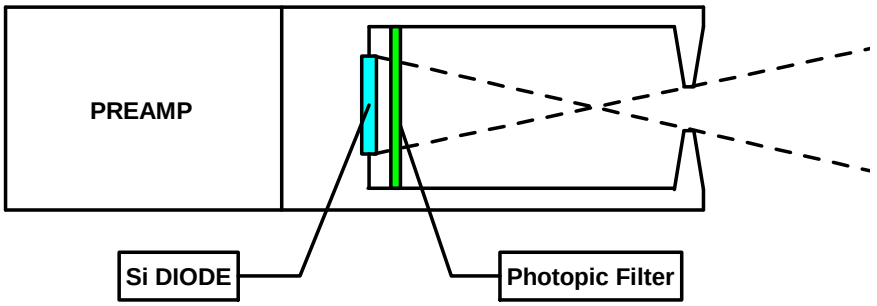

Figure 3.4: IL1700 Light Radiometer with collimator.

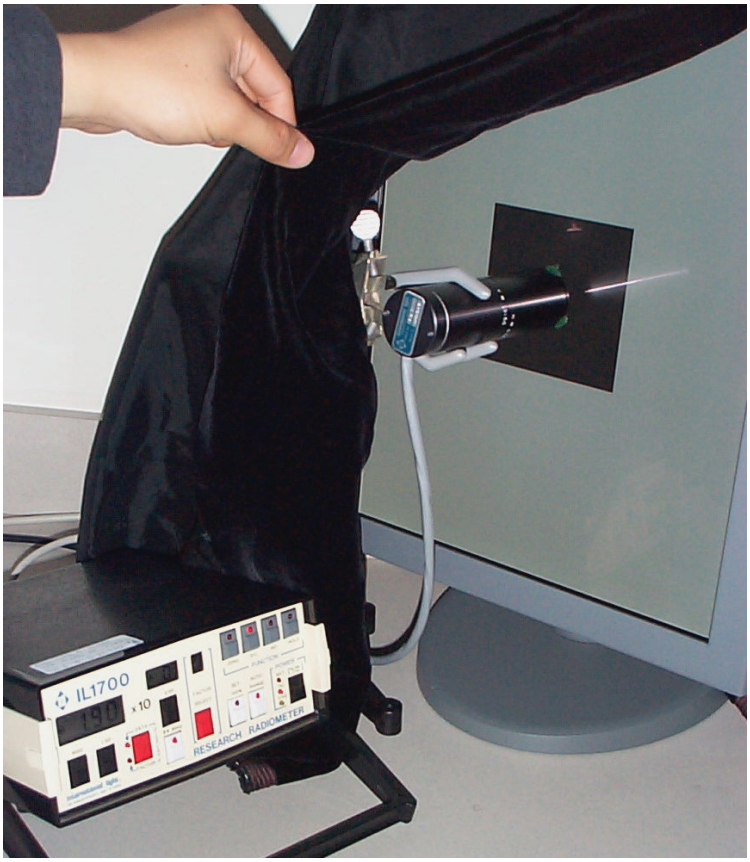

Figure 3.5: Luminance measurements setup with IL1700.

The precision of the probe and radiometer for measurements of small light differences was evaluated by using an integrating sphere light source with a halogen lamp powered by a constant potential voltage source. Four sets of 128 luminance measures were made using the automated software. The average luminance values for each of the sets were 0.482, 7.12, 93.5, and 1111 cd/m^2. The measured data were converted to values of the relative light change between each sequential measurement, $\Delta L/L$. These 127

values of $\Delta L/L$ were used to determine the root mean squared error (RMSE). A line was then fitted to the four RMSE values to find a function to represent the error in the luminance measurements. The resulting error equation was

$$\mathrm{RMSE}(L) = (-7.1102E-05)\ln(L) + 7.0427E-04. \qquad (3.2)$$

The RMSE values were found to be very small within the luminance range that the experiments covered. It was therefore concluded that the precision of the IL1700 was sufficient and any error in the device's measurements would be negligible.

3.2.3 Generating a calibration LUT

The uLR results obtained from LumResponse can be used to produce a calibration LUT for each display evaluated. Another pacsDisplay utility, LUTGenerate, was developed for this purpose. LUTGenerate obtains the parameters needed (maximum luminance, luminance ratio, ambient luminance), reads the display luminance data from a specified uLR text file, and computes the LUT by finding the palette entries that most closely match the 256 desired luminance values. The desired values are determined from the polynomials reported in DICOM PS 3.14 to compute the GSDF (see Chapter 2)[3]. LUTGenerate then records the LUT in a text file for reference and installation. Figure 3.6 shows a diagram illustrating this process.

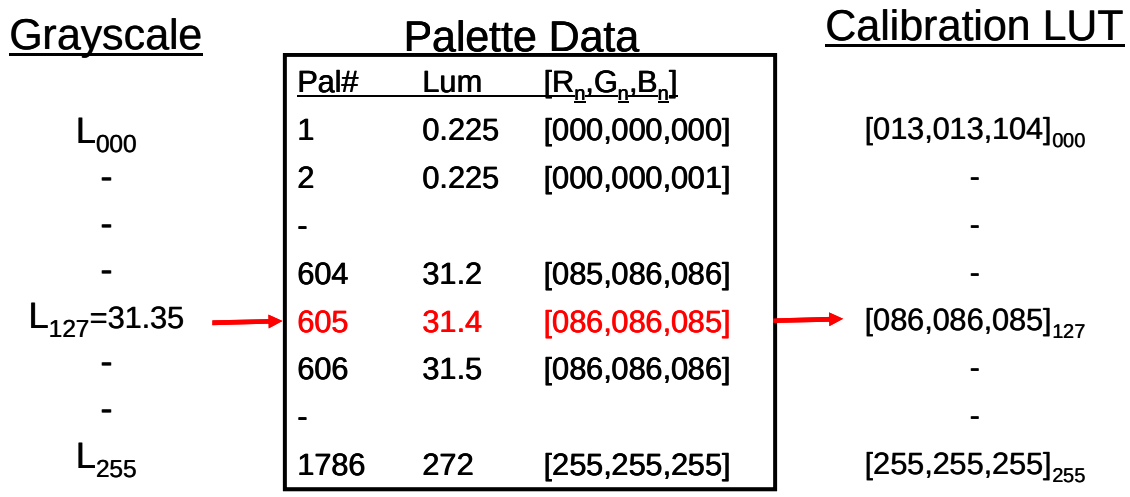

Figure 3.6: Steps for generating a calibration LUT. The left side represents the desired grayscale luminance values divided into 256 steps. The available palette entries are displayed in the center, and are selected to provide the best match with the grayscale entries. The right side represents the 256 RGB entries, selected from the palette data, that make up the calibration LUT.

Before building the LUT, the palette entries must be adjusted for ambient luminance so that they correctly represent the perceived brightness in a given environment. The ambient luminance value is added to all of the measured values, thus shifting the palette upwards. Incorrect specification of the ambient luminance can alter the desired response and change contrast significantly in the dark portions of images. For this work, ambient luminance measurements were taken with a Minolta LS-110 spot luminance meter. Measurements were taken at a distance of 1 meter from the surface of each display with the room lights on and the display turned off.

The maximum luminance of an LCD display depends on the backlight used. For medical monitors, the backlight is typically stabilized so that a constant maximum luminance is maintained over the usable life of the device. For unstabilized monitors, generation of a new LUT is required as the backlight dims with age. For this work, the maximum calibrated luminance was set the same as the maximum measured luminance.

As mentioned in Chapter 2, the human visual system adapts to the overall brightness when viewing an image and perceives contrast within a relatively narrow luminance range[8]. For devices with similarly calibrated gray scales, if the luminance ratio (L_{max}/L_{min}) is the same, an equivalent appearance will be observed even if the maximum luminance of each display is different[9]. The LUT generation program requests the desired luminance ratio for the cLR. A luminance ratio of 250 is well-matched to the human visual system. For traditional radiological films, this corresponds to a film with an optical density range of 2.4. A somewhat higher luminance ratio, 350, is often employed to achieve darker black luminance with modest contrast reduction in the dark region. This value will typically result in a minimum luminance that is higher than the device minimum. While displayed images may lack the dark black often preferred in graphic arts systems, the dark portions of a displayed image will be able to depict subtle contrast. Additionally, since the viewing angle response of LCD displays is particularly poor at very low luminance values, elevation of the minimum luminance improves the performance of the monitor with respect to the viewing angle response.

For large networks with many displays to maintain, calibrating each LCD display individually can be prohibitively time-consuming. However, testing of recent LCDs has shown similar luminance response characteristics among displays of the same model by

the same manufacturer. By taking an average luminance response from a sample set of a particular make and model of display, a "generic LUT" can be constructed and applied to all systems with that display type. A generic LUT could be easily deployed on a large scale, requiring far less time and effort than building custom calibrations for each system. It would not provide the same level of calibration as a custom LUT, but this may be acceptable for some display systems depending on their function. AAPM OR-03 defines two categories of displays, primary and secondary, based on usage.

- "Primary display systems are those used for the interpretation of medical images. They are typically used in radiology and in certain medical specialties such as orthopedics."

- "Secondary systems are those used for viewing medical images for purposes other than for providing a medical interpretation. They are usually used for viewing images by general medical staff and medical specialists other than radiologists and utilized after an interpretive report is provided for the images."

Use of a generic LUT would likely not be acceptable for primary radiological display systems, but it could provide sufficient calibration for secondary reading displays.

A pacsDisplay utility, ULRstats, was developed to facilitate the creation of generic LUTs. ULRstats reads multiple uLR files and averages the luminance response data. This average uLR data can then be input to LUTGenerate, producing a generic LUT for that display type. Care must be taken to choose a properly representative sample set of displays. The uLRs of the entire sample should be compared to identify those with luminance responses that deviate widely from the rest. Such displays should be removed from the sample set, as they may disproportionately affect the average. They are also likely to require a custom calibration in order to produce an acceptable contrast response, even if the generic LUT proves acceptable for the rest of the displays.

3.2.4 Applying a calibration LUT

The LUTGenerate software computes a calibration LUT, but does not apply it to the system. Initially, this step had to be done by accessing the Windows registry and loading the LUT into specific locations depending on the combination of graphics adapter and display software installed on the system. Early evaluation of the calibration process (described in section 2.4) used this method, but it was unwieldy and highly variable

between different system configurations. It was later learned that one could apply a LUT without having to manually edit the registry by using standard commands from the Microsoft Windows Graphic Device Interface (GDI) library, particularly the SetDeviceGammaRamp function call. A pacsDisplay utility called LoadLUT was developed to take advantage of this library.

LoadLUT was designed to allow users to easily apply and change the LUT for all modern Windows-based systems. Installation options were included to help automate the process of applying a calibration. This was accomplished without the need to tie up system resources with background processes and it can be configured to load a calibration whenever a user logs in. Configuration options were included to allow for the calibration of multiple displays with different LUTs on the same system, along with a configuration utility to assist in the process.

The LoadLUT software also features the ability to read the extended display identification data (EDID) included with all displays, providing information about the monitor such as its model name, serial number, pixel size, and timing and resolution settings. This information, along with a library of collected display calibration data, can be used with special search functions to automatically load LUTs that match the displays attached to a system without having to manually configure for each one. For this purpose, a library of LUT files for a variety of displays was compiled and included in the pacsDisplay package. The library is organized using model names and serial numbers, allowing LoadLUT to use EDID information to guide its search functions and locate appropriate calibration LUTs for a system's monitors. Generic LUTs were also built and are included in the library, providing default LUTs for each of the listed models in case a specific match cannot be found.

3.2.5 Evaluation of the calibrated response

For systems that have a calibration LUT properly applied, the display controller automatically transforms the gray levels to their newly prescribed RGB values. These 256 calibrated gray levels are referred to by DICOM as "P-values". The cLR is obtained by setting LumResponse to 256 mode and measuring the luminance response for each of the P-values. The luminance response data can then be compared to the DICOM and AAPM display standards to evaluate the quality of the calibration.

Regarding the evaluation of the luminance response of a system, AAPM OR-03 recommends the following for the advanced analysis of luminance response[9]:

- "...The measured contrast associated with the luminance difference between each sequential gray level available from the display controller, $\Delta L'_p/L'_p$, should be compared to the expected contrast per JND associated with the DICOM GSDF..."
- "...The contrast noise can then be described by the maximum deviation and the root mean squared error of the observed JNDs per luminance interval values."

Following the AAPM guidelines, the 256 calibrated luminance values are obtained from the cLR (L), adjusted for ambient luminance (L'), then used to compute the contrast response for sequential grey levels ($\Delta L'/L'$). The maximum and minimum luminance values (L'_{max} and L'_{min}) are then used as inputs to the DICOM GSDF to produce the predicted luminance (L'_{DICOM}), contrast ($\Delta L'/L'_{DICOM}$), and JND values. The mean JND per P-value ($<\Delta JND/\Delta p>$) will be of particular use.

$$< \Delta JND / \Delta p > = \frac{JND_{\max} - JND_{\min}}{P_{\max} - P_{\min}} = \frac{JND_{\max} - JND_{\min}}{255} \qquad (3.3)$$

These predicted values can then be compared to the measured values. Figure 3.7 shows a sample plot of $\Delta L'/L'$ compared to $\Delta L'/L'_{DICOM}$. By taking the ratio of these two responses and multiplying them by $<\Delta JND/\Delta p>$, the observed JNDs per luminance interval ($\Delta JND/\Delta p$) can be obtained. A sample plot is shown in Figure 3.8.

The AAPM OR-03 uses $<\Delta JND/\Delta p>$ as a primary benchmark for evaluating calibration quality. The following diagnostic criteria are set down in AAPM OR-03:

- "For primary class display devices, [$<\Delta JND/\Delta p>$] should not be greater than 3.0 to prevent visible discontinuities in luminance from appearing in regions with slowly varying image values."
- "The maximum deviation of the observed JNDs per luminance interval should not deviate from [$<\Delta JND/\Delta p>$] by more than 2.0."
- "The root mean square deviation relative to [$<\Delta JND/\Delta p>$] should not be larger than 1.0. …"

These criteria will be used to evaluate calibrated displays in section 3.3.

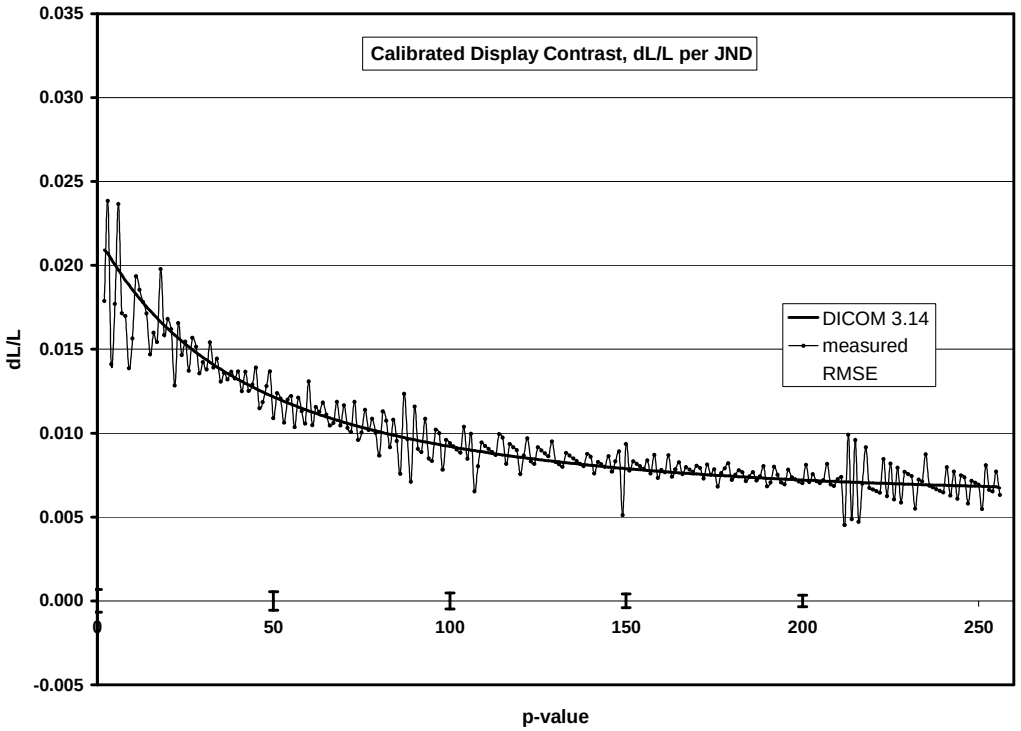

Figure 3.7: Contrast response for a calibrated NDS QXGA Color LCD compared to the contrast predicted by the DICOM GSDF.

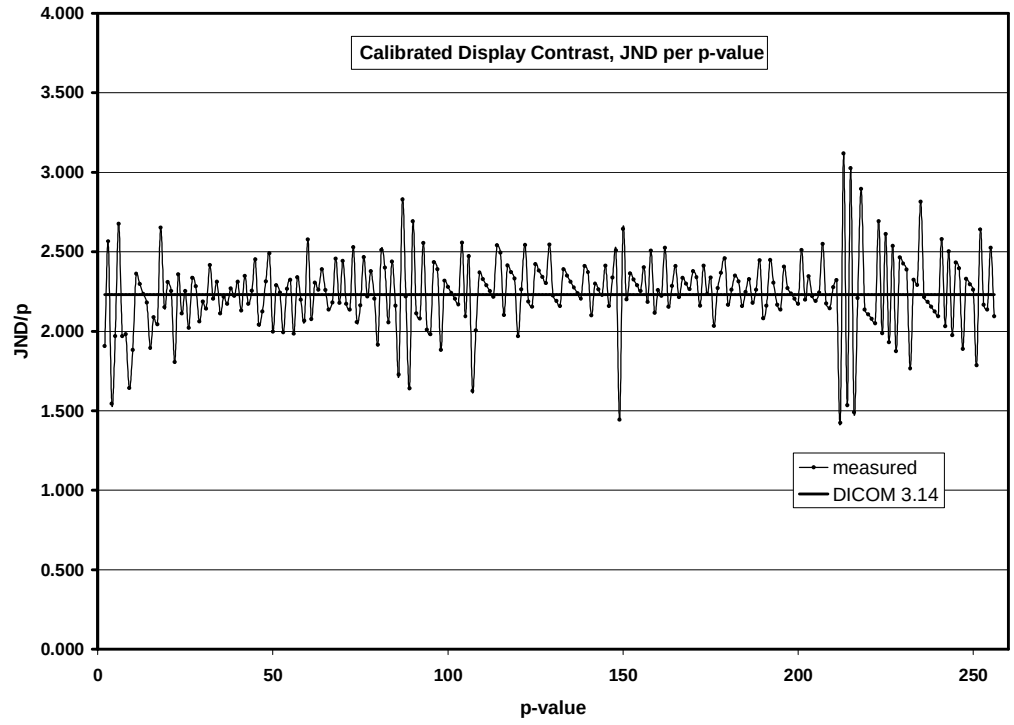

Figure 3.8: Observed JND per P-value for a calibrated NDS QXGA Color LCD compared to the JND per P-value predicted by the DICOM GSDF.

3.3 Results

3.3.1 Calibrated grayscale characteristics of monochrome medical LCDs

Measurements of the uLR, generation of a calibration LUT, and evaluation of the cLR using the pacsDisplay software were done for 6 monochrome LCD monitors. These displays are listed in Table 3.2, along with the video controllers used. For this work, LumResponse was set to measure 766 gray levels. Ambient luminance values of around 0.3 cd/m^2 were found for all systems. The highest luminance that each display could produce was chosen for the calibrated maximum, up to a limit of 500 cd/m^2, and the luminance ratio was set to 250. After installation of the LUT, the cLR was measured and compared to the expected DICOM GSDF. An example is shown in Figure 3.9. Statistics for all displays are given in Table 3.3 as per the evaluation guidelines and criteria set by AAPM OR-03. All of the calibrations were within the AAPM recommended limits.

Table 3.2: Monochrome medical LCDs and controllers used in the pacsDisplay study.

Designation	Video Controller	Display	Resolution
NDS-3MP (x2)	Matrox Med3MP	NDS Nova 3MP (IBM)	1536 x 2048
NDS-2MP	Matrox G200	NDS Nova 2MP (Sharp)	1200 x 1600
Image-3MP	Image Systems Image-3MP	Image Systems FP2080HBMAX	1536 x 2048
Image-2MP	Image Systems Image-2MP	Image Systems FP1900HBMAX	1200 x 1600
Dome-3MP	Dome DX	Dome C3	1536 x 2048

Table 3.3: Display evaluation and diagnostic criteria for the monochrome medical LCDs.

	NDS 2MP	NDS 3MP (1)	NDS 3MP (2)	Img Sys 2MP	Img Sys 3MP	Dome 3MP	AAPM Limits
<ΔJND/Δp>	2.37	2.71	2.31	2.34	2.34	2.53	3.0
ΔJND/Δp max error	1.52	1.42	1.21	1.48	1.48	0.87	2.0
ΔJND/Δp RMSE	0.59	0.47	0.48	0.41	0.37	0.25	1.0

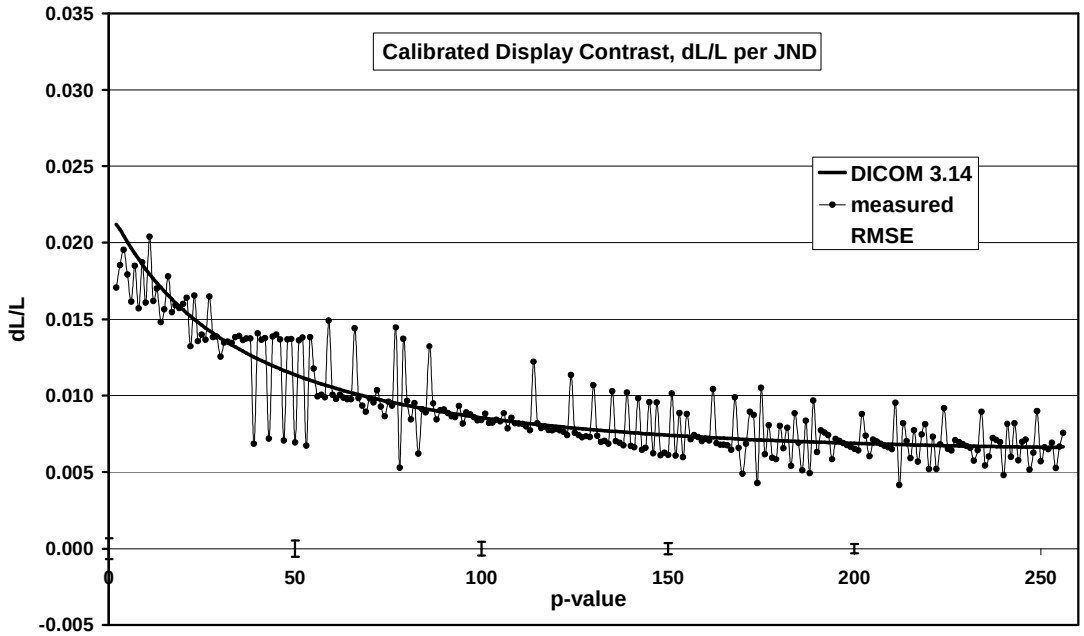

Figure 3.9: Contrast response for a calibrated NDS 3MP compared to the DICOM GSDF.

3.3.2 Luminance response characteristics of LCD consumer color monitors

The luminance characteristics for 60 color LCDs of 4 consumer model types are reported in this section. The uLRs are shown in Figures 3.10-3.13, separated by display model. The first set of measurements is from 2003 for 20 DELL 1900FP monitors that were evaluated for the possible use of a generic LUT. However, as can be seen in Figure 3.10, the DELL 1900FP displays exhibited wide luminance variation. As a result, the use of a generic LUT was deemed inappropriate for these displays and they were instead calibrated individually. The second set of measurements comes from 2005 for 18 DELL 2001FPs, the third from 2006 for 10 DELL 2007FPs, and the fourth from 2007 for 11 DELL 3007WFPs. All of these sets show only minor luminance variability within the same display models. Each of these uLR sets have been averaged to generate generic LUTs for their respective model types, all of which are currently in use.

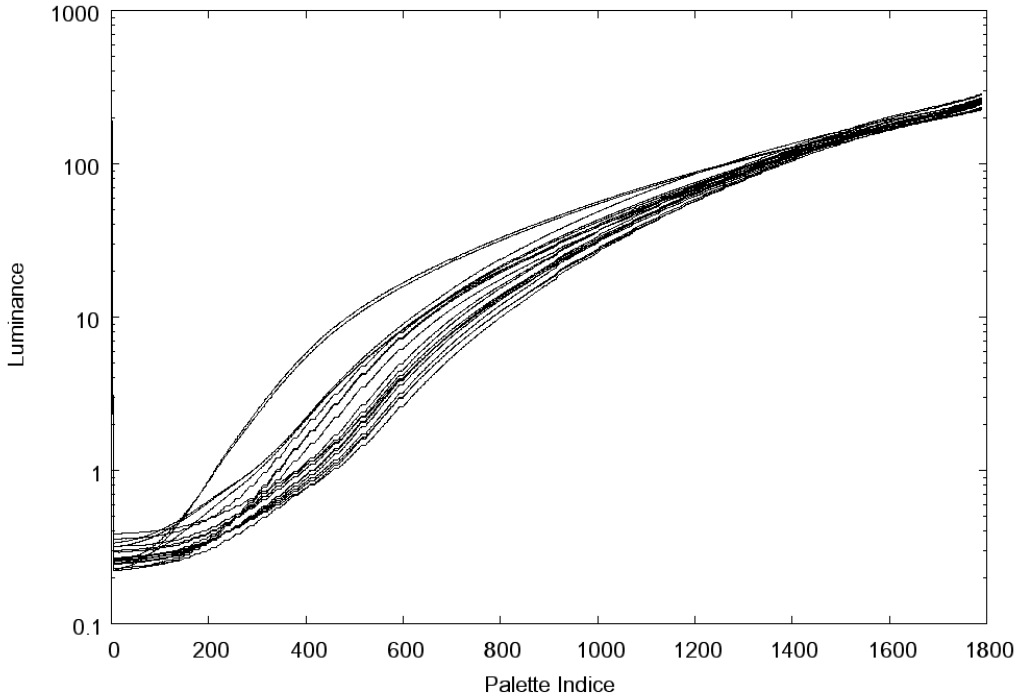

Figure 3.10: uLRs from 20 uncalibrated DELL 1900FP monitors showing wide variation.

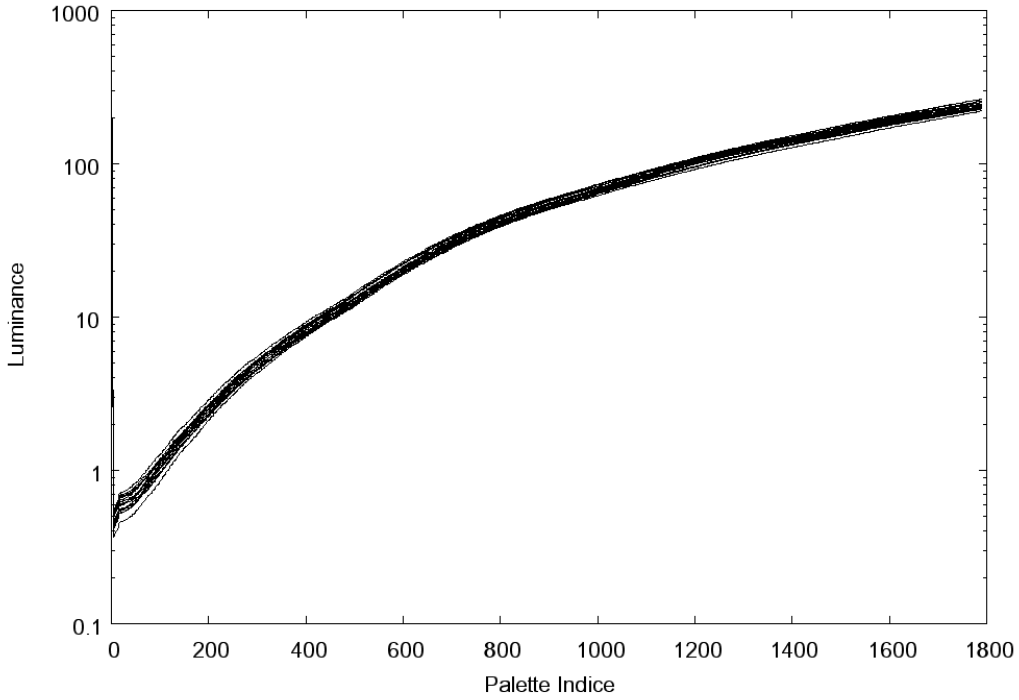

Figure 3.11: uLRs from 18 uncalibrated DELL 2001FP monitors showing minor variation.

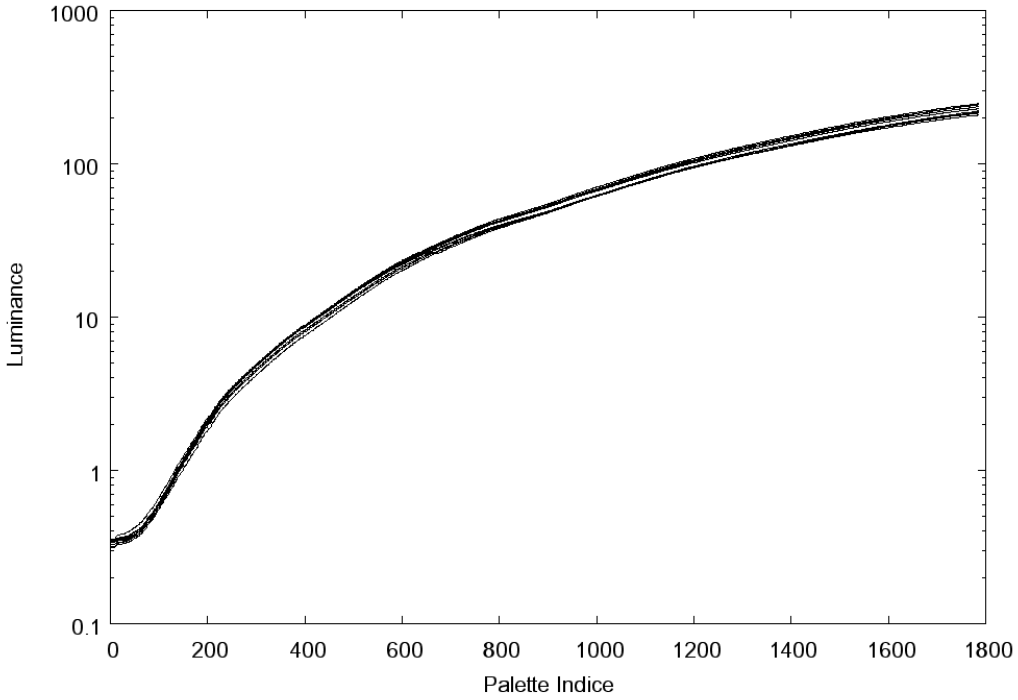

Figure 3.12: uLRs from 10 uncalibrated DELL 2007FP monitors showing minor variation.

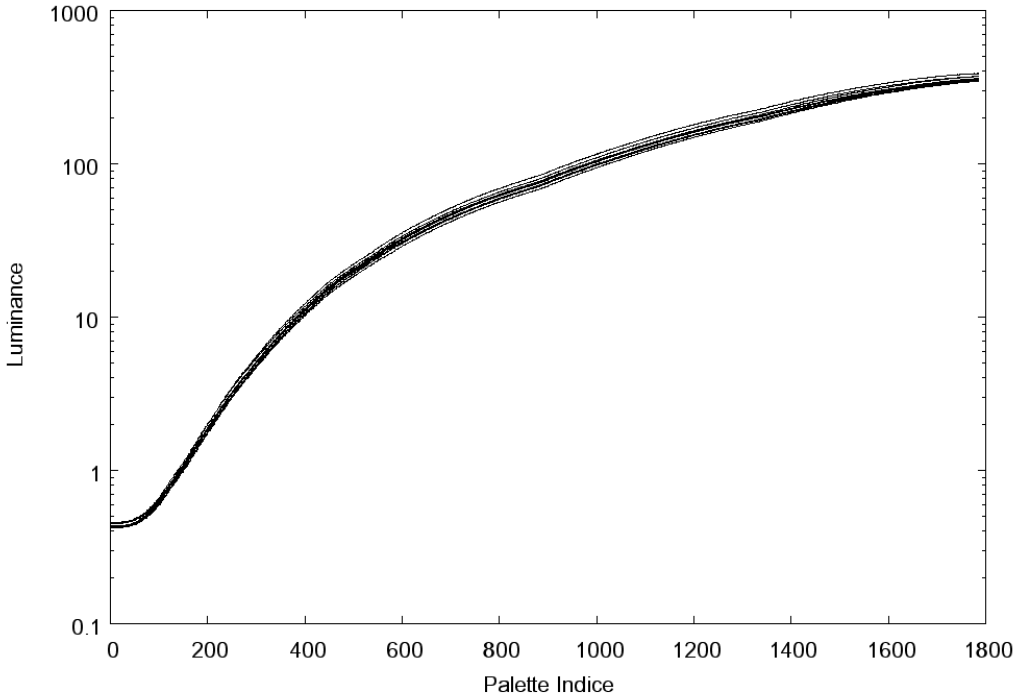

Figure 3.13: uLRs from 11 uncalibrated DELL 3007WFP monitors showing minor variation.

3.3.3 Calibrated grayscale characteristics of monitors with a common LUT

Luminance responses were measured for 29 DELL 1905FP monitors using a common calibration LUT. The LUT was built from a single DELL 1905FP monitor that was considered prototypical. The cLRs showed little variation between monitors, with the exception of one, which exhibited very atypical calibration characteristics, including much lower L_{min} and L_{max} values.

The calibration statistics for the DELL 1905FP monitors are shown in Table 3.4. The <ΔJND/Δp> and root mean squared deviation values were found to be well within the AAPM limits. However, the maximum error was higher than prescribed. These statistics indicate that the common LUT did not fully account for some of the variations that existed among the displays, but the majority of deviations were small, resulting in good overall performance. Given that the common LUT was derived from a single display, significant improvements could be achieved with a generic LUT built from a larger sample set.

Table 3.4: 2006 DELL 2001FP display evaluation statistics and diagnostic criteria.

	Dell 2001FP (average from 29)	AAPM Limits
<ΔJND/Δp>	2.12	3.0
ΔJND/Δp max error	2.69	2.0
ΔJND/Δp RMSE	0.53	1.0

3.4 Discussion

For the calibration evaluations shown in Figures 3.7-3.9, the measured results appear to have significant noise. This is due to the fact that each measurement represents the very small luminance difference between sequential gray levels, as shown in Figure 3.14. When converted to the relative contrast response, $\Delta L/L$, these deviations become much more apparent. This method of reporting is sensitive enough to depict all of the small fluctuations that are present even with monitors that are considered to have excellent grayscale calibration. For monochrome monitors calibrated from a palette of 766, the RMSE of measured $\Delta JND/\Delta p$ values is found to be about 0.5. For color monitors calibrated from a palette of 1786, the RMSE can be improved by a factor of about 2 [10]. Similar improvement might be possible for monochrome monitors if the

sub-pixel luminance is not equal. However, to date it has not been established that exceeding a specific RMSE limit for a system will negatively impact the interpretation tasks associated with viewing a medical image. At present, monitors with a RMSE for all $\Delta JND/\Delta p$ values less than the AAPM limit are assumed to be acceptable for medical use.

The pacsDisplay software package is public domain and is intended as a utility for both enterprise and desktop use. It is capable of providing precise grayscale calibration for both monochrome and color LCDs and can apply them using standard Windows function calls. Such tools will be of emerging importance as 3MP color LCDs using consumer graphic cards become available for use in primary diagnosis. The ability to build a LUT from an average luminance response was developed to address the need for effective generic LUTS due to the use of PACS systems with large networks of displays systems for many different medical imaging professionals. At HFHS, pacsDisplay is loaded by the computer supplier as part of the standard software used by the IT group. To date over 2000 workstations have been deployed with DICOM calibrated LUTs and an image quality verification application called iQC, another utility of pacsDisplay.

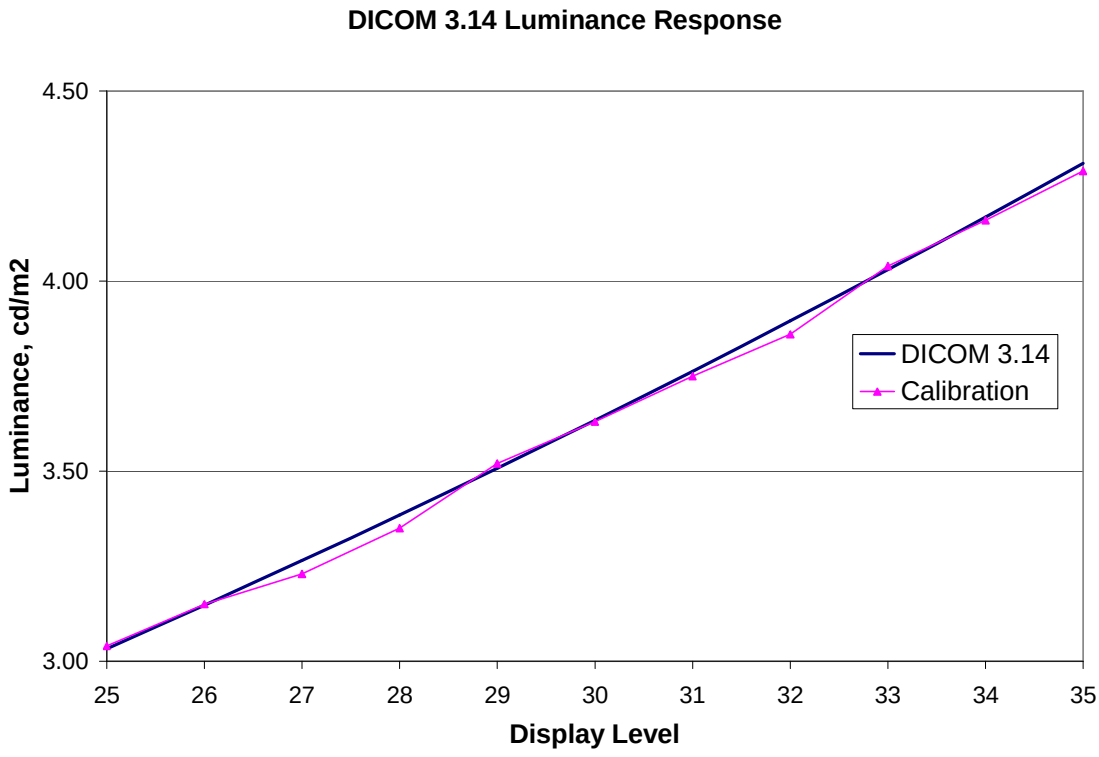

Figure 3.14: Calibrated luminance response for display levels 25-35 demonstrating the small luminance differences between the DICOM GSDF and an actual calibration.

In summary, a method has been developed for the precise measurement of luminance from an LCD device. The pacsDisplay software package has been developed in order to implement these methods. It has been used to calibrate several different LCD displays, both monochrome and color, and can be easily modified to calibrate a wide range of medical display devices. These calibrations successfully met the standards set by the DICOM GSDF and the AAPM OR-03. This work followed the AAPM guidelines with respect to reporting luminance response as a contrast result, $\Delta L/L$, and summarizing the statistical error of the observed JND changes. The results for all monitors calibrated fell within the AAPM recommended limits. In addition, the luminance responses of monitors of the same model were found to be similar, with the exception of the DELL 1900FP. Newer models that were tested showed only minor variability, making them candidates for the use of a generic LUT. Those calibrated using a common LUT demonstrated good overall contrast performance.

This chapter has demonstrated that the methods implemented by the pacsDisplay software can be used to reliably measure luminance and provide grayscale calibration. In later chapters, this method will be extended beyond calibration and applied to the design and implementation of observer performance tests. Knowledge of the uLR is used in Chapter 4 to produce very low-contrast images in order to measure the contrast threshold. In Chapter 6, the calibration process is used to generate a very narrow luminance range LUT for the production of low-contrast noise patterns. In addition to the significance of pacsDisplay for routine medical image presentation, the tools described here provide an experimental foundation for this dissertation research.

3.5 References

1. Klompenhouwer, M.A. and G. de Haan, *15.3: Color Error Diffusion: Accurate Luminance from Coarsely Quantized Displays.* SID Symposium Digest of Technical Papers, 2001. **32**(1): p. 214-217.
2. Tyler, C.W., et al., *Bit stealing: how to get 1786 or more gray levels from an 8-bit color monitor.* 1992. **1666**(1): p. 351-364.
3. NEMA, *Digital Imaging & Communications In Medecine (DICOM), Part 14: Grayscale Standard Display Function*. 1998, National Electrical Manufacturers Association.
4. Flynn, M. and P. Tchou, *Accurate Measurement of Monochrome Luminance Palettes for the Calibration of Medical LCD Monitors.* Proceedings of SPIE, 2003. **5029**(Medical Imaging 2003: Visualization, Image-Guided Procedures, and Display): p. 438-448.
5. Muka, E., H.R. Blume, and S.J. Daly. *Display of medical images on CRT soft-copy displays: a tutorial*. in *Medical Imaging 1995: Image Display*. 1995. San Diego, CA, USA: SPIE.
6. Flynn, M.J., K.D. Compton, and A. Badano. *Luminance response calibration using multiple display channels*. in *Medical Imaging 2001: Visualization, Display, and Image-Guided Procedures*. 2001. San Diego, CA, USA: SPIE.
7. Fetterly, K., et al., *Introduction to Grayscale Calibration and Related Aspects of Medical Imaging Grade Liquid Crystal Displays.* Journal of Digital Imaging, 2007.
8. Flynn, M.J., et al., *High-fidelity electronic display of digital radiographs.* Radiographics, 1999. **19**(6): p. 1653-69.
9. Samei, E., et al., *Assessment of display performance for medical imaging systems: executive summary of AAPM TG18 report.* Med Phys, 2005. **32**(4): p. 1205-25.
10. Averbukh, A.N., D.S. Channin, and M.J. Flynn, *Assessment of a novel, high-resolution, color, AMLCD for diagnostic medical image display: luminance performance and DICOM calibration.* J Digit Imaging, 2003. **16**(3): p. 270-9.

CHAPTER 4

TWO-ALTERNATIVE FORCED CHOICE MEASUREMENT OF CONTRAST THRESHOLD FOR BI-LEVEL BAR PATTERNS PRESENTED ON A MEDICAL LCD MONITOR

4.1 Introduction

Medical liquid crystal display (LCD) systems are quickly replacing film systems for the interpretation of radiological images. However, relatively few observer performance experiments have been done with LCD monitors. Such experiments are important for evaluating the effectiveness of LCDs for medical imaging. However, these evaluations are made more difficult by the luminance characteristics of these devices. As reported in Chapter 3, uncalibrated LCD devices have an irregularly-shaped luminance response. Significant discontinuities occur in the relative luminance change associated with each change in the LCD controller output value. These discontinuities complicate visual experiments requiring precise image presentation and contrast control. The luminance measurement methods detailed in Chapter 3 are used here to solve this issue and permit observer performance to be evaluated with an LCD monitor.

The DICOM gray scale display function, the calibration standard used for medical displays, is based on the contrast threshold (C_T) of the human visual system. Specifically, it uses the contrast threshold derived by Barten's model (C_{BM}), which describes the contrast sensitivity ($1/ C_{BM}$) of an observer viewing a DICOM Standard Target (as referenced in Chapter 2). The Barten model has been compared to a large number of published studies measuring the contrast sensitivity of the eye and has been found to agree well. However, these studies used either CRT or projection devices. No prior studies have measured the contrast threshold using LCD devices.

The goal of this experiment was to establish that the contrast threshold measured using a monochrome medical LCD monitor and graphics card agrees with the Barten

Model under the conditions defined by the DICOM standard. An underlying purpose for this was to establish an experimental method that could be used in future research on observer performance in digital radiography using medical LCD monitors. By demonstrating that the LCD device can present patterns at or below the threshold for detection, the display can be considered appropriate for medical perception research.

This chapter reports on a novel method to generate very low contrast bi-level bar patterns by using the full palette of available gray values. The method is used to determine contrast threshold in observer experiments. Section 4.2 begins by reviewing two-alternative forced choice (2AFC) testing, including the mathematical analysis of the results. Section 4.3 describes the methods to generate low contrast bar patterns in a computer application for 2AFC testing. Section 4.4 reviews the results of an experimental observer study and section 4.5 discusses the results in relation to the DICOM standard. Much of this chapter has been previously reported in a conference proceeding article[1].

4.2 Two-alternative Forced Choice Contrast Threshold Testing

4.2.1 Comparison of methods for contrast threshold experiments

Various techniques have been used in the past to measure the contrast threshold. The most commonly used technique has been the method of variable adjustment[2], where observers manipulate the contrast level until they find what they believe is their detection threshold. Historically, the method of adjustment has been the dominant technique for perception research. The majority of the data that the Barten model was based on used the method of adjustment. A list of the studies and the methods they used is shown in Table 4.1. A few of the studies used the 2AFC method, but most of those were done by the same research group[2]. The Barten model was, in turn, the basis for the DICOM standard; thus, the DICOM standard is also based on these same studies using the method of adjustment.

When measuring the contrast threshold with the method of adjustment for a Standard Target as per the DICOM standard, a value of $C_T/C_{BM} = 1$ can be expected. This method is quick, but the threshold criterion is not well defined and the confidence of the observer plays a significant role. A method that has been favored in more recent

work is the “alternative forced choice” (AFC) method[3], in which the observer is forced to choose the location of a target from a set of options. While AFC experiments are more time consuming than adjustment experiments, they have been used extensively in psychometric research. This research uses the 2AFC method, where the observer must decide which of two image regions contains the signal. If a distinction cannot be made, then the observer must make a guess. Since the determination of the threshold level is not subject to observer confidence, the measured value of the threshold is expected to be more consistent than that found using the method of adjustment. According to a contrast threshold study by Legge, a 2AFC experiment with a continuous target background and background adaptation found contrast thresholds to be approximately 2/3 of those found using the method of adjustment[4]. Thus, when measuring the contrast threshold with a 2AFC method for a Standard Target as per the DICOM standard, a value of C_T/C_{BM} of about 0.66 can be expected.

Table 4.1: Methods for measuring the contrast threshold in experimental studies contributing to the Barten model.

author	test type
DePalma & Lowry (1962)	method of adjustment
Patel (1966)	method of adjustment
Robson (1966)	method of adjustment
van Nes & Bouman (1967)	method of adjustment
Campbell & Robson (1968)	method of adjustment
Watanabe et al. (1968)	method of adjustment
Howell & Hess (1978)	method of adjustment
Carlson (1982)	method of adjustment
Virsu & Rovamo (1979)	2AFC
Rovamo et al. (1992)	2AFC
Rovamo et al. (1993a)	2AFC
Rovamo et al. (1993b)	2AFC
van Meeteren & Vos (1972)	~2AFC
Sachs et al. (1971)	???

4.2.2 Numerical analysis for a 2AFC contrast threshold experiment

In a contrast threshold experiment, observers interpret scenes having object contrast levels that are distributed above and below the threshold for detection. In the analysis of the responses, it is assumed that the HVS performance can be described by a psychometric function for which a cumulative Gaussian, as given by Eq. (4.1), is often used[2, 5, 6]. For the case of a DICOM standard target, the signal strength s is equal to the grating pattern contrast, $C = (L_{max}\text{-}L_{min})/L_{avg}$, where L is luminance. The signal strength for a 50% probability of detection, s_0, would be equal to the contrast threshold, C_T. For a 2AFC experiment, given that guessing has a 50% chance of being correct, the chance of a correct response for a target at the contrast threshold (50% detection probability) is 75%[2].

Most prior 2AFC experiments have been designed using a small number of distinct signal strengths, with many scenes repeated for each strength level and randomly distributed[5]. The percent correct is then computed for each level from the observer's responses and fit to the assumed psychometric function. Often, the percent correct values are first transformed by using an inverse cumulative normal distribution function, also referred to as a "z-score", to obtain a detectability index[6, 7] (see Eq. 4.1 and 4.2). The detectability index is a convenient mathematical transformation that scales the percent-correct linearly with contrast. The combined z-scores can then be fit to a linear function of signal strength and extrapolated to find C_T at 75%.

$$P(z) = \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\left(x^2/2\right)} dx \tag{4.1}$$

$$z = \Phi^{-1}(P) \tag{4.2}$$

P = percent correct

z = z-score (detectability index)

4.3 Observer Performance Test Methods

The experimental method used in this study presented a large number of images (306) with a uniform background to human observers. A test target was placed in one of two positions and the observers were asked to identify the position that contained the target. The display used was an NDS Nova 2MP LCD monochrome medical monitor

with a Matrox Med3MP (G200 MMS) graphics controller. Each observer's results were analyzed to determine their psychovisual performance relative to target contrast.

4.3.1 Contrast control

In order to measure the contrast threshold using an LCD system, a method for generating near-threshold contrast test patterns was needed. This was accomplished by measuring the full palette for the luminance response of the LCD monitor (as described in Chapter 3) and choosing display levels that could be combined to produce very fine contrast differences. Since a monochrome display was used, luminance measurements were made for 766 DDLs. These gray levels could be combined in pairs to produce various contrast changes,

$$C = \frac{|L_1 - L_2|}{(L_1 + L_2)/2} = \frac{\Delta L}{L}. \qquad (4.3)$$

From the palette of 766 luminance measurements, a range of DDLs were selected that produced luminance levels around 348 +/- 28 cd/m^2 (as shown in Figure 4.1), with C_{BM} in this range being approximately 0.0068 ($\Delta L/L$ per JND). These levels were chosen because they could be paired to produce both sub-threshold and super-threshold contrast relative to the threshold predicted by Barten. These contrast changes were made by choosing pairs of DDLs that were 1-3 steps apart. These "steps" could be grouped into four categories of similar relative contrast: C_T/C_{BM} = 0.43, 0.86, 1.3, and 1.7 +/- 4%. This method for producing unusually low-contrast targets on a monochrome LCD had not been used before in medical imaging. 36, 37, 37, and 28 pairs were chosen from each category, respectively. Each contrast pair was shown twice during the 2AFC testing, once in the right target region and once in the left target region. In addition, 30 zero-contrast states were included to provide data at the "false alarm" rate of success. This provided a total of 306 contrast scenes to present to the observers during the test. The distributions of the image contrast values in these bins are listed in Tables 4.2 and 4.3.

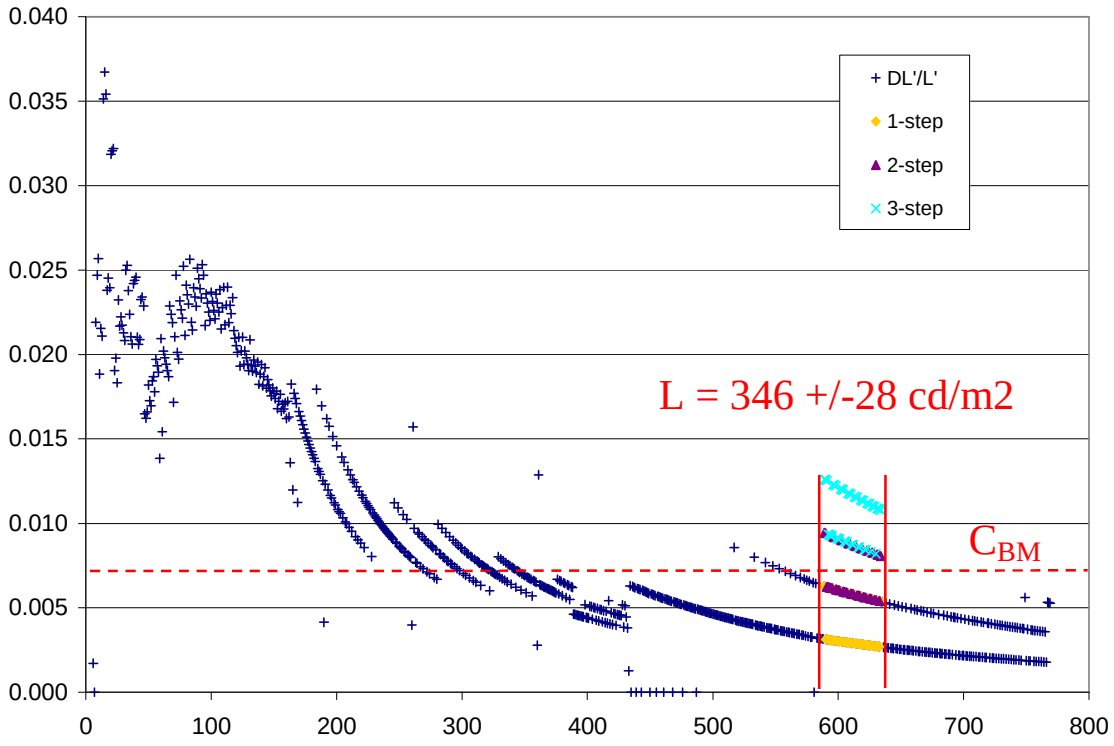

Figure 4.1: Relative luminance change of sequential states, ΔL/L, plotted in relation to the sequence (palette) number for 766 gray values (NDS Nova 2MP LCD medical monitor). The vertical bars mark the luminance range used in the 2AFC contrast threshold experiment. The 3 sets of contrast steps used in the test are plotted in between. These were produced by combining pairs of luminance responses from the sequential states. The dashed line (C_{BM}) is the approximate contrast threshold for this range of luminance according to the Barten model.

Table 4.2: ΔL/L bin distributions.

Bin	Count	Mean	Min	Max	StdDev
1	30	0.00000	0.00000	0.00000	0.00000
2	72	0.00291	0.00270	0.00313	0.00013
3	74	0.00583	0.00537	0.00628	0.00027
4	74	0.00876	0.00807	0.00944	0.00040
5	56	0.01158	0.01077	0.01261	0.00056

Table 4.3: C_T/C_{BM} bin distributions.

Bin	Count	Mean	Min	Max	StdDev
1	30	0.00000	0.000	0.000	0.00000
2	72	0.42767	0.397	0.457	0.01806
3	74	0.85543	0.792	0.919	0.03723
4	74	1.28570	1.189	1.380	0.05592
5	56	1.70071	1.587	1.843	0.07825

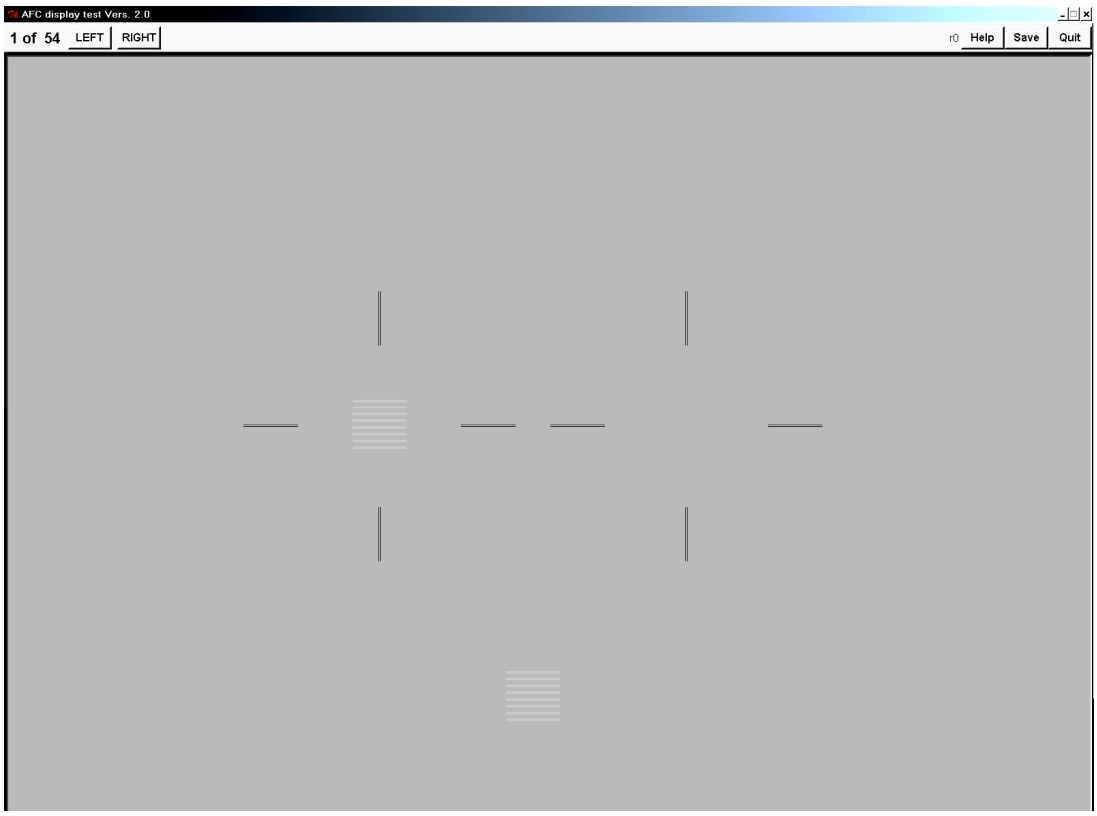

Figure 4.2: Screenshot of the 2AFC contrast threshold test (training session).

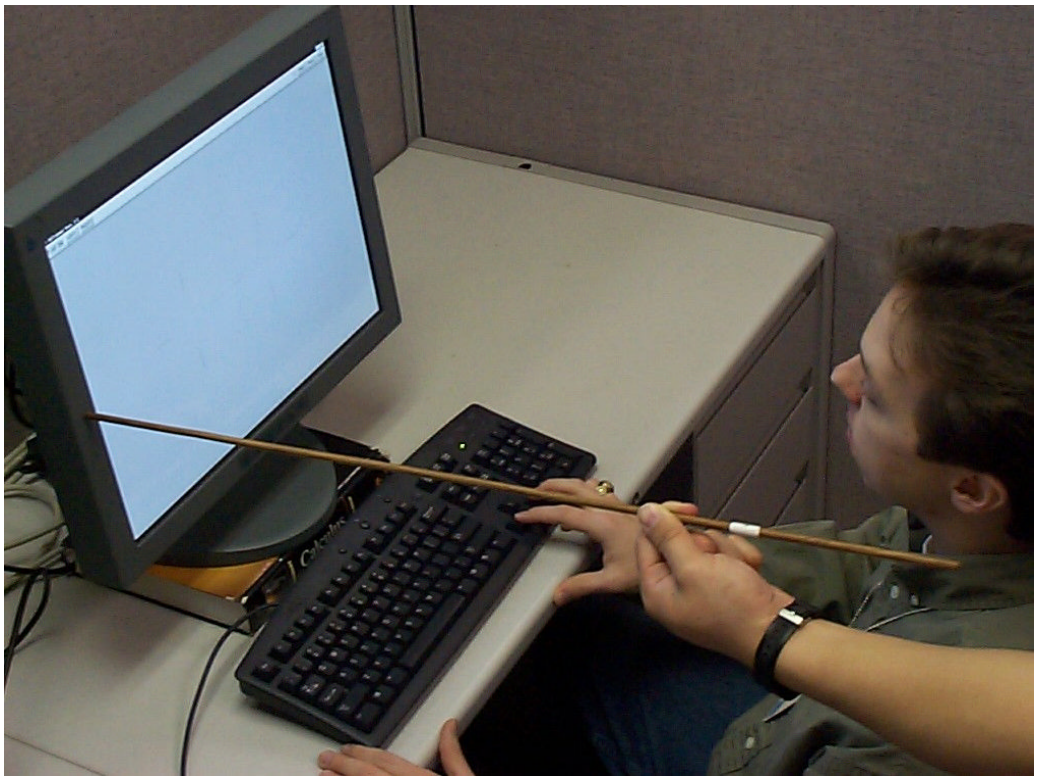

Figure 4.3: Test subject being positioned to take the 2AFC contrast threshold test.

4.3.2 Measuring the contrast threshold

A TCL/TK graphic application was created to generate the 2AFC test. Two target regions were presented on the NDS Nova 2MP monochrome LCD with a uniform background luminance. A low-contrast test pattern was randomly positioned in the center of one of the two target regions (see Figure 4.2). The test patterns followed the DICOM definition of a Standard Target when viewed at a 60cm distance. Ambient lighting was taken into account and kept to a minimum. Participants were positioned so that their eyes were level with the center of the screen and 60cm from the surface (see Figure 4.3). A sample target was placed near the bottom of the screen for participants to use as a reference. The program then presented a large number of images of different contrast levels with a uniform background to the observers. A test target was placed in one of two positions, left or right, and the observers were asked to identify the position that the target was in. If they were uncertain, then they were instructed to guess as best they could. Selections were made via keyboard. Once a selection was made the screen was updated with a new background and target. 306 images were viewed for each test session, with breaks inserted every 51 images during which the screen was black. A training sequence was provided for participants to acclimate themselves with the test process. The training sequence used 54 images starting with a very high contrast target and decreasing to zero contrast.

The contrast images were presented to the observer in random order using the TCL/TK random number generator function (“rand”). Prior to the test, the images were divided evenly into 6 cases, with 51 images per case. The test was administered by presenting one image from each case in a set order and repeating until all of the images had been presented. Each time the test was administered, a new random order was first established for the order in which the cases were presented. The case sequence was then divided equally into group A and group B. One presentation per case was made from group A followed by one presentation per case from group B. The case sequence associated with group A and with group B was newly randomized before the next sequence of presentations. Observers thus scored images in varying order as presentations for each case were repeated, but the same case would not appear twice in a row.

Several aspects of the graphic scene presentation used in this work deserve comment. Rather than fix the background luminance at one value and place new targets on the background, each new scene involves a small change in luminance within a range of +/- 8%. This results because of the need to use different combinations of background and target luminance to achieve the desired contrast. An important side benefit is that the observer is required to make small adjustments in adaptation from scene to scene as the overall scene luminance changes. Without this shift, the new contrast patterns appear with a rapid temporal change that can contribute to detection by virtue of the excellent temporal response of the human visual system. This is sometimes referred to as a "flicker response" and is particularly strong in the peripheral field of view[8, 9].

It should also be noted that the graphic scenes used in this work involve bar patterns where the bars are composed entirely of a single luminance level with the background luminance between and around them. This differs from many prior psycho-visual studies that used targets with a sinusoidal modulation of luminance about the background value. Because of the digital nature of LCDs and the limits of the test software at the time, a simple square wave grid pattern was used as opposed to the sine wave patterns that the Barten model assumes. The difference in perceived contrast between the two patterns can be found by looking at the fundamental wave modulation, and a correction factor can be applied to the contrast values. For a square wave pattern, the equivalent sinusoidal modulation is $4/\pi$ times the square wave contrast[2]. This correction factor of $4/\pi$ has been applied to all contrast values from the 2AFC test results presented in section 4.4.

4.3.3 Test groups and statistical analysis

The 2AFC contrast threshold tests were first given to a group of 14 observers with radiology research experience. Each observer was tested once. Three observers from the first group, Philip Tchou (PMT), Michael Flynn (MJF), and Allen Seifert (HAS) were then tested multiple times over several months in order to examine intra-observer variability.

The test program kept track of each participant's responses for analysis at the end of the session. The observer results were binary in nature, i.e., correct or incorrect. These results were evaluated using the z-score method to obtain the C_T parameter of the

psychovisual functions. The mean and standard deviation of C_T were also computed. Results from the intra-observer group were similarly analyzed. The statistical differences between the results of the intra-observer group were examined using a t-test analysis (Microsoft Excel data analysis, t-Test: Two-Sample Assuming Unequal Variances).

4.4 Results

4.4.1 Initial observer test results

Figure 4.4 shows a histogram of the results for the initial 2AFC tests with 14 observers, including the first test for each of the three extended observers. The mean contrast threshold was 0.68 (C_T/C_{BM}) with a standard deviation of 0.202. The minimum was 0.36 and the maximum was 1.09. The spread of C_T/C_{BM} values between the maximum and minimum were relatively uniform and did not appear to follow a Gaussian distribution. Note that all results include the $4/\pi$ correction mentioned in section 4.3.2.

4.4.2 Intra-observer test results

Figure 4.5 shows the contrast threshold results for the three extended observers, MJF, HAS, and PMT. This group showed noticeable improvement after three test sessions, suggesting a learning process. Based on this, the first three sessions were not included in the statistical analysis for the intra-observer group. The test results were otherwise relatively stable over the test period, with some possible indication of improvement over time.

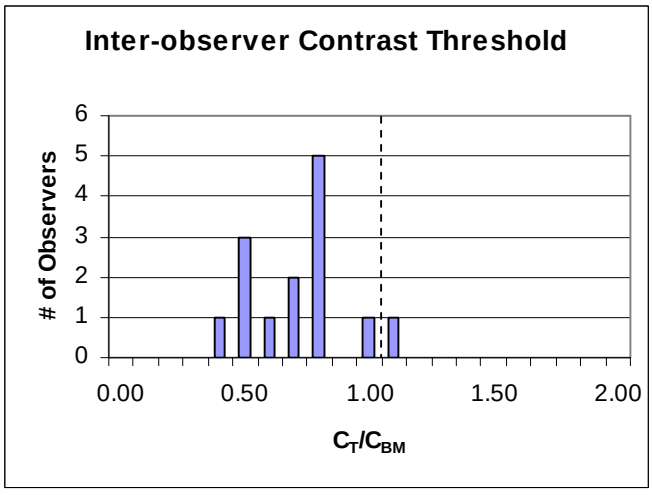

Figure 4.4: Histogram of the inter-observer contrast threshold results.

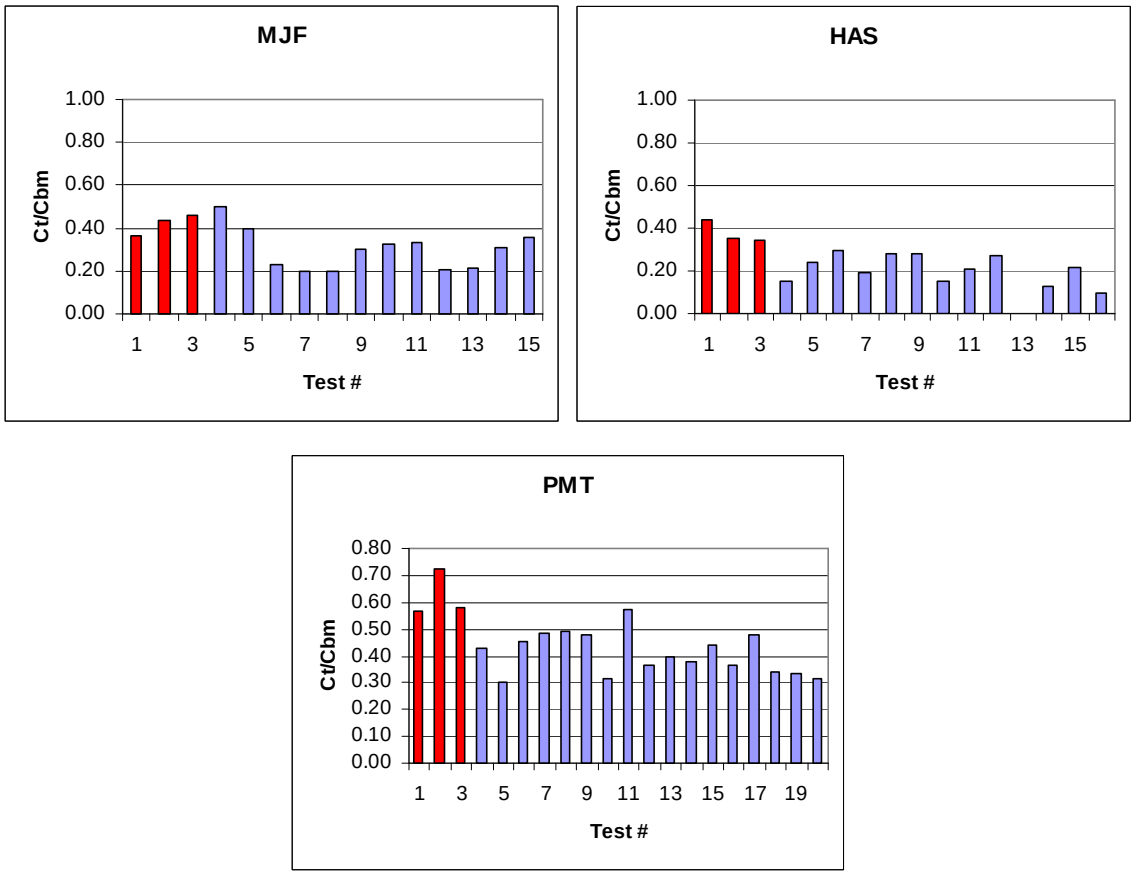

Figure 4.5: Intra-observer contrast threshold measurements.

Table 4.4 Intra-observer contrast threshold measurements.

Observer	**Mean T (minutes)**	**Mean (C_T/C_{BM})**	**STDEV (C_T/C_{BM})**
MJF	14.7	0.37975	0.11902
HAS	19.0	0.26632	0.08534
PMT	12.6	0.52010	0.09824

Table 4.5 Intra-observer t-test results.

Observers	**P(T<=t) two-tail**
MJF and HAS	0.014302433
MJF and PMT	0.002987029
HAS and PMT	0.0000000731

HAS, in particular, exhibited extremely good contrast perception that improved to the point that he was receiving nearly perfect scores on the tests. In one instance he managed to correctly identify targets at all available contrast levels with the exception of the zero contrast scenes. As a result, this test was unable to produce any meaningful sub-threshold data for him, and the software could not effectively calculate his C_T value. This test was not included in the analysis.

The mean and standard deviation of the contrast thresholds for the intra-observer tests are shown in Table 4.4. A t-test analysis was performed on the intra-observer results and confirmed that the mean contrast thresholds were statistically different. The results of this analysis are shown in Table 4.5.

4.5 Discussion

On average, the 2AFC experimental results were consistent with the Barten model. As stated in section 4.2.1, a result of ~0.66 (C_T/C_{BM}) is predicted for a 2AFC experiment compared to the variable adjustment method[4]. This largely explains the mean value of 0.68 (C_T/C_{BM}) measured in this work, which is well within one standard deviation (0.202) of the predicted result. Therefore, the 2AFC method and the monochrome LCD and medical graphic card used for this experiment resulted in a C_T/C_{BM} that matched the values predicted by the Barten model under the conditions laid out by DICOM 3.14. In addition, the experiment was able to successfully measure contrast thresholds levels well below the expected values. The mean contrast thresholds for the intra-observers were significantly less than those predicted by the Barten model, with individual test results of 0.2 C_T/C_{BM} and lower. All of these results positively demonstrate that the experimental methods and equipment are both valid and exceptional for use in medical perception research.

The Barten model has been used as the basis for the DICOM GSDF[10], and the experimental results were intentionally referenced to this model so that differences could be easily observed. However, we do not wish to suggest that DICOM calibrated monitors may need to have grayscales with more finely spaced levels, since the scenes used in this experiment are much different from those seen in typical medical images.

4.5.1 Variability in performance

There was wide variation in performance amongst individual observers. Further testing of a small subset of the original observers showed that these differences were statistically significant. One observer (HAS) even improved to the point that the test could not adequately characterize his performance in one instance. Taken overall, these results show differences between observers that do not reflect systematic noise. This may suggest that optical, neurological, or cognitive differences exist between observers that are significant factors in visual performance. Clinical conditions such as myopia or astigmatism may affect outcomes as well, even with corrective lenses or surgery[11, 12]. Age may also play a role, as contrast sensitivity has been shown to change throughout adulthood[13]. However, variability amongst observers is of potential medical imaging importance only if it is shown to affect the perception of medically important patterns.

4.5.2 Adaptive testing

AFC testing can be very time consuming. Enough data must be collected over a broad range of signal values to allow for a good estimate of the psychometric curve. Enough time must also be given for the subjects to make rational and consistent decisions. Extensive testing runs the risk of subject fatigue, possibly leading to misjudgments or hurried responses. As such, it is often not feasible to measure observer performance over all available signal levels [14, 15]. One possible solution is to use adaptive testing.

Adaptive testing is a method of progressively choosing which signal values to test based on the subject's earlier responses. This allows for fewer trials in order to get the specific data one is looking for. For example, an adaptive staircase method used in a psychovisual threshold test would increase or decrease the signal strength based on the observer's previous responses, narrowing the range of values to only those that are close to the observer's threshold. This adaptive reevaluation process is often repeated after every trial until specific stopping criteria have been met, usually based on the level of precision or a fixed maximum number of trials. Adaptive testing strategies are discussed further in Chapter 5 in relation to simulated 2AFC results.

4.5.3 Issues with the z-score analysis

The z-score method used with this experiment to analyze the results is commonly used in 2AFC studies. However, it relies on the binning of results and the use of a linear

function to represent a non-linear process. Both of these processes can cause distortion by improperly weighting the data. For the experiment described in this chapter, binning shifts contrast levels to match one set value, in this case the bin average, distorting the contrast associated with each response. This is compounded by the fact that the contrast levels being tested are not evenly distributed, due to the erratic contrast response of the LCD. This distorts the linear fitting process by giving each bin (converted to a z-score) equal weight in determining the fit regardless of how many responses they contained or how they were distributed. The fitting process itself is also an issue because it forces data from a non-linear model to a linear one. Chapter 5 will examine these issues with the z-score method and present a solution in the form of a maximum likelihood estimation method, which determines the parameters of the psychometric function without the use of binning or linearization.

4.5.4 Issues with the bi-level test pattern

The Barten model predicts contrast sensitivity for observers viewing a sinusoidal pattern on a flat background. The experiment presented in this chapter uses a bi-level bar pattern instead. Although an approximate correction was made using the fundamental wave modulation, performing these tests with a sinusoidal pattern would provide a better comparison with the Barten model and the DICOM standard. However, producing a sinusoid having a particular contrast requires a series of smaller contrast steps that are subtle enough to be perceived as smoothly varying. The software used in this experiment was not capable of this. In Chapter 6, a method is reported to present complex patterns with luminance variation and new software is developed to conduct new visual performance experiments.

4.6 References

1. Tchou, P., M. Flynn, and E. Peterson, *2AFC assessment of contrast threshold for a standardized target using a monochrome LCD monitor.* Proceedings of SPIE, 2004. **5372**(Medical Imaging 2004: Image Perception, Observer Performance, and Technology Assessment): p. 344-352.
2. Barten, P.G.J., *Contrast sensitivity of the human eye and its effects on image quality*. 1999, Bellingham, Wash.: SPIE Optical Engineering Press. xix, 208.
3. Macmillan, N.A. and C.D. Creelman, *Detection theory : a user's guide*. 2nd ed. 2005, Mahwah, N.J.: Lawrence Erlbaum Associates. xix, 492.
4. Legge, G.E., *A power law for contrast discrimination.* Vision Res, 1981. **21**(4): p. 457-67.
5. Wichmann, F.A. and N.J. Hill, *The psychometric function: II. Bootstrap-based confidence intervals and sampling.* Percept Psychophys, 2001. **63**(8): p. 1314-29.
6. Klein, S.A., *Measuring, estimating, and understanding the psychometric function: a commentary.* Percept Psychophys, 2001. **63**(8): p. 1421-55.
7. Burgess, A.E., *Comparison of receiver operating characteristic and forced choice observer performance measurement methods.* Medical Physics, 1995. **22**(5): p. 643-55.
8. Jarvis, J.R., N.B. Prescott, and C.M. Wathes, *A mechanistic inter-species comparison of flicker sensitivity.* Vision Research, 2003. **43**(16): p. 1723-1734.
9. Tyler, C.W., *Analysis of visual modulation sensitivity. II. Peripheral retina and the role of photoreceptor dimensions.* J. Opt. Soc. Am. A, 1985. **2**: p. 393-.
10. NEMA, *Digital Imaging & Communications In Medecine (DICOM), Part 14: Grayscale Standard Display Function*. 1998, National Electrical Manufacturers Association.
11. Woods, R.L., N.C. Strang, and D.A. Atchison, *Measuring contrast sensitivity with inappropriate optical correction.* Ophthalmic and Physiological Optics, 2000. **20**(6): p. 442-451.
12. Lombardo, A.J., et al., *Changes in contrast sensitivity after Artisan lens implantation for high myopia.* Ophthalmology, 2005. **112**(2): p. 278-285.
13. Owsley, C., R. Sekuler, and D. Siemsen, *Contrast sensitivity throughout adulthood.* Vision Res, 1983. **23**(7): p. 689-99.
14. Park, S., et al., *Efficiency of the human observer detecting random signals in random backgrounds.* J Opt Soc Am A Opt Image Sci Vis, 2005. **22**(1): p. 3-16.
15. Linschoten, M.R., et al., *Fast and accurate measurement of taste and smell thresholds using a maximum-likelihood adaptive staircase procedure.* Percept Psychophys, 2001. **63**(8): p. 1330-47.

CHAPTER 5

A MAXIMUM LIKELIHOOD ESTIMATION METHOD FOR ANALYSIS OF TWO-ALTERNATIVE FORCED CHOICE PSYCHOMETRIC TESTS

5.1 Introduction

Chapter 4 described a 2AFC psychovisual contrast threshold test. Most 2AFC experiments are designed using a small number of signal strengths with many scenes for each strength level[1]. The percent correct is computed for each level and then is transformed using an inverse cumulative normal distribution function to obtain z-scores, which are measures designed to scale with signal strength [2, 3]. The z-scores are then fit to a linear function for analysis. To date, the majority of 2AFC vision studies conducted for medical imaging research use a z-score or similar method, as was done in Chapter 4. However, this process has several disadvantages.

First, the small number of signal strengths corresponds to a small number of data points on the psychometric curve, thus limiting how well the curve shape can be estimated. Since many trial responses are obtained for each of the signal strengths in order to estimate the percent correct, increasing the number of signal strengths used results in a multiplicative increase in the number of trials. Too many trials can lead to very long testing periods and possible observer fatigue, which can adversely affect results. The number of trials per signal strength could be reduced, but at the cost of precision in determining the percent correct. Taken to the extreme, if only a few trial responses are obtained for each of the signal strengths, the percent correct cannot be effectively determined.

Secondly, an issue arises when one does not have full control of the signal strength. The signal strengths must then be grouped into intervals, or “bins”. Each response is assigned to the average signal strength within the bin and percent correct determined for each interval. However, binning introduces error because the signal

strengths of individual responses are shifted to match the value of each bin. Thus, while binning provides a means to estimate the percent correct locally, the averaging is a potential source of error. This is of particular concern if the signal strengths are variably distributed within the bins. These binned values are then converted to z-scores that are equally weighted for determining the linear fit, even though each bin contains a different distribution of values. This can potentially affect the accuracy of fitting techniques used to determine the parameters of the psychometric function.

Finally, linear regression is often used to fit a line to the z-score values, which are non-linear transformations of percent correct. Such transformations can distort the experimental error if the standard deviation is assumed to be the same at all signal strengths. In general, linear fitting methods should not be used to characterize non-linear systems.

This chapter derives a method using a “maximum likelihood estimation" (MLE) statistical technique that addresses these problems. The method computes the statistical likelihood that the observer’s binary responses agree with an estimate of the psychometric function describing the observer’s performance. The psychometric function is modeled by an equation with two variables, and the best fit for the function is determined by finding the combination of parameters that produce the maximum likelihood. Such an approach has been reported previously for non-visual psychophysical stimuli experiments[4, 5]. For psychovisual experiments, Park used MLE methods to analyze percent correct for a limited number of signal strengths[6], which eliminates the problem associated with a non-linear function, but it does not address the issue of signal strength variability. The method reported in this chapter analyzes the binary 2AFC observer responses directly using MLE techniques and can thus be used for experiments with variably distributed signal strengths.

The method is first demonstrated using the experimental results described in Chapter 4. In that experiment, numerous contrast levels were binned into five groups and the average signal strength computed. The percent correct was measured for each group, converted into z-scores, and related to signal strength using linear regression. The available contrast levels were determined by the discrete states of the LCD monitor with significant contrast variability for each group. As a result, the original analysis of the

experimental results was subject to all of the limitations of the z-score method that were described above.

In the following, section 5.3 reevaluates the experimental results from Chapter 4 using the originally recorded binary data and the MLE method. The new observer contrast thresholds are slightly larger than reported in Chapter 4, but are still consistent with the Barten model. Potentially important additional information is gained on the shape of the psychovisual response curve.

Section 5.4 examines the statistical precision of the MLE method in relation to the number and distribution of trials. The standard deviations are assessed using simulated response data to mimic the expected psychovisual function for a typical observer. The results of these simulations are used to discuss 2AFC testing strategies to minimize error, including the possibility of using adaptive testing.

5.2 Maximum Likelihood Estimation

Given a single observer's set of test results from a psychophysical experiment, the "likelihood" is defined as the product of the probabilities of the individual responses. The probability associated with each response is determined from a mathematical expression chosen to describe the psychometric response function. The expression that maximizes likelihood can be assumed to be the best estimate of the psychophysical function. A maximum likelihood estimate can be determined by computing a likelihood matrix for discrete values of the variables of the response expression and selecting the combination with the largest result. However, this limits the possible solutions to combinations from a chosen set of values, and it can require an excessive number of computations unless a very good initial estimate is made to limit the variables. A more efficient method is to use a directed search algorithm to find the maximum likelihood.

5.2.1 The logistic model (logit) and the false alarm rate

Most directed search algorithms use the partial derivatives of the likelihood expression to guide their search. In general, an estimate is provided as a starting point and the algorithm tests the derivatives at that point to determine the direction of increasing likelihood. It then "steps" in that direction, shifting each variable by a prescribed amount and testing the derivatives again. This process is repeated until a local

maximum is found. Various parameters can be set to customize the search, from the size and number of the steps to the conditions that define a local maximum.

While the cumulative Gaussian function is often used to represent the psychometric function, the partial derivatives of the Gaussian function cannot be easily evaluated. Other functions that are similar to the cumulative Gaussian function but are more mathematically manageable can be used instead, such as the logistic function (logit) or the Weibull function [5, 6]. The logistic function was chosen for use with the 2AFC experiments. Eq. (5.1) shows the logit model as described by Bierens[1]. A sample logistic function is shown in Figure 5.1 along with a similar cumulative Gaussian function.

$$p(x) = \frac{1}{1 + e^{-(a+Bx)}} \tag{5.1}$$

p = probability of a correct response (success)

x = contrast level

a = coefficient on the constant term

B = coefficient on the independent variable ("slope" coefficient)

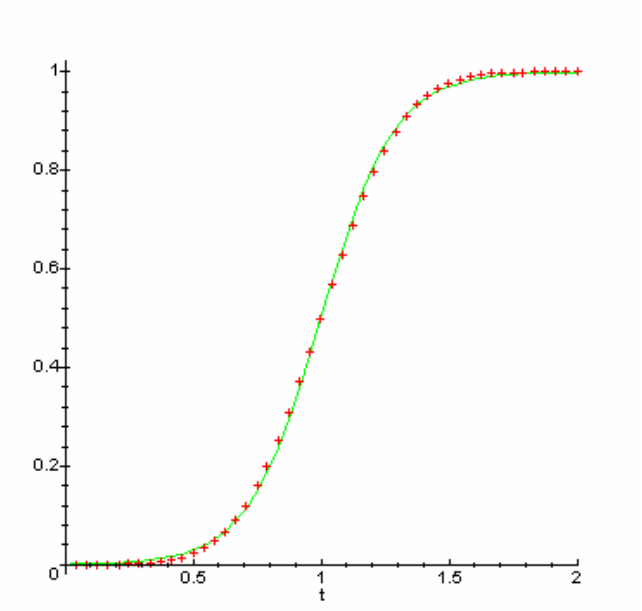

Figure 5.1: Sample plot of the cumulative Gaussian function (crosses) and the logistic function (line), demonstrating their similar shapes.

[1] Detailed derivations of the logit likelihood equations by Bierens can be found on a Pennsylvania State University webpage, http://econ.la.psu.edu/~hbierens/ML_LOGIT.PDF.

For low-amplitude signals in an AFC experiment, there is a minimum probability that the observer will choose correctly. This value is commonly referred to as the "guess rate", γ [5, 6]. It is also referred to by Barten as the "false alarm" rate (*FA*), and by Baker, who reported a modified logit expression for use in AFC experiments [4, 7] as shown below.

$$p(x) = \gamma + \frac{(1-\gamma)}{1+e^{-(\alpha+\beta x)}} \tag{5.2}$$

p = probability of a correct response (success)

x = contrast level

α = coefficient on the constant term

β = coefficient on the independent variable ("slope" coefficient)

γ = "guess rate" (minimum probability of success)

5.2.2 Adapting the logit model for a 2AFC experiment

For a 2AFC experiment, the probability of success for a signal at the contrast threshold level is 75% and the minimum probability of success, γ, is 50%. For this work, the variables (α, β) in Eq. (5.2) have been rearranged so that they relate to the contrast threshold, C_T, of the response function, along with a transition width variable, W. For a 2AFC experiment, C_T corresponds to x for $p(x) = 75\%$. W corresponds to the contrast range over which the probability of success goes from 63%-87%, or 26%-74% probability of detection. This translates to

$$C_T = -\alpha/\beta \tag{5.3}$$

and

$$W = 2/\beta. \tag{5.4}$$

This definition of W was chosen primarily to simplify the probability equation and provide an easily visualized parameter. However, it can also be related to the standard deviation, σ, of the psychometric function. Other works consider the width of the psychometric function in terms of the signal to noise ratio, such as

$$k = s_o/\sigma, \tag{5.5}$$

from Eq. (2.4) in Barten [7]. These variables can be approximated by C_T and W as

$$k = 1.28\left(\frac{C_T}{W}\right) \tag{5.6}$$

and

$$W = 1.28\sigma. \tag{5.7}$$

The derivation of these values is given in the Appendix, Section 5.6.1. Replacing α, β, and γ in the logit equation for the case of a 2AFC experiment gives the following equation.

$$p(x) = (0.5)[1 + \frac{1}{1 + e^{-2\left(\frac{x - C_T}{W}\right)}}] \tag{5.8}$$

p = probability of a correct response (success)

x = contrast level

C_T = "threshold" coefficient ($-\alpha / \beta$ = 75% probability of success)

W = "width" coefficient ($2 / \beta$ = width of the function from 63% to 87%)

Eq. (5.8) can be used to compute the likelihood where the conditional probability of a true response for a signal with amplitude x_i is $p(x_i)$ and the probability of a false response is $(1- p(x_i))$. The maximum likelihood based on the logit expression can be shown to be unique [8]. This allows one to use common search methods, such as the steepest descent and gradient search algorithms. In addition, finding the maximum of the log-likelihood can be shown to be equivalent to finding the maximum of the likelihood [9]. One of the advantages of using the logit expression in log-likelihood form is that the partial derivatives can be easily calculated. The log-likelihood and its partial derivatives have been derived using the chosen version of the logit model [see Eq. (5.9-5.11)]. The complete derivation for the log-likelihood equation can be found in the Appendix, Section 5.6.2.

Log-likelihood, $\ln(L_n)$:

$$\ln[L_n(C_T, W)] = \sum_{j=1}^{n} \{Y_j \ln[1 + 0.5e^{-2(x_j - C_T)/W}] + (1 - Y_j)[\ln(0.5) - \frac{2}{W}(x_j - C_T)] - \ln[1 + e^{-2(x_j - C_T)/W}]\} \tag{5.9}$$

Partial derivative, dL_n/dC_T:

$$\frac{d \ln[L_n(C_T, W)]}{dC_T} = \sum_{j=1}^{n} \{\frac{2}{W}[(1 - Y_j) + \frac{0.5Y_j e^{-2(x_j - C_T)/W}}{1 + 0.5e^{-2(x_j - C_T)/W}} - \frac{e^{-2(x_j - C_T)/W}}{1 + e^{-2(x_j - C_T)/W}}]\} \tag{5.10}$$

Partial derivative, dL_n/dW:

$$\frac{d\ln[L_n(C_T,W)]}{dW}=\sum_{j=1}^{n}\{\frac{2}{W^2}(x_j-C_T)[(1-Y_j)+\frac{0.5Y_j e^{-2(x_j-C_T)/W}}{1+0.5e^{-2(x_j-C_T)/W}}-\frac{e^{-2(x_j-C_T)/W}}{1+e^{-2(x_j-C_T)/W}}]\} \quad (5.11)$$

C_T = "threshold" coefficient (75% probability of success)

W = "width" coefficient (63% to 87%)

n = total number of observations

Y = observation binary result (0 = incorrect, 1 = correct)

Many numerical search algorithms can use the partial derivatives to help find the maximum quickly. For this work, a conjugate gradient search algorithm was used. This technique typically converges on the maximum faster than the method of steepest descent[10]. Specifically, we used the DUMCGG function from the IMSL Fortran 90 MP Library with an initial C_T of 1.0, an initial W of 0.4, an expected increment of 0.5, and a convergence value of 0.00000001.

It should be noted that traditional logistic regression functions, such as CTGLM from IMSL Fortran 90, are not applicable to this problem because they do not include a guess rate variable similar to γ or FA. The inclusion of the guess rate significantly alters the log-likelihood equation and its partial derivatives, which in turn affect the outcome of the MLE search algorithm.

5.3 Evaluation of the 2AFC-MLE Method Using Human Observer Data

The MLE method described above was first applied to the 2AFC human observer binary response data from the contrast threshold experiment detailed in Chapter 4. The experiment presented a large number of small bar patterns with varying contrast on a uniform background. A bar pattern was placed in one of two positions, left or right, and the observers were asked to identify which position the target was in. The observer results were evaluated as correct or incorrect. The order that the patterns of different contrast were displayed was random. Each contrast level was shown twice, once on the left and once on the right, thus ensuring that there were an equal number of targets shown on either side. A total of 306 images were shown to each observer during a test session.

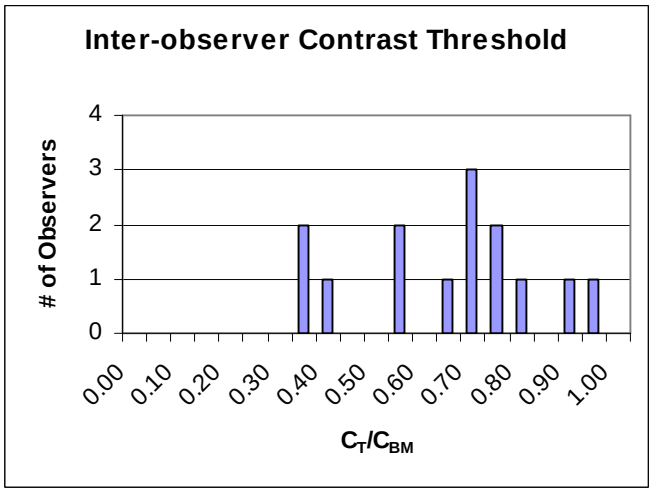

Figure 5.2: Histogram of the contrast threshold for 14 observers using the MLE method.

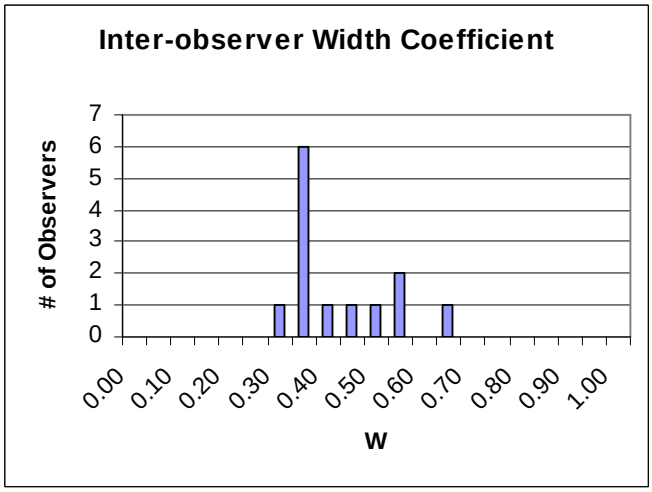

Figure 5.3: Histogram of the width parameter results for 14 observers using the MLE method.

5.3.1 Inter-observer results

Figure 5.2 shows a histogram of the results for the 2AFC contrast threshold for 14 observers who were tested once after having a training session. The mean contrast threshold was 0.79 (C_T/C_{BM}) with a standard deviation of 0.243. The minimum was 0.39 and the maximum was 1.157. The spread of C_T/C_{BM} values between the maximum and minimum were relatively uniform and did not appear to follow a Gaussian distribution.

Figure 5.3 shows a histogram of the observers' width parameters. The mean width (W) for the 14 initial observers was 0.45 cd/m^2 with a standard deviation of 0.23 cd/m^2. The minimum was 0.27 cd/m^2 and the maximum was 1.15 cd/m^2. The spread of

W values between the maximum and minimum were most densely concentrated around 0.35 cd/m^2, with the remaining values mostly being higher.

5.3.2 Intra-observer test results

For two of the 14 observers, referred to as PMT and MJF, the 2AFC test was repeatedly administered over a period of about one month.[2] Figures 5.4-5.5 show normalized log-likelihood contour plots for three tests of each observer. The contour lines represent equal changes in the log-likelihood. The results show that the log-likelihood computed for this experimental data has a single well-formed peak with a similar shape from test to test for the same observer. The results for all of the experiments for both observers were similar to the examples reported here. However, it should be noted that the location of the peaks changed slightly from one test to another.

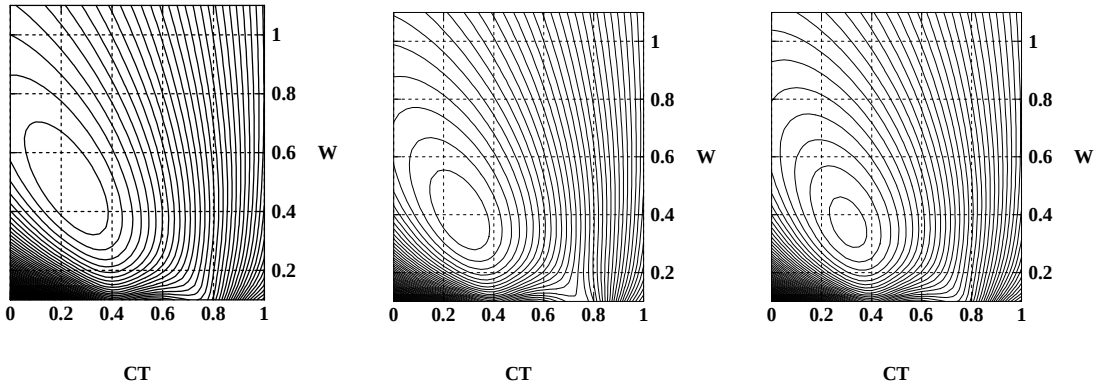

Figure 5.4: Normalized log-likelihood contour plots for MJF.

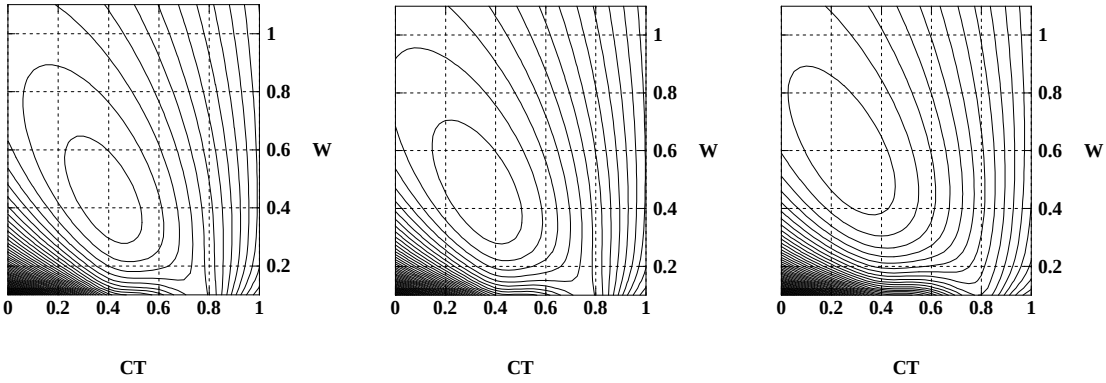

Figure 5.5: Normalized log-likelihood contour plots for PMT.

[2] One additional observer, referred to as HAS, was also part of the extended testing group but exhibited extremely low contrast threshold. This resulted in very few incorrect responses during several test sessions, causing the MLE software to return errors or produce unrealistic values. Thus, his contrast threshold response was too low to be effectively characterized by this experiment.

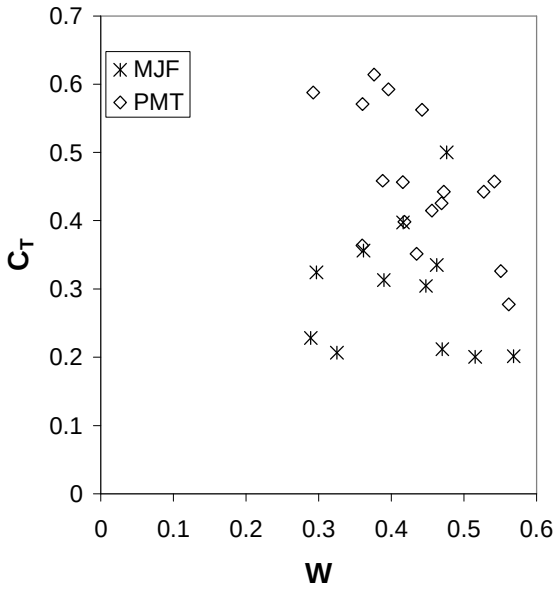

Figure 5.6: Contrast threshold vs. width for repeated tests administered on two observers.

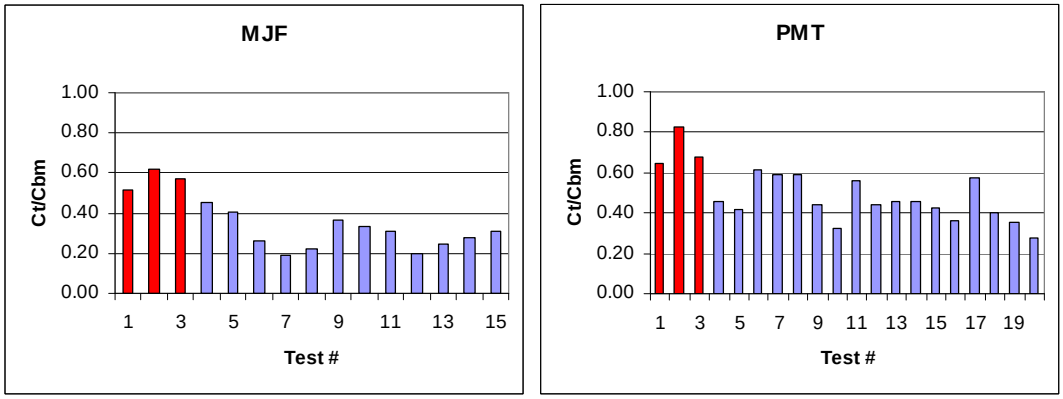

Figure 5.7: Intra-observer contrast threshold measurements.

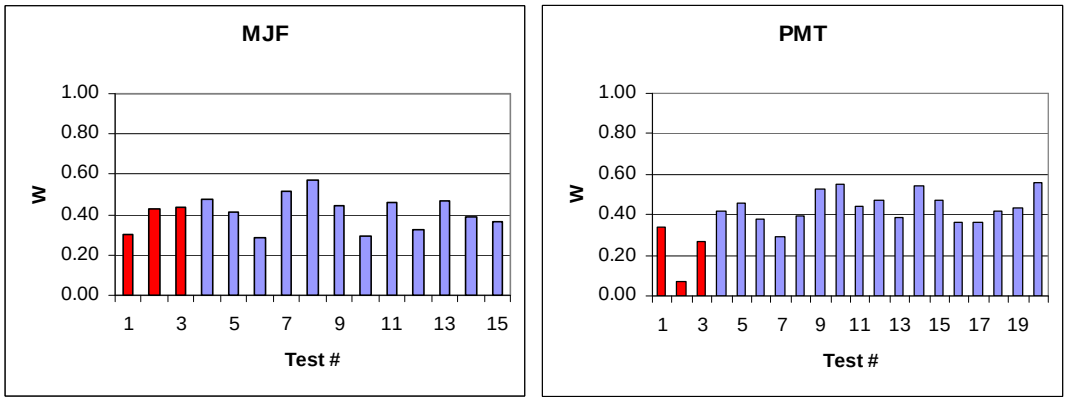

Figure 5.8: Intra-observer width measurements.

Table 5.1: Comparison of the MLE and Z-score contrast threshold results.

	Mean (C_T/C_{BM})	Standard Deviation
Z-score:		
Inter-observers	0.67581	0.20210
MJF	0.37975	0.11902
PMT	0.52010	0.09824
HAS	0.26632	0.08534
MLE:		
Inter-observers	0.79241	0.24324
MJF	0.37801	0.10492
PMT	0.57992	0.12742
HAS	indeterminate	indeterminate
DIFFERENCE: (MLE - Z-score)		
Initial Observers	0.1166	0.04114
MJF	-0.00174	-0.0141
PMT	0.05982	0.02918
HAS	indeterminate	indeterminate

For each test, both the contrast threshold and width were estimated. Figure 5.6 plots W versus C_T over all of these tests. This illustrates how the two variables are distributed and suggests little correlation between W and C_T. The distributions of points for the two observers are noticeably different along the C_T axis and similar for W.

Figures 5.7-5.8 show the contrast threshold and width results for the two observers using the MLE method. Evidence for a learning effect was seen in the first three tests and these were not used to determine the average or standard deviation. For MJF, the mean contrast threshold was 0.38 (C_T/C_{BM}) with a standard deviation of 0.105. For PMT, the mean contrast threshold was 0.58 (C_T/C_{BM}) with a standard deviation of 0.127. A t-test analysis performed on these results gave a probability of 0.00416 that the thresholds were from similar distributions. For the width, MJF had a mean of 0.42 (cd/m^2) with a standard deviation of 0.0880. PMT had a mean width of 0.44 (cd/m^2) with a standard deviation of 0.0756. The t-test analysis gave a probability of 0.880 that the widths were from similar distributions.

5.3.3 Comparison of the MLE and z-score results

Table 5.1 compares the inter-observer and intra-observer results obtained using the MLE method to the z-score analysis results reported in Chapter 4. The mean C_T/C_{BM} result for the inter-observers is larger with the MLE method by approximately 0.12 and

the standard deviation is larger by about 0.04. For intra-observer PMT, C_T/C_{BM} increased by 0.06 and the standard deviation by 0.03. For intra-observer MJF, the C_T/C_{BM} and standard deviation values showed little change.

As stated before in Chapter 4, a result of ~0.66 (C_T/C_{BM}) is expected for a 2AFC experiment [11]. For the inter-observer MLE result (0.79), this is still well within one standard deviation (0.243) of the predicted result. Therefore, the overall conclusions remain the same. The 2AFC method and the monochrome LCD and medical graphic card used for this experiment result in a C_T/C_{BM} consistent with the Barten model under the conditions laid out in DICOM 3.14. In addition, these results continue to support the use of LCD monitors for medical perception research.

5.4 Evaluation of the 2AFC-MLE Method Using Simulations

5.4.1 Simulation methods

Simulated 2AFC experiments were used to determine the variation in the estimates of C_T and W as a function of the number of test responses and the distribution of signal amplitudes relative to the threshold. A random number generator[3] was used to generate simulated responses for a 2AFC test. The logistic expression given in Eq. (5.8) was used for the response probability function. The resulting data sets were then analyzed using the maximum likelihood method to determine the estimated threshold and width parameters.

The first series of simulations examined the error in identifying the contrast threshold as a factor of the number of observations, N_{obs}. The contrast parameter in Eq. (5.8) was set to 1.0 and the width parameter was set to 0.4. These values were close to those obtained from human observer tests, which will be discussed in Chapter 4. The signal amplitudes were evenly spaced between values of 0 and 2. The MLE search program was constrained to signal values from 0 to 2 and used [1,0.4] for its starting values. Simulated observer response experiments were generated for N_{obs} = 10, 30, 100, 300, 1000, 3000, 10000, and 30000. For each value of N_{obs}, 100 simulations experiments were used to determine the mean and standard deviation of C_T and W.

[3] RANMAR, proposed by Marsaglia and Zaman in report FSU-SCRI-87-50, modified by F. James, 1988 and 1989, and adapted by Aldo Badano, 1/99.

The second series of simulations examined how the error changed with the signal range of the observations and the width parameter. Each test used 300 observations. The range of the observations was varied in its width and in its position relative to the contrast threshold. Three positions were tested: centered about the threshold, 75% to the left of the threshold, and 75% to the right of the threshold. The width of the signal range over which the observations were taken varied from 0.1 to 4, in increments of 0.1. Those ranges that would require negative signal values were tossed out. As a result, the centered data sets had a maximum width of 2, and the mostly left data sets had a maximum width of 1.3. The width parameter was set to 0.2, 0.3, 0.4, and 0.5.

5.4.2 Simulation results

Figure 5.9 shows the results of the simulations and the measured error for different numbers of observations (N_{obs}) during the 2AFC contrast threshold test. The standard deviation was calculated for the 75% and 85% success levels, as well as the width coefficient (W). The standard deviation values for all three exhibited nearly identical behavior, decreasing logarithmically at an approximate rate of $N_{obs}^{-0.5}$. A non-linear regression to the contrast threshold data indicates a standard deviation of $\sigma_c = 1.5201 N_{obs}^{-0.5236}$. The fact that this MLE method follows the $N_{obs}^{-0.5}$ relationship down to a small sample size is an indication that it is robust.

Figures 5.10-5.13 show the standard deviations of the contrast threshold for the centered, left-shifted, and right-shifted distributions as the signal range over which the observations were taken increased. All of the cases dropped off quickly at first as the range increased, and then began to rise again before asymptotically approaching a value around 0.1. As W increased, the standard deviations appeared to shift upward and to the right. The centered distributions generally had the lowest standard deviation values, except for the case of W=0.2. The minimum standard deviations for the centered distributions were found for signal ranges near or slightly larger than W. The left and right shifted simulations show some erratic behavior for small ranges. These represent test sets that do not include many points at either very high or very low probabilities of success.

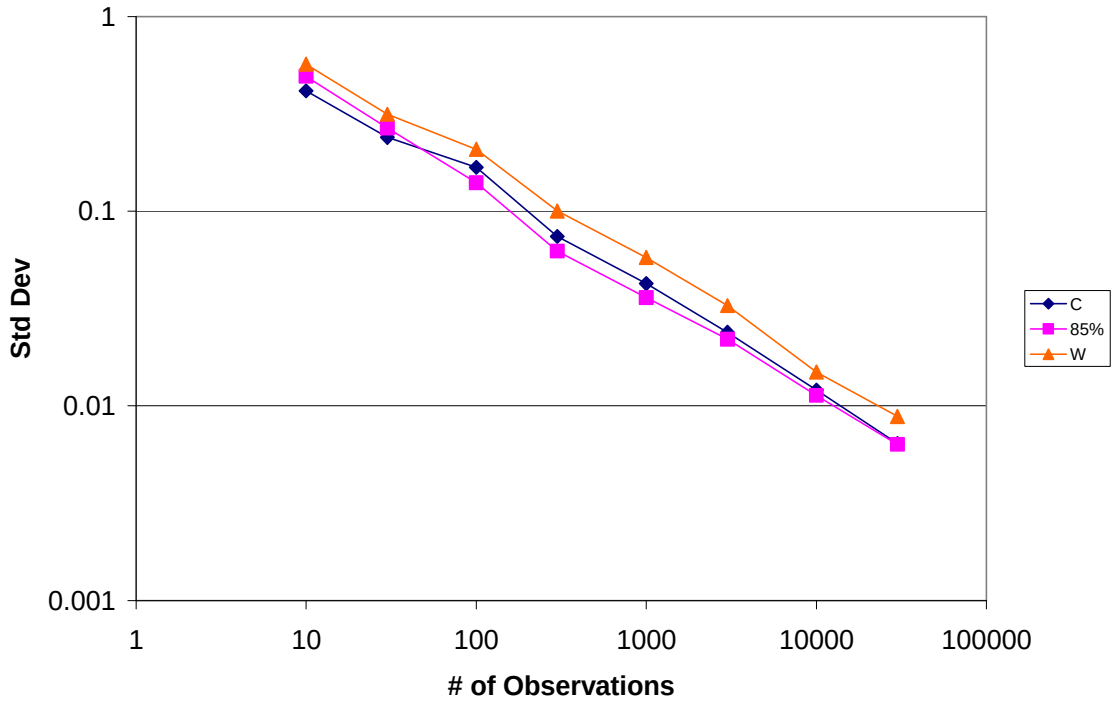

Figure 5.9: Standard deviation vs. number of observations for C_T=1 and W=0.4.

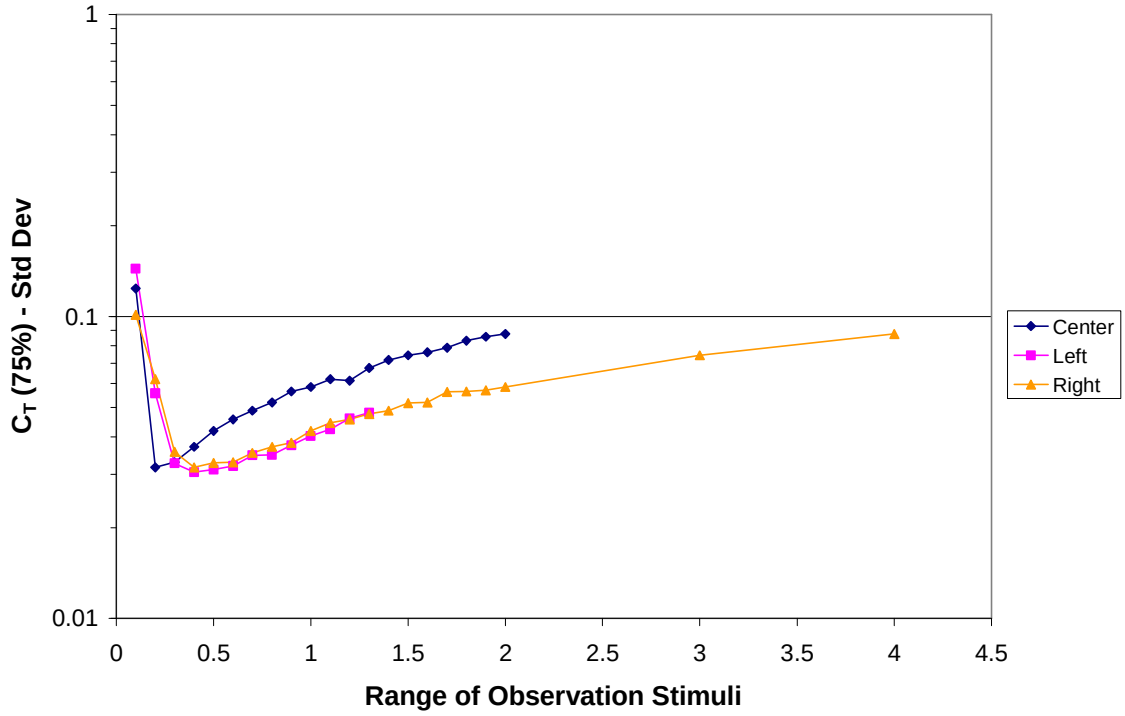

Figure 5.10: Standard deviation vs. range of observations for C_T=1 and W=0.2.

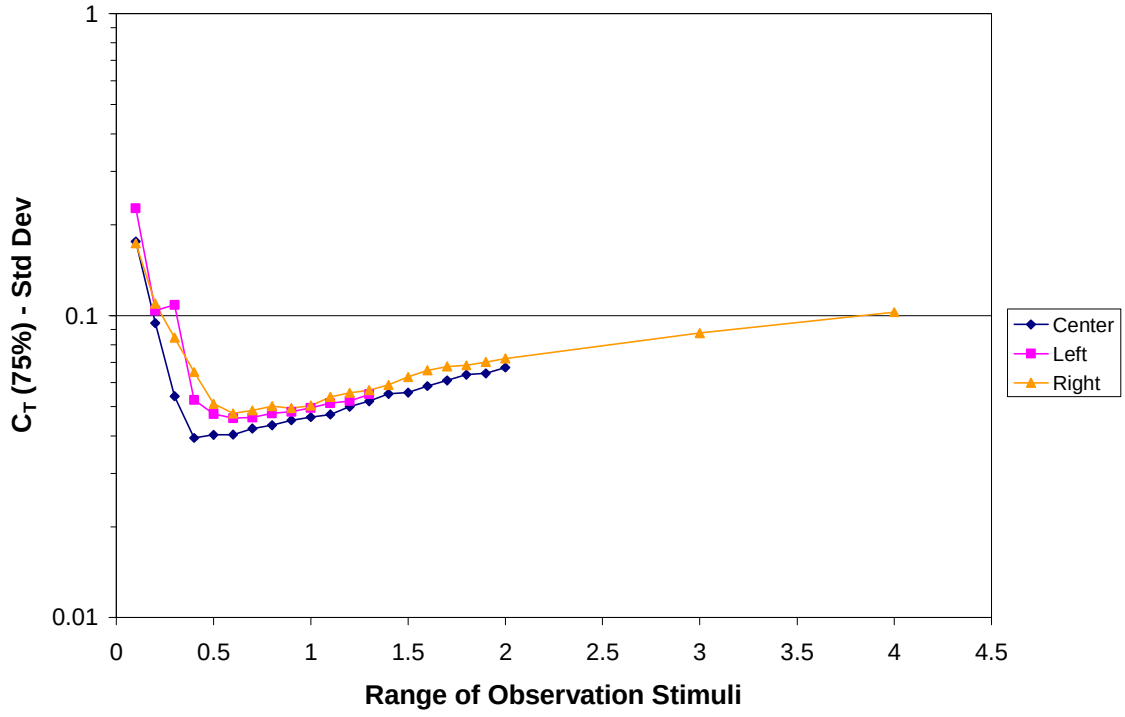

Figure 5.11: Standard deviation vs. range of observations for C_T=1 and W=0.3.

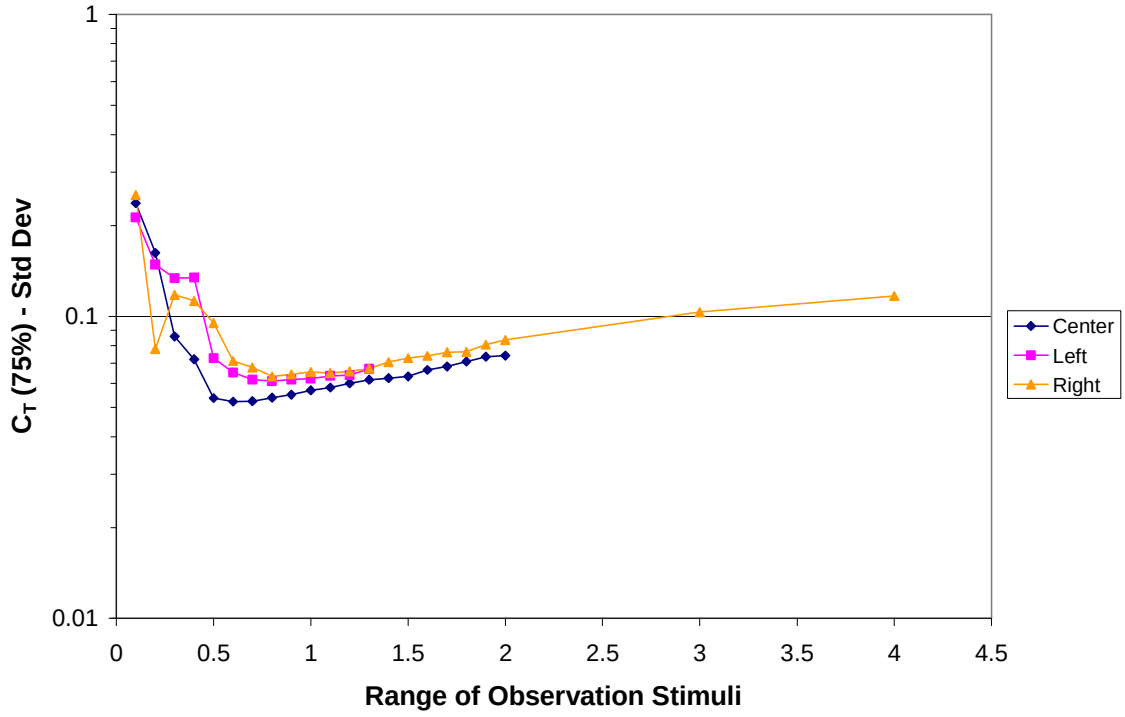

Figure 5.12: Standard deviation vs. range of observations for C_T=1 and W=0.4.

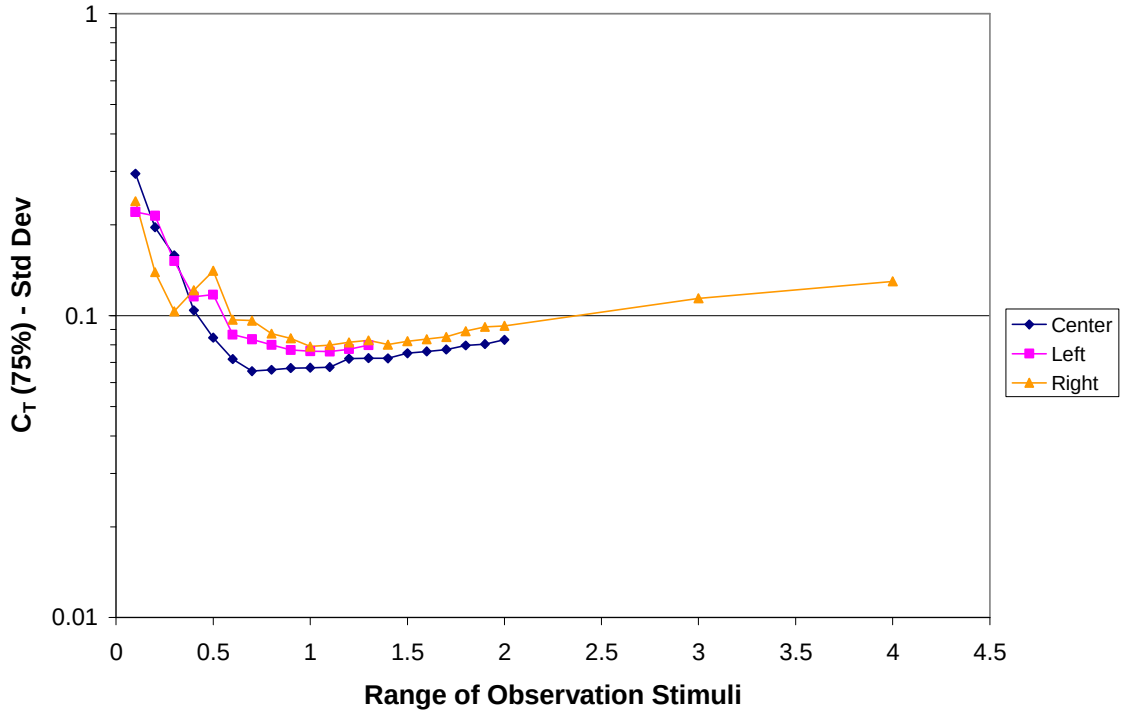

Figure 5.13: Standard deviation vs. range of observations for C_T=1 and W=0.5.

5.4.3 Adaptive testing

AFC testing can be very time consuming. Enough data must be collected over a broad range of signal values to allow for a good estimate of the psychometric curve. Enough time must also be given for the subjects to make rational and consistent decisions throughout the testing session. Extensive testing runs the risk of subject fatigue, possibly leading to misjudgments or hurried responses. As such, it is often not feasible to measure observer performance over all available signal levels [5, 6]. One possible solution is to use adaptive testing.

Adaptive testing is a method of progressively choosing which signal values to test, based on the subject's earlier responses. This allows for fewer trials in order to get the specific data one is looking for. For example, an adaptive staircase method used in a psychovisual threshold test would increase or decrease the signal strength based on the observer's previous responses, narrowing the range of values to only those that are close to the observer's threshold. This adaptive reevaluation process is often repeated after

every trial until specific stopping criteria have been met, usually based on the level of precision or a fixed maximum number of trials.

Adaptive testing was not used during the initial 2AFC contrast threshold tests (see Chapter 4) since such methods had not yet been considered. They were also not implemented in the later experiments discussed in Chapter 6, as it was decided that further research was needed. However, the results from the simulations in Section 5.4, discussed in the context of the contrast threshold tests, are intended to help develop adaptive test procedures for later experiments.

5.5 Discussion

Limitations of the z-score method have been presented and an MLE method proposed as an alternative. The concepts and equations behind the MLE method have been explained and used to model a 2AFC experiment. Finally, the results of contrast threshold measurements previously analyzed using a z-score approach (see Chapter 4) have been re-evaluated using the MLE method. Some changes have been noted, but the overall conclusions have remained the same.

It has been common to analyze the responses from an AFC experiment by determining the percent correct for identical stimuli, transforming percent correct values to z-scores, fitting them to a linear function versus signal (amplitude, energy, or signal to noise ratio), and deducing a threshold signal from the linear fit. This method is subject to the same error encountered whenever non-linear data is transformed in order to use a linear regression. The transformation distorts the experimental error. Secondly, this method is not applicable to experiments where the signal stimuli are broadly distributed over a large range of values, since grouping responses to determine percent correct causes a shift in the signal amplitude of the individual responses. The maximum likelihood method developed using the logit model and a conjugate gradient search algorithm addresses these problems. Since each individual response is analyzed as a binary (true/false) result in relation to the signal without binning, the set of signals used in an experiment can be arbitrarily distributed. The MLE method is used to directly fit a non-linear psychometric function to the binary data and thus avoids error distortion from

linearization. No previous 2AFC medical imaging studies have used the MLE method, and most new studies still use the z-score method.

The MLE method reported here determines the width of the assumed psychometric function, W, in addition to the contrast threshold, C_T. Figure 5.6 suggests that there is little correlation between these two values. Similar behavior was observed for the simulation data. Thus, for purposes of the MLE method, constraining W to an assumed value may not significantly improve the variability in the estimate of C_T. While the values for C_T were significantly different for the two human observers, the values of W equal to 0.42 and 0.44 were not. Other work considers differences in the width of the psychometric function in terms of the minimum signal to noise ratio required for detection, $k = s_o/\sigma$, as in Eq. (2.4) of Barten [7]. The ratio C_T/W in this work is approximately equivalent to $0.78k$ (or $0.99k$ if correcting for a square wave pattern). For the simulations, the C_T/W ratio of 2.5 corresponds to a signal to noise ratio of 3.2. For the two extended human observers, the observed C_T/W ratios were in the range of about 1.0 to 1.3. These relatively low values may be due to the noiseless character of the visual scene, which had a mostly flat background and very limited features.

The standard deviation values for PMT and MJF were close to the standard deviation values predicted by the simulations, although the simulations used $C_T = 1$, while the human observer threshold was closer to 0.4. An additional simulation was run to examine this difference using characteristics similar to the observer tests. For $C_T = 0.4$, $W = 0.4$, and 306 observations, the standard deviation was 0.079294 for C_T/C_{BM} and 0.087190 for W. These results were also similar to those of the human extended observers. The predicted standard deviations for $C_T = 0.4$ were not much different than for the simulations at $C_T = 1$, indicating that translation of the psychometric function does not significantly effect the predicted variance. These results provide some evidence that this simulation method can predict the response of human observers and can provide a guide for the design of new studies.

The initial simulations show that the standard error for C_T and W decreases with the inverse of the square root of the number of observations made. The fact that this MLE method follows the $N_{obs}^{-0.5}$ relationship down to a small sample size is an indication that it is robust. The simulations of various ranges of observations indicate that centering

the contrast levels about C_T produces the least error when the range of contrast is near or slightly larger than W. Since the exact contrast threshold of an observer is unknown, adaptive testing could be used to determine the signal values to focus on. However, since a range of values is desired to best determine C_T rather than simply focusing on it as quickly as possible, it would be best not to continuously reevaluate the parameters after every trial, as most adaptive strategies do. Instead, a number of initial contrast measurements would be made to obtain a preliminary MLE result for C_T around which the remaining trials would be centered. This strategy for adaptive testing has not been seen in other psychovisual contrast threshold tests. Those that have used adaptive testing have typically used the staircase method, which increases or decreases the range of contrast based on a specific number of correct/incorrect responses in sequence.

Similar measurements could be made to estimate W, providing a range of signal values to be tested around C_T. Alternatively, if it is assumed that W is similar between observers, it could simply be set to a fixed value. This has been done in other psychovisual experiments [5]. However, one can speculate that the slope of the psychovisual curve may be associated with noise in the vision system and may vary with different backgrounds or visual tasks.

5.6 Appendix

5.6.1 Relating C_T and W to the Barten Model

As described in Eq (1.1), the Barten Model defines the psychometric function as a cumulative Gaussian probability function, shown here again for reference:

$$p(s) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{s} e^{-\frac{(x-s_0)^2}{2\sigma^2}}\, dx \tag{5.12}$$

p = detection probability

s = signal strength

s_0 = signal strength for a detection probability of 50%

σ = standard deviation of the Gaussian distribution

We can simplify the equation by inserting

$$z = \frac{s - s_0}{\sigma}. \tag{5.13}$$

This gives us a Gaussian equation that is much easier to evaluate,

$$p(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}} dx . \tag{5.14}$$

By referencing standard mathematical tables, we can solve for *z* at the more positive edge of *W*, or $p(z) = 74\%$, getting a value of approximately 0.64. We can then combine this with Eq. (5.5) and Eq. (5.13) to solve for *k* in terms of C_T and *W*, with a result of

$$k = 1.28\left(\frac{C_T}{W}\right). \tag{5.15}$$

We can also put Eq. (5.5) back in for *k* and get

$$W = 1.28\sigma. \tag{5.16}$$

These expressions appears earlier in the paper as Eq. (5.6) and Eq. (5.7).

5.6.2 Derivation of the 2AFC-MLE equations

This section details the derivation of the maximum log-likelihood equations for a 2AFC experiment using the logistic regression model (logit). This derivation is a modified form of that found in "*The Logit Model: Estimation, Testing, and Interpretation*", by Herman J. Bierens [8]. For 2AFC experiments, the difference is the inclusion of a minimum probability (i.e. "false alarm") value of 0.5 to account for the chance of guessing the correct answer in a 2AFC experiment. We begin with the probability equation for the logistic model under these conditions. This equation is the same as Eq. (5.2) given above.

$$P(Y_j = 1 \mid X_j) = F(C_T, W, x) = (0.5)[1 + \frac{1}{1 + e^{-2(x_j - C_T)/W}}] \tag{5.17}$$

Y = binary result X = independent variable

C_T = "threshold" coefficient W = "width" coefficient

Next, given the sample set (Y_1,X_1), ...,(Y_n,X_n), we define the conditional probability function

$$f(y \mid X_j, C_T, W) = F(C_T, W, X_j)^y [1 - F(C_T, W, X_j)]^{1-y} \tag{5.18}$$

the conditional likelihood function

$$L_n(C_T, W) = \prod_{j=1}^{n} f(Y_j \mid X_j, C_T, W) \tag{5.19}$$

and the conditional log-likelihood function

$$\ln[L_n(C_T,W)] = \sum_{j=1}^{n} \ln[\, f(Y_j \,|\, X_j, C_T, W)] \tag{5.20}$$

$$= \sum_{j=1}^{n} \ln[F(C_T,W,X_j)^{Y_j}] + \sum_{j=1}^{n} \ln\{[1-F(C_T,W,X_j)]^{1-Y_j}\}$$

$$= \sum_{j=1}^{n} Y_j \ln[F(C_T,W,X_j)] + \sum_{j=1}^{n} (1-Y_j) \ln[1-F(C_T,W,X_j)]$$

$$= \sum_{j=1}^{n} Y_j \ln\{(0.5)[1+\frac{1}{1+e^{-2(x_j-C_T)/W}}]\} + \sum_{j=1}^{n} (1-Y_j) \ln\{1-(0.5)[1+\frac{1}{1+e^{-2(x_j-C_T)/W}}]\}$$

$$= \sum_{j=1}^{n} Y_j \ln[(0.5)\frac{2+e^{-2(x_j-C_T)/W}}{1+e^{-2(x_j-C_T)/W}}] + \sum_{j=1}^{n} (1-Y_j) \ln[(0.5)\frac{e^{-2(x_j-C_T)/W}}{1+e^{-2(x_j-C_T)/W}}]$$

$$= \sum_{j=1}^{n} Y_j \{\ln[1+0.5e^{-2(x_j-C_T)/W}] - \ln[1+e^{-2(x_j-C_T)/W}]\} + \sum_{j=1}^{n} (1-Y_j) \{\ln[0.5e^{-2(x_j-C_T)/W}] - \ln[1+e^{-2(x_j-C_T)/W}]\}$$

$$= \sum_{j=1}^{n} Y_j \{\ln[1+0.5e^{-2(x_j-C_T)/W}]\} + \sum_{j=1}^{n} (1-Y_j) \ln[0.5e^{-2(x_j-C_T)/W}] - \sum_{j=1}^{n} \ln[1+e^{-2(x_j-C_T)/W}]$$

$$= \sum_{j=1}^{n} Y_j \{\ln[1+0.5e^{-2(x_j-C_T)/W}]\} + \sum_{j=1}^{n} (1-Y_j)[\ln(0.5) - 2(x_j-C_T)/W] - \sum_{j=1}^{n} \ln[1+e^{-2(x_j-C_T)/W}]$$

$$= \sum_{j=1}^{n} \{Y_j \ln[1+0.5e^{-2(x_j-C_T)/W}] + (1-Y_j)[\ln(0.5) - 2(x_j-C_T)/W] - \ln[1+e^{-2(x_j-C_T)/W}].$$

This expression appears earlier in the paper as Eq. (5.4). The partial derivatives of this expression with respect to C_T and W, Eq. (5.10) and (5.11) respectively, are otherwise straightforward to derive. They are repeated here for reference:

$$\frac{d\ln[L_n(C_T,W)]}{dC_T} = \sum_{j=1}^{n} \{\frac{2}{W}[(1-Y_j) + \frac{0.5Y_j e^{-2(x_j-C_T)/W}}{1+0.5e^{-2(x_j-C_T)/W}} - \frac{e^{-2(x_j-C_T)/W}}{1+e^{-2(x_j-C_T)/W}}]\} \tag{5.21}$$

$$\frac{d\ln[L_n(C_T,W)]}{dW} = \sum_{j=1}^{n} \{\frac{2}{W^2}(x_j - C_T)[(1-Y_j) + \frac{0.5Y_j e^{-2(x_j-C_T)/W}}{1+0.5e^{-2(x_j-C_T)/W}} - \frac{e^{-2(x_j-C_T)/W}}{1+e^{-2(x_j-C_T)/W}}]\} \tag{5.22}$$

5.7 References

1. Wichmann, F.A. and N.J. Hill, *The psychometric function: II. Bootstrap-based confidence intervals and sampling.* Percept Psychophys, 2001. **63**(8): p. 1314-29.
2. Klein, S.A., *Measuring, estimating, and understanding the psychometric function: a commentary.* Percept Psychophys, 2001. **63**(8): p. 1421-55.
3. Burgess, A.E., *Comparison of receiver operating characteristic and forced choice observer performance measurement methods.* Medical Physics, 1995. **22**(5): p. 643-55.
4. Baker, R.J., S. Rosen, and A. Godrich, *No Right Ear Advantage in Gap Detection.* Speech, Hearing and Language: work in progress, 2000. **12**: p. 57-69.
5. Linschoten, M.R., et al., *Fast and accurate measurement of taste and smell thresholds using a maximum-likelihood adaptive staircase procedure.* Percept Psychophys, 2001. **63**(8): p. 1330-47.
6. Park, S., et al., *Efficiency of the human observer detecting random signals in random backgrounds.* J Opt Soc Am A Opt Image Sci Vis, 2005. **22**(1): p. 3-16.
7. Barten, P.G.J., *Contrast sensitivity of the human eye and its effects on image quality*. 1999, Bellingham, Wash.: SPIE Optical Engineering Press. xix, 208.
8. Bierens, H.J., *The Logit Model: Estimation, Testing and Interpretation*. 2004, Pennsylvania State University, Department of Economics.
9. Bierens, H.J., *Introduction to the Mathematical and Statistical Foundations of Econometrics*. Themes in Modern Econometrics, ed. P.C.B. Philips. 2004: Cambridge University Press.
10. Press, W.H., *Numerical recipes in C : the art of scientific computing*. 2nd ed. 1992, Cambridge [England] ; New York: Cambridge University Press. xxvi, 994.
11. Legge, G.E., *A power law for contrast discrimination.* Vision Res, 1981. **21**(4): p. 457-67.

CHAPTER 6

TWO-ALTERNATIVE FORCED CHOICE MEASUREMENT OF CONTRAST THRESHOLD FOR COMPLEX PATTERNS PRESENTED ON A MEDICAL LCD MONITOR

6.1 Introduction

Medical radiographic images typically have complex spatial characteristics and a wide range of luminance values. Observer performance experiments examining patterns found in these images are needed to evaluate the suitability of LCDs for medical interpretation. In order to do this, methods need be developed to present complex patterns with precise luminance and subtle contrast. In this chapter, a method is reported to use a 1786 grayscale palette and a narrow luminance calibration to present complex low-contrast patterns on a medical LCD system. The performance of the method is validated using two-alternative forced choice (2AFC) experiments employing patterns with sinusoidal modulation. New results are reported for observer thresholds measured for sinusoid patterns having different background luminance and for noise patterns emulating radiographic quantum mottle. These results are analyzed relative to the Barten model and the contrast of the DICOM Gray Scale Display Function.

Chapter 4 described a 2AFC test to measure the contrast threshold of the human visual system when viewing bi-level bar patterns. The target patterns were based on the DICOM Standard Target[1], and the observer results were found to coincide with the Barten model of contrast sensitivity[2]. The software used in Chapter 4 could produce extremely small contrast for patterns with two gray levels, but it did not have the ability to generate patterns with many gray levels. Producing complex patterns having a particular contrast on an LCD system requires that they be shown using a set of gray display states that are perceived as smoothly varying in brightness. In this chapter, methods are reported to conduct 2AFC contrast threshold tests using complex test

patterns. The method is demonstrated using sinusoidal patterns and patterns emulating the quantum noise seen in digital radiographs.

The general method used is described in section 6.2 below. A color medical LCD was used to obtain fine contrast steps from a palette of 1786 gray values. A calibration with a narrow luminance range was used to produce small luminance changes between each gray level. However, because of the 8-bit grayscale associated with display devices using Windows operating systems, the minimum contrast that could be achieved was not as small as for the bi-level patterns used in Chapter 4. To validate the method, a 2AFC experiment was done using test patterns with sinusoidal modulation of a large area, medium gray (9 cd/m^2) background. The results agreed with the larger contrast threshold that is expected for the human vision system at lower brightness.

While bar and sinusoid patterns are traditional for measures of contrast threshold, they are not representative of image features seen in medical radiographs. Radiographic images include complex luminance patterns from anatomic structures, disease features such as lesions and nodules, and noise from various stages of the imaging process. Section 6.2.2.2 describes the generation of simulated radiographic quantum mottle, a type of noise caused by the random distribution of the finite number of x-ray quanta that form the image[3]. A “white noise” pattern simulates the appearance of quantum mottle seen in direct digital radiography systems, where the x-rays are converted directly into an electric charge. White noise is defined as having equal power at all spatial frequencies[4]. A “filtered noise” pattern has reduced high-frequency components. This gives the noise a more structured or “clumped” look, which is typical of images from indirect digital radiography where the x-rays are first converted to light by a scintillator before being converted to electric charge. A previous 2AFC study of noise perception was conducted by Luijendijk in 1994[5]. However, that experiment considered noise in a radiograph with anatomic structures, and the images were viewed on a CRT monitor. No prior perception studies have measured the contrast threshold for noise targets using medical LCDs.

The use of uniform backgrounds equal to the average target luminance is also not typical of the conditions seen in medical radiographs. Medical radiographs contain a wide range of luminance, with contrast features appearing in areas of local brightness that

can vary significantly from the average luminance of the image. Observers adapted to the average luminance of the image will perceive contrast in the brighter and darker regions differently[6]. As presented in Chapter 2, when viewing an image, the human visual system adapts to the average quantity of light falling on the retina. This is referred to as "fixed adaptation"[7]. The adaptation process results in a shift in the neuronal response to input luminance[8]. Baxter[9] reported a relationship for photo-receptor sensitivity in relationship to luminance for particular levels of adaptation that was based on Hecht's photochemical theory of photo-receptor response[10] and Baylor's experiments[11]. Flynn differentiated this expression to describe a biological contrast transfer function[12]. Figure 6.1, also presented in Chapter 2, shows how this biological contrast transfer affects contrast threshold for a particular adaptation level[13]. Line "B" shows the contrast threshold as a function of luminance, given that the eye has had time to adapt to each luminance level. Line "A" shows the expected contrast threshold for fixed adaptation. After adaptation, the human visual system's perception of contrast drops off in brighter and darker regions, constraining the useful luminance range.

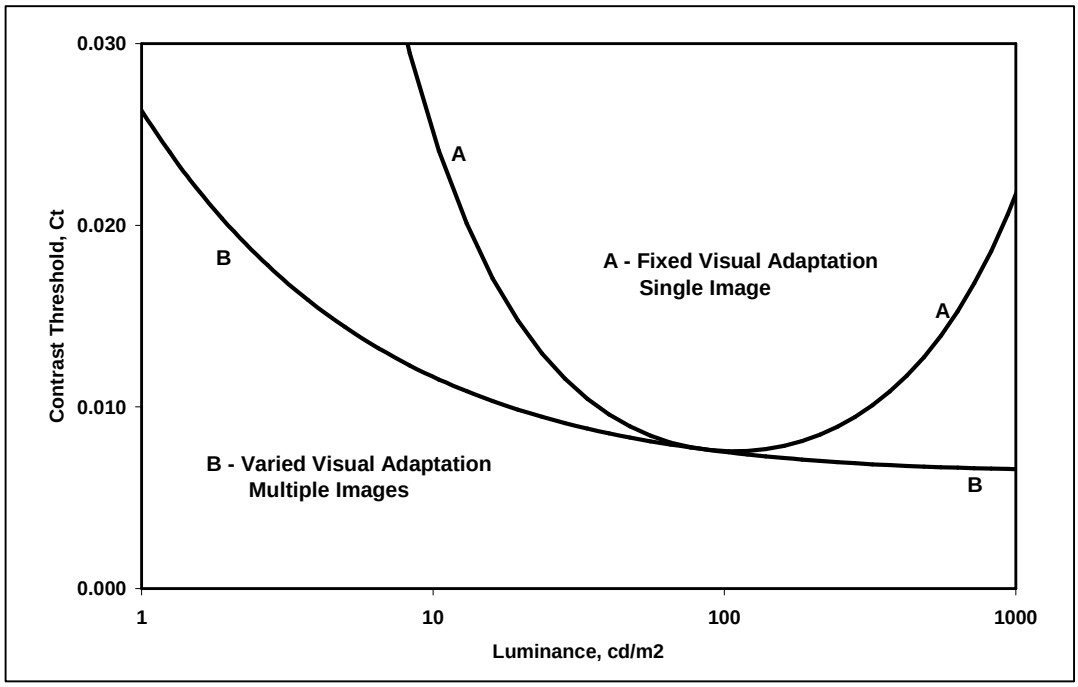

Figure 6.1: Contrast threshold predicted for conditions where the observer is globally adapted at 100 cd/m2 to a single image (A). This response is compared to the contrast threshold for an observer who is variably adapted to a series of images with different luminances (B).

In the medical literature, the only experimental results confirming the effect of fixed adaptation used projected images with the test pattern contrast varied until the observer reported target visibility[9]. This chapter provides experimental evidence using medical LCD devices and a 2AFC study design.

6.2 Methods

The techniques used in this experiment include a combination of methods presented in Chapters 3-5. The calibration methods from Chapter 3 were used to provide a narrow calibrated luminance range to control the contrast levels within the test images. The 2AFC testing procedures were carried over directly from Chapter 4 to measure human visual perception. However, a color monitor with smaller pixel pitch was used (color, 2048 x 1536) and new software was developed to create the image patterns and present them in 2AFC experiments. Finally, the maximum likelihood estimation (MLE) method presented in Chapter 5 was used to analyze the results and compute the contrast threshold and width parameters. These methods are applied in new observer experiments to measure the contrast threshold for complex test patterns with a full scale of gray levels.

Seven separate observer tests were conducted, each one focused on measuring a particular perception threshold under different conditions. Each test presented a large number of images (~200) with test patterns on uniform backgrounds to human observers. A sequence of images was presented with the test pattern placed in one of two positions and the observers asked to identify the position that contained the target. The tests were conducted using sinusoid, white noise, and filtered noise patterns. Three of the tests, one for each type of pattern, used background luminance that was kept equal to the average luminance of the target pattern. For the four remaining tests, the sinusoid patterns were repeated, but the image backgrounds outside of the target regions were shifted in luminance away from the target average. The display used was an NDS AC-QX21-AC9300 color LCD monitor with an NVIDIA Quadro FX 4400 graphics controller. Each observer's results were analyzed using an MLE algorithm to determine their psychovisual performance relative to target contrast. These values were also compared to the responses predicted by the Barten model and the DICOM standard.

6.2.1 Low contrast calibration

In Chapter 4, the graphic software generated the test images while each observer took the test. Images were drawn directly by specifying two RGB values for the background and the bars of the pattern. While this method was effective for bi-level bar patterns, the software was not designed to quickly present complex patterns using a calibrated grayscale. Instead, new software was developed to rapidly present images stored in a directory and record observer scores. For each test given to an observer, images are presented in a new randomized order. A specialized calibration LUT was applied in advance to control the range of luminance values and to ensure that changes in contrast were smoothly varying throughout.

Images with medium gray background regions of about 10 cd/m^2 were used because the contrast threshold predicted by the Barten model increases as luminance decreases. For the test patterns used to derive the DICOM GSDF, the contrast threshold at 10 cd/m^2 is 1.66 times larger than at 300 cd/m^2. By using a darker image, more gray level changes could be obtained that were below the contrast threshold. Since human vision systems perceive brightness in approximate proportion to the log of the luminance, this appears on the screen as a medium gray.

Medical monitors are typically calibrated using a luminance ratio of 350, for which the relative luminance change, $\Delta L/L$, for sequential gray levels is more than twice that of the $\Delta L/L$ associated with the just noticeable difference (JND) used in the DICOM Grayscale Standard Display Function (GSDF). In part 3.14 of the DICOM standard, this ratio is referred to as JNDs per driving level, JNDs/DL. For psychovisual performance tests, it is desirable to have the JNDs/DL be less than 1. This was accomplished by calibrating the LCD display for a very small luminance ratio using the procedures detailed in Chapter 3. The relation between JNDs/DL and luminance ratio is shown in Figure 6.2. Since a color display was used, luminance measurements were made for 1786 digital driving levels (DDLs). The maximum luminance was set to 50 cd/m^2 with a luminance ratio of only 10. This produced 256 calibrated gray levels (or 255 DLs) with very fine contrast differences. This calibration produced a contrast range of 226 JNDs with an average of approximately 0.887 JNDs/DL.

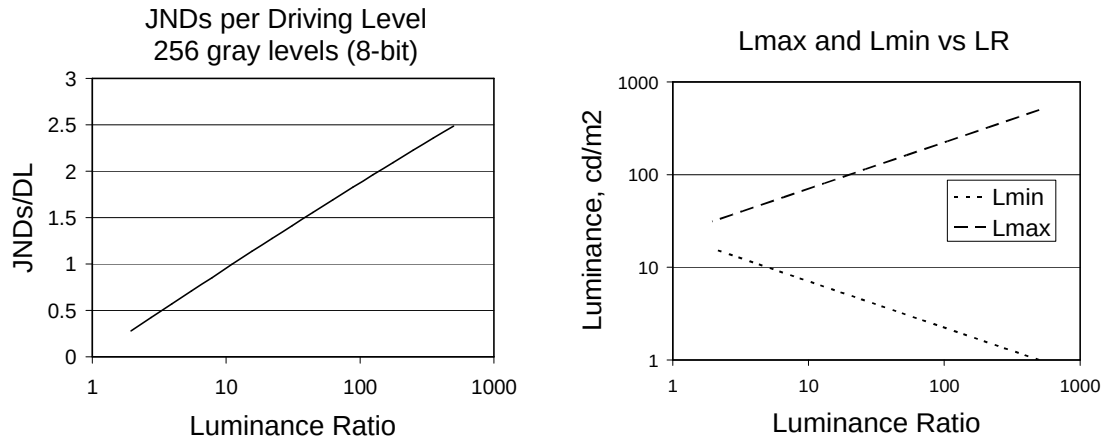

Figure 6.2: The plot on the left shows the JNDs per driving level in relation to the luminance ratio. The maximum and minimum luminance were respectively increased and decreased by factors of $2^{1/4}$ to vary the luminance ratio, as shown in the plot on the right.

With the 2AFC LUT applied, the luminance response was experimentally measured for all 256 gray levels and the JND/DL evaluated for the 255 gray level transitions. Over the full range, the JNDs/DL was as expected equal to 0.88. However, the standard deviation of all JND/DL values was relatively large, 0.46. The measured luminance response was examined and a region of 19 sequential gray levels with small deviations in JND/DL was identified in the luminance range from 8.7 to 10.6 cd/m^2. The standard deviation of the JND/DL values in this range was 0.15. This is less than the value of about 0.25 typically found in medical LCD monitors calibrated to a 350 luminance ratio. These gray levels were used to produce the sinusoid test patterns. The noise test patterns were also centered in this region, but they required a somewhat wider range of gray levels.

It should be noted that the ambient luminance was assumed to be 0.30 cd/m^2 when the calibration LUT was built. The ambient luminance, when measured for the final experimental setup, ranged from 0.05 to 0.15 cd/m^2. The original calibration LUT was still used for the experiment, and the ambient luminance was corrected during the analysis of the 2AFC data. The analysis involved extracting luminance values from the cLR of the calibrated LCD and, since the cLR measurements do not include ambient luminance, an ambient luminance value of 0.1 cd/m^2 was simply added to each

luminance value afterwards. This adjustment had a minimal effect on the results of the experiment.

6.2.2 Generation of complex test patterns

Grayscale test images were produced using software developed to add patterns to an existing image. For this study, the patterns were added to images of 2000 by 1500 pixels in size. A set of images was then generated with the added pattern and desired background value. The image values were then converted to grayscale display values (0-255) using a window value of 255 and a level value of 1000, as described in Chapter 2. The window value was chosen such that a change in image values would correspond to an equal change in display values. Image values outside the window (i.e. less than 872 or greater than 1125) would result in a display value of 0 or 255, respectively. The display values were saved as images in Tagged Image File Format (TIFF) files. These images were then read in by the 2AFC software for each test and converted to gray levels according to the calibration LUT loaded in the graphic adapter.

The target patterns for these experiments were all limited to circular regions of a given radius, rather than the square target regions used in Chapter 4. In addition, a function similar to a Hann filter was applied to remove sharp transitions at the edge of the target that the human visual system might detect. This function, shown in Eq. (6.1), was applied starting at a distance of half the target pattern's radius and continuing outwards, smoothing the edges and blending them in with the background.

$$HF(r) = 0.5\left(1 + \cos\left(\pi \frac{r - 0.5r_{tar}}{0.5r_{tar}}\right)\right) \tag{6.1}$$

HF = Hann factor (applied for $r > 0.5r_{tar}$)

r = radial distance from center of target region

r_{tar} = radius of the target region

Three types of target patterns were used in this experiment – a sinusoid pattern, a white noise pattern, and a filtered white noise pattern. The sinusoid pattern was similar to the grating pattern from the previous experiment, but with intermediate transition states instead of simply switching between two gray levels. Also, whereas the grating

pattern only had a positive modulation relative to the background, the sinusoid pattern varied both above and below the background value.

6.2.2.1 Sinusoid test patterns

Figure 6.3 gives an example of the sinusoid test pattern used in this experiment. The characteristics of the sinusoid pattern were controlled by an image background (I_{sin}) and modulation (mod_{sin}) given in image values, and a target radius (r_{tar}) and spatial period (P_{sin}) given in pixels. The background was simply the starting image value of the target region, before pattern modulation was applied. The modulation variable was an integer equal to the peak-to-peak amplitude of the sinusoid. The image software used a cosine function to modulate the background, as shown in Eq. (6.2). Odd modulation values resulted in the positive amplitude being larger than the negative amplitude by one image value. This caused a slight shift in the target's average value away from I_{bkgd}, but the resulting change in luminance was considered small enough to not affect the visibility of the sinusoid significantly.

$$I_{\sin}(x) = I_{bkgd} + \frac{\mathrm{mod}_{\sin}}{2}\cos\left(2\pi\frac{x}{P_{\sin}}\right) \tag{6.2}$$

I_{sin} = image value within the sinusoidal target region

x = horizontal pixel position within the target region

I_{bkgd} = background image value

mod_{sin} = peak-to-peak modulation in image values

P_{sin} = spatial period in pixels

For the adaptation tests, the image background away from the sinusoid target region was replaced with I_{adapt}, a value significantly different from I_{sin}. The intent was to alter the overall average luminance of the image, thus causing the observer's visual system to adapt to a luminance different from the average in the target region. The area just outside the target was kept equal to I_{sine}, out to a radius of $2r_{sine}$, in order to keep the edges of the test pattern from becoming more visible when seen against the new background. Figure 6.4 shows a sample of one of these off-adaptation test targets. In order to produce adaptation luminance outside of the available calibrated gray levels, specialized LUTs were created with the RGB values of a single gray level modified to

produce a selected luminance. This modified gray level was chosen to correspond with I_{adapt} and was taken from levels that were not utilized by the test pattern.

Figure 6.3: Close-up image of a sinusoid pattern used in the 2AFC contrast threshold test. The actual target size was 103.5mm in diameter.

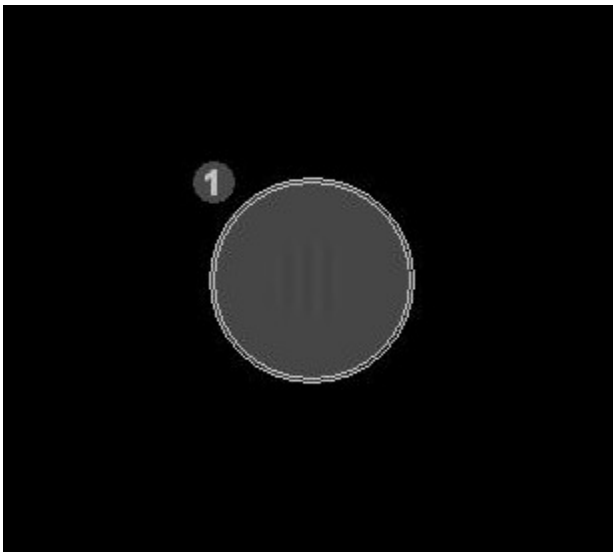

Figure 6.4: Screenshot of a sinusoid pattern with an altered background used in the 2AFC off-adaptation contrast threshold test. The target was 103.5mm in diameter. The circular region around the target was 207mm in diameter.

6.2.2.2 Noise test patterns

For digital radiographs, image noise is traditionally called quantum mottle. This arises from the statistical fluctuations associated with the limited number of x-rays per unit area incident on the detector. For a well designed system, instrument noise is minimal and the image noise derives from a Poisson distribution. In this work, the

quantum mottle is simulated using numbers randomly sample from a Gaussian distribution.

Similar to the sinusoid pattern, the characteristics of the noise pattern were controlled by an image background (I_{bkgd}) and modulation (mod_{noise}) given in image values, and a target radius (r_{tar}) given in pixels. The deviation from the background for each pixel was chosen randomly from a Gaussian distribution with an average of 0 and a standard deviation of 1 (i.e. a standard normal distribution), which was then multiplied by mod_{noise} to control the strength of the noise signal, as shown in Eq. (6.3). These values were then added to the original pixel values to get a white noise pattern with an average image value equal to I_{bkgd} and a standard deviation equal to mod_{noise}. The image software used the GASDEV_S function from "*Fortran Numerical Recipes*" by Press[14] and a random number generator[1] to select image values from a standard normal distribution. Multiple white noise patterns were produced for this experiment, each with a different random appearance. Figure 6.5 shows a sample of the white noise patterns.

$$I_{noise}(x, y) = I_{bkgd} + \mathrm{mod}_{noise}\, G_{0,1}(x, y) \tag{6.3}$$

I_{noise}	= image value within the noise target region
x	= horizontal pixel position within the target region
y	= vertical pixel position within the target region
I_{bkgd}	= background image value
mod_{noise}	= standard deviation in image values
$G_{0,1}$	= random value from a Gaussian distribution with an average of 0 and a standard deviation of 1

For all radiographic systems, the signal is created first with continuous dependence on position and then periodically sampled to create the digital image. Direct digital radiography systems convert absorbed x-ray energy directly to charge in a photoconductive layer. The charge under small square regions is then measured to produce the image pixel values. Indirect digital radiography systems detect x-rays using a scintillation phosphor. For storage phosphor systems, stored energy is released by a

[1] RANMAR, proposed by Marsaglia and Zaman in report FSU-SCRI-87-50, modified by F. James, 1988 and 1989, and adapted by Aldo Badano, 1/99.

scanned laser and measured in relation to position along each scanned line. For flat panel systems, prompt scintillation light is measured using an array of photo-diodes. For both storage phosphor and indirect flat panel digital radiography systems, the signal is significantly blurred by light scattering before the sampled measurement is made.

To simulate these systems, patterns of image noise were first computed at sample positions that were $1/4^{th}$ of the image pixel spacing. This approximates the pre-sampled signal. To simulate direct digital radiography systems, the noisy signal of each pixel was computed as the simple average of 16 sub-pixels. This approximates the square signal collection occurring in these flat panel detectors. To simulate indirect digital radiography systems, the sub-sampled image values were filtered by a function representing the modulation transfer function (MTF) of a typical storage phosphor system. The 16 sub-pixels were then similarly averaged. This approximates the pre-sampled effect of light blur in these types of systems.

For the patterns simulating direct radiography systems, the noise pattern (see Figure 6.5) has a white power spectrum typical of that measured for these devices[15]. For the patterns simulating indirect digital radiography systems, the blur caused by light spread results in a reduction of high spatial frequencies. This increases the relative strength of low frequency structures, giving the noise a lumpy or clumped appearance in comparison to white noise (see Figure 6.6). In both cases, the noise amplitude was determined by the standard deviation of the pre-sampled sub-pixel data. Thus, for the filtered patterns, the noise modulation in the final image is reduced, since the filtering process reduces the area under the noise power spectrum, but it does not change the zero frequency value that is associated with the noise equivalent quanta (NEQ) of the image.

The noise patterns were filtered using a fast Fourier transform (FFT) method similar to that described by Saunders and Samei[16]. The Gaussian distributed noise was converted to a spatial frequency spectrum using a two-dimensional FFT and multiplied by an MTF. An inverse FFT function was then used to transform the results back to the spatial domain. The MTF was modeled on an expression from Barrett and Swindell[17],

$$MTF(f) = \left(1 - e^{(M \cdot f)}\right) / \left(M \cdot f\right), \tag{6.4}$$

where M was set to 1.23 based on evaluations of computed radiography (CR) systems by Flynn and Samei[18]. This filter function was slightly modified by a half-Gaussian filter

function to include a low-frequency drop seen in CR systems and primarily influenced by the optical design of the reader for the storage phosphor[18]. Figure 6.7 shows a plot of the final MTF.

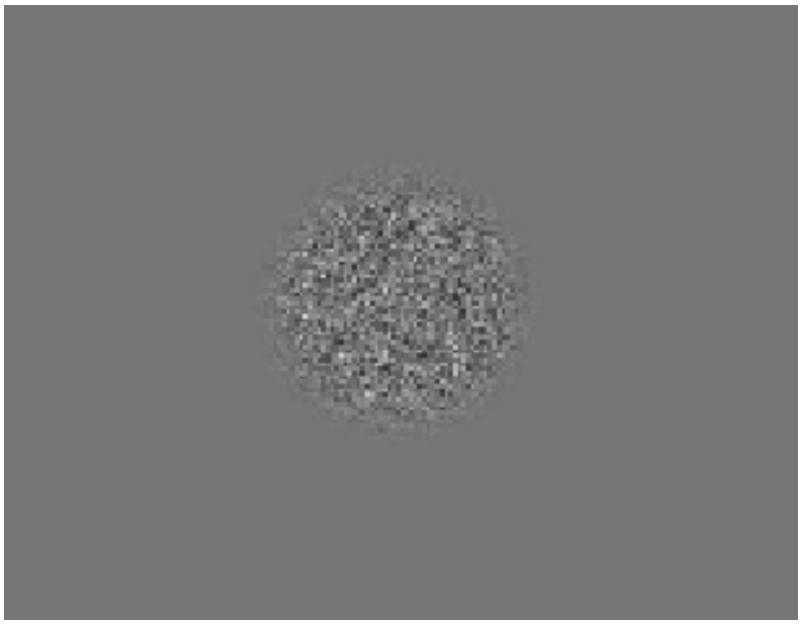

Figure 6.5: Close-up image of a white noise pattern used in the 2AFC contrast threshold test. The actual target size was 145mm in diameter. The pattern simulates quantum mottle in a direct digital radiography system.

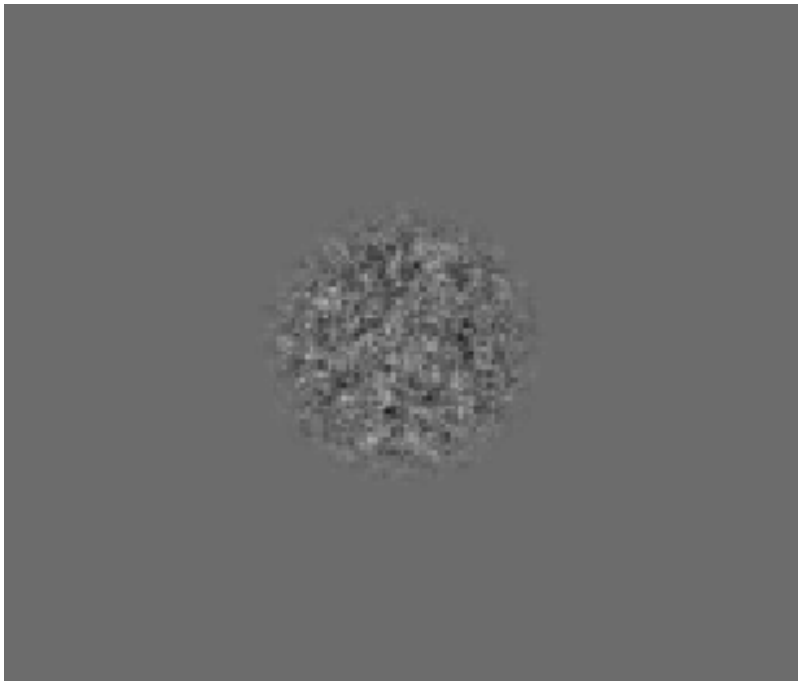

Figure 6.6: Close-up image of a filtered noise pattern used in the 2AFC contrast threshold test. The actual target size was 145mm in diameter. The pattern simulates quantum mottle in an indirect digital radiography system.

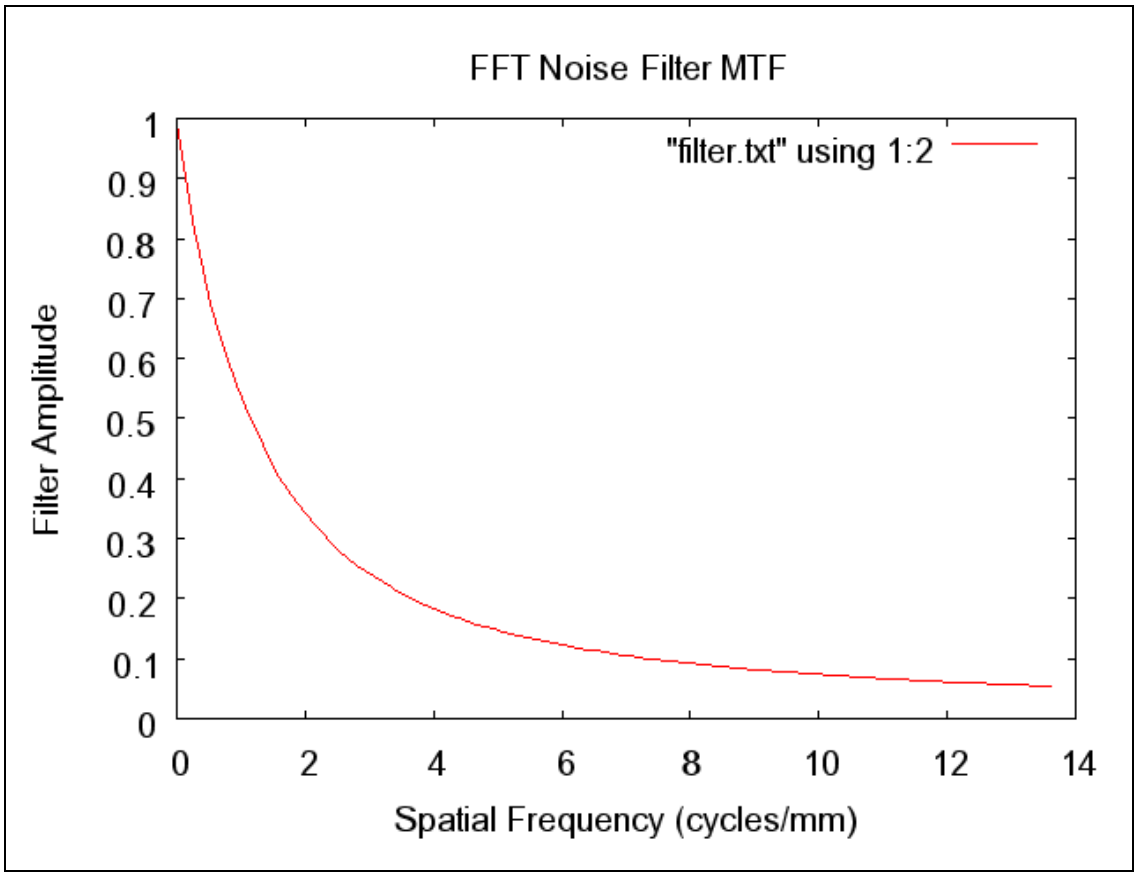

Figure 6.7: Plot of the MTF used to produce the filtered noise patterns.

6.2.3 Test design

Images were built for seven perception tests. The first three were contrast threshold tests for the basic sinusoid, white noise, and filtered noise patterns. These will be referred to as the SINE, WHITE, and FILTER tests. The next four tests used sinusoid patterns with altered backgrounds to measure the effect of adaptation on contrast perception. These will be referred to as the ADAPT tests, with individual designations of A through D. Image parameters were selected for each test to provide a wide range of contrast levels both above and below the expected contrast threshold. The contrast steps available were determined by the calibration described in section 6.2.1. A calibrated luminance response (cLR) was measured to verify the contrast values. The LCD display used was an NDS AC-QX21-AC9300 color LCD with a pixel size of 0.207mm. The nominal viewing distance was set to 1.0 meters in order to insure that the Nyquist frequency associated with the pixel pitch was well above the limiting frequency of the

human visual system. This insured that observers could not detect any patterns associated with the detailed shape of LCD pixels and their red, green, and blue sub-pixels. This also allowed the pixels per cycle for the sinusoid patterns to be sufficiently large.

Each test was constructed using one of the patterns described in section 6.2.2. For the SINE test, the Barten model was used to predict the threshold and image patterns created with both larger and smaller contrast. For the remaining tests, preliminary testing was done to determine image values near the threshold. Testing was also done for all seven tests to select the number of images and the number of forced observer pauses that would minimize observer fatigue while still providing statistically significant results. The size of the patterns was set differently for each test. For the SINE and ADAPT tests, r_{tar} was set to 25 pixels and P_{sin} was set to 10 pixels. For the NOISE and FILTER tests, r_{tar} was set to 35 pixels.

For all tests, I_{bkgd} was initially set to 938, which corresponded to a gray level of 67 and a luminance of 9.26 cd/m^2 (assuming an ambient luminance of 0.10 cd/m^2). The mod_{sin} and mod_{noise} values were initially set to 0 (i.e. no contrast). Additional images with different contrast levels were then created by incrementing the modulation value followed by incrementing the background value and repeating the modulations. The background values were incremented at a rate of 1 image value, as were the modulation values for most of the tests. Some test patterns that required more contrast than others had their modulation values incremented at a rate of 2 image values. The number of increments varied between the seven tests, but each had a total of ~100 images. Each image combination was used twice, once for each of the two target regions in the 2AFC test. Each set of images was repeated 2 or 3 times in order to have ~200 images for each test. The final parameters chosen for the seven tests are summarized in Table 6.1. It includes the I_{adapt} values chosen for the ADAPT tests and the corresponding luminance values, L_{adapt} (cd/m^2).

6.2.4 2AFC testing

A TCL/TK graphic application was created to present the sequence of 2AFC images used in a test. A typical test window is shown in Figure 6.8. Seven test types, SINE, WHITE, FILTER, and ADAPT A-D were included, each consisting of about 200 images and requiring approximately 10 minutes to complete. Three training tests were

provided to let observers become acclimated with the test process and the sinusoid, white noise, and filtered noise patterns they would be viewing. The test process, including data storage and analysis of individual results, was automated so that observers could take the tests without supervision. Users were allowed to take the tests at their convenience and were not required to complete all seven tests in one sitting. Users were also encouraged to take the tests multiple times.

The program presented a large number of images of different contrast levels with a uniform background to the observers. A test target was placed in one of two positions, left or right ("1" or "2"), and the observers were asked to identify the position that the target was in. If they were uncertain, then they were instructed to guess as best they could. Selections were made via keyboard or mouse. Once a selection was made, the screen was updated with a new background and target. Approximately 200 images were viewed for each test session, with breaks inserted every 35 images during which the screen was set to a mid-gray level. Training sequences using sinusoid, white noise, and filtered noise patterns were provided for participants to acclimate themselves with the test process. The training sequences used only 30 images with higher contrast than the actual test images.

Table 6.1: Image target and background parameters for the 2AFC tests.

Test Name	r_{tar}	P_{sin}	**mod incr**	**# mods**	**# bkgds**	**Initial** I_{bkgd}	L_{bkgd} **(cd/m^2)**	I_{adapt}	L_{adapt} **(cd/m^2)**	**Repeat**
SINE	25	10	1	10	5	938	9.26	N/A	N/A	x2
WHITE	35	N/A	1	12	4	938	9.26	N/A	N/A	x2
FILTER	35	N/A	2	10	5	938	9.26	N/A	N/A	x2
ADAPT A	25	10	1	7	5	938	9.26	0	1.477	x3
ADAPT B	25	10	1	10	5	938	9.26	971	26.0	x2
ADAPT C	25	10	2	10	5	938	9.26	1070	155.4	x2
ADAPT D	25	10	2	10	5	938	9.26	1121	339.1	x2

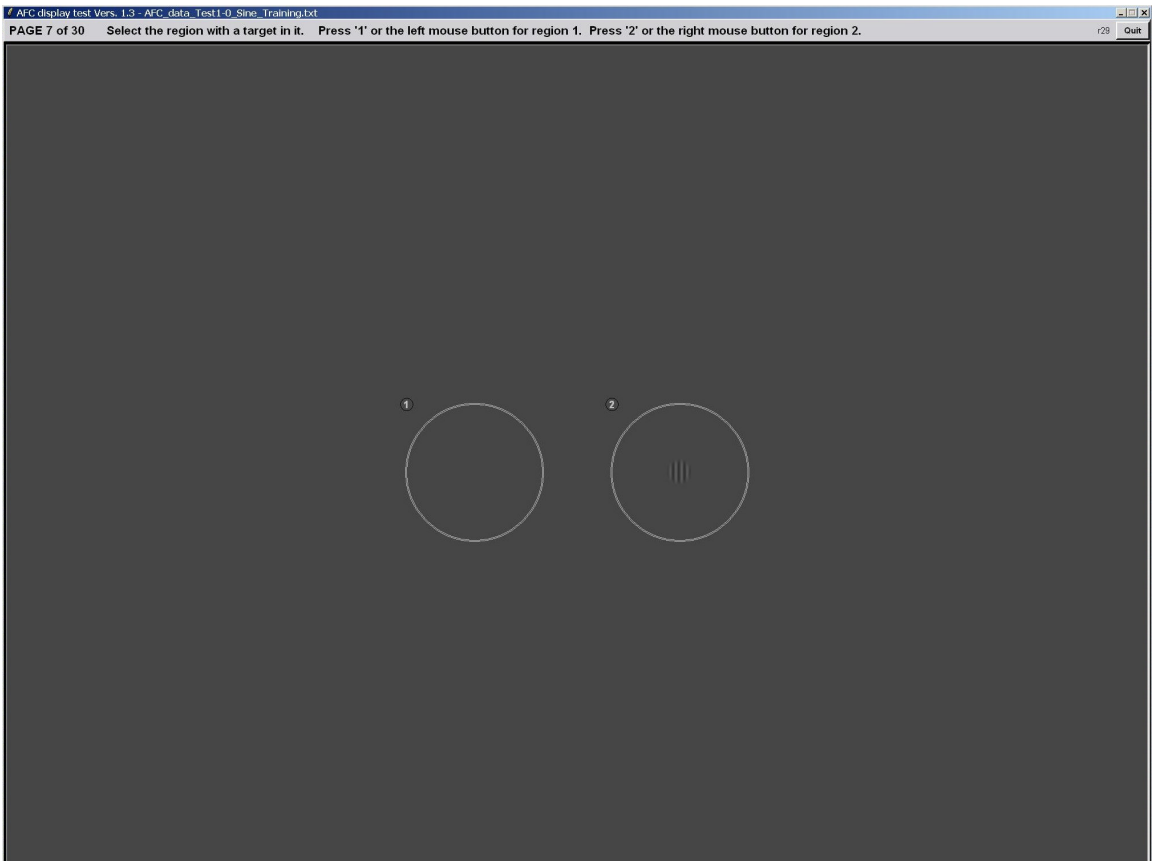

Figure 6.8: Screenshot of the 2AFC contrast threshold test (training session).

The contrast images were presented to the observers in random order using the TCL/TK random number generator function ("rand"). For each test, the images were divided evenly into cases. The number of cases for a particular test was equal to the number of backgrounds, multiplied by the number of times the image set was repeated. The number of images was equal to twice the number of modulations, multiplied by the number of times the image set was repeated. Each time a test was administered, a random case sequence was first determined. The case sequence was then divided equally into a group A and a group B. One presentation per case was made from group A, followed by one presentation per case from group B. An image from each case in the sequence was then randomly selected and presented. This was repeated until all of the images had been presented. Thus, observers scored images in varying order as presentations for each case were repeated, but images from the same case never appear twice in a row. Since the background was slightly different between most cases, this

randomization method helped to reduce the "flicker response" described in Chapter 4[19, 20]. However, the repetition (x2 or x3) of each image set was done in such a way that background levels were also repeated in 2 or 3 cases. Thus, there were some instances where the background did not change between images. This should be corrected in future testing.

Normalization of the signal contrast was done differently for each type of test pattern. For the SINE and ADAPT tests, the contrast threshold (C_T) was normalized relative to the contrast threshold predicted by the Barten model (C_{BM}), as was done in Chapter 4. C_{BM} was calculated from the sinusoid frequency, viewing distance, target size and mean brightness. For the WHITE and FILTER tests, the image patterns are not sinusoidal, and so it is inappropriate to use C_{BM} for normalization. Instead, contrast was calculated for each noise target by treating the pre-filtered standard deviation as the luminance amplitude (half the peak-to-peak modulation). This value was then normalized to the $\Delta L/L$ associated with a JND of the DICOM GSDF for the average brightness of the presented image. The normalized noise contrast is referred to as C_{STD}/C_{JND}. Although the Barten model and DICOM standard are based on sinusoidal contrast patterns, they still provided a useful frame of reference for evaluating the contrast of other types of patterns viewed on medically calibrated devices. The C_{STD}/C_{JND} metric represents the number of 'JNDs' needed for a noise pattern to be just noticeable. For a medical display system calibrated to the DICOM GSDF, C_{STD}/C_{JND} can then be used to infer the number of gray levels needed.

The MLE software described in Chapter 5 was integrated into the 2AFC software and used to analyze each observer's results after each test was completed. The initial contrast threshold (C_T/C_{BM} or C_{STD}/C_{DICOM}) and width (W) values for the psychometric function were set to 1.0 and the maximum C_T/C_{BM} was set to 1000. The convergence was 0.00000001 and the expected increment was 0.5. With these settings, the MLE function was able to find the minimum of the likelihood function effectively for all of the test types. For the software used in this test, the MLE analysis is done immediately after the score for the last image is obtained, and the normalized contrast result is presented in a window to the observer that has just completed the test. Additionally, two plots are

generated that provide verification that the test results were reasonably distributed about the measured contrast threshold.

The set of tests was administered to 10 persons in a work carrel located in a radiology reading room routinely used for medical interpretation of digital radiographs. The participants were radiologists and medical physicists with experience in viewing images on medical LCD monitors. Environmental lighting was taken into account and kept to a minimum. The light striking the screen, or illuminance, was approximately 9.2 lux (lumens per m^2). The ambient luminance was approximately 0.1 cd/m^2. Participants were allowed to adjust their seat to whatever positions and angles were most comfortable as long as the 1.0 meter viewing distance was maintained. The monitor was adjusted such that the observer viewed the screen at an angle normal to the surface.

Several potential participants required corrective optics in order to interpret images on LCD monitors. For a few persons, a viewing distance of 0.8 meters was used to be within the range of the corrective optics. Data analysis parameters were adjusted to take this into account when calculating the contrast thresholds for these observers. In other cases, the observers' optical requirements could not be met, either because they would have had to sit extremely close to the display, or their conditions made it difficult for them to perceive the types of patterns used, regardless of distance. These observers did not participate in further testing.

6.3 Results

Table 6.2 shows the contrast threshold results, individually averaged for those observers that took the tests multiple times. Initial test results were not included for observers that took early versions of the tests or exhibited significant difficulty due to visual conditions or fatigue. These observers repeated the tests after adjustments were made. An overall mean and standard deviation is listed for each of the seven tests. Since the test used a sinusoid pattern instead of a bar pattern, there was no need for a $4/\pi$ correction, as was done in Chapters 4 and 5. Table 6.3 shows the intra-observer results for observer PT. The relatively low standard deviation values provide evidence of intra-observer consistency. Figure 6.9 shows a plot of the SINE and ADAPT test results for those observers that took all four ADAPT tests. The C_T/C_{BM} values of each observer

were normalized to the result for the SINE tests for that observer. These are plotted against the average scene luminance relative to the initial average target luminance (L_{adapt}/L_{bkgd}). This shows the relative effects of adaptation when the average scene luminance is different from the average target luminance.

Table 6.2 Inter-observer contrast threshold measurements, averaged for each observer.

Observer	**SINE**	**WHITE**	**FILTER**	**ADAPT A**	**ADAPT B**	**ADAPT C**	**ADAPT D**
(units)	C_T/C_{BM}	C_{STD}/C_{JND}	C_{STD}/C_{JND}	C_T/C_{BM}	C_T/C_{BM}	C_T/C_{BM}	C_T/C_{BM}
AJ	1.263						
DB	0.469	8.157	14.906	0.610	0.833	1.246	1.938
DP	0.701	8.588	12.116	0.802	0.793	1.207	1.574
JB	1.244				1.363		
MF	0.865	10.482	14.869	0.947	0.963	1.138	1.460
MH	0.895	9.196	15.671				
MP	0.554			0.576	0.583	1.213	1.525
PT	0.727	9.910	12.480	0.746	0.756	1.254	1.771
PR	0.501	6.024	10.088	0.658	0.888	0.875	1.939
SL	0.933	8.811	14.409	0.873	0.746	1.357	1.787
MEAN	**0.815**	**8.885**	**13.753**	**0.745**	**0.866**	**1.184**	**1.710**
STDEV	**0.282**	**1.132**	**1.567**	**0.138**	**0.230**	**0.152**	**0.199**

Table 6.3 Intra-observer contrast threshold measurements for observer PT.

	SINE	**WHITE**	**FILTER**	**ADAPT A**	**ADAPT B**	**ADAPT C**	**ADAPT D**
(units)	C_T/C_{BM}	C_{STD}/C_{JND}	C_{STD}/C_{JND}	C_T/C_{BM}	C_T/C_{BM}	C_T/C_{BM}	C_T/C_{BM}
N	5	6	4	4	4	5	5
MAX	0.782	11.226	14.963	0.880	0.909	1.454	1.885
MIN	0.694	8.731	6.708	0.539	0.494	1.124	1.658
MEAN	**0.727**	**9.910**	**12.480**	**0.746**	**0.756**	**1.254**	**1.771**
STDEV	**0.0353**	**1.026**	**3.897**	**0.151**	**0.187**	**0.134**	**0.0926**

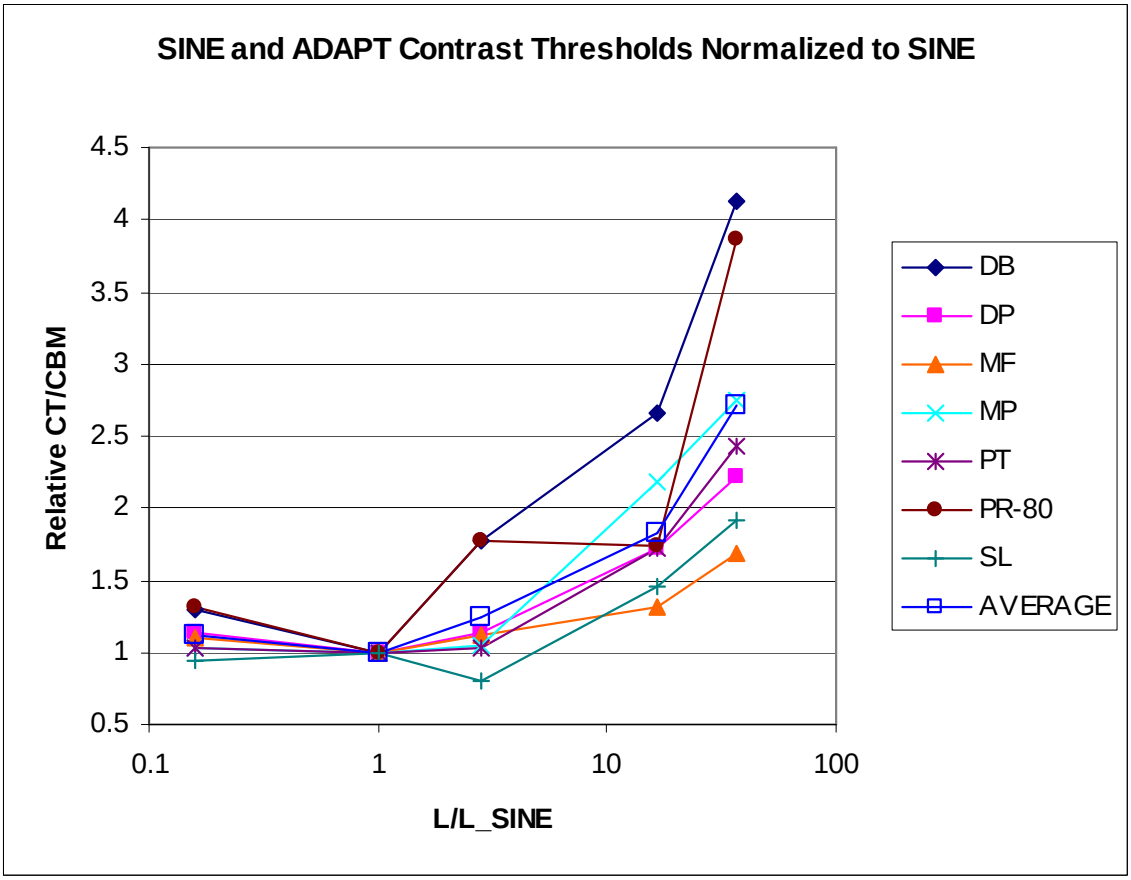

Figure 6.9: Contrast threshold results for observers that took both the SINE and ADAPT tests, normalized to the SINE results and plotted against the average scene luminance relative to the initial average target luminance. This plot shows the relative effects of adaptation when the average luminance of the image is different from the average luminance of the target sinusoid pattern.

A set of sample images were created to analyze the post-filtered characteristics of the noise patterns used in the experiments. The MOD values for these images were 5, 10, 15, 15.5, and 20. The patterns in these images were created using the same procedures used for the 2AFC noise patterns, with the mean image value set to 938. Unlike the 2AFC test patterns, these were spread over single a large square area (800 x 800 pixels). The image values were converted to luminance values based on the display conditions used in the 2AFC observer tests (i.e. a display window /level of 255/1000 and a display luminance response of 0.887 JNDs/DL based on the measured response of the calibrated display). The input standard deviations of 10 and 15.5 correspond to 8.87 and 13.75 in units of C_{STD}/C_{JND}. These are near the average contrast threshold results for the WHITE

and FILTER tests. The images values were then analyzed to determine the final standard deviations.

Table 6.4 shows the pre-filtered standard deviations and analysis results for these images. The final standard deviations for the WHITE noise patterns are all nearly 0.4 image values less than the input MOD value. This systematic decrease is due to the image values being converted from floating point values to integer values during the image creation process. The post-filtered standard deviations for the FILTER noise patterns were all approximately 1/3 of the MOD values. Adjusting for these observations, the post-filtered contrast threshold values were roughly 8.5 and 4.6 (C_{STD}/C_{JND}) for the WHITE and FILTER patterns, respectively.

A noise power spectrum (*NPS*) analysis was done to examine the frequency content of the white and filtered noise patterns. The *NPS* was computed in units of mm^2 using software developed by Flynn and Samei, 1999[21]. The software computed the *NPS* values for each spatial frequency using radial averages. However, the noise power values for spatial frequencies at or very near zero are reduced due to de-trending. Instead, the noise power at zero-frequency, *NPS*(0), was estimated from the measured just noticeable input standard deviations, 8.885 and 13.753 (C_{STD}/C_{JND}), using Eq. (6.5),

$$NPS(0) = a_p C_{STD}^2, \tag{6.5}$$

where a_p is the pixel area of $0.0428mm^2$ and C_{JND} is 0.01154 +/- 0.00006 for the noise patterns. The noise power values at 0.075 and 0.151 cycles/mm were interpolated from the neighboring values.

Table 6.4 Image value standard deviations for the white and filtered noise patterns before and after image generation.

MOD (input STDEV)	**5**	**10**	**15**	**15.5**	**20**
WHITE (output STDEV)	4.62	9.62	14.62	15.11	19.61
FILTER (output STDEV)	1.51	3.36	5.23	5.41	7.09
FILTER / MOD	0.302	0.336	0.349	0.349	0.355
FILTER / WHITE	0.327	0.349	0.358	0.358	0.362

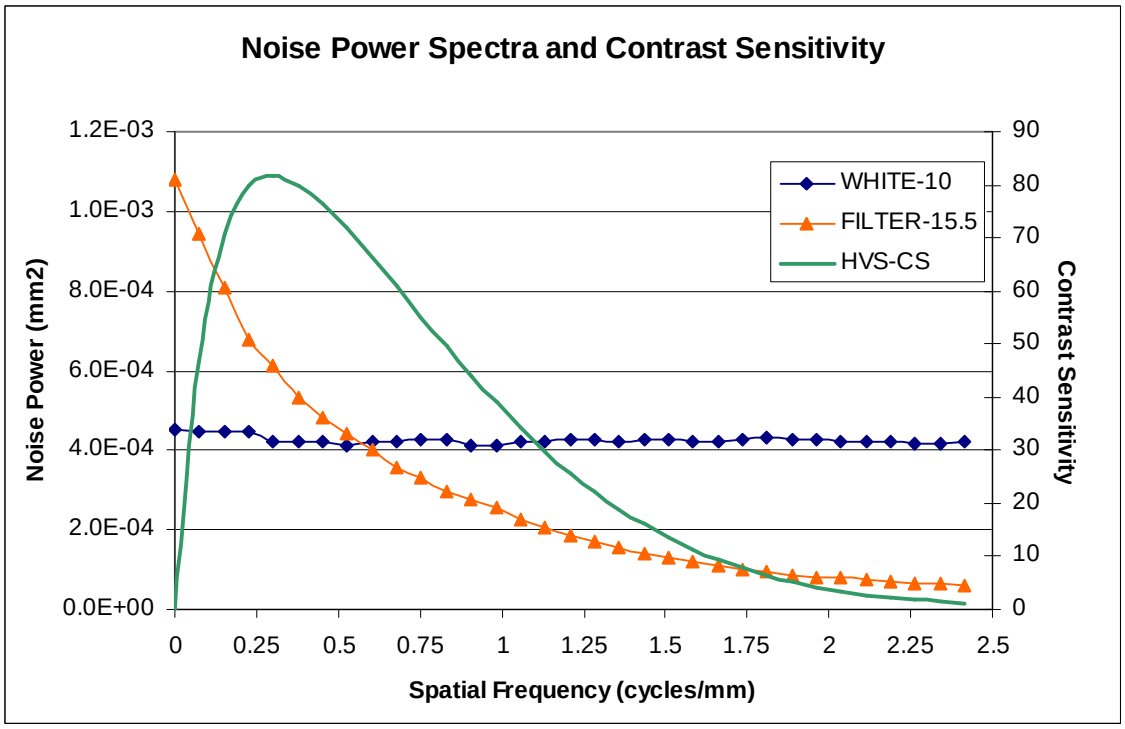

Figure 6.10: Noise power spectra for the luminance values of two displayed images containing white and filtered noise patterns. Images parameters were chosen to produce just visible noise based on the contrast thresholds observed for the WHITE and FILTER tests. Human contrast sensitivity is also plotted.

Figure 6.10 shows the noise power spectra for a white noise pattern with a MOD value of 10 and a filtered noise pattern with a MOD value of 15.5. A plot of the contrast sensitivity as a function of spatial frequency is also included. The contrast sensitivity was calculated assuming a viewing distance of 1.0 meters, a background luminance of 9.26 cd/m^2, and a target size of 14.5mm (70 pixels wide at 0.207mm per pixel). As the spatial frequency increases, both the filtered noise power and the contrast sensitivity are reduced.

6.4 Discussion

6.4.1 Comparing the just noticeable sinusoid and noise patterns

The mean contrast threshold for the SINE test was 0.815 (C_T/C_{BM}) with a standard deviation of 0.282 for the results of 10 observers. These results are similar to those found using bi-level bar patterns that were detailed in Chapter 4 and analyzed with the MLE method in Chapter 5. As stated in Chapter 4, a result of ~0.66 (C_T/C_{BM}) is predicted for a

2AFC experiment, compared to the variable adjustment method used in most of the studies on which the Barten model was based[22]. The results from the SINE test are thus consistent with the Barten model, reinforcing the conclusions drawn from Chapters 4 and 5. It is taken here as an indication that calibrated medical LCD devices can be used for high fidelity psychovisual experiments that use patterns with varying gray levels.

The ADAPT tests demonstrated the effects of adaptation on contrast perception. For those observers that took both the SINE and the ADAPT tests, the results indicate an elevated contrast threshold (i.e. degradation of observer performance) for tests in which the background luminance was different than the mean in the region of the modulated sine pattern. One observer, SL, exhibited similar performance for L/L_SINE levels at 0.1, 1, and 10, although this observer still showed degradation in performance at the highest background level. This observer only took each test once. This behavior may have been due to a learning effect similar to that seen in the intra-observer tests results from Chapter 4. For high background luminance values, the observed degradation is consistent with that suggested in Figure 6.1. For the results shown in Figure 6.9, an L/L_SINE value of 10 corresponds to a point on curve "A" of Figure 6.1 at 10 cd/m^2, since that curve is for an observer adapted to 100 cd/m^2. However, the amount of degradation was much different for different observers. Also, the degradation measured for low background luminance values was not as significant as it was for high background values. The results might be due to light scattering in the lens or vitreous humor that would produce an effect analogous to flare in a camera lens. Additional experiments at different luminance levels are needed to further verify and expand on this aspect of the human visual system.

In terms of the pre-filtered standard deviations, the mean contrast threshold for the WHITE noise test was 8.885 (C_{STD}/C_{JND}, N = 7) with a standard deviation of 1.132. The mean contrast threshold for the FILTER noise test was 13.753 (C_{STD}/C_{JND}, N = 7) with a standard deviation of 1.567. The input standard deviation is associated with the zero-frequency noise power, *NPS*(0), and is important for observer detection of large area, low contrast objects. This value is easily deduced from Eq (6.5) and was used for the *NPS*(0) values in Figure 6.10, which shows estimated *NPS* values for just noticeable white and filter patterns. For the "WHITE-10" noise pattern, *NPS*(0) was 4.50E-04 mm^2. For the "FILTER-15.5" noise pattern, *NPS*(0) was 10.8E-04 mm^2 or 2.4 times larger.

These differences indicate that neither the pre-filtered standard deviation nor the *NPS*(0) value are effective at predicting the just noticeable condition.

In terms of the post-filtered standard deviations, the mean contrast threshold for the WHITE and FILTER patterns were 8.5 and 4.6 (C_{STD}/C_{JND}), respectively. Based on Parseval's theorem, the standard deviation is a reflection of the area under the noise power curve[21]. The ratio of these standard deviations, 1.8, is not far from the ratio of the area of their corresponding *NPS* curves, 1.4, after extrapolation to zero frequency. While better than *NPS*(0), the post-filtered standard deviation and the area under the *NPS* curves are also not effective at predicting the just noticeable condition.

In order to evaluate the ability of an observer to detect a noise pattern, the human visual system must be taken into account. Figure 6.10 shows the contrast sensitivity curve predicted by the Barten model under the conditions of our experiment. The two noise power spectra are equal at a spatial frequency that is slightly higher than the frequency of maximum contrast sensitivity. As a measure of noise visibility (*NV*), the product of the *NPS* and the square of the contrast sensitivity (C_S) is considered as a function of the spatial frequency,

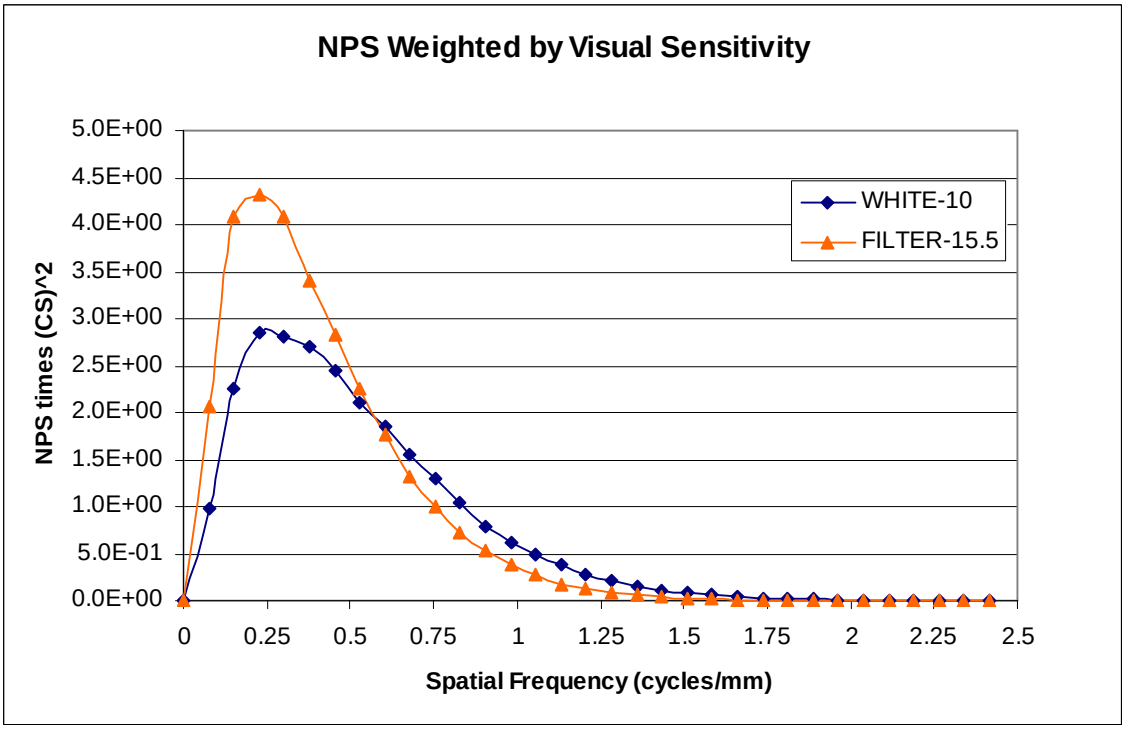

Figure 6.11: WHITE-10 and FILTER-15.5 noise power curves from Figure 6.10 weighted by the square of the HVS-CS curve.

$$NV(\omega) = NPS(\omega)CS^2(\omega). \tag{6.6}$$

In Figure 6.11, the extended high frequencies of the WHITE-10 *NPS* are markedly reduced, as are the elevated low frequency values for the FILTER-15.5 *NPS*. Since *NV(ω)* is similar for the WHITE-10 and FILTER-15.5 patterns that are both just visible, this suggests the use of the area under the weighted *NPS* as a noise visibility index (*NVI*). In order to properly integrate the radial averages for the entire 2-D frequency space, the *NVI* was calculated as shown in Eq. (6.7).

$$NVI = \pi \int NV(\omega)\omega d\omega \tag{6.7}$$

Since *NV(ω)* has units of mm^2 and $\omega d\omega$ has units of mm^{-2}, *NVI* is non-dimensional. Numeric integration of this expression results in *NVI* values for the WHITE-10 and FILTER-15.5 images of 2.95 and 2.74, respectively, with a ratio of 1.08. These similar values further indicate that the *NVI* may be effective for predicting the just noticeable condition. The fact that a just visible *NVI* for a noise pattern is 2.5-3.0 rather than 1.0 may be due to the use of a contrast sensitivity function, $C_S(\omega)$, that represents the visibility of sine patterns rather than noise patterns as is further discussed below.

For purposes of comparison, the results of the SINE test were recalculated in terms of C_T/C_{JND}. The mean contrast threshold was 2.376 (C_T/C_{JND}) whereas the post-filtered values of C_{STD}/C_{JND} for the WHITE and FILTER tests were 8.5 and 4.6, respectively. These values do not account for the frequency dependence of the human visual system, but it is difficult to relate the area of the weighted *NPS* to the contrast associated with the pattern used in the SINE test. It is noted, however, that the spatial frequency of the pattern in the SINE test, 0.48 cycles/mm, is near the spatial frequency where the *NPS* of the two noise patterns are equal. It seems clear that the noise patterns require significantly more contrast to be perceived than the sinusoid patterns. This may be due to the lack of extended signal correlation in the noise patterns. The sinusoid patterns vary in only one dimension, providing the human visual system with extended signal correlation in the other dimension that facilitates cognitive recognition by the human visual system. In comparison, the noise patterns randomly vary in two-dimensions such that signal correlation will rarely be present beyond the range of a few pixels in any direction.

The $NV(\omega)$ values effectively represent ratios of the noise power of an image and the noise power produced by just visible contrast (i.e. the inverse square of the contrast sensitivity) at specific spatial frequencies. The *NVI* represents the 2D integration of these ratios. Since the contrast sensitivity is based on sinusoid contrast patterns, a just noticeable sine pattern would be expected to produce an *NVI* value of about 1. Noise patterns, which have been shown in this chapter to require significantly more contrast to be visible, would be expected to require higher *NVI* values for the just noticeable condition. The values of 2.95 and 2.74 found for the WHITE-10 and FILTER-15.5 patterns support this. Other aspects of these patterns may play a role in this difference, such as the complex two-dimensional variations of noise compared to simple one-dimensional variations in sinusoids. There is also the possibility that the $C_S(\omega)$ function may change, depending on the frequency content of the noise[23]. How exactly these and other image factors influence the *NVI* and how it can be modeled to predict the visibility of other patterns requires further study.

6.4.2 Comparison of actual direct and indirect imaging systems

In order to understand the magnitude of simulated just noticeable noise compared to noise presented by typical radiography systems, actual medical studies from two Kodak radiography systems were reviewed. Twelve AP abdomen images were selected with typical radiographic factors leading to an exposure index in the range of 1900-2100, where the nominal target exposure index is 2000. Noise measures were taken in regions where no anatomic structures were visible and that had a mid-gray presentation luminance. The images used a window and level (W/L) of 4096/2048 to convert to display levels (DL). Assuming a maximum luminance of 500 cd/m^2, a luminance ratio of 350, and an ambient luminance of 0.1 cd/m^2, the DLs were converted to luminance and used to calculate C_{STD}/C_{JND} values. Table 6.5 lists six cases from a Kodak DR 7100 direct radiography system with a mean C_{STD}/C_{JND} of 10.51. Table 6.6 lists six cases from a Kodak CR 800 storage phosphor (indirect radiography) system with a mean C_{STD}/C_{JND} of 15.36. The differences seen come from a multitude of factors, including the area of the pixels, detector efficiency, and detector blur. Additionally, Image processing is implemented differently for the direct and indirect digital radiography systems considered.

Figures 6.12 shows samples of the noise produced in images from another direct radiography system, a Shimadzu Safire II, and the CR 800 indirect radiography system.

Table 6.5 Noise measurements for six abdomen studies taken by a Kodak DR 7100 direct radiography system with a pixel size of 139 microns. Image values are converted to display levels (DL) using a W/L of 4096/2048, luminance (*L'*), and contrast relative to the contrast threshold of a DICOM Standard Target with the same mean luminance (C_{STD}/C_{JND}). A maximum luminance of 500 cd/m^2, a luminance ratio of 350, and an ambient luminance of 0.1 cd/m^2 were assumed.

	Image Mean	Image STDEV	DL Mean	DL STDEV	*L'* Mean	C_{STD}/C_{JND}
Case 1	2000	25	126	1	51.37	4.86
Case 2	1966	43	123	3	48.41	14.55
Case 3	2089	36	131	2	56.65	9.71
Case 4	2026	39	127	3	52.39	14.56
Case 5	1961	34	123	2	48.41	9.71
Case 6	2022	34	127	2	50.37	9.71
MEAN	**2011**	**35**	**126**	**2**	**51.60**	**10.51**
STDEV	**47.06**	**6.05**	**2.99**	**0.75**	**3.07**	**3.65**

Table 6.6 Noise measurements for six abdomen studies taken by a Kodak CR 800 indirect radiography system with a pixel size of 168 microns. Image values are converted to display levels (DL) using a W/L of 4096/2048, luminance (*L'*), and contrast relative to the contrast threshold of a DICOM Standard Target with the same mean luminance (C_{STD}/C_{JND}). A maximum luminance of 500 cd/m^2, a luminance ratio of 350, and an ambient luminance of 0.1 cd/m^2 were assumed.

	Image Mean	Image STDEV	DL Mean	DL STDEV	*L'* Mean	C_{STD}/C_{JND}
Case 1	2008	64	126	4	51.37	19.39
Case 2	1983	58	124	4	49.38	19.39
Case 3	2005	44	126	3	51.37	14.55
Case 4	2028	57	127	4	52.39	19.39
Case 5	1968	45	124	2	49.38	9.71
Case 6	1941	41	122	2	47.46	9.71
MEAN	**1989**	**52**	**125**	**3**	**50.22**	**15.36**
STDEV	**31.37**	**9.35**	**1.83**	**0.98**	**1.81**	**4.56**

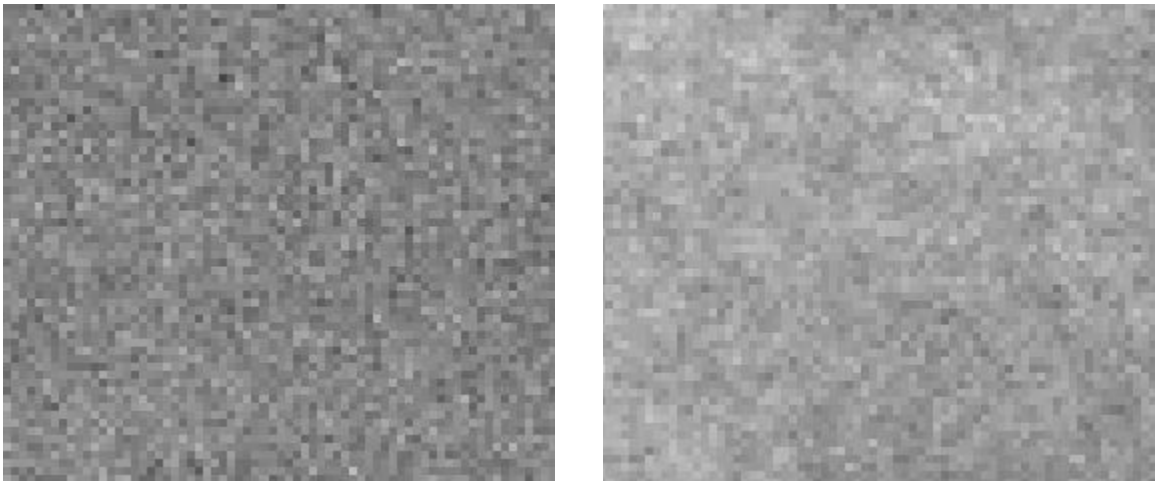

Figure 6.12: Example of image noise at 4x magnification from a Shimadzu Safire II direct radiography system (left) and a Kodak CR 800 indirect radiography system (right).

The noise values seen in the Kodak studies were up to 2-4 times greater than those seen for the just noticeable noise patterns seen in the 2AFC experiments. This suggests that quantum mottle is noticeable in typical radiographic images. In addition, the noise for the CR 800 indirect radiography system was larger than for the DR 1700 direct radiography system, whereas the 2AFC experiments predicted the reverse. This was likely due to reduced detective efficiency in the CR 800. These results indicate that further reduction of the quantum mottle could benefit observers. This might be possible with advanced noise reduction methods applied during the image processing used to prepare the image for presentation. However, some barely visible quantum mottle is a good indication that excessive exposure has not been used, and radiation exposure should not be increased without evidence that diagnostic performance is justifiably improved. In addition, some studies have suggested that small amounts of noise may actually facilitate the detection of other image features[24].

6.4.3 Limitations of medical LCD systems

In most medical centers in developed countries, diagnostic radiology is a filmless operation with images typically interpreted on monochrome medical LCD devices. These present images using computer operating systems with 8-bits per color channel to display 256 gray levels. The monitors typically have an L_{max} of 500 cd/m^2 and a luminance ratio of 350. The JNDs/DL for these parameters is about 2.4. It is not

possible to display a sine pattern of the type used in the experiments of this chapter with a just noticeable contrast, since the total modulation is about 2.5 JNDs (i.e. C_T/C_{JND} = 2.550). In order to obtain smaller contrast steps, a special calibration had to be used involving a very narrow luminance range for which the JNDs/DL was about 0.9. Even using a low luminance ratio, it was still difficult to produce sub-threshold sinusoid targets from the available gray levels. In order to reduce observers' contrast sensitivity and increase the number of sub-threshold gray levels, the sinusoid target dimensions were changed from the DICOM Standard Target parameters used in Chapter 4. If the sinusoid target parameters had not been changed, the just noticeable contrast would have been about 0.8 JND, which could not have been presented using the narrow luminance range.

In addition, the narrow luminance range calibration LUTs were prone to repeated luminance values among the 256 gray levels, resulting in zero contrast for some grayscale steps. Special provisions were made to avoid use of these gray levels. Further examination of other possible calibrations may produce a more consistent contrast response and should be considered for future experiments. Difficulties in producing low contrast sinusoid targets were also reported in a 2AFC contrast perception study by Sund suggesting that increased bit depth would improve the options for perception testing considerably[25].

Flynn et al, 1999[12] defines a set of requirements for high-fidelity medical displays. One of the requirements is that the number of available gray levels be sufficient to produce an average JNDs/DL of 1.0 or less. In general, the results of this chapter support this requirement. To achieve this will require systems that can display about 1024 gray levels using an operating system that supports 10-bits per color channel. It is otherwise understood that the digital video interface (DVI) and LCD monitor internal control drivers can support 10-bit graphic display.

While the 8-bit grayscale currently used with most medical LCD devices is insufficient for achieving high-fidelity presentation of medical images, the limitation is heavily influence by a requirement based on sine patterns, and these are not typical of the patterns encountered in medicine. For the noise patterns studied here, an 8-bit system with 2.5 JNDs/DL is probably just sufficient to display the just noticeable FILTER noise pattern with C_T/C_{JND} of about 4.5. Thus, 8-bit LCD systems may be sufficient for

displaying the complex features seen in typical radiographs, and the requirements for high-fidelity may not be necessary for achieving good diagnostic performance in general medical imaging. This is supported by a recent study by Krupinski et al.[26] that looked at performance differences using 8-bit and 11-bit LCD devices to examine medical images. Observers trying to locate pulmonary nodules using LCDs showed no significant performance differences due to bit-depth, although there was a notable difference in viewing times that favored the 11-bit systems. Additional experiments involving complex image features and radiographic detection tasks will be required to examine this issue further.

The pixel structures within LCD panels present another possible limitation on performance. Most medical LCDs use 3 sub-pixels for each pixel and each sub-pixel is comprised of 2 domains, as seen in Figure 6.14. The widths of the individual domains are about 20 microns. For structures this small, manufacturing variations are highly likely, resulting in fixed pattern luminance noise in LCD devices. Figure 6.13 and Figure 6.14 display images of the domains and sub-pixels in two types of LCD panels, with some possible defects and non-uniformities highlighted. These spatial variations in the pixel and sub-pixel structures of LCD devices create a noise pattern that could interfere with interpretations if the contrast is large enough to be visible. Two recent studies have examined this type of noise, one by Badano et al.[27] and another by Fan et al.[28]. Both studies measured the coefficient of variance (*CV*) of the pixel luminance (i.e. – the standard deviation relative to the mean luminance) for an LCD device at different average luminance levels. However, the two studies each looked at a different LCD devices and found different levels of noise. The study by Badano obtained *CV* values of about 4-5% whereas the values seen in the study by Fan were around 1-2%. For comparison, the just noticeable noise patterns used in the 2AFC noise tests had *CV* values of roughly 5-10%. Further research is needed to determine if pixel uniformity is a significant factor in LCD performance.

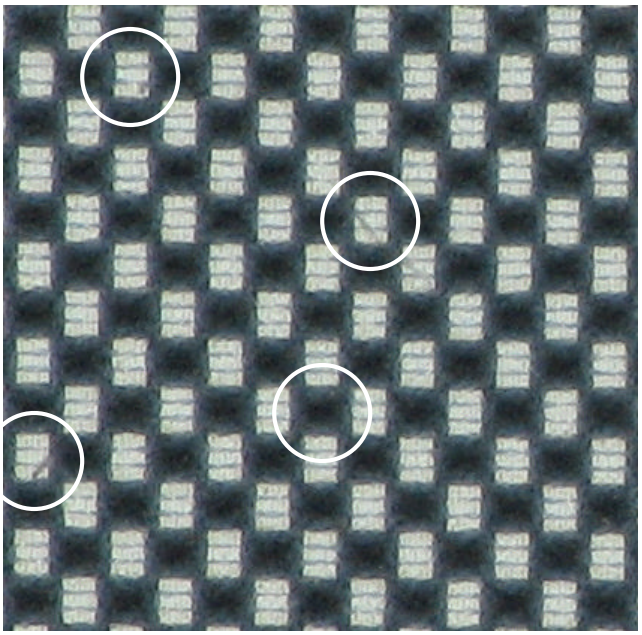

Figure 6.13: Pixel structures from an NEC MD21GS monochrome monitor (1536 x 2048, 212 micron pitch). This panel uses 3 sub-pixels with 1 domain each. A checkerboard pattern was displayed, highlighting the spreading of light from surrounding pixels into regions that are supposed to be dark. Some possible defects and areas of non-uniformity have been marked.

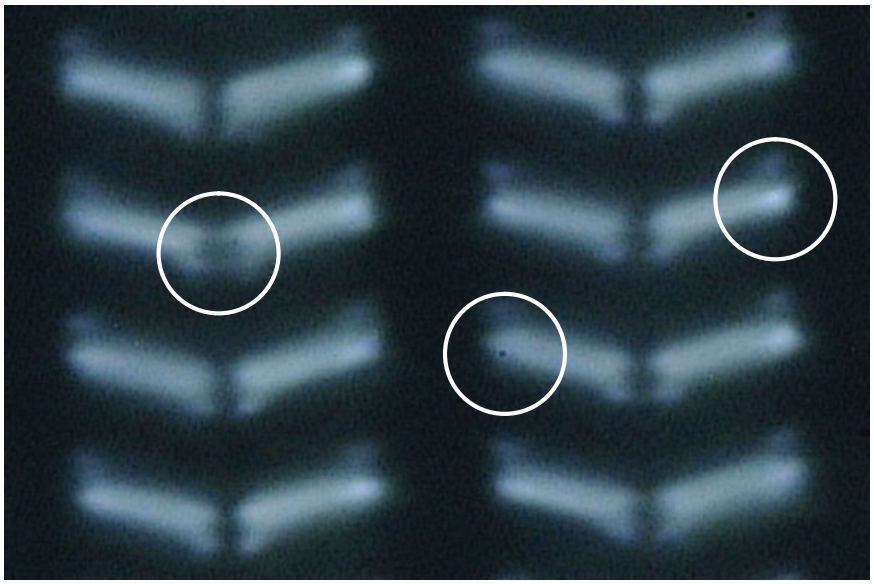

Figure 6.14: Dual-domain pixel structures from an IDTech ITQX21 monochrome LCD panel (1536 x 2048, 207 micron pitch) used in monitors made by Planar, Barco, Totoku, NDS, and others. One pixel contains 3 sub-pixels with 2 domains for each. Some possible defects and areas of non-uniformity have been marked.

6.4.4 Observer variability

Other factors need to be taken into consideration in future experiments. In particular, the difficulties in adjusting the tests for participants with significant optical corrections need to be considered. Although changes in distance may improve an observers' contrast perception to the point that useful data can be collected, it also alters many features of the image. At closer distances, the target size in degrees increases, the spatial frequencies in cycles/degree decreases, and the LCD's pixel structure becomes more visible. One possible solution would be to adjust the image parameters based on the viewing distance in order to maintain perceived size and frequency. This would require that the images be generated dynamically at the beginning of the test or that multiple image sets be produced to accommodate alternate viewing distances. Preliminary testing of participants may be needed to determine an appropriate viewing distance. One possibility might be to use the method of adjustment to estimate the distance at which an observer can just notice a target with $C_T/C_{BM} = 1$ or equivalent. A minimum viewing distance could be set to limit the visibility of the LCD's pixel structure, although this may disqualify some potential participants. These cases may also indicate a need to examine whether corrective optics used in medical imaging are properly optimized for certain visual tasks.

Wide variability in observer performance was seen amongst the participants in this experiment. A similar variability was seen in the contrast threshold experiment in Chapter 4, and it was suggested that that optical, neurological, or cognitive differences exist between observers that are significant factors in visual performance. The contrast sensitivity function has been shown to change significantly due to various clinical conditions[29, 30] or simply due to age[31]. More in-depth studies are needed to determine if these factors have a significant effect on the perception of medically important patterns.

6.5 References

1. NEMA, *Digital Imaging & Communications In Medecine (DICOM), Part 14: Grayscale Standard Display Function*. 1998, National Electrical Manufacturers Association.
2. Barten, P.G.J., *Contrast sensitivity of the human eye and its effects on image quality*. 1999, Bellingham, Wash.: SPIE Optical Engineering Press. xix, 208.
3. Samei, E., M.J. Flynn, and W.R. Eyler, *Detection of subtle lung nodules: relative influence of quantum and anatomic noise on chest radiographs.* Radiology, 1999. **213**(3): p. 727-34.
4. Johns, H.E. and J.R. Cunningham, *The physics of radiology*. 4th ed. 1983, Springfield, Ill., U.S.A.: Charles C. Thomas. xix, 796 p.
5. Luijendijk, H.A. *Practical experiment on noise perception in noisy images*. in *Medical Imaging 1994: Image Perception*. 1994. Newport Beach, CA, USA: SPIE.
6. Barten, P.G.J. *Formula for the contrast sensitivity of the human eye*. in *Image Quality and System Performance*. 2003. San Jose, CA, USA: SPIE.
7. Samei, E., et al., *Assessment of display performance for medical imaging systems: executive summary of AAPM TG18 report.* Med Phys, 2005. **32**(4): p. 1205-25.
8. Normann, R.A. and I. Perlman, *The effects of background illumination on the photoresponses of red and green cones.* J Physiol, 1979. **286**: p. 491-507.
9. Baxter, B., H. Ravindra, and R.A. Normann, *Changes in lesion detectability caused by light adaptation in retinal photoreceptors.* Invest Radiol, 1982. **17**(4): p. 394-401.
10. Hecht, S., *The relation between visual acuity and illumination.* J. Gen. Physiol., 1928. **11**(3): p. 255-281.
11. Baylor, D.A. and M.G. Fuortes, *Electrical responses of single cones in the retina of the turtle.* J Physiol, 1970. **207**(1): p. 77-92.
12. Flynn, M.J., et al., *High-fidelity electronic display of digital radiographs.* Radiographics, 1999. **19**(6): p. 1653-69.
13. Flynn, M., *Visual requirements for High-Fidelity Display.* Advances in Digital Radiography: RSNA Categorical Course in Diagnostic Radiology Physics 2003, 2003: p. 103-107.
14. Press, W.H., *FORTRAN numerical recipes*. 2nd ed. 1996, Cambridge [England] ; New York: Cambridge University Press. v.
15. Samei, E., et al. *DQE of direct and indirect digital radiography systems*. in *Medical Imaging 2001: Physics of Medical Imaging*. 2001. San Diego, CA, USA: SPIE.
16. Saunders, J.R.S. and E. Samei, *A method for modifying the image quality parameters of digital radiographic images.* Medical Physics, 2003. **30**(11): p. 3006-3017.
17. Barrett, H.H. and W. Swindell, *Radiological imaging : the theory of image formation, detection, and processing*. 1981, New York: Academic Press. 2 v. (xxiii, 693 p.).
18. Samei, E. and M.J. Flynn, *An experimental comparison of detector performance for computed radiography systems.* Medical Physics, 2002. **29**(4): p. 447-459.

19. Jarvis, J.R., N.B. Prescott, and C.M. Wathes, *A mechanistic inter-species comparison of flicker sensitivity.* Vision Research, 2003. **43**(16): p. 1723-1734.
20. Tyler, C.W., *Analysis of visual modulation sensitivity. II. Peripheral retina and the role of photoreceptor dimensions.* J. Opt. Soc. Am. A, 1985. **2**: p. 393-.
21. Flynn, M.J. and E. Samei, *Experimental comparison of noise and resolution for 2k and 4k storage phosphor radiography systems.* Med Phys, 1999. **26**(8): p. 1612-23.
22. Legge, G.E., *A power law for contrast discrimination.* Vision Res, 1981. **21**(4): p. 457-67.
23. Levi, D.M., S.A. Klein, and I. Chen, *What is the signal in noise?* Vision Res, 2005. **45**(14): p. 1835-46.
24. Blackwell, K.T., *The effect of white and filtered noise on contrast detection thresholds.* Vision Res, 1998. **38**(2): p. 267-80.
25. Sund, P., et al. *A comparison between 8-bit and 10-bit luminance resolution when generating low-contrast sinusoidal test pattern on an LCD*. in *Medical Imaging 2007: Image Perception, Observer Performance, and Technology Assessment*. 2007. San Diego, CA, USA: SPIE.
26. Krupinski, E.A., et al. *Influence of 8-bit versus 11-bit digital displays on observer performance and visual search: a multi-center evaluation*. in *Medical Imaging 2007: Image Perception, Observer Performance, and Technology Assessment*. 2007. San Diego, CA, USA: SPIE.
27. Badano, A., et al., *Noise in flat-panel displays with subpixel structure.* Med Phys, 2004. **31**(4): p. 715-23.
28. Fan, J., et al., *Evaluation of and compensation for spatial noise of LCDs in medical applications.* Medical Physics, 2005. **32**(2): p. 578-587.
29. Woods, R.L., N.C. Strang, and D.A. Atchison, *Measuring contrast sensitivity with inappropriate optical correction.* Ophthalmic and Physiological Optics, 2000. **20**(6): p. 442-451.
30. Lombardo, A.J., et al., *Changes in contrast sensitivity after Artisan lens implantation for high myopia.* Ophthalmology, 2005. **112**(2): p. 278-285.
31. Owsley, C., R. Sekuler, and D. Siemsen, *Contrast sensitivity throughout adulthood.* Vision Res, 1983. **23**(7): p. 689-99.

CHAPTER 7

CONCLUSIONS

This thesis has examined the contrast performance of liquid crystal display (LCD) devices for use in medical imaging. Novel experimental tools and computational methods were used to experimentally measure the ability of medical LCD devices to produce patterns whose contrast is at or below the just noticeable limit of the human visual system. It was demonstrated that medical LCD devices are capable of high performance in medical imaging tasks and are suitable for conducting psychovisual research experiments.

Novel methods for measuring and controlling the luminance response of an LCD were presented in Chapter 3. A set of software tools, referred to as pacsDisplay, was developed to obtain a palette of 766 or 1786 gray levels and then employed to produce a calibrated 8-bit grayscale LUT. This software was used to apply DICOM GSDF calibrations to several medical LCD systems, demonstrating that the methods used could reliably measure luminance and manipulate fine contrast. In addition, it was shown that luminance responses averaged from a sample of LCD devices of the same make and model can be used to produce generic LUTs that are sufficient for secondary display devices. The pacsDisplay software has been used by two major medical centers, Henry Ford Health System and the Mayo Clinic, and its techniques have been adopted by National Display Systems for their calibration software. Further development of these methods may allow for the measurement of larger palettes, although the current 8-bit grayscale limits their effectiveness.

In order to evaluate the effectiveness of LCD devices for medical imaging tasks, human observer experiments are needed to determine the performance achieved with LCD devices in relation to human observer performance previously documented in

psychovisual experiments. However, relatively few observer performance experiments have previously been done using LCD monitors. In addition, such evaluations are made difficult by the irregular luminance responses of LCD devices. Chapter 4 reported on a novel method to generate very low contrast bi-level bar patterns by using the full palette of available gray values. The method was used in a two alternative forced choice (2AFC) psychovisual experiment to measure the contrast threshold of human observers. A z-score method was used to analyze the results, which were found to be consistent with the Barten model of contrast sensitivity. Notably, the experiment was able to successfully measure contrast thresholds levels for certain observers that were well below the expected mean value.

Chapter 5 examined error distortion associated with using the z-score method to analyze a non-linear process such as the psychometric function. An MLE statistical method was presented as an alternative to using z-scores, and an algorithm was derived to analyze 2AFC binary data. The method was used to re-evaluate the experimental results described in Chapter 4. The new contrast thresholds results were slightly different, but were still consistent with the Barten model. Simulations were conducted to evaluate the statistical precision of the MLE method in relation to the number and distribution of trials. The results of these simulations provided information that could be used in designing future 2AFC testing strategies to minimize error, including the possibility of adaptive testing.

New experiments to measure the human contrast threshold for complex patterns with varying luminance were presented in Chapter 6. These experiments built upon the methods and results from Chapters 3-5. A narrow luminance response calibration was used to achieve very small contrast steps in order to build sub-contrast patterns containing a range of gray levels. Processes were developed to insert patterns into images and were used to build 2AFC tests examining contrast thresholds for sinusoid, white noise, and filtered noise patterns under varying background luminance conditions. The results of the sinusoid tests using background luminance equal to the target luminance were consistent with the results from Chapters 4 and 5. For the sinusoid tests where the background luminance was allowed to differ, the average contrast threshold increased significantly, demonstrating the reduced contrast sensitivity due to adaptation

to luminance levels that are shifted away from the target luminance. The results of the noise pattern tests showed that quantum mottle patterns require more contrast to be just noticeable when compared to sinusoid patterns. Compared to white noise, noise patterns that were filtered to model images from indirect radiography required significantly higher zero-frequency noise power to be visible, but also exhibited lower overall noise power in the final image. These results indicated that the input and output standard deviations were insufficient for determining the just noticeable condition. A noise visibility index (NVI) was introduced to describe the noise power in a target pattern weighted by contrast sensitivity as a function of spatial frequency. White noise and filtered noise patterns exhibited similar NVI values, indicating that this measure may be used for determining the visibility of complex patterns.

These experiments also demonstrated the limits of medical LCD systems associated with 8-bit grayscales. The results showed that current medical LCD devices are capable of high performance in presenting contrast for psychovisual testing over a limited luminance range. However, 256 gray levels are insufficient for displaying typical medical images with high fidelity. 10-bit grayscales can provide the number of gray levels needed to achieve high fidelity performance for a full 350 luminance ratio, and can allow this experimental model to be extended to radiographic images. Should greater bit-depth become more widely available in display systems, then it may be possible to improve the calibrated grayscale to include richer contrast variations for medical imaging applications. However, it is not known whether such changes would affect diagnostic performance in radiology.

Optimization of medical image presentation requires an in-depth understanding of the capabilities of medical LCD devices and the limits of human observer performance. The work detailed in this dissertation provides the initial concepts and tools by which to accomplish this. The studies conducted in this dissertation help to further validate the sub-threshold contrast capabilities of medical LCD devices and establishes the techniques used as a high-performance experimental vehicle for psychovisual testing in medical imaging. The results of these studies have helped to establish basic concepts regarding the visualization of contrast and noise, and have laid the groundwork for future studies to expand on these and other similar issues of importance to medical imaging. For example,

further research is needed to examine the effects of high fidelity and other presentation techniques on observer performance in diagnostic medical imaging tasks. Additional research is also needed to model the NVI, or to develop other measures of human visual detection for the viewing of more complex target patterns and backgrounds. These include the simulation of disease features, such as nodules and lesions, as well as the use of realistic radiographic backgrounds with anatomic structures. Improvements in the experimental methods or equipment used are also of significant interest. Better procedures or increased bit-depth for calibrating medical LCD devices for psychovisual testing may provide reduced measurement noise, making it easier to produce low-contrast target patterns. Improved test design, such as incorporating adaptive testing techniques or accounting for possible sources of observer variability, may provide more efficient testing, reduced observer fatigue, and more accurate observer performance measurements.

APPENDIX

INSTRUCTION MANUAL FOR PACSDISPLAY

[The information below is taken from "README-HFHS_pacsDisplay.txt", distributed with pacsDisplay. Those interested in the software can request a copy by sending an email to pacsDisplay@rad.hfh.edu.]

HFHS-pacsDisplay

Microsoft Windows programs for generating and installing
DICOM grayscale look up tables (LUTs). Includes applications
to display grayscale test patterns.

IMPORTANT: Only use with LUTs intended for the make and model
of monitor used on your workstation. Requires a configuration
file specific to the computer that these programs are used on.

M. Flynn & P. Tchou
Dec, 2006

GENERAL PUBLIC LICENSE:
HFHS-pacsDisplay

This program is free software; you can redistribute it and/or modify
it under the terms of the GNU General Public License as published by

the Free Software Foundation; either version 2 of the License, or
(at your option) any later version.

This program is distributed in the hope that it will be useful, but
WITHOUT ANY WARRANTY; without even the implied warranty of
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.
See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License
along with this program; if not, write to the Free Software
Foundation, Inc., 51 Franklin Street, Fifth Floor, Boston, MA
02110-1301, USA.

The full license can be found in the GNU-GPL.txt document in the
HFHS/pacsDisplay directory.

CONTACT INFORMATION:

1. Michael Flynn, Henry Ford Health System
One Ford Place - Suite 2F
Detroit, MI 48202
mikef@rad.hfh.edu

2. Philip Tchou, University of Michigan
One Ford Place - Suite 2F
Detroit, MI 48202
pmtchou@umich.edu

CONTENTS:

---

1. Installation

HFHS-pacsDisplay may be installed by a simple Microsoft batch file that invokes an installation tcl script. Instructions for installation using this are summarized in section 1.1 below. Alternative manual installation methods are described in section 1.2.

- No registry entries are made or changed by this install process and no system background services are installed.
- Adjustments to the display LUTs are made by a call to the graphic driver that is executed for each user upon login.

IMPORTANT:
Once installation is complete, you must setup look-up tables (LUTs) and configuration files for the pacsDisplay software to function properly. See Section 3 for details.

1.1 Quick Installation:

IMPORTANT - The installation described below will fully install executable programs and the distributed LUT-library. However, the current system described is distributed with a LUT based on a Dell 1905FP monitor. You will be able to view the effects of this LUT in relation to a linear LUT to verify that the loading of a LUT is operational. However, achieving a properly calibrated display will require configuration with the proper LUT file.

Note: These instructions assume that you have just unzipped the pacsDisplay distribution files and are currently viewing the .../HFHS directory which includes this document and the pacsDisplay folder.

Steps for installing the HFHS-pacsDisplay package:

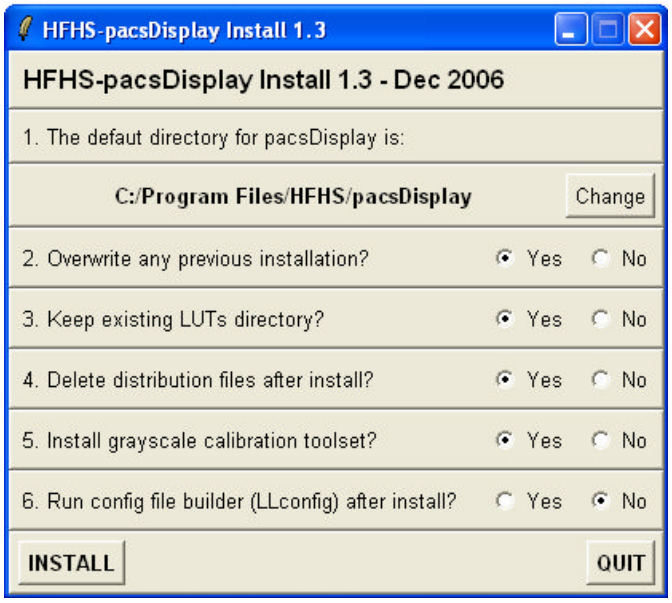

Figure A.1 pacsDisplay installation utility

Step 1 - Open the 'pacsDisplay' folder and run the 'pacsDisplay_install.bat' file.
Step 2 - Review the terms of the license agreement.
Step 3 - Select the installation options you want.

1. "The default directory for pacsDisplay is:" (Change)

   Default: 'C:/Program Files/HFHS/pacsDisplay'

   It is recommended that the default directory for installation of the pacsDisplay files be used. Shortcuts will always be installed to specifc folders in the Start Menu regardless of how this option is set. However, the shortcut targets will need to be changed if a non-default directory is specified.

2. "Overwrite any previous installation?" (Yes/No)

   Default: Yes

   This option should be set to "Yes" if you want to overwrite a previous pacsDisplay installation. A "No" response will abort installation if a prior version is encountered.

3. "Keep existing LUTs directory and configuration files?" (Yes/No)

   Default: Yes

   Answering "Yes" to this option will save the LUTs directory from a previous installation (including older versions) and use it in the new installation. This only applies if it was previously intalled in the default location.
   - If the "Linear" directory is found at the same directory

level as "Current System", it will be moved into the "Current System" directory.

- If the "LUT-Library" is not found, it will be installed.

NOTE: If you want to update the "LUT-library" but keep the "Current System", select Yes and manually update the "LUT-library".

4. "Delete distribution files after install?" (Yes/No)

Default: Yes

Answering "Yes" to this option will cause the installation files to be deleted once the install process is complete. All of these files will be copied to the target installation directory beforehand (see option 1). This option is ignored if the working directory and the target installation directory are the same, in which case the files will not be deleted.

5. "Install grayscale calibration toolset?" (Yes/No)

Default: Yes

Answering "Yes" to this option will install shortcuts to the Start Menu for the various utilities that come with the pacsDisplay program. These tools are intended for IT/physics support.

Answering "No" will install only the enterprise shortcuts, which provide tools for applying and verifying the calibration.

6. "Run the config file builder (LLconfig) after install?" (Yes/No)

Default: No

When set to "Yes", this option will run the LLconfig program after installation. LLconfig is a tool to help build configuration files for pacsDisplay applications. It is recommended that you not use this option unless you are familiar with LLconfig. Instructions for its use can be found in the pacsDisplay directory in the Readme file (README-HFHS_pacsDisplay.txt). There will be an option to view the Readme file at the end of the install process.

Step 4 - Review your selections and press the "INSTALL" button when ready.

Step 5 - If the installation completes successfully, an option will be given to view the Readme file (README-HFHS_pacsDisplay.txt) for further details and instructions.

1.2 Manual Installation

The following instructions are intended for manual installation.
The files and directories involved only need to be placed in the required locations. No registry entries need to be made or changed.

This package is intended only for installation on a system having a 'C:\Program Files' directory.

Unzip the HFHS-pacsDisplay.zip file to 'C:\Program Files'.
You should see the following path:

C:\Program Files\HFHS\pacsDisplay\..

Do not change the organization of files under this path.

In the above path is a 'shortcuts' directory. It contains three directories with shortcuts in them which need to be copied to:

C:\Documents and Settings\All Users\...

where '...' depends on the name of the source directory. Each of these directories is named so as to indicate where their contents should be copied to.

There are three locations indicated in the directory names:

* allUsers_startMenu => \Start Menu
  A single shortcut to iQC that places an icon at the top of the Start=>Programs menu that starts the test pattern application.

* allUsers_startMenu_programs_startup => \Start Menu\Programs\Startup
  A single shortcut to loadLUT-dcm that will load a DICOM grayscale LUT to the graphic card whenever a user logs into the system.

* allUsers_startMenu_programs => \Start Menu\Programs
  There are two folders for which one or both can be moved to the ..All Users\programs directory:

  * HFHS ePACS grayscale
    Two shortcuts to start the test pattern application and to start an application that allows the user to change between a DICOM grayscale and a LINEAR grayscale.

* HFHS Grayscale Tools
  A full set of shortcuts to applications for
  measuring the luminance response, generating
  calibrated LUTs, and installing LUTs.
  This should only be installed for qualified users.

It is possible to manually install the application on a drive other than C:\Program Files\... If this is done, the shortcut targets summarized above need to be modified for the correct path. Each shortcut must have the 'read-only' state removed, the paths modified including the icon paths, and the shortcut saved. It is suggested that the shortcuts be set to 'read-only' after these changes are made.

---

2. Application Summary

Below are brief descriptions of each of the applications included in the pacsDisplay package. Further details on these programs and how to use them can be found in sections 3.

2.1 HFHS ePACS Grayscale applications

---

iQC:

* Presents an image quality test pattern that demonstrates the grayscale characteristics of a display.
* The base image size is 756 wide by 792 high.

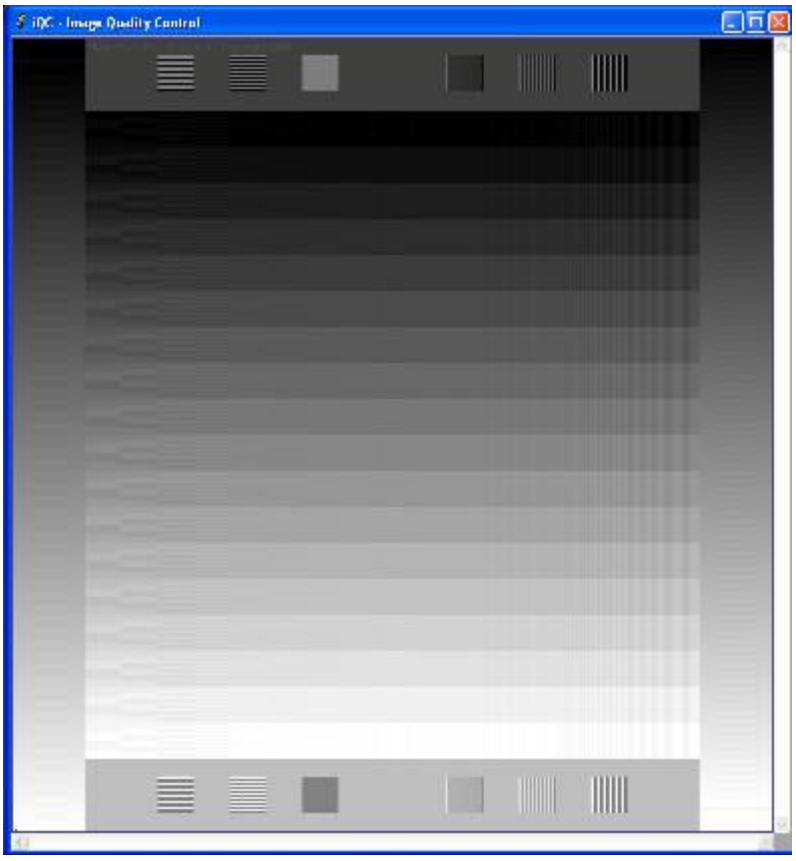

Figure A.2 pacsDisplay iQC test pattern

* The window size can be made smaller than the image, in which case the image can be panned using either the scroll bars or 'click and drag' using the left mouse button.
* An alternate image of 2X size can be loaded using the <UP> arrow key. The window size will stay the same. The regular size image is then obtained using the <DOWN> arrow key.
* An alternative window size of 2X that displays the 2X image can be obtained using the <RIGHT> arrow key. The normal window size is obtained using the <LEFT> arrow key.

ChangeLUT:

* Presents a small window with button to load the DICOM Grayscale currently configured or an alternative LINEAR Grayscale.

* When used with the iQC test pattern displayed, this provides an effective way to see the difference in image appearance between the normally loaded grayscale (LINEAR) and the DICOM grayscale.

loadLUT-dcm:

* Only put in the All Users startup directory for the ePACS installation.
* Loads the DICOM grayscale at startup.

HFHS Grayscale Toolkit applications (additions to the above)

gtest:

* Application to present uniform regions with adjustable gray levels to support macro images of lcd pixel structures. All controls are implemented with key bindings.

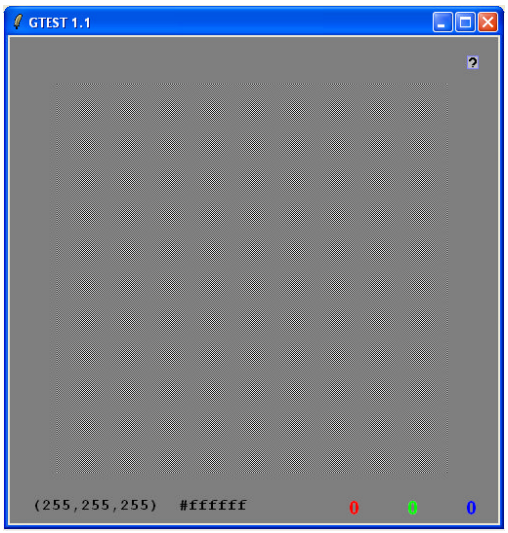

Figure A.3 pacsDisplay gtest utility

iQC-2x

* Same as iQC but opens in the 2X image/window size.

LLconfg

* This program provides a form-fillable interface to help build a config file for LoadLUT.
* Supports up to 4 displays and provides access to the model and serial number information from the EDID.

loadLUT-dcm(lin)

* Depending on the argument in the shortcut, this loads the LUTs in the system/linear directory. LoadLUT.exe is executed from the execLoadLUT.exe program to catch and report errors.
* The error messages reported can be changed in the execLL-messages.txt file in the loadLUT directory.

loadLUT-demo

* Provides a utility tool to find a LUT file and load it to a specific monitor number.

lumResponse

* Application to generate a test pattern that steps through a palette of gray levels so that luminance can be measured using an IL1700 luminance meter connected using a serial line interface.
* This is used in 256 mode to measure the calibrated response of a monitor having DICOM grayscale LUTs installed.
* This is used in 766 or 1786 mode to measure the uncalibrated response of a monochrome or color display having a LINEAR LUT installed.
* Make, model, and serial number of display normally used for identification.
* A plot of the luminance vs p-values is provided at the end of a measurement.

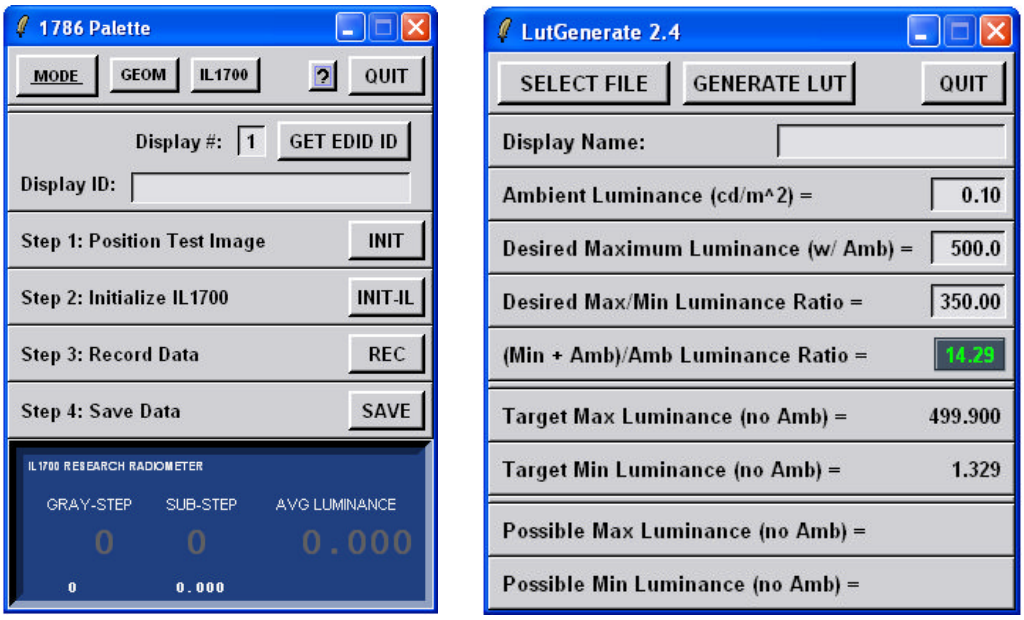

Figure A.4 pacsDisplay LumResponse and LutGenerate utilities

LutGenerate

* Generates a LUT for installation with loadLUT.
* Reads luminance response measured by lumResponse.
* Requires specification of maximum luminance, luminance ratio and ambient luminance.

Uninstall pacsDisplay

* Uninstall program that removes all files and folders from the 'C:/Program Files/HFHS/pacsDisplay' directory.
* An option is provided to save the LUTs directory.

---

3. Creating the LUT File

Once intallation is complete, the first step in calibrating your display is to create a lookup table (LUT). The LUT is a list of 256 RGB values

used to replace the standard grayscale values (R=G=B) in order to match with the DICOM grayscale standard. Creating the LUT requires measuring the full range of gray values that the display is capable of and then taking those values that are closest to the desired DICOM grayscale.

Two tools are provided for this purpose. The first is LumResponse, which measures the uncalibrated luminance response (uLR) of a display. This is followed by LutGenerate, which takes a uLR file and builds a LUT to match your specifications.

3.1 LumResponse

LumResponse is used to measure the gray pallette for a display and record the luminance values in a uLR text file.

This program will put up a large window on the screen with a grey background and a central target region. This pattern is intended for assessing display luminance response. The grey intensity of the target region is cycled through increasing intensity values to measure luminance versus grey value. Currently the measurements are made by an IL1700 luminance meter connected using a serial line interface.

The 'Mode' button at the top of the utility selects the type of measurement to be made. The current mode is displayed in the window title. The various modes are as follows:

1786 LEVELS - This mode is used for measuring the uncalibrated luminance response of a color display. Make sure the display is using a linear LUT (i.e - no calibration) before proceeding.

766 LEVELS - This mode is used for measuring the uncalibrated luminance response of a monochrome display. Make sure the display is using a linear LUT (i.e - no calibration) before proceeding.

256 LEVELS - This mode is used for measuring the luminance response of a display after a LUT has been installed in order to verify that a DICOM calibration is in place.

DEMO MODE - This mode is for demonstration purposes. It will quickly cycle from black to white but no measurements will be taken. The IL1700 does not need to be connected for the demo to run.

OTHER - This option allows the user to use a customized perturbation series for the luminance measurements. For advanced users only.

The "GEOM" and "IL1700" buttons provide access to advanced options for the test image window and the IL1700 photometer, respectively. The "GEOM" options can be used to customize the geometry and position of the test image. The "IL1700" options include settings for communicating with the photometer and controlling the way display luminance is measured. Further information is given byt "?" buttons next to each setting.

Once a mode has been selected, a Display ID should be created to identify the measurement results once they are saved. The name that is selected will be used to name the output file.

The "Display #" box is used to select the display that you wish to identify. This number is the same as that used by Windows and listed in the Display Properties window. Pressing the "GET EDID ID" button will then retrieve the model and serial number (S/N) information from the EDID for that display and

use it to build the Display ID. The result is displayed in the text box. A custom ID can be entered into the text box if so desired. It is suggested that the model and serial number of the display be included.

The remaining portion of the lumResponse window provides the steps for the luminance response measurement and a button to begin each step. They are described here:

Step 1: Position Test Image

This opens up a test image window that needs to be centered on the screen of the display to be measured. The window should be about the same size as the screen. You may want to set the display to be measured as the primary display in order to get the right test window size. Otherwise, you can change the display size manually using the GEOM settings.

Step 2: Initialize IL1700

This button begins communication with the IL1700 luminance meter. It activates the IL1700 monitor at the bottom of the window, providing information regarding the measurement. The details of this monitor are explained further below.

At this point, the luminance meter should be setup in front of the display, centered on the square target in the middle of the test window. The end of the photometer should be positioned close to the screen, but not touching it in order to prevent distortions. A dark cloth should be placed over the display and photometer to block ambient light. It is suggested that the cloth not cover any vents in the back of the display as this can cause it to heat up quickly and may affect the measurement.

Step 3: Record Data

This starts the measurement process. The luminance meter should be properly

setup before hitting this button. Pressing this button once a measurement has begun will pausse the measurement and provide an option to abort or continue.

Step 4: Save Data

Press this button to save the luminance data once a measurement is complete. You will be asked where to save the output file. An option to plot the luminance vs p-values is provided after the data is saved.

IL1700 Research Radiometer:

GRAY-STEP - The current stage of the luminance measurement. This number represents the standard graylevel (1-256). Negative values represent the colored frames that are measured at the beginning before the gray levels.

SUB-STEP - This second number represents the perturbation steps in the graylevel sequence. There are 7 steps in 1786 mode and 3 in 766 mode. The 256 and Demo modes do not use sub-steps.

AVG LUMINANCE - The average luminance value measured for each gray step.

Additional indicators are displayed in small print at the bottom of the monitor, without labels. They are as follows, from left to right:

(Measurement #) - The number at the bottom left is the number of accepted measurements that have been made.

(IL1700 value) - The luminance value from the IL1700, updated in real-time.

(OUTLIER) - This indicator only appears when an outlier is detected, and remains lit while the outlier is being resolved.

3.2 LutGenerate

LutGenerate takes a uLR file (output by LumResponse) and builds a LUT file based on the parameters you specify. Directions are listed here:

1. Select a uLR File

   Start by pressing the "SELECT FILE" button and choosing the appropriate uLR file for the display to be calibrated. LutGenerate will read the uLR file and automatically update the fields throughout the window as appropriate. This includes the 'Desired Maximum Luminance' field, which will be set to the maximum luminance value found in the uLR file. Both the display name and desired maximum luminance may be changed manually after loading a uLR.

2. Determine Calibration Parameters

   Three parameters must be specified before generating the calibration LUT, the ambient luminance expected for the display, the desired maximum luminance, and the desired luminance ratio.

3. Verify Calibration Parameters

   Hitting the 'ENTER' key after changing one of the calibration parameters will update the other values presented below:

   (Min + Amb)/Amb Luminance Ratio - The AAPM TG-18 report calls for this value to be no smaller than 2.5 and recommends a value of at least 5. The text changes to yellow if the value is below 5 and red if it falls below 2.5.

Target Max Luminance (no Amb) - This value is equal to the desired maximum luminance minus the ambient luminance. The text will turn red if it goes above the possible maximum luminance indicated by the uLR.

Target Min Luminance (no Amb) - This value is equal to the desired minimum luminance minus the ambient luminance. The text will turn red if it goes below the minimum luminance indicated by the uLR.

Possible Maximum Luminance (no Amb) - The largest luminance value found in the selected uLR.

Possible Minimum Luminance (no Amb) - The smallest luminance value found in the selected uLR.

Check to be certain that these values are correct before continuing.

4. Generate the LUT

Click on the "GENERATE" button to build the calibration LUT. You will be asked where to save the LUT file. It is recommended that you save it within the '..pacsDisplay\LUTs\Current System\' directory.

---

4. loadLUT

Usage: loadLUT.exe [(working directory)]

Command Line Options:

(working directory) - If a working directory is specified, then all input files will be read from that directory. The log file will also

be written to that directory. If no directory is specified, the current directory will be used.

loadLUT is the program in the pacsDisplay package responsible for applying a calibration LUT to a display. It reads from a configuration file, 'configLL.txt', which it looks for in its starting directory or working directory (if specified). ChangeLUT, execLoadLUT, and loadLUTdemo provide expanded interfaces for running loadLUT. Further details on using loadLUT and related utilities are given in this section.

4.1 The Configuration File

---

loadLUT requires a configuration file, 'configLL.txt', to designate which displays are to be calibrated and which LUTs to load. It will look for this file either in the same directory as loadLUT.exe or in a specified directory, as mentioned in the usage instructions above.

For a standard pacsDisplay installation, there are configLL.txt files in two strategic directories that need to be properly configured relative to the make and model of each monitor installed on the system:

C:\Program Files\HFHS\pacsDisplay\LUTs\Current System

C:\Program Files\HFHS\pacsDisplay\LUTs\Linear

Below is the standard layout for the configLL.txt file, configured for a two-monitor system:

```
# First 2 lines reserved for comments.
#
/LUTsearch [dir]   - Optional line, invokes model/SN LUT search
/LDTsearch         - Optional line, invokes dated LUT search
/noload            - Optional line, checks LUT but does not load
/noEDID            - Optional line, bypasses EDID checks
/nolog             - Optional line, prevents writing of log file
2                  - Number of displays to be calibrated
1                  - Display number
"DLL MMMM1"        - Model descriptor (or "*")
"SNSNSNSN1"        - Serial number (or "*")
"calLUT1.txt"      - Default calibration filename
2
"DLL MMMM2"
"SNSNSNSN2"
"calLUT2.txt"
```

Comments:

The first two lines are reserved for comments and will not affect how loadLUT performs.

Options:

Options, if present, are included right after the comment lines and must start with a '/' character. The following options are currently available:

/noload - when this option is set, loadLUT will save the current display LUTs in backup files without loading new LUTs.

/noEDID - this option prevents loadLUT from searching the registry for EDID information. This may allow

loadLUT to avoid errors with some display configurations, but also disables loadLUT's ability to verify display information.
THIS OPTION IS NOT RECOMMENDED.

/nolog - when this option is set, loadLUT will not attempt to generate a log file. This may be needed if loadLUT is run under an account that has limited access privelages.

/LUTsearch [dir (optional)] - if this option is set, loadLUT gets the LUT filename from the config file, but does not immediately use it. It first checks the registry for a monitor's EDID values and gets the model name and serial number. If it cannot find them, it will instead get these values from the config file. It then uses them to find the appropriate LUT file from the ...\LUT-Library\<model_name*>\LUTs directory. The "*" denotes a wildcard. If the folder name has characters beyond <model_name>, it will still be accepted.

If this search fails, loadLUT will look in the ...\LUT-Library\<model_name*> directory for a generic LUT file. If that also fails, it will use the default file.

LUTsearch can take a directory string as an argument, directing where it should search for the <model_name> directories. Long directory

names (those with spaces) should be enclosed in quotes ("). If no directory is specified, it defaults to the standard LUT-Library directory:

"C:\Program Files\HFHS\pacsDisplay\LUTs\LUT-Library"

If the /noEDID option is set, then the model name and serial number are taken from the config file.

LUTsearch file format:
LUT_<model_name>_<S/N>_*
where <S/N> can be either the 4 digit VESA EDID number or the extended VESA EDID number

LUTsearch generic file format:
LUT_<model_name>_GENERIC*

Note: Previous versions used the ...\Current System\LUT directory for the LUTsearch option. This directory will no longer be accessed unless loadLUT is instructed to do so manually as detailed above.

/LDTsearch [#] - this option is similar to the LUTsearch option. It uses the year and week of manufacture, along with the model name from the EDID to find the appropriate LUT file from the ..\Current System\LDT\ directory. The "#" indicates the date tolerance, i.e. - the number of weeks before or after the specified date that the search will accept. The default is 3 weeks.

If the /noEDID option is set, then the search fails.

file format:

LDT_<model_name>_<year(xxxx)><week(1-52)>_*

Note: If both LUTsearch and LDTsearch are set, LUTsearch takes priority. If LUTsearch fails, loadLUT will still run LDTsearch. If LDTsearch also fails, then loadLUT will use the default LUT file.

Number of Displays:
Following the options is the number of displays to be calibrated. Each display is listed below this line and four lines must be present for each display.

Display Number:
The reference to display number is for the number that Windows reports for each display under Display Properties => Settings.

Model Descriptor (or "*"):
The model descriptor is checked against what is listed in the EDID. A mismatch will cause loadLUT to output an error message and will not load a LUT. It will then continue on to the next display. The Model can be replaced with a "*" in order to bypass the check for that line. If there are spaces in the model descriptor, it should be enclosed in quotes. This is generally recommended even if there are no spaces.

Serial number (or "*")

The serial number is checked against what is listed in the EDID. A mismatch will cause loadLUT to output an error message and will not load a LUT. It will then continue on to the next display. The Model can be replaced with a "*" in order to bypass the check for that line. If there are spaces in the serial number, it should be enclosed in quotes. This is generally recommended even if there are no spaces.

In general, the EDID will contain a 4-digit serial number. Some EDIDs also include an extended serial number longer than 4 digits. If either one matches the serial number given in the config file, then this check will be successful.

Default Calibration Filename:
The next line is the default calibration filename. This is the LUT file that the progarm will use to adjust the display. If any of the search options are in place, then they will take precedence in selecting a LUT file. Should the search options be unsuccessful, then loadLUT will use this filename by default.

LUT files are expected to be in the working directory

For '..\LUTs\Current System\Linear\configLL.txt', the filenames should be 'linearLUT.txt'.

For '..\LUTs\Current System\configLL.txt', the DICOM grayscale LUT for each monitor can vary.

These LUT files are typically generated using the LutGenerate program after measuring the intrinsic grayscale response using lumResponse.

For this distribution, the LUT-library contains collections of

uncalibrated luminance response files, uLR files, along with derived average response and calibration LUT files for these monitors:

DELL_1905FP_(3g)

DELL_2001FP_(21g)

DELL_2007FP_(10g)

where the number in parentheses indicates the number of uLR files. The "g" indicates that a generic file also exists. Additionally, directories with uLR files but no generic LUT are included for seven (7) other monitor models.

A 'linearLUT.txt' file is included in the Linear directory. This LUT is identical to the standard R=G=B pattern that Windows uses when no custom LUT is being asserted.

For this distribution, the Current System LUT directory has a configLL.txt file that asserts the \LUTsearch option with the model_name and S/N set to "*". If an S/N match or GENERIC match is not found in the LUT-library, the default LUT is assigned to that for a specific Dell 1905FP monitor. This is done so that a change can be observed immediately after installation when the iQC test pattern is shown and changeLUT is used to switch between DICOM and Linear grayscales.

IMPORTANT - The 'Current System' configLL.txt file must be properly configured for the monitors that are being used as is illustrated in the examples below.

4.2 Sample Config Files

Example 1:

This configuration is for a single display, identified in Windows as display "1", with no model name or S/N verification.

```
# First 2 lines reserved for comments.
#
1
1
"*"
"*"
"<LUT filename for display #1>"
```

Example 2:

This configuration extends to display "3". The EDID information for this display will be checked for a matching model descriptor. If the model name given here is different from what is found in the EDID, an error will occur and the program will abort. No check will be made to match a serial number.

```
# First 2 lines reserved for comments.
#
2
1
"*"
"*"
"<LUT filename for display #1>"
3
"DELL 1905FP"
"*"
"<LUT filename for display #3>"
```

Example 3:

This configuration includes the option to search for the LUT files in the '..\LUTs\LUT-Library' directory. For display #1, the filename will be based on the model name and serial number given in the EDID. For display #3, the model name "DELL 1905FP" will be checked against that in the EDID, while the serial number will be whatever is found in the EDID. These model names and serial numbers will be used to build the filenames for LUTsearch. If the search for the specific files and generic files are unsuccessful, then the listed default LUT files will be used instead.

```
# First 2 lines reserved for comments.
#
/LUTsearch
2
1
"*"
"*"
"<LUT filename for display #1>"
3
"DELL 1905FP"
"*"
"<LUT filename for display #3>"
```

Example 4:
This configuration includes a search based on the date of manufacture of the display. It will search through the files in '..\Current System\LDT\', choosing the one that is closest to the date of manufacture, but not going beyond 52 weeks from that date.

```
# First 2 lines reserved for comments.
#
/LDTsearch 52
```

```
2
1
"*"
"*"
"<LUT filename for display #1>"
3
"DELL 1905FP"
"*"
"<LUT filename for display #3>"
```

Example 5:
Here we have a 4 display system with LUTs being drawn from the '..\LUTs\LUT-Library' directory. However, since the /noEDID option is being used, LUTsearch will instead build the intended filenames using the monitor descriptors and serial numbers given below. Those that have only "*" for both the model and serial number will automatically fail the file search and will instead default to the given LUT filename. Display #3 gives a model name, but no serial number. LUTsearch will not be able to build a specific LUT filename for display #3, but it will still search for a generic file in the <model_name*> directory before going to the default LUT file.

```
# First 2 lines reserved for comments.
#
/noEDID
/LUTsearch
4
1
"*"
"*"
"<LUT filename for display #1>"
```

2
"*"
"*"
"<LUT filename for display #2>"
3
"DELL 1905FP"
"*"
"<LUT filename for display #3>"
4
"DELL 1905FP"
"T61164A5ABYU"
"<LUT filename for display #4>"

4.3 LLconfig Tool

LLconfig is a tool to help build a LoadLUT configuration file for a particular display setup. The format for the configuration file is described in the README-HFHS_pacsDisplay.txt document in the pacsDisplay directory.

REQUIREMENTS:
- The EDID functions require that getEDID.exe be in the same directory as the LLconfig executable.

USAGE:
1. Options - The top bar lists the possible LoadLUT options that can be set. Click on the boxes next to the option names to select (or de-select) them.

2. Number of Displays - Selecting a number means that entries for all of the

displays up to and including that number will be included. For example, if "3" is selected, entries for displays 1, 2, and 3 will be generated in the final configuration file. These display numbers correspond to the numbers in the "Display Properties" window.

3. Display Entries - Text entry bars are provided to list the configuration information for each individual display.

Monitor Descriptor - This is the name of the display as given by the EDID information. If left blank, a wild card character ("*") will be inserted into the configuration file instead.

Serial Number - This is either the 4-digit or extended serial number found in the EDID. If left blank, a wild card character ("*") will be inserted into the configuration file instead.

Default LUT file - This is the LUT file that will be applied to the display if LUTsearch and LDTsearch are not selected or if the file being searched for is not found. If left blank, "linearLUT.txt" will be inserted as the LUT file for that display.

GET EDID - This button runs getEDID and looks for the monitor descriptor and serial number entries for that display. If found, these items will be copied to the appropriate text entry bars. If both a 4-digit and extended serial number are found, then the user will be asked to select one.

SELECT LUT - This button opens a file-select window so that the user can browse for the LUT file they want for that display.

4. BUILD CONFIG FILE - This button will take the current form information and

build a configLL.txt file formatted for LoadLUT. The user will be asked where to save the file.

5. RESET FORM - This button resets the form to its initial state.

6. QUIT - This button exits the program. Form information will not be saved.