Calibration of Medical Displays

Effects of Luminance Conditions and the Limitations of the Human Visual System

Patrik Sund

UNIVERSITY OF GOTHENBURG

Calibration of medical displays

Effects of luminance conditions and the limitations of the human visual system

Patrik Sund

Faculty of Science

Department of Radiation Physics

UNIVERSITY OF GOTHENBURG

Gothenburg 2015

Cover: The photograph illustrates a limited dynamic range with a loss of contrast in dark and bright areas. In this case, the luminance range of the digital camera was insufficient to capture all details in the scene, but the same effect is also present in the human visual system. The graph overlay shows some of the possible display calibration curves developed in this work. Each curve compensates for any loss of contrast due to the limited range of the human visual system at a specific adaptation level.

Photo by Hillevi Jenslin Sund

"In order for the light to shine so brightly, the darkness must be present."

Francis Bacon

ABSTRACT

Calibration of medical displays is important in order for images to be displayed consistently. A consistent appearance ensures that all images are always perceived in the same way regardless of display device, location and time. Since a true consistent appearance requires displays with equal luminance ranges, which is neither practically achievable nor desirable, the aim of a calibration is rather to obtain a consistent distribution of perceived contrast throughout the gray-scale. An inconsistent contrast distribution may lead to an increased workload when reviewing and possibly an erroneous diagnosis. It is also important that a display calibration utilizes the luminance range of a display as efficiently as possible by avoiding gray-scale regions with low contrast. The widely used DICOM part 14 calibration method, the grayscale standard display function (GSDF), meets these necessary requirements for a large range of display settings. However, the GSDF does not account for the limited range of the human visual system (HVS) to detect low-contrast objects when adapted to a certain luminance level, so called fixed adaptation. The luminance range of modern displays is increasing, which is beneficial since the overall contrast increases, but when calibrated to the GSDF, an increasing luminance range compromises the intention of consistent contrast distribution and an effective use of the gray-scale.

The main aim of this thesis was to determine the properties of the HVS under conditions of fixed adaptation, and to use this information to derive a new calibration method that compensates the GSDF for fixed adaptation, thereby extending the original intentions of the GSDF for displays with a large luminance range. In order to study contrast properties of the HVS on medical displays under realistic conditions for a radiologist, a method using an extended image bit-depth together with a sub-pixel modulation technique, was developed to display sinusoidal test patterns close to the detection threshold. These patterns were used in several observer studies with different display luminance ranges and ambient lighting conditions.

The results show how the ability to detect low-contrast patterns, using equipment and viewing conditions typical for a radiology department, decreases when the difference between the luminance of a pattern and the adaptation luminance increases. A new calibration method is presented that compensates the GSDF for fixed adaptation, provided that the adaptation luminance is known. The new calibration method was, in an evaluation study, found to distribute the perceived contrast more consistently on high luminance range displays than the GSDF. Other results show that only the average luminance in an image, not the luminance distribution, affects the adaptation level. Also, light from outside the display may reduce the ability of an observer to detect low-contrast objects, especially when the luminance surrounding the display is greater than the average luminance of the display.

POPULÄRVETENSKAPLIG SAMMANFATTNING

Nästan all granskning av diagnostiska bilder inom den moderna sjukvården sker idag på bildskärmar. Dessa bildskärmar kan vara av oerhört låg kvalitet, dåligt inställda samt placerade i olämpliga ljusmiljöer vilket påverkar de visade bilderna negativt. Kalibrering av diagnostiska bildskärmar är viktigt för att bilder ska kunna visas på ett konsekvent sätt. En konsekvent bildvisning innebär att alla bilder alltid uppfattas likadant, oberoende av bildskärm, plats och tid. Eftersom en fullständigt konsekvent bildvisning kräver bildskärmar med samma omfång i ljusstyrka, vilket varken är praktiskt genomförbart eller önskvärt, är målet med en kalibrering snarare att åstadkomma en jämn fördelning av uppfattad kontrast i hela gråskalan. En ojämn fördelning av kontrasten medför att en och samma bild kan få olika utseenden på olika bildskärmar, vilket försvårar granskningsarbetet. Jämförelser av bilder blir svårare och kräver ständiga justeringar av varje granskare, vilket är både är tidskrävande och svårt att genomföra korrekt. Konsekvenserna blir ett försämrat arbetsflöde och troliga feldiagnoser. Det finns en standard som beskriver hur bildskärmar ska kalibreras för att visade bilder ska presenteras med en jämn fördelning av kontrasten, även om bildskärmarnas egenskaper varierar och de är placerade i olika ljusmiljöer. Denna standard gäller monokroma bilder och är främst avsedd för röntgendiagnostiken.

Teknikutvecklingen inom området går fort och på senare år har kvaliteten på bildskärmar höjts avsevärt. Moderna bildskärmar för diagnostiskt bruk är väldigt ljusstarka, vilket är positivt eftersom detta möjliggör en bildvisning med hög kontrast. Den existerande standarden är däremot inte anpassad för stora skillnader i ljusstyrka mellan mörka och ljusa områden i visade bilder. Orsaken till att bilder uppfattas olika beroende på det använda ljusomfånget ligger hos det mänskliga synsystemet. Även om ögat kan uppfatta ljus i ett oerhört stort område från nästan mörker till direkt solljus, så är förmågan att urskilja detaljer väldigt begränsad när ögat är anpassat (adapterat) till en viss ljusnivå, t.ex. vid bildgranskning. Då bildskärmar med olika ljusomfång används kommer fördelningen av kontrast inte längre vara jämn om den existerande standarden följs, då denna inte tar hänsyn till synsystemets begränsningar.

I denna avhandling presenteras en metod för att genomföra studier av synsystemets begränsningar under förutsättningar relevanta för klinisk granskning på bildskärmar. Genom att använda testmönster med en kontrast på gränsen till det synbara utfördes ett antal observatörsstudier med syfte att klarlägga hur bildskärmarnas ljusinställningar samt omgivningsljuset påverkar förmågan att upptäcka objekt med låg kontrast. Då en normal bildskärm endast klarar att visa mönster som är klart synliga krävdes en metod där testmönstren visades med ett utökat bitdjup. Dessutom användes de individuella färgkanalerna för rött, grönt och blått för att ytterligare förfina detaljerna hos testmönstren.

Resultaten från de olika observatörsstudierna i denna avhandling visar bl.a. hur mycket förmågan att detektera små kontraster minskar då testmönstret visas vid en annan ljusstyrka än den som ögat är anpassat till. Denna information har använts för att kompensera den existerande kalibreringsstandarden för egenskaperna hos det mänskliga synsystemet och på så sätt åstadkomma en jämn fördelning av uppfattad kontrast, även hos bildskärmar med stora ljusomfång.

LIST OF PAPERS

This thesis is based on the following papers, referred to in the text by their Roman numerals.

I.   Båth M, Sund P, Ungsten L, Månsson LG
      Calibration of diagnostic monitors: Theoretical determination of optimal luminance settings
      J Soc Inf Disp. 2006;14(10):905–911

II.   Sund P, Båth M, Ungsten L, Månsson LG
      Generation of low-contrast sinusoidal test patterns on a high-brightness display
      J Soc Inf Disp. 2006;14(10):913–919

III. Sund P, Båth M, Månsson LG
      Investigation of the effect of ambient lightning on contrast sensitivity using a novel method for conducting visual research on LCDs
      Radiat Prot Dosimetry. 2010;139(1-3):62–70

IV.   Sund P, Månsson LG, Båth M
      Development and evaluation of a method of calibrating medical displays based on fixed adaptation
      Med Phys. 2015;42(4):2018-2028

All publications are reprinted by permission of the copyright holders.

RELATED PRESENTATIONS

Sund P, Båth M, Ungsten L, Månsson LG
A Comparison between 8-bit and 10-bit luminance resolution when generating low-contrast sinusoidal test pattern on an LCD.
SPIE Medical Imaging, San Diego, USA, February 2007

Sund P, Båth M, Månsson LG
Detection of low contrast test patterns on an LCD with different luminance and illuminance settings.
SPIE Medical Imaging, San Diego, USA, February 2008

Sund P, Båth M, Månsson LG
Investigation of the effect of ambient lightning on contrast sensitivity using a novel method for conducting visual research on LCDs
Optimisation in X-ray and Molecular Imaging, Malmö, Sweden, June 2009

Sund P, Månsson LG, Båth M
The effect of fixed eye adaptation when using displays with a high luminance range.
SPIE Medical Imaging, San Diego, USA, February 2012

Sund P, Månsson LG, Båth M
The effect of fixed adaptation on the calibration of medical displays.
Optimisation in X-ray and Molecular Imaging 2015, Gothenburg, Sweden, May 2015

TABLE OF CONTENTS

ABBREVIATIONS

2AFC	Two alternative forced choice
AUC	Area under the curve
CI	Confidence interval
CIE	International commission on illumination (commission internationale de l'eclairage)
CRT	Cathode ray tube
CSF	Contrast sensitivity function
CT	Computed tomography
d'	Detectability index
DDL	Digital driving level
DICOM	Digital imaging and communications in medicine
FOM	Figure of merit
GSDF	Grayscale standard display function
$GSDF_{FAC}$	Fixed adaptation compensated GSDF
HVS	Human visual system
JND	Just noticeable difference
LCD	Liquid crystal display
lsb	Least significant bit
LED	Light emitting diode
LUT	Look-up-table
MAFC	Multiple alternative forced choice
msb	Most significant bit
MRI	Magnetic resonance imaging
PC	Percent correct
p-value	Presentation value
ROC	Receiver operating characteristics
SDK	Software development kit
VGC	Visual grading characteristics

DEFINITIONS IN SHORT

Brightness   Apparent luminance

DICOM standard target   A 2° by 2° 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.

Illuminance   Light incident on a surface [lx]

Lambertian   Diffuse surface with a cosine-shaped angular light distribution

Lightness   Apparent reflectance

Luminance   Light emitted from a surface in a specified direction [cd/m^2]

Photometry   Measurement methods for the properties of visible light

Psychophysics   Relationships between physical stimuli and perceived sensations

Radiometry   Measurement methods for the properties of radiant energy

Standard colorimetric observer CIE description of human spectral sensitivity to visible light

1 GENERAL INTRODUCTION

X-ray imaging technology has been available for more than a century. Historically, the most common way of displaying images has been by using analogue film on light-boxes. Later inventions, such as computed tomography (CT) and magnetic resonance imaging (MRI) were natively digital, and although the possibility of reviewing digital images on computer displays existed, the display quality was usually too low and images were therefore printed on transparent film and viewed on light boxes together with traditional X-ray images. In the mid 1990´s, the technology to replace the traditional analogue equipment became available. Image intensifying screens and films were replaced by storage phosphor plates and digital detectors using cesium iodide or selenium. Film archives were replaced by local networks and computer hard drives. The old light-boxes were no longer needed and computer displays took their place.

Light-boxes definitely had their drawbacks such as mechanical problems and broken light sources. Film sheets were lost in them and mounting/removing images required a lot of manual work. Although the displayed image quality was far from ideal, it was fairly consistent. The characteristics of the X-ray film varied very little and film processing was monitored carefully. Bright light-boxes lit up the viewing room, thereby reducing the fluctuations in perceived contrast due to variations in room illumination. Computer displays are much more heterogeneous as a group. There are large variations in luminance range, reflective properties and grayscale transformation, causing large variations in image rendering. Also, since image processing and display properties both determine the appearance of an image, different display properties require different image processing in order to achieve equal appearance. If a radiographer applies image processing to an image displayed at the imaging modality workstation and sends it to a review workstation where it is viewed on another display, the image will not be rendered as intended if the two displays are not calibrated in the same way. One and the same digital image may be perceived differently depending on factors like, for example, luminance range, display calibration and room lighting. Radiologists seldom have time to adjust the image processing for every image, and even if they do, the outcome will be dependent on the radiologist. Many medical images are compared to images taken at previous occasions in order to detect physiological differences in the patient, and it is important that differences in displayed images are not caused by differences in image processing and/or display properties. An incorrectly calibrated display can also cause details to be displayed with such a low contrast that they are invisible, possibly resulting in an erroneous diagnosis. However, displays together with a digital workflow can significantly enhance image quality by using advanced image processing. Small details in an image can be enhanced by digital magnification. Details with little contrast can be enhanced by software alteration of the grayscale transformation. In order to fully benefit from these digital advantages, the problem with inconsistent rendering must be addressed.

The topic of this thesis is calibration of displays used for viewing medical monochrome images. Digital images consist of pixels, where the value of each pixel represents the recorded intensity of the imaged object. Depending on imaging technique, pixel values represent completely different physical properties. Pixel values in images obtained using imaging modalities like, for

example, MRI, CT, ultrasound or photographic camera can only be understood correctly in the context of the used imaging technology. Display calibration only includes the perception of pixel values and not the production of pixel values, but in order to illustrate what pixel values in a medical image represents, and also to explain some of the terminology used, an example is given below where pixel values are obtained using traditional X-ray imaging.

The relationship between properties of an X-rayed object and the perceived light intensity in the displayed X-ray image is schematically described in Figure 1. The first quadrant shows the attenuation of exposed homogenous objects. As the object thickness increases, the detector signal decreases exponentially. (This simplified explanation is fully valid only for narrow-beam geometry without scattered radiation and with monoenergetic photons.) A logarithmic transformation is then usually applied to the detector signal, as shown in quadrant II, and stored as pixel values. By using a logarithmic transformation, the pixel values will be linearly related to object thickness. The pixel values in quadrant II are reversed in order for thick objects to be represented by high pixel values. This is natural for ordinary X-ray imaging since high input values to computer displays represents white. For fluoroscopy, where dense objects are displayed as black, it would have been better to use non-reversed pixel values. Some systems use a linear transformation between detector signal and pixel value, although this requires a higher number of bits per pixel in order to reduce quantization effects.
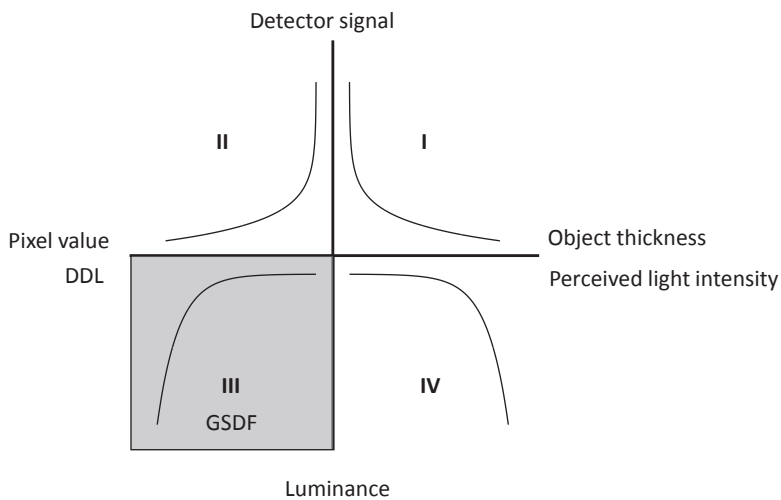
Figure 1: Schematic view of the relationship between properties of an X-rayed object and the perceived light intensity in the displayed X-ray image. Display calibration is represented by quadrant III. All four axes represent increasing values with increasing distance from the origin. DDL (digital driving level) is the input signal to a display device and GSDF is the grayscale standard display function.

Apart from any detector corrections, such as offset or pixel gain corrections, the pixel values represent unprocessed image data, at least according to the terminology used by the DICOM standard (Digital Imaging and Communications in Medicine).[1] The pixel values can then be altered by image processing, such as window/level adjustments or edge enhancement, either

by the imaging modality or by attributes (a.k.a. tags) stored together with the pixel data in the DICOM format. The DICOM standard contains attributes for various grayscale transformations, such as window/level adjustments and corrections for several pixel value encodings (relationships in quadrant II). Note that the implied assumption of linearity between object thickness and pixel value is a simplified model that is even less valid today than it was when using X-ray film technology. The traditional S-shaped film characteristic curve was non-linear and today modern image processing such as dynamic range compression has an even larger influence on image appearance. The processed pixel values intended for display are called p-values (presentation values), which are device independent values in a perceptually linear grayscale space intended for display on devices calibrated according to part 14 in the DICOM standard: grayscale standard display function (GSDF),[2] see section 3.4. The p-values are sent to the graphics board of the computer where they are converted to digital driving levels (DDLs) and used as input to a display device. This takes us to the lower part of Figure 1 and into the field of display calibration, the topic of this thesis.

The perceived light intensity is, more or less, logarithmically related to the displayed luminance (quadrant IV), and in order to obtain linearity between DDLs and perceived light intensity, an inverse to this relationship is needed in quadrant III. The relationship in quadrant III between DDL and luminance is determined by the display calibration (or lack thereof). A close match between the relationships in quadrant III and IV is necessary in order to display images consistently between different display devices without the need for adjustments of the grayscale transformation for every image. Unfortunately, the perceived light intensity as a function of luminance is complex and depends on numerous factors such as luminance level, luminance range and observed object.

The GSDF is probably the most used standard for calibration of displays in the field of medical imaging. This part of the DICOM standard defines the display grayscale transformation between the endpoints, which are the maximum and minimum luminance of a display. The endpoints determine the luminance range of the display, but are not defined by the standard. This means that all displays can be calibrated although their luminance outputs and ranges vary greatly. Obviously, images displayed on different devices will not appear exactly the same. This is intentional and the purpose of the standard is rather to distribute the available contrast evenly throughout the grayscale. In this way, a difference of a given number of pixel values will appear with the same contrast in all parts of the grayscale. However, a display with a large luminance range will still show this pixel difference with a higher contrast. Unfortunately, since the GSDF is defined only by the extreme luminance values of a display, while the relationship in quadrant IV is rather complex, a close match between the two functions is not always possible. As a result, the perceived contrast will in many cases not be evenly distributed throughout the grayscale.

The GSDF is based on several studies[3–16] compiled by Barten.[17, 18] Although their results are still very useful, most of these studies were performed in the 1960's and 1970's, so their primary objective was obviously not to improve the quality of digital radiography. New studies are needed that investigate how contrast is perceived on modern displays in an environment typical for a radiology reading room. Today, medical grade displays are very bright and capable of displaying images with a high luminance range, an effect that is not accounted for by the

GSDF.[19, 20] Images displayed on devices with vastly different luminance ranges will not be perceived as having similar contrast distributions, even though they are calibrated according to the GSDF. As a result, consistent image appearance across different display devices will be difficult to obtain. There is also a possibility that subtle details in an image will potentially be missed. Another aspect to consider is the lighting level in radiology reading rooms. Lighting conditions in radiology reading rooms can vary substantially and light originating from outside the display could possibly influence the contrast sensitivity of a radiologist. More studies are needed in order to obtain knowledge of how contrast perception is influenced by effects like display luminance range and ambient illumination. Contrast perception is preferably studied using observer studies where patterns with a contrast close to the detection threshold are displayed and evaluated on digital displays. By knowing how the contrast sensitivity varies with different luminance conditions, it is then possible to derive an alternative display calibration that enables images to be perceived with a more uniform contrast distribution on a wider range of display settings than the GSDF.

2 AIMS

The overall aim of the present thesis was to quantify changes in perceived contrast due to different display luminance settings and different illuminance conditions in order to derive a new calibration method that enables images to be perceived as having equally distributed contrast for a wide range of lighting conditions.

The specific aims of the separate studies were:

I.   to theoretically quantify changes in perceived contrast when using displays with various luminance ranges.
II. to develop a method to determine the contrast thresholds using sinusoidal test patterns on modern displays under clinically relevant conditions in observer studies.
III. to investigate the effects of ambient lighting on the perceived contrast.
IV. to develop a display calibration method that enables images to be perceived as having equally distributed contrast for a wide range of lighting conditions.

3 THEORETICAL BACKGROUND

The science of display calibration requires knowledge of light – how it measured (section 3.1) and perceived (section 3.2) by the human visual system (HVS). As displays are part of larger digital display systems with computers and graphic boards (section 3.3), knowledge of technical details about signal transfer and contrast properties in all parts of the imaging chain is important in order to calibrate a display. The principle behind the GSDF (section 3.4) must also be understood together with the technical limitations imposed by the display system. The limitations of the validity of the GSDF (section 3.5) must also be known in order to understand the potential benefits of a new calibration. A new calibration method requires information about the response of the HVS to various luminance conditions, and this response is studied by using test patterns close to the detection threshold in observer studies (section 3.6). The new calibration method can be evaluated in subjective observer studies by using visual grading characteristics (VGC) (section 3.7).

3.1 Photometry and radiometry

While radiometry describes the physical properties of radiant energy (including light from infrared to ultraviolet), photometry describes the human perception of visible light. A description of the response of the HVS to wavelengths within the visible spectrum is given by the International Commission on Illumination, CIE (Commission Internationale de l'Eclairage).[21] Photometric quantities are obtained by applying the CIE standard colorimetric observer[21] (Figure 2) to the radiometric quantities.

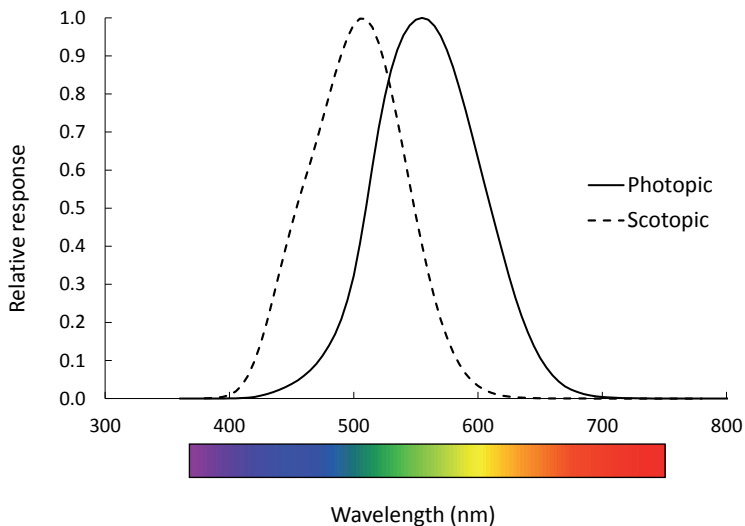*Figure 2: CIE standard colorimetric observers for photopic vision (light adapted eye, >3 cd/m^2) and scotopic vision (dark adapted eye, <0.01 cd/m^2).*

The radiometric and photometric quantities with corresponding units are shown in Table 1. Energy-saving light bulbs are usually described by their luminous power in lumen, which is a measure of the total amount of emitted visible light per second. In the field of display calibration, the most commonly used quantities are illuminance (incident light on a surface) and luminance (light from an area emitted in a specified direction).

Table 1: Radiometric and photometric quantities and units. The candela (cd) is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency 540 x 10^{12} hertz and that has a radiant intensity in that direction of 1/683 watt (W) per steradian (sr). The unit for luminous power is lumen (lm).

Radiometry	Source power	Power per unit area
Source power	Radiant power [W]	Irradiance [W/m^2]
Power per unit solid angle	Radiant intensity [W/sr]	Radiance [W/(m^2×sr)]

Photometry	Source power	Power per unit area
Source power	Luminous power [lm]	Illuminance [lx=lm/m^2]
Power per unit solid angle	Luminous intensity [cd=lm/sr]	Luminance [cd/m^2=lm/(m^2×sr)]

Since a full solid angle is 4π steradians (sr), a light source with 4π lumen will emit 1 lumen of luminous power within a solid angle of 1 steradian. The amount of luminous power within the solid angle is constant, but the cross-sectional area defined by the solid angle increases with distance from the light source, causing a decrease in illuminance, see Figure 3. The unit for illuminance is lm/m^2, or lux (lx).
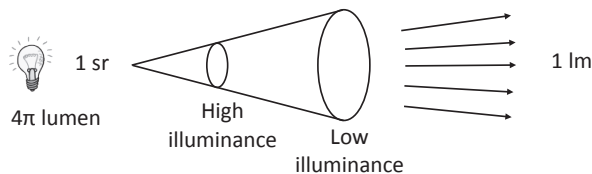
Figure 3: Illustration of illuminance.

A Lambertian surface (see section 3.3.3 for a more detailed description) with a luminous emittance of π lm/m^2 will emit 1 lm of luminous power per square meter into a solid angle of 1 steradian, see Figure 4. Thus, the luminance is 1 lm/(m^2×sr), or 1 cd/m^2. When the viewing distance is increased, the sampled area covered by the solid angle increases, and the larger area emits more light into the solid angle. Since luminance is the amount of light emitted within a solid angle per unit area, these two effects cancel each other out, making luminance independent of distance.

Figure 4: Illustration of luminance.

3.2 Perception of light

Although properties of the human visual system are included in the definition of luminance, it is still a well-defined quantity. The CIE standard colorimetric observer that is included in the definition of luminance describes the response of a typical human observer, and luminance meters can therefore be properly calibrated. Unfortunately, at least from a scientific viewpoint, the human visual system is far from an ideal luminance me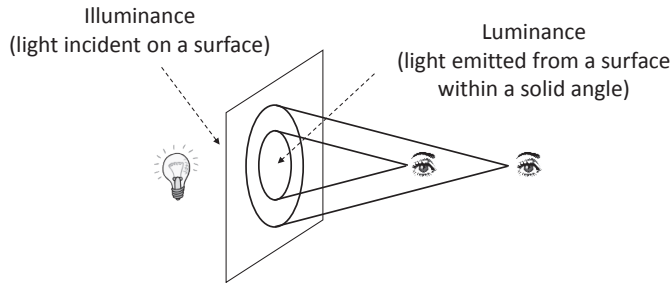ter. Our perception of light is complex and very difficult to describe by mathematical models. The human eye is capable of adapting itself to 14 decades of light intensity,[22] but once adapted to a certain level, the absolute luminance level has very little influence on our perception. The fact that the luminance difference between an indoor office environment and a sunny day outside can be several decades usually goes unnoticed once the eye has adapted itself to the new luminance level. Instead of being an accurate luminance meter, the human visual system is very capable of detecting low-contrast objects within a limited luminance range. The psychophysic quantities most commonly used to describe our perception of light are brightness and lightness, but unfortunately they are not strictly defined.

Brightness is defined by the CIE as "The attribute of a visual sensation according to which a given visual stimulus appears to be more or less intense, or according to which the area in which a visual stimulus is presented appears to emit more or less light".[23, 24] Historically, it has been used to describe both luminance and perceived luminance. The brightness of a source varies with the adaptation level, and the flame of a lit candle, for example, can be experienced as being bright in a dark room but almost invisible in sunshine. Lightness, on the other hand, is defined by the CIE as "The attribute of a visual sensation according to which the area occupied by the visual stimulus appears to emit more or less light in proportion to that emitted by a similarly illuminated area perceived as a 'white' stimulus".[23, 24] Lightness is a description of the property that makes a white piece of paper appear white and a black piece of paper appear black, regardless of luminance level. A black piece of paper exposed to sunlight is still perceived as black, and a white piece of paper in a dim indoor lighting is still perceived as white, even though the reflected luminance from the black piece of paper may be several magnitudes higher. Lightness is independent of luminance level. Other commonly used definitions of brightness and lightness are "apparent luminance" and "apparent reflectance", respectively.[25]

Many display calibration models, such as CIELab,[21] are based on the relationship between relative luminance and lightness and several descriptions of this relationship has been presented during the last century. Albert Munsell, creator of the Munsell color system[26] more than one hundred years ago, was indecisive whether to use a square root approximation or a logarithmic curve as proposed by the earlier works in psychophysics by Weber and Fechner.[27-29] The relationship between relative luminance and lightness (Eq. 1 and Figure 5) is presently, according to CIELab, described by a function close to a cube-root curve, consistent with the general power-law descriptions of psychophysics formulated by Stevens.[30] However, close to black the relationship is linear:[31]

$$L^* = 116 f\left(\frac{Y}{Y_n}\right) - 116, \tag{1}$$

where

$$f\left(\frac{Y}{Y_n}\right) = \begin{cases} \frac{841}{108}\frac{Y}{Y_n} + \frac{4}{29} & for \ \frac{Y}{Y_n} \leq \left(\frac{6}{29}\right)^3 \\ \left(\frac{Y}{Y_n}\right)^{1/3} & for \ \frac{Y}{Y_n} > \left(\frac{6}{29}\right)^3 \end{cases}, \tag{2}$$

where L^* is the lightness and Y/Y_n is the relative luminance.
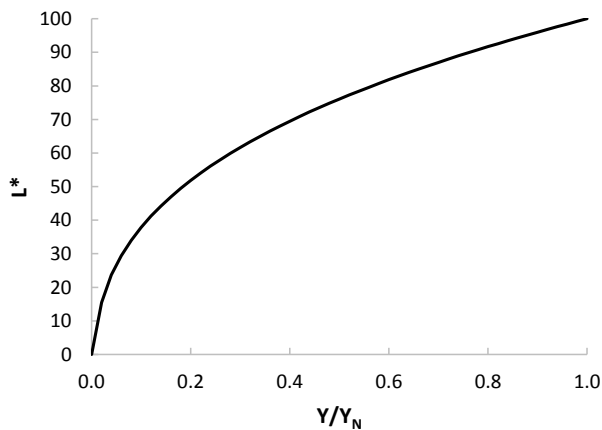
Figure 5: CIE relationship between relative luminance and lightness as described by Eq. 1 and Eq. 2.

Although the lightness can be described by Eq. 1, it is only valid for large homogenous surfaces. Lightness is also affected by the spatial luminance distribution. For example, two areas having the same luminance but placed in differently bright backgrounds will not be perceived as having the same lightness. Parameters that influence lightness include shadows, reflections and angle of illumination. Plenty of images describing lightness effects are available by searching the internet. Recommended keywords are, for example, *checker shadow illusion* and *Koffka ring illusion*. Also, a paper by Anderson describing how the lightness of an object is affected by interference with the background is highly recommended.[32]

3.3 Display systems

Displays are available in a large variety of sizes and resolutions. There are also numerous other properties than size and resolution that will influence the behavior of a display, some of which are explained below.

The spatial resolution is determined by the pixel density of the display, which in turn is determined by the physical size of the display and the number of pixels. Each pixel usually consists of three subpixels that, in a color display, control the amount of red, green and blue.

The luminance range is determined by the white point (maximum luminance) and the black point (minimum luminance). A liquid crystal display (LCD) uses some sort of backlight, such as fluorescent tubes or light emitting diodes (LEDs), to produce light and modulates the light output using liquid crystals.[33, 34] The power of the backlight largely sets the white point while the black point is limited by the light blocking properties of the liquid crystals. Also, the black point is very dependent on room illumination level and screen reflections.

The luminance resolution is determined by the luminance range together with the number of unique luminance levels that the display is capable of producing. Lower-end displays are usually limited to 256 (8 bits) luminance levels while higher-end displays rarely has less than 1024 (10 bits).[35, 36]

The mere fact that a display is capable of producing a large number of luminance levels does not mean it can display all of these simultaneously. The number of simultaneously displayed luminance levels is limited by the input side of the display. In most cases, this is limited to 256 DDLs (8 bits). In order to display 1024 gray levels (10 bits) at the same time, it is necessary to also have a display capable of 10-bit input together with 10-bit capable display cable, graphics board and image display software.

The relationship between the DDLs and the output luminance levels is, on higher-end displays, determined by a look-up-table (LUT). By adjusting this LUT it is possible to calibrate the display according to some standard, like the GSDF. This is also where the high number of possible luminance levels becomes beneficial; for every DDL, the available output luminance is closer to the ideal luminance, see section 3.4.3 "Quantization effects".

3.3.1 LUT

A LUT is basically a simple table where every input value has a corresponding output value. Technically, it can be described as an array of output values where the input values are determined by the positions in the array. The size of a LUT is important as it affects the accuracy of a display calibration. By using a LUT of insufficient size, the number of grayscale levels may even be reduced. It is possible, and quite common, for the input and output side to be of different sizes. On the input side, the size is determined by the number of positions in the array where every position contains a value within a given range. This range determines the output size of the LUT. Since computers use binary numbers (base 2) instead of decimal numbers (base 10), a range of numbers is usually expressed by the number of bits where each

bit can take the value 0 or 1. For example, a LUT with a size of 8 bits on the input side and 10 bits on the output side will be an array with 256 (2^8) positions where each value is in the range from 0 to 1023 (2^{10} values). See Figure 6 for an example.

Position 0										
Position 1	0	0	0	0	0	0	0	1	0	0
Position 2										
…					…					
Position 255										

Figure 6: Illustration of a LUT with a size of 8 bits on the input side and 10 bits on the output side. The binary value in position 1 corresponds to the decimal value 4.

LUTs only contain integer values but are typically used to describe complicated functions. The ideal output value may be a fractionated value but has to be truncated to the nearest integer value. In Figure 7, a LUT with 8-bit size on both the input and output side is used to represent a function with a slope of 0.5. Since fractions are truncated, the 256 unique input values have been reduced to 128 and a large part of the original information has been lost. In this case, a LUT with 9 or more bits on the output side would have increased the resolution and kept all 256 output values unique.

8 bit in ; 8 bit out ; Slope 1.0   8 bit in ; 8 bit out ; Slope 0.5

0	0
1	0
2	1
3	1
4	2
5	2
…	…
255	127

Figure 7: Example of truncation errors in LUTs. The two LUTs represent linear functions with different slopes. The output values are expressed in decimal notation.

In a display system, LUTs are used in order to modify the contrast properties of displayed images and can be found in different parts of the imaging chain. High-end medical displays have internal LUTs used for display calibration, but LUTs are also found in most computer graphic boards and, sometimes, LUTs are present in imaging software. An illustration of the LUTs available in a display system is shown in Figure 8. Every pixel in an image contains a pixel value that can be altered, by image processing and/or by LUTs. Pixel values and output values from the software LUT and Graphic board LUT are numerical values without any physical meaning. The output values from the display LUT, however, correlate to specified luminance outputs. The graphic board LUT on computers running Windows operating system has a size of 8 bits on the input side and 16 bits on the output side. The large output size enables smooth LUT functions with minimal truncation errors. It is also possible to use advanced display devices that utilize more than 8 bits as input.
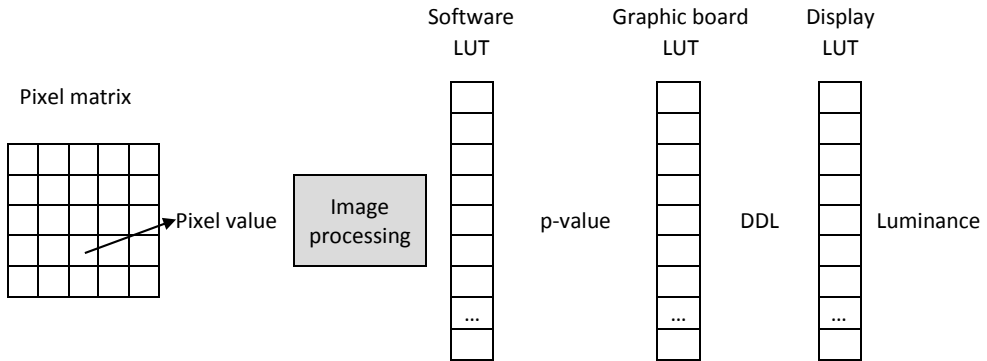
Figure 8: Illustration of the LUTs used in a display system. The original pixel value may be changed by image processing but also by any of the LUTs available in the display system.

The output value of one LUT is used as input value to the next LUT. If the output size of a LUT is different from the input size of the next LUT, the LUT output values have to be scaled accordingly, see Figure 9. Downscaling is performed by removing the surplus number of least significant bits. Upscaling is performed by adding the missing number of bits in the least significant position.
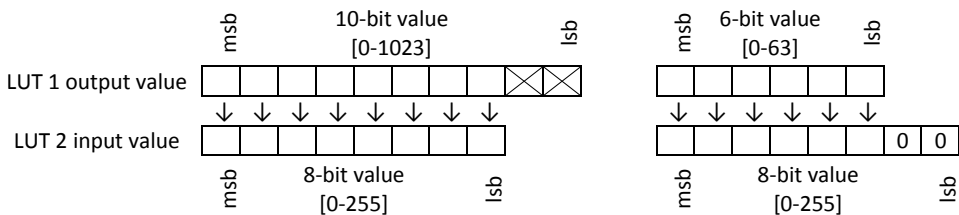

Figure 9: Downscaling of 10-bit binary numbers to 8 bits (left) and upscaling of 6-bit binary numbers to 8 bits (right). msb is the most significant bit and lsb is the least significant bit in respective number.

Color graphic boards have one LUT for each color channel (red, green and blue). Color displays with internal LUTs usually also have one LUT for each color channel, although it is possible that only one LUT exists that applies to all color channels simultaneously.

3.3.2 Displayed image luminance resolution

By increasing the luminance resolution, more grayscale and color levels can be displayed simultaneously. A display with 8-bit DDLs can display 256 grayscale levels and more than 16 million colors, and for a display with 10-bit DDLs these numbers increase to 1024 grayscale levels and approximately 10^9 colors. For a display system to be able to show 10-bit images with full luminance resolution, a number of factors must be fulfilled. The image itself must contain the desired bit depth and 10-bit support must also be available through the whole imaging chain, including imaging software, operating system, graphics board, display cable and the display itself. Images displayed by standard methods will be limited to 8 bits since standard graphic boards only accept 8-bit input values. In order to display 10-bit images with

full luminance resolution, the graphic board must be bypassed by using a direct display mode, such as DirectX[37] or OpenGL,[38] with 10-bit support.

3.3.3 Surface angular luminance distribution

A Lambertian surface is a diffuse surface that emits light with a cosine-shaped angular distribution (Figure 10).[39] Most light is emitted perpendicular to the surface (0°) and the light emission decreases with increasing angle until it reaches zero at a direction parallel to the surface (90°). Many light-emitting surfaces like cathode ray tubes (CRTs) or diffuse projection screens are close to Lambertian surfaces, while light emitted from an LCD tends to decrease more rapidly with increasing angle. When luminance is measured on a Lambertian surface, the measured value is independent of the angle of measurement. As the angle of measurement increases, the increase in surface area covered by a solid angle will exactly compensate the decrease in light emission in that angle, and the two effects cancel each other out.

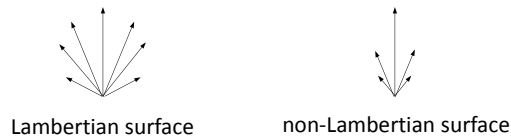Lambertian surface   non-Lambertian surface

Figure 10: Angular luminance distribution from a Lambertian surface (left) and from a surface with more forward directed light, such as an LCD (right)

3.4 DICOM PART 14 (GSDF)

The GSDF is a standard that addresses visual consistency in image display.[2] It was included as a separate part of the DICOM standard in 1996 and is now used in many radiology departments worldwide.[40] The primary objective of the GSDF is to define a single mathematical function that relates display input values to luminance output levels over a wide luminance range. The secondary objective is to provide some degree of similarity of grayscale rendition on display systems of different luminance range and also to use the number of available gray levels efficiently. The GSDF provides luminance levels that will be neither too close, where two neighboring values might be indistinguishable, nor too far apart, which could cause contour lines in the image.

A common interpretation of the GSDF is that it will distribute the contrast uniformly throughout the luminance range. However, although GSDF is defined for a very large luminance range, calibrated displays with large differences in luminance range will not have similar contrast distributions. The GSDF is best suited for displays with a low luminance range. For increasing luminance ranges, the secondary objective becomes increasingly less valid.

3.4.1 Background

The GSDF is based on studies where the smallest changes in contrast detectable by the human visual system were studied.[17, 18] The contrast thresholds at different luminance levels were determined by using patterns with contrasts close to the detection thresholds in observer studies. In Figure 11, results similar to those obtained in the observer studies are shown, although the exact values in the left part of the figure are calculated from the GSDF (Eq. 3 and Eq. 4). The contrast required in order for the DICOM standard target[*] to be visible varies with luminance. For example, at 1 cd/m^2, a contrast of 2.5% is necessary, while a contrast of 0.7% is sufficient at 1000 cd/m^2. One such increase in luminance is called a just noticeable difference (JND) and each luminance level (separated by one JND) is called a JND index, e.g. if one JND index is at 1.0 cd/m^2, the next JND index will be at 1.025 cd/m^2, and if one JND index is at 1000 cd/m^2, the next JND index will be at 1007 cd/m^2.
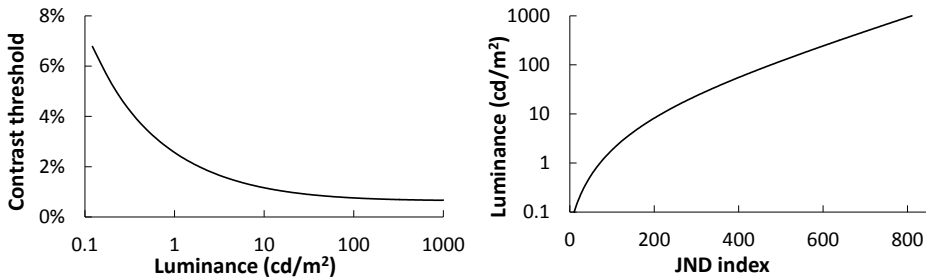

Figure 11: The contrast threshold as a function of luminance (left) and luminance as a function of JND index (right). The distance in luminance between two JND indexes correspond to the contrast threshold at that luminance.

The GSDF describes the relationship between JND index and luminance. By changing the JND index by one, the luminance will change just enough to make a noticeable difference to an observer.

The mathematical relationship between JND index (j) and luminance (L) is:[2]

$$\log_{10} L(j) = \frac{a + c*\ln(j) + e*(\ln(j))^2 + g*(\ln(j))^3 + m*(\ln(j))^4}{1 + b*\ln(j) + d*(\ln(j))^2 + f*(\ln(j))^3 + h*(\ln(j))^4 + k*(\ln(j))^5}, \tag{3}$$

where

a = -1.3011877   b = -2.5840191*10^{-2} c = 8.0242636*10^{-2}
d = -1.0320229*10^{-1} e = 1.3646699*10^{-1} f = 2.8745620*10^{-2}
g = -2.5468404*10^{-2} h = -3.1978977*10^{-3} k = 1.2992634*10^{-4}
m = 1.3635334*10^{-3}.

The inverse relationship is described by the function[2]

---

[*] The DICOM standard target is a 2° by 2° 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.

$$j(L) = A + B * \log_{10}(L) + C * (\log_{10}(L))^2 + D * \left(\log_{10}(L)\right)^3 + E * \left(\log_{10}(L)\right)^4 + F *$$
$$\left(\log_{10}(L)\right)^5 + G * \left(\log_{10}(L)\right)^6 + H * \left(\log_{10}(L)\right)^7 + I * \left(\log_{10}(L)\right)^8, \quad (4)$$

where

A = 71.498068	B = 94.593053	C = 41.912053
D = 9.8247004	E = 0.28175407	F = -1.1878455
G = -0.18014349	H = 0.14710899	I = - 0.017046845.

The luminance range covered by the GSDF spans from 0.05 to 4000 cd/m^2 but the standard does not specify any requirements on what luminance range should be used for a particular display. When calibrating a display, only the part of the GSDF that falls within the luminance range of the display is utilized. An example is shown in Figure 12, where the luminance ranges for two displays determine the corresponding ranges of JND index values. The available ranges of JND index values are then distributed linearly over the DDLs. Equal increments in DDL will then correspond to equal increments in JND index value, which corresponds to equal increments in perceived contrast. As mentioned above and also discussed further in section 3.5, the JND is a quantity that is derived under simplified conditions where e.g. luminance range is not considered – although the GSDF can be applied to any luminance range within the defined luminance limits.

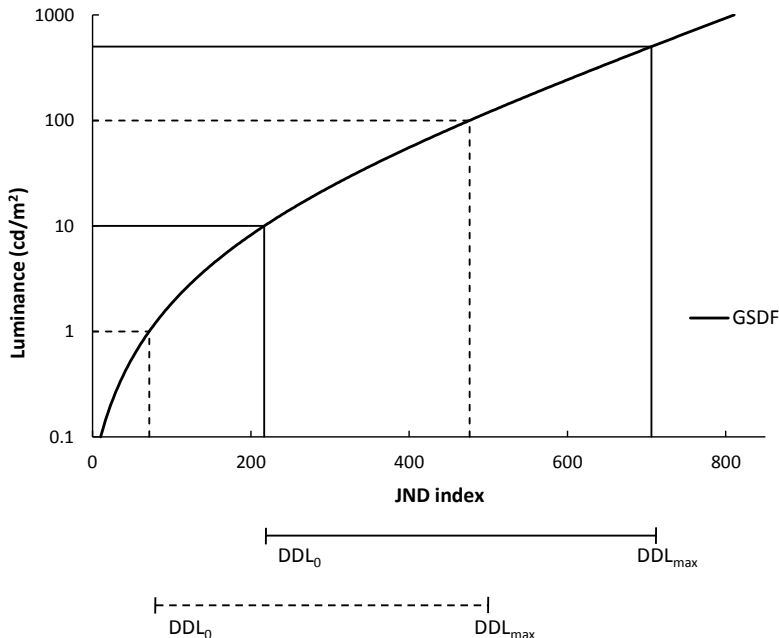*Figure 12: Range of JND index values covered by two displays with different luminance ranges. One display covers the luminance range from 1 cd/m^2 to 100 cd/m^2 (dashed lines) while the other display covers the luminance range from 10 cd/m^2 to 500 cd/m^2 (solid lines). The range of JND indexes covered by the luminance range of a display are evenly distributed over all DDLs.*

3.4.2 Display calibration according to the GSDF

For a display to be calibrated according to the GSDF, the calibration software needs to display the specified test pattern (see Figure 13) for a number of DDLs and record the corresponding luminance output for each DDL using a luminance meter. The minimum luminance is then achieved for DDL_0 and the maximum luminance is achieved for DDL_{max}. All measured luminance values must include reflections from the display surface, since this is the luminance perceived by an observer. The minimum luminance determines the minimum JND index value used (Eq. 4), which is associated with DDL_0, see Figure 12. The same method is then applied for the maximum luminance so that a JND index value is associated with DDL_{max}. The intermediate DDLs are associated with JND index values in such a way that every increment in DDL between DDL_0 and DDL_{max} results in equal changes in JND index value. Every DDL then corresponds to a JND index, which in turn corresponds to the ideal luminance as specified by the GSDF (Eq. 3), but unless the inherent characteristic curve of the display equals the GSDF, the actual measured luminance will be different from the ideal luminance. A display calibration is therefore necessary, in which the output luminance for every DDL is corrected.
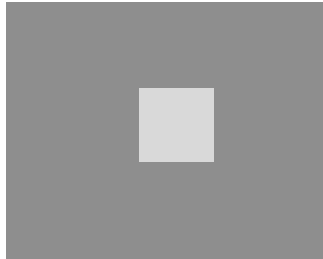

Figure 13: GSDF calibration pattern[2]. The luminance is measured in the small square for all of the DDLs. The size of the small square covers 10% of the display surface and the luminance surrounding the square is 20% of the maximum luminance.

Three things are needed in order to calibrate a display.

1. A luminance meter.
2. A software that compares the measured values to the ideal luminance values specified by the standard and calculates a calibration LUT. The software should also be able to implement the calibration LUT somewhere in the display system.
3. An accessible LUT in the display system that affects displayed images.

3.4.2.1 Luminance meters

Since the GSDF only utilizes luminance values for calibration, the luminance meter only has to be able to measure luminance – any color information can be disregarded. However, it is still important that the luminance meter measures luminance from colored surfaces correctly and compliance to the CIE standard colorimetric observer is important.

Luminance meters can be divided into two groups: telescopic and near-range.[41, 42] Telescopic luminance meters operate at a distance from the light source and for a display this means that all reflections are included in the measurements. It is also possible to use this type of

luminance meter when measuring the luminance from reflecting surfaces, such as projection screens. Near-range luminance meters are attached to the display surface, thereby effectively blocking all incoming light and excluding reflections from the measurements.

Another important aspect of a luminance meter is its opening aperture, i.e. the angle from which light is collected and measured. A narrow aperture angle will measure light in a specified direction while measurements with a large aperture angle will be an average over a wide field-of-view.[43] The size of the opening aperture is important when measuring since light emitted from the calibration surface and the measured surface may have different angular luminance distributions.
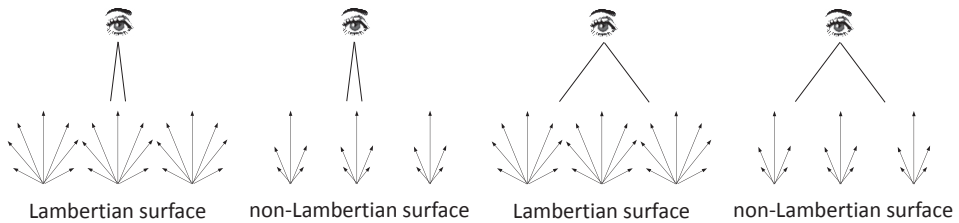

Lambertian surface   non-Lambertian surface   Lambertian surface   non-Lambertian surface

Figure 14: Luminance measurement geometries with different luminance meter opening apertures and different surface angular luminance distributions.

When a luminance meter with a large opening aperture is calibrated, not only light perpendicular to the display surface is included but also light emerging from the display at other angles, see Figure 14. Assume that the luminance meter is calibrated for a certain angular luminance distribution. If the luminance meter is used at another surface with the same perpendicular luminance but with a different angular luminance distribution, the contribution from light at other angles will be higher or lower compared to the calibration conditions and the measured luminance will therefore be over- or underestimated. A luminance meter with a narrow opening aperture measures light in a specified direction and measurements perpendicular to the two surfaces described above would yield the same value. Measurements at other angles would result in different values, which is correct since the luminance at these angles differs between the two surfaces. So, when using a luminance meter, it is important that the opening aperture is known together with the angular luminance distributions for the measured surface and the calibration surface.

3.4.2.2 Calibration software

The calibration software displays test patterns, as specified by the GSDF, for a number of DDLs and retrieves the corresponding measured values from the luminance meter. If a near-range luminance meter is used, the amount of reflected light must be estimated and entered into the calibration software in order to be added to the measured values. The obtained relationship between DDLs and luminance is called characteristic curve and should be measured with every LUT set to unity, i.e. the LUT contains a linear distribution of the output values, from 0 to the maximum output value. In this way, the maximum number of luminance levels is achieved.
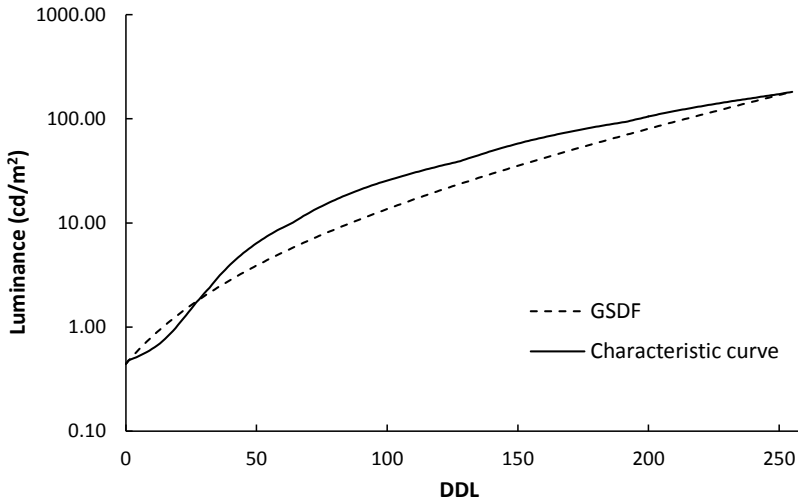
Figure 15: An example of a measured characteristic curve for an LCD together with the ideal luminance values according to the GSDF.

For every DDL, the ideal luminance according to the GSDF can be determined by using the luminance values for DDL_0 and DDL_{max} and the method described in section 3.4.2. However, unless the inherent characteristic curve of the display equals the GSDF, the actual measured luminance will be different from the ideal luminance. See Figure 15 for an example of a measured characteristic curve. By measuring the characteristic curve, the display LUT output value and corresponding luminance is known for every measured DDL. The display LUT output value that best corresponds to the ideal luminance is obtained for all DDLs and stored in a calibration LUT. If the calibration software only measured the luminance for some of the DDLs or if the display LUT output size is larger than the input size, it is possible to obtain a LUT output value that better represents the ideal luminance by interpolation. See Table 2 for an example of the calibration procedure for six DDLs. When the display LUT output values are replaced by the calibration LUT output values, every DDL will correspond to a display LUT output value that represents a luminance as close as possible to the ideal luminance. The calibration software now has to load the calibration LUT values into the hardware LUT in the display for it to take effect.

Note that it is also possible to use the graphic board LUT, or a software LUT, instead of the display LUT in the same way as described above. The used LUT output value will then correspond to the DDL that best represents the ideal luminance, according to the measured characteristic curve. The output size of the used LUT will inevitably be scaled to the input size of the display LUT, usually 8 bits.

Table 2: Procedure for deriving the calibration LUT. The example data in this table are taken from the measurement shown in Figure 15. Note that the display LUT in this table is positioned after the DDL, i.e. inside the display, but the same procedure applies when the LUT is positioned before the display, for example in the graphics board.

DDL	10-bit display LUT output value	Measured luminance		Ideal luminance according to GSDF	Calibration LUT output value
29	116	1.97		1.91	?
30	120	2.11		1.98	116
31	124	2.25		2.06	119
32	128	2.39		2.14	121
33	132	2.57		2.22	123
34	136	2.75		2.30	125
35	140	2.95		2.38	128

3.4.2.3 Display system LUTs

In a display system, one or more hardware LUTs are available and can be used to store the calibration LUT values.[44, 45] The best option is to use a LUT inside the display device. In this case, the display device will stay calibrated even if the computer attached to it were to be replaced,[45] provided the LUT on both computers graphic boards are set to unity.

Another option is to load the values from the calibration LUT into the computer graphics board, but care must then be taken to ensure that the values stay there. The LUT on the graphics board is not only accessible to the calibration software, but also to every other software that wants to use it and no warning is issued if it is overwritten. This LUT also returns to unity when the computer is turned off which requires a LUT-loading software to start automatically every time the computer starts. Ideally, this software also regularly verifies that the calibration values have not been overwritten. Many graphic boards support more than one display to be used simultaneously and although it is possible that there exists one hardware LUT for each connected display it is also likely that there is only one hardware LUT that affects only one of the connected displays – or possibly all of the connected displays. If more than one display are to be calibrated, a separate LUT is required for each individual display.

It is also possible for the image displaying software to implement the calibration LUT, although this is not a common solution. In this case, the calibration would only be valid for this particular software.

3.4.3 Quantization effects

The display LUT output size usually corresponds to the number of unique luminance levels that the display is able to produce, although it is possible that the display LUT output size is excessively large and that only some of the most significant bits actually affect the displayed luminance. Figure 16 exemplifies the available luminance levels for a display with 8-bit luminance resolution (solid lines) and a display with 10-bit luminance resolution (solid and dashed lines). In the 8-bit case there are 256 available luminance levels distributed over the entire luminance range while there are four times as many for the 10-bit case. The ideal GSDF

luminance value is usually not available in the display and has to be rounded to the nearest available luminance level. In the 8-bit case, DDL 31 and DDL 32 are mapped to the same luminance level (2.11 cd/m^2) and DDL 33 and DDL 34 are both mapped to the next available luminance level (2.25 cd/m^2). In the 10-bit case, every DDL is mapped to a unique luminance level, but the ideal luminance is still unavailable. With a higher luminance resolution, a smoother calibration is possible to obtain whereas a too low luminance resolution will result in a loss of grayscale levels due to quantization effects when neighboring DDLs are mapped to the same luminance level.
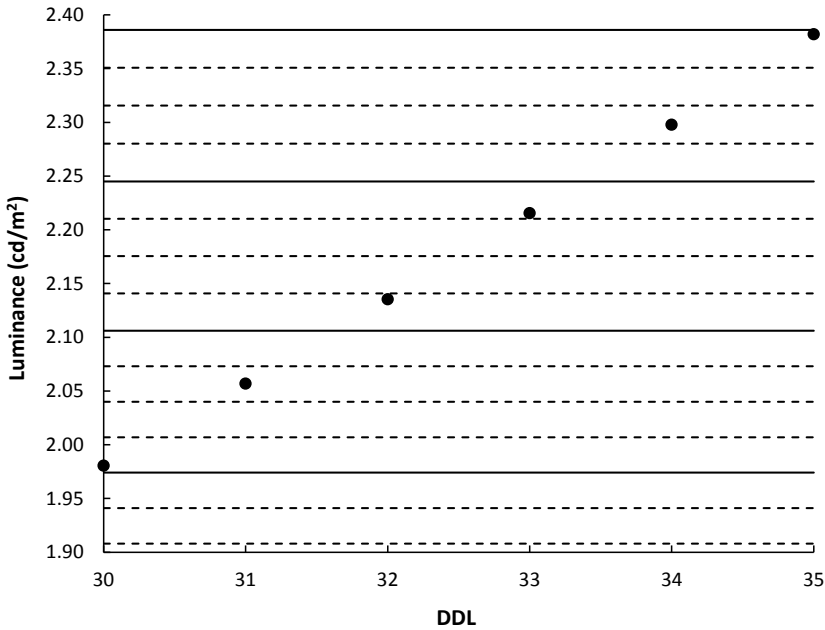

Figure 16: Available luminance levels for an 8-bit display (solid lines) and a 10-bit display (solid and dashed lines). The dots indicate the ideal luminance according to the GSDF. The example data in this figure are taken from the measurement shown in Figure 15 and the luminance values are also found in Table 2.

Even if the graphic board LUT with an output size of 16 bits is used for calibration, the number of simultaneously displayed gray levels will still be limited by the number of DDLs of the display. A display limited to 8-bit DDLs will only use the 8 most significant bits from the graphic board output.

An example of changes in luminance and corresponding changes in JND index values between neighboring DDLs for two GSDF calibrated displays with 8-bit and 10-bit luminance resolutions is shown in Figure 17. Ideally, the change in JND index value for each DDL would be constant, resulting in a horizontal straight line. The changes in luminance would be a smooth, gradually increasing curve. Due to quantization effects, the ideal luminance is not available and unwanted variations are introduced between neighboring DDLs. The lower the luminance

resolution, the higher the variation. In the 8-bit case, only 221 unique luminance levels were still present in this display after calibration, while all 256 luminance levels were maintained in the 10-bit case. The loss of luminance levels and the unwanted variations are caused by the fact that the characteristic curve differs from the ideal luminance values, as shown in Figure 15. Between DDLs 30 and 35 the slope of the characteristic curve is much steeper than that of the GSDF, which means that the available luminance levels are further apart than desired. This is also illustrated in Table 2 and in Figure 16 where six GSDF luminance levels cover the same luminance range as four luminance levels in the 8-bit display.
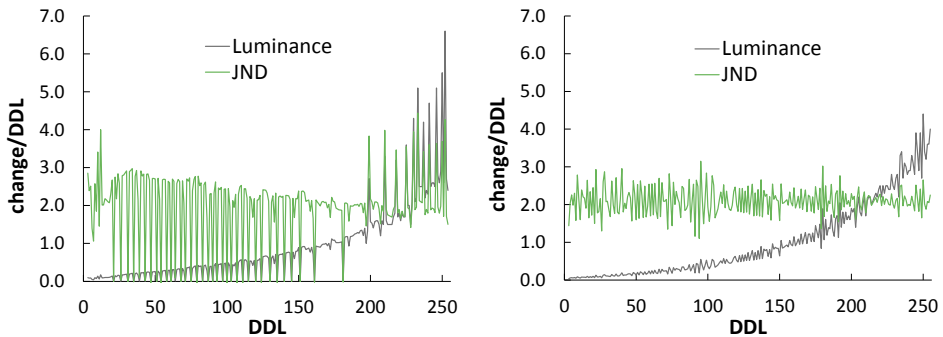

Figure 17: Changes in luminance and JND index values between neighboring DDLs for a display with 8-bit luminance resolution (left) and for a display with 10-bit luminance resolution (right).

3.4.4 Ambient light

A calibration is only valid as long as the LUTs are unaltered and the minimum and maximum luminance levels are not changed. The maximum luminance level decreases over time but this is a relatively slow process, although some displays may lose a considerable amount of luminance output during the first months of their lifetime. Keeping the minimum luminance level constant is a more difficult task and also more important as it affects the contrast properties of the display considerably. The minimum luminance level is determined by the luminance output of the display as well as reflections caused by the room illuminance. A calibration to GSDF specifications includes display reflections and the calibration is therefore only valid at the illuminance level used at calibration. Any changes in illuminance will also change the reflected light by the same amount for all luminance levels. Since an additive change in luminance is relatively higher for low luminance levels, this is also where the greatest changes in contrast properties occur.[46, 47]
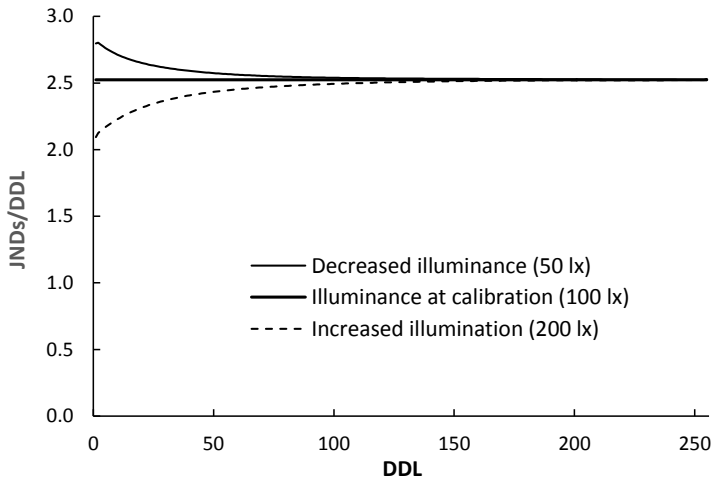
Figure 18: The number of JNDs for each DDL for three illuminance levels (50, 100 and 200 lx) where the display is calibrated at 100 lx. The luminance range of the display is 1 cd/m^2 to 400 cd/m^2 and the reflection coefficient is 0.005 cd/m^2 per lx.

Figure 18 is an illustration of a theoretical calculation, based on the GSDF (Eq. 3 and Eq. 4), of what will happen to the perceived contrast if the illuminance is increased or decreased by a factor of two from the calibration illuminance. When the illuminance is increased, the contrast in the darkest parts of the image will be reduced by approximately 17%. A reduction in illuminance will instead cause an increase in the perceived contrast. Although an increased contrast is not as serious, the principle of equally distributed contrast is violated. A factor of two may seem like a large change in illuminance but, unfortunately, larger changes are likely to occur when lights are turned on or off, or when lighting conditions outside changes the amount of light entering through windows.

Radiology reading rooms with digital displays have traditionally been quite dark. By keeping the illuminance at a low level, screen reflections are kept at a minimum and the total contrast of the display increases. When using CRTs, where screen reflections are high and the maximum luminance is low, a low illuminance is an important consideration. Another important reason is to maintain an acceptable contrast in the dark parts of an image. Uncalibrated displays, or displays calibrated with a near-range luminance meter where the reflected luminance has been omitted, are best viewed in complete darkness. If such a display is used in a brighter environment, contrast in the dark parts will be reduced as described above. However, reading rooms do not have to be dark if displays are calibrated correctly. An increased illumination makes it easier to perform other non-display related tasks, but it can also have other positive effects. Ideally, the illuminance should be kept at a constant level at all times, but since this requires reading rooms without windows, radiologists tend to reject such solutions and favor rooms with windows where the illumination level varies with weather conditions and time of day. By calibrating the displays at a relatively high minimum luminance,

contrast reductions due to rapid changes in illuminance can be reduced, although not eliminated.
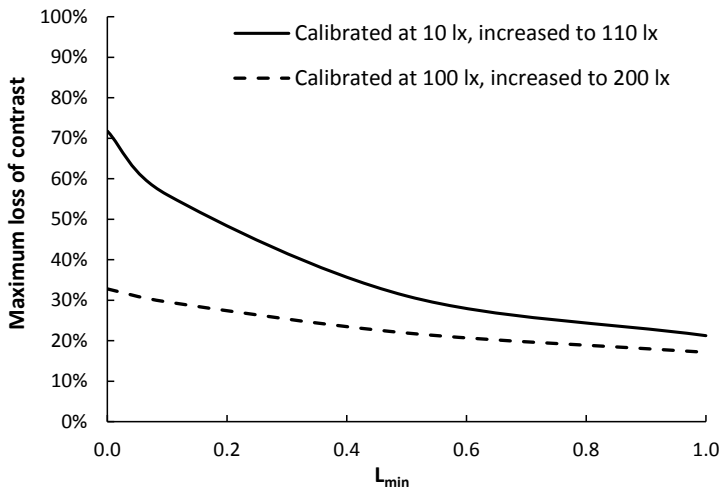

Figure 19: Maximum loss of contrast at different calibration conditions when the illuminance is increased by 100 lx compared to the calibration illuminance. L_{max} is 600 cd/m² and L_{min} is as indicated in the figure. L_{min} and L_{max} do not include reflections. The calculations are based on a reflection coefficient of 0.005 cd/m² per lx.

The data in Figure 19 are calculated in the same way as for Figure 18 and illustrates the maximum loss of contrast for a display calibrated at a certain illuminance but used at a 100 lx higher illuminance. When a display is calibrated in a very dark room (10 lx) and with a low L_{min} of 0.1 cd/m², a 100 lx increase will cause a contrast reduction of up to 56% in the darkest parts of a displayed image. When the display is calibrated at 100 lx, the corresponding reduction is 30%. However, by calibrating the display at a higher L_{min}, the display will be less sensitive to variations in illuminance. By setting L_{min} to 1 cd/m² in the example above, the reduction in contrast will be approximately 20%, regardless of illuminance at calibration. The disadvantage of calibrating a display at a high illuminance and a high L_{min} is a general reduction in display total contrast. A display calibrated at 10 lx with L_{min} set to 0.1 cd/m² will have an average contrast of 2.8 JNDs/DDL. For a display calibrated at 100 lx with L_{min} set to 1 cd/m², the average contrast will be 2.5 JNDs/DDL. One thing to note is that the reflection coefficient on modern LCDs is relatively low. If it could be eliminated completely, the illuminance level would no longer be an issue as the displayed image would not be affected.

3.4.5 Verification of a calibration

One way of verifying a calibration is to use the calibration software once again to make sure that the luminance response agrees with the luminance response specified by the GSDF. However, the calibration software and the software used to display medical images may use

different techniques. Also, if a near-range luminance meter is used, any changes in room illumination will not be detected.
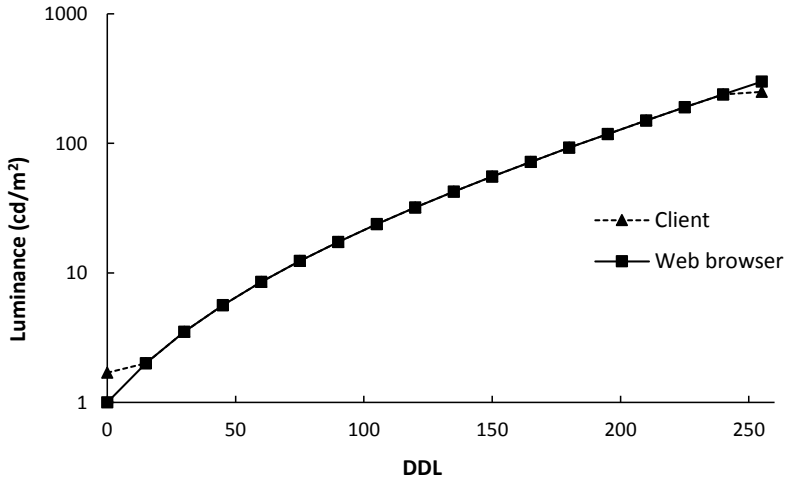

Figure 20: Luminance responses for two different imaging softwares using the same computer and the same display. The contrast in the darkest and brightest regions is clearly worse in the stand-alone client interface than in the web browser.

An example of the luminance responses for two different imaging softwares from the same manufacturer using the same computer and the same display are shown in Figure 20. One of the softwares was a stand-alone client while the other was a web browser interface. The display system was perfectly calibrated according to the calibration software, but images displayed in the stand-alone client had considerably lower contrast in the darkest and brightest parts. The characteristic curves were determined by displaying patterns similar to the GSDF calibration pattern (TG18-LN, see Figure 21) at 18 DDLs in respective software and measuring the luminance output with a telescopic luminance meter. The difference in luminance output between the two softwares was probably caused by alternative ways of dealing with colors reserved by the operating system. This example emphasizes the fact that test images should be displayed as similarly as possible to medical images. By using test images like TG18-QC, any grayscale abnormalities in image display can easily be found.
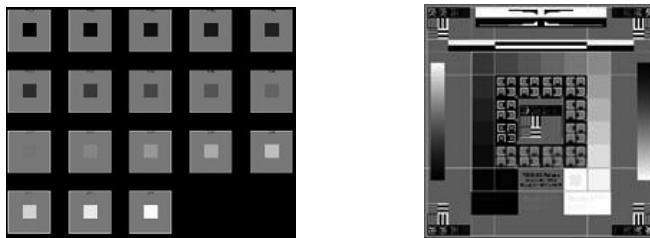

Figure 21: Some of the test patterns available from AAPM TG18.[48] *Left: TG18-LN is a series of 18 images from black to white. Right: TG18-QC is a multi-purpose test pattern intended for quality control of displays.*

3.4.6 Calibration standards and guidelines

There are a number of standards and guidelines dealing with display calibration, some of which are listed below. The AAPM TG18 guidelines are based on DICOM Part 14, but also include a comprehensive tutorial on display systems as well as test patterns intended for display quality control. Most of the other standards and guidelines are referring to DICOM part 14 or AAPM TG18, sometimes with modifications or additional requirements.

DICOM part 14: Grayscale Standard Display Function[2]
AAPM TG18[20]
DIN V 6868-57[49]
JESRA X-0093: Quality Assurance (QA) Guideline for Medical Imaging Display Systems[50]
IPEM Report 91[51]
Qualitätssicherungs- Richtlinie (QS-RL)[52]
European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis[53, 54]
ACR–AAPM–SIIM Practice parameter for determinants of image quality in digital mammography[55]
PAS 1054[56]
IAEA human health series No.17: Quality assurance programme for digital mammography[57]
Guidelines for Quality Assurance in Mammography Screening[58]
Guidelines for quality control testing for digital (CR DR) mammography[59]

3.5 GSDF limitations

Perceptual linearization[60–62] is a concept in which the perceived contrast is equally distributed at all luminance levels and a similar appearance can thereby be obtained on all types of displays. Perceptual linearization is only possible to obtain for well-defined circumstances and will depend on the observers used, studied patterns, etc. The GSDF uses perceptual linearization as a tool for meeting the secondary objective (i.e. to provide some degree of similarity of grayscale rendition on display systems of different luminance range and also to use the number of available gray levels efficiently) and uses the JND as a quantity of equally perceived contrast. A JND varies with the luminance level and is the contrast required to obtain a visible change by the *adapted* eye, i.e. the luminance difference to be detected is a small deviation from the adaptation level of the observer and thereby close to peak contrast sensitivity.[63] In a display calibrated according to the GSDF, the bright parts of the display are calibrated as if the observer were adapted to this high luminance level and the dark parts are calibrated as if the observer were adapted to a low luminance level. This is often referred to as variable adaptation.[20] For a display with a low luminance range, this is not a large problem since all details in an image will be displayed with a luminance relatively close to the adaptation level.
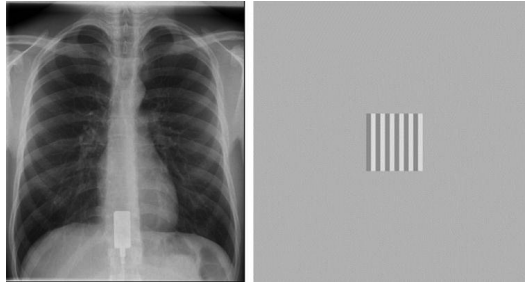
Figure 22: Left image – A chest x-ray demonstrates fixed adaptation where the observer is exposed to the entire luminance range of the display but is adapted to the average luminance in the image. Right image – The GSDF is derived from studies in which the observers viewed test patterns with a luminance close to the background luminance.

As pointed out in the AAPM TG18 report,[20] one of the limitations with the GSDF is that the JND unit is derived using variable adaptation while actual images are viewed with fixed adaptation, i.e. the adaptation level of the observer is determined by the average luminance in the image.[64] Figure 22 demonstrates the difference between fixed and variable adaptation. As the luminance range of the display increases, more details in the image will be at a luminance far from the adaptation level. Since the contrast sensitivity decreases when the difference between object luminance and adaptation level increases, the ability to detect low contrast details in bright and dark areas will be reduced as compared to details in the mid gray. As a consequence, images displayed on GSDF calibrated devices will only have a similar appearance when displayed with approximately the same luminance ratio.

A description of the human contrast sensitivity under the assumption of fixed adaptation has been presented by Barten.[63] His model is based on data by Rogers and Carel[65] where observers detected 2-dimensional waffle shaped gratings for a number of surround luminance levels. This correction, called the f-factor, was used in most of the studies included in this thesis. In Paper I, the f-factor was used to theoretically determine the optimal luminance settings of high luminance range displays, and in Papers II-IV it was used for comparison with the obtained results. The f-factor is given by

$$f = e^{-\dfrac{\ln^2\left(\frac{L_s}{L}(1+\frac{144}{X_o^2})^{0.25}\right) - \ln^2\left((1+\frac{144}{X_o^2})^{0.25}\right)}{2\ln^2(32)}}, \tag{5}$$

where f is the correction factor by which the contrast sensitivity has to be multiplied, L is the luminance of the object, L_s is the surround luminance, and X_o^2 is the object area in square degrees of visual angle.

An example of the f-factor for an object area covering a visual angle of 2 degrees and a luminance range of 1 to 1000 cd/m^2 is shown in Figure 23. The surround luminance (adaptation luminance) was calculated as the logarithmic average of the minimum and maximum luminances. As can be seen, the f-factor peaks at a luminance slightly higher than the surround luminance and decreases with increasing distance from the peak luminance.
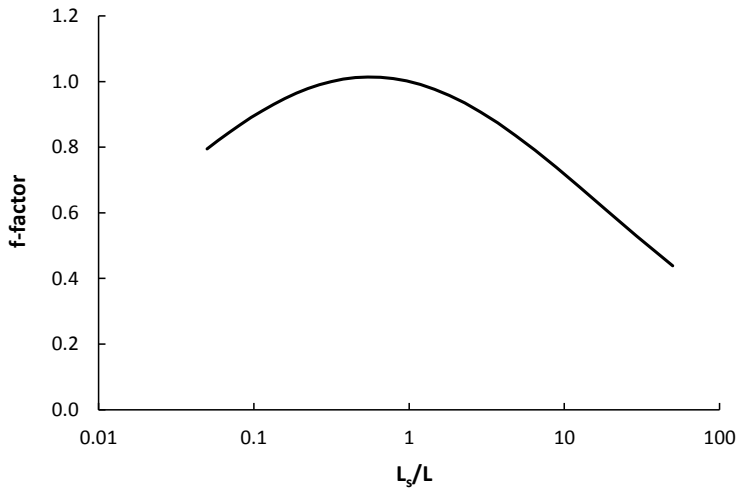
*Figure 23: The f-factor for a visual angle of 2 degrees. The luminance range spans from 1 to 1000 cd/m²
and the surround luminance level is calculated from the logarithmic average of the minimum and
maximum luminance.*

Another description of fixed adaptation has been presented by Flynn et al.[33] which is based
on Hecht and Hsia's[66] photochemical theory of photoreceptor response. This response is
approximately described by the expression[67] P=L/(L+S) where P is the photoreceptor
response, L is the retinal luminous intensity, and S is a constant that conforms with the state
of adaptation. Other studies[68, 69] have confirmed this expression by measuring the voltage
response from cones on the retina on turtles. Flynn et al. obtained the biologic contrast
response of the human visual system by differentiating the expression for photoreceptor
response. Since the expression for photoreceptor response only included the signal
properties, any influence of photoreceptor noise was not included in the contrast response.

3.6 Observer studies close to the detection threshold

In order to determine the contrast threshold, i.e. the contrast at which an observer is barely
capable of detecting a given pattern, one must first decide what kind of pattern to detect and
also what method to use.[70]

Psychometric studies of physical stimuli are performed by varying the magnitude of the stimuli
close to the limits where they become perceivable.[71] In the present work, the used stimuli
were low-contrast sinusoidal test patterns displayed on computer displays. If the contrast is
too low, the pattern will never be detected and if the contrast is too high it will always be
detected. Between the levels of never detected and always detected, detection is a matter of
probability. Patterns with a contrast in this transition range will be perceived some proportion
of all observations. A higher contrast corresponds to a greater probability for detection and
the relationship between contrast and detection probability depends on the characteristics of

the used patterns. This relationship is called the psychometric function and is typically a sigmoid function, such as the Weibull, logistic, cumulative Gaussian, or Gumbel distribution.[72] Some of the possible evaluation methods are described below.[73]

In the **method of adjustment**, the observer controls the contrast of the pattern and adjusts it until it becomes perceivable, as judged by the observer. Since the adjusted pattern ends up perceivable, the obtained contrast threshold is likely to correspond to a detection probability close to 100%. In the **method of limits**, the contrast of the pattern is alternately decreased until not perceivable and then increased until perceivable. The contrast threshold is determined by the average of the two limits. The **staircase method** usually starts at a contrast that is easily perceivable and the observer must then decide whether the pattern is present or not. A correct answer will decrease the contrast and an erroneous answer will increase the contrast until the contrast converges at a certain level. Several variants of this method exist where a consecutive number of correct/incorrect answers are required before the contrast changes, resulting in convergence at different contrast thresholds.[73] Also, more than the latest response can be utilized by using Bayesian and maximum-likelihood procedures. In the **method of constant stimuli**, the contrast of a pattern is chosen randomly and presented to the observer who must decide whether the pattern is present or not. This method allows for a full sampling of the psychometric function but requires a lot of observations relatively far from the detection threshold.

The method of adjustment and the method of limits are both based on observer preference, i.e. the obtained contrast thresholds are based on subjective opinions. The staircase method and the method of constant stimuli determine contrast thresholds objectively and both are well suited to use in alternative forced choice (AFC) experiments.[74] In a 2AFC study where only 2 alternatives exist, the possible responses are "pattern present" and "pattern not present". As an alternative procedure, the pattern may be presented at one of two different locations where the observer must decide which one contains the pattern. In a multiple AFC (MAFC; M ≥ 2) study, the observer must decide in which one of the M possible alternative locations the pattern is present. For a number of repeated observations, the percentage of correctly identified patterns (PC) is determined for a number of contrast levels close to the contrast threshold. Since the possibility of choosing the correct alternative is 1/M by chance alone, the relationship between pattern contrast and PC will depend on the evaluation method used, see Figure 24. The contrast threshold can be set to any arbitrary PC but a level halfway between pure chance (1/M) and 100% correct is commonly used.[75] In order to translate the PC for all MAFC studies to a common scale linear to pattern contrast, a detectability index (d') can be calculated by using analytical expressions (Eq. 6) or tabulated values, as presented by Elliott.[76] The relationship between PC and d' is:

$$PC = 1 - (M - 1) \int_{-\infty}^{\infty} \Phi(t)^{M-2} \Phi(t - d') G(t) dt, \tag{6}$$

where

$$G(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \tag{7}$$

is the standard normal distribution (average at t=0 and a variance of 1), and

$$\Phi(t) = \int_{-\infty}^{t} G(x)dx \tag{8}$$

is the cumulative normal distribution. See Figure 24 for an example. For a 2AFC experiment, the detectability index (d') is calculated by[74]

$$d' = \sqrt{2}\ \Phi^{-1}(PC), \tag{9}$$

where Φ^{-1} is the inverse cumulative normal distribution.
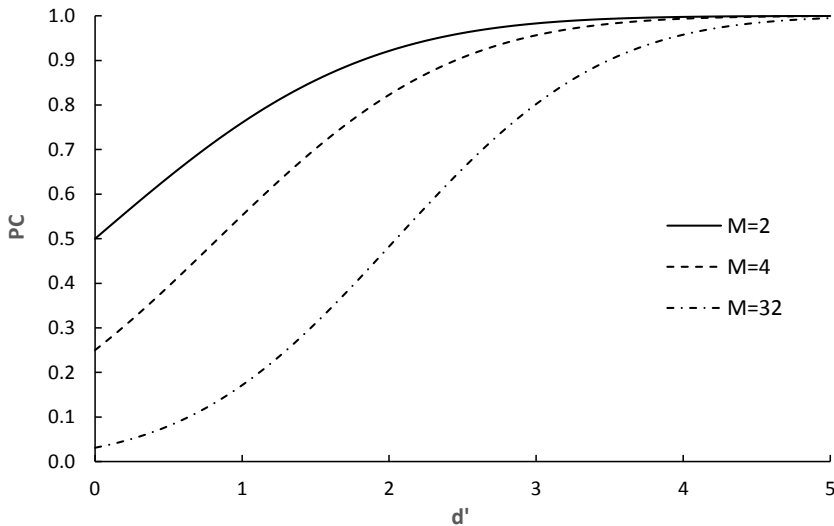

Figure 24: Relationships between the detectability index (d') and the percentage of correctly identified patterns (PC) for three MAFC studies. Data were obtained by using Eq. 6.

3.7 Visual grading characteristics

While MAFC studies are useful for determining contrast thresholds, other methods are necessary when the contrast of clearly visible objects is to be rated subjectively. In visual grading studies, the perceived contrast of an object is rated on an ordinal scale, like [1, 2, 3, 4, 5] or [bad, neutral, good]. Visual grading characteristics (VGC) analysis is a method of analyzing rating data from subjective observer studies that do not violate the properties of ordinal data and that also acknowledges that the ratings given by different observers belong to different ordinal scales.[77, 78] In VGC analysis, the ratings belonging to two classes of observations are compared by creating a VGC curve, similar to how the ratings on normal and abnormal cases in receiver operating characteristics (ROC) analysis are used to create an ROC curve. The VGC curve is obtained by plotting the cumulative distribution of the ratings for one

class of observations (the evaluated class) against the same distribution for the second class (the reference class). The VGC curve thus ranges from (0, 0) to (1, 1). The deviation from the diagonal is a measure of how much the two classes of ratings are separated. Normally, the area under the VGC curve (AUC_{VGC}) is used as a scalar measure of this separation.[79–82] An AUC_{VGC} of 0.5 reflects that, on average, there is no separation between the two classes. The ratings are usually analyzed in such a way that an AUC_{VGC} above 0.5 indicates a better result for the evaluated class and an AUC_{VGC} under 0.5 indicates a better result for the reference class.

3.8 Previous studies

The unsatisfactory condition of variable adaptation when calibrating medical displays according to the GSDF has previously been pointed out by e.g Muka and Whiting,[19] but apart from the previously mentioned studies by Barten[63] and Flynn et al.[33], not much has been made to address the problem. There are however several studies that have investigated how the combination of room illuminance and monitor luminance influences observer performance.[47, 83–93] The influence on observer performance caused by the number of simultaneously displayed gray levels has also been investigated.[35, 94] Another study verified the validity of the GSDF by estimating the contrast thresholds for disc-shaped objects in complex backgrounds.[95] The study was performed under conditions close to variable adaptation. The contrast threshold in GSDF calibrated displays has also been studied by using a location-unknown methodology.[96] Currently there is no standard that includes color calibration in medical displays, but several studies have debated and investigated the possibility.[97–105] The change in color temperature with ageing displays and its effects on observer performance has also been studied.[106] Another reason for contrast degradation when viewing medical images is veiling glare, which can originate from light sources outside the display as well as from bright parts in an image. A number of studies have focused on quantifying the effects caused by veiling glare.[85, 107–115] Since LCDs are non-Lambertian, the effects caused by different viewing angles have been documented in several papers.[116–121]

4 MATERIALS AND METHODS

4.1 Overview of the papers

Paper I describes a theoretical study that investigated the effects of combining the GSDF (variable adaptation) with the f-factor describing fixed adaptation. The purpose of the study was to quantify the perceived contrast for all DDLs when using displays with different luminance ranges under the assumption of fixed adaptation.

In **Paper II**, sinusoidal test patterns were constructed using 8-bit DDLs together with a technique called subpixel modulation. The patterns were used in a 2AFC study in order to determine the contrast thresholds and to compare the results to both the GSDF and the f-factor.

The pattern generation technology was improved further in **Paper III**, where 10-bit DDLs were used together with subpixel modulation. Contrast thresholds were determined for both variable adaptation and fixed adaptation. Also, the influence of luminance originating from outside the display was studied.

In **Paper IV**, a new calibration method that compensates GSDF for fixed adaptation was developed and evaluated. In order to develop the new calibration method, contrast thresholds for fixed adaptation had to be determined with greater precision than in previous papers.

4.2 Equipment and locations

The room used for all display measurements and observer studies was a small room without windows where the walls, ceiling and door were painted in a matte black color in order to reduce light reflections. Light emerging from the edges of the door was covered using black plastic sheets. The display was positioned on a desktop and behind the display there was a lightbox with adjustable light output. The size (width by height) of the lightbox and the display were 125 by 80 cm and 48 by 36 cm, respectively.

Displays from Eizo (Study II & III: RadiForce R31; Study IV: RadiForce RX320; Nanao Corporation, Ishikawa, Japan) were used together with a graphics board from RealVision (VREngine/SMD3-DUL; RealVision Inc., Yokahama, Japan). The displays were capable of using 10-bit input resolution and the graphic board was able to produce a 10-bit output signal if the image display software used certain routines in the software development kit (SDK) provided by RealVision.

The softwares used for display measurements and observer studies were developed using Visual studio 2002 (Study II & III) and further improved using Visual studio 2010 Express (Study IV), both softwares from Microsoft (Microsoft Corporation, Redmond, Washington, USA).

Luminance measurements were performed using a telescopic luminance meter from Minolta (Minolta LS-100; Konica Minolta optics, Inc., Osaka, Japan) with a circular viewing aperture of 1 degree and a smallest focusing distance of 100 cm, resulting in a smallest measurement area of diameter 1.75 cm.

4.3 Calibration software

A calibration software was developed since commercially available softwares did not fulfil all of the necessary requirements. The developed software was able to display different types of test patterns and automatically retrieve luminance readings from the luminance meter through a serial interface (RS232). The software was also capable of deriving GSDF calibration corrections and reading/writing LUTs to the display and/or graphic board, where applicable. Also, every possible display setting, such as brightness, display mode, etc., could be controlled by using a SDK provided by Eizo. It was also possible to change the settings of the luminance meter using this software. Other features were: measurements of luminance stability over time, option to restrict measurements to a certain pixel value interval, repeated measurements with automatic saving of the results, corrections for close-up lenses and automatic shutdown of computer and display after completed measurements.

Three different test patterns were available for display characterization and/or calibration:

1. **DICOM**
   Standard DICOM calibration procedure where the pattern area covered 10% of the surface. The luminance of the background area was set to 20% of the maximum luminance.
2. **Homogenous**
   The pattern area covered the entire screen.
3. **Fixed size mode**
   The pattern area was as small as possible. The smallest measurement area possible with the used luminance meter, considering the shortest focal distance, was circular with a diameter of 1.75 cm and the pattern area was somewhat larger to reduce veiling glare from the background area, which was set to a user defined pixel value. In this mode, the pattern area was similar in size to the patterns used in the detection studies, although somewhat larger.

For each mode the characteristic curve could be measured in a number of equidistant DDLs. In order to increase the number of unique luminance levels, a subpixel modulation technique was used[43] (see section 4.5 for an explanation). The number of possible steps implemented in the software was:

For 8 bit displays:	4, 18, 86, 256
For 8 bit displays using subpixel modulation:	766 (3 subpixel levels) and 1786 (7 subpixel levels)
For 10 bit displays:	1024
For 10 bit displays using subpixel modulation:	3070 (3 subpixel levels) and 7162 (7 subpixel levels)

4.4 Measurements of the characteristic curves

Careful measurements of the characteristic curves were necessary in order to obtain the luminance level corresponding to every (sub-) pixel value. Figure 25 shows the measurement setup and the three luminance areas that were altered in Study II-IV: **Pattern area:** The area of the test patterns in observer studies or the area displayed when measuring. **Background area:** Entire display surface excluding the pattern area. **Ambient area:** Entire lightbox area.

When changing the luminance levels at the background and ambient areas, room illuminance also changed. Even though the room was designed to scatter as little light as possible, display reflections could not be completely avoided. Therefore, the characteristic curve had to be measured for every combination of background and ambient area luminance level. The characteristic curves were measured in fixed size mode under luminance conditions as similar as possible to the conditions in the observer studies. In order to find the pixel value corresponding to the desired background area luminance, a characteristic curve was first obtained by measuring in homogenous mode.
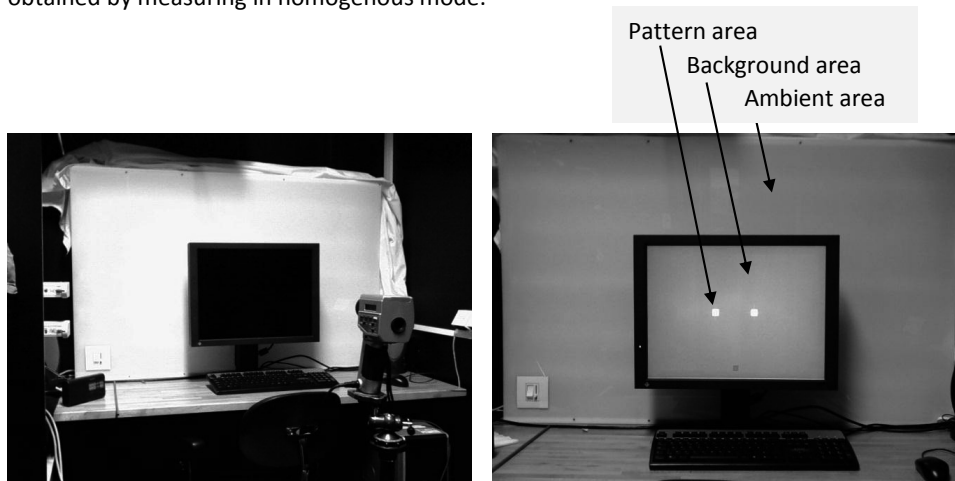

Pattern area
Background area
Ambient area

Figure 25: Measurement setup and used luminance areas.

A measurement of the characteristic curve involved the following steps:

1. Adjustment of the ambient area luminance to the desired level.
2. Setting the luminance range of the display.
3. Verification that the LUT in the graphic board and display were set as intended.
4. Measurement of the characteristic curve in homogenous mode in order to find the pixel value corresponding to the desired background luminance.
5. Measurement of the characteristic curve in fixed size mode with the background area set to the desired pixel value.

The measured relationship between pixel value and luminance could then be used to generate test patterns. When generating low contrast test patterns, it is important to have a display that is stable in output over time. An example of repeated measurements on a display with

1024 (10 bit) DDLs using subpixel modulation is shown in Table 3. Measurements from black to white of all 7162 gray levels showed that the variation in luminance for any subpixel combination was usually smaller than the distance to any of the neighboring subpixel combinations. Even if the luminance output might have drifted somewhat, it is likely that this drift was more or less linear and influenced all pixel values for a particular pattern to an equal amount.

Table 3: Repeated measurements of the 10-bit characteristic curve between pixel values 22 and 23 including subpixels. Every measurement of luminance was performed from pixel value 0 to pixel value 1023 before the next measurement started. R, G and B represent pixel values for red, green and blue, respectively.

Pixel Value			Measurement of luminance (cd/m^2)				
R	G	B	1	2	3	4	5
22	22	22	1.484	1.484	1.483	1.483	1.482
22	22	23	1.487	1.486	1.485	1.485	1.485
23	22	22	1.489	1.489	1.488	1.488	1.487
23	22	23	1.492	1.490	1.491	1.490	1.490
22	23	22	1.498	1.497	1.497	1.496	1.496
22	23	23	1.501	1.500	1.498	1.499	1.498
23	23	22	1.504	1.504	1.502	1.502	1.501
23	23	23	1.505	1.505	1.504	1.503	1.504

4.5 Generation of test patterns

Patterns can be of any type but the ones most commonly used in detection threshold studies are sinusoidal patterns and bar patterns of varying size and spatial frequency. High-contrast bar patterns are easy to generate on a display, but most studies before the digital era were performed using low-contrast sinusoidal patterns, which are a challenge to generate on a digital display due to the finite number of available luminance levels. Assume a display with a luminance range of 1 cd/m^2 to 600 cd/m^2. This range covers 662 JNDs. When using an 8-bit display system this corresponds to 662/255 = 2.6 JND per change in DDL. In other words, the average change in luminance is 2.6 times the detection threshold. Obviously, any pattern displayed on this system would be easily detected. By using a 10-bit display system the number of available luminance levels is increased by a factor of 4, and the change in output luminance is now less than one JND per change in DDL. Still, it would be difficult generating a complex pattern with a precise shape and contrast. In order to create a low-contrast sinusoidal test pattern, an even higher luminance resolution is necessary. One way of achieving this is to use the subpixels on a color display. A grayscale value is characterized by equal values for all three subpixels; red, green and blue. By allowing any of the subpixels to deviate by one from the other two, the number of available luminance levels is increased by a factor of seven. As an example, observe the following pixel values [red, green, blue]:

All equal:
[99,99,99] ; [100,100,100]

Deviation of one:
[99,99,99] ; [99,99,100] ; [99,100,99] ; [99,100,100] ; [100,99,99] ; [100,99,100] ; [100,100,99] ; [100,100,100]

By using the subpixels in this way on a 10-bit display system, the total number of available luminance levels is 7162, resulting in an average change of 0.09 JND per change in input value. The small color tint is not perceivable.

The DICOM standard test target is an 8-period sinusoidal pattern covering a viewing angle of 2° × 2°. Considering the display pixel resolution together with the desired viewing distance, eight pixels (sampling points) were used for each period. Also, by using eight equidistant sampling points per period, five luminance levels were sufficient since some sampling points were mapped to the same luminance level, see Figure 26. The pixel values corresponding to the five luminance levels were determined using the measured characteristic curve and were assigned to respective sampling point. Since the available luminance levels were not equally distributed, a computer program was developed that searched the available luminance levels for the best possible sinusoidal fit to a pattern with a specified luminance and contrast. The search was made within a limited luminance range and a limited contrast range. The contrast was defined as the difference in JND index between the maximum and minimum luminance of the sinusoidal fit.
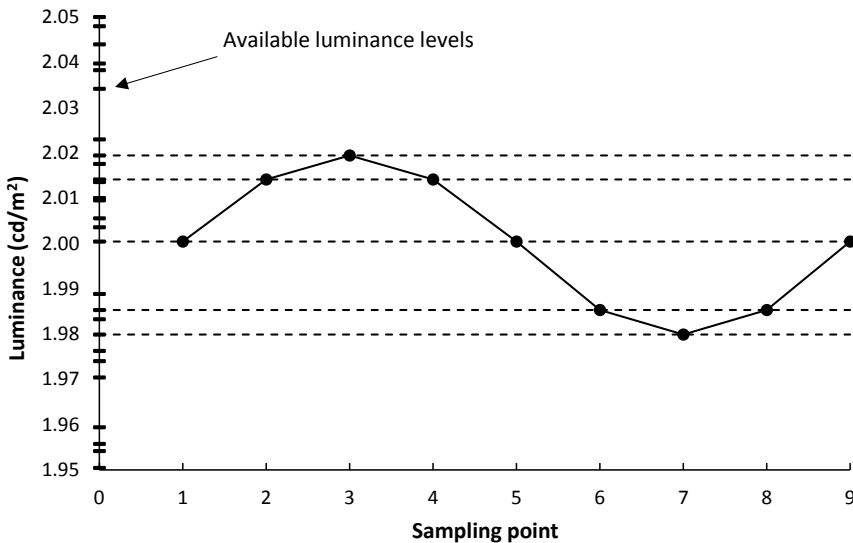

Figure 26: Sampling points and corresponding luminance levels for a sinusoidal pattern. The available luminance levels in the display are marked by ticks on the y-axis.

4.6 Observer studies

All observer studies were performed in the same room and under the same lighting conditions as the display measurements. The viewing distance used was approximately 40 cm which corresponded to a viewing angle of 2° x 2° as specified by the DICOM standard test target. In the 2 AFC studies, the observers were forced to decide whether the patterns were present to the left or to the right. Each pattern was present on one of the sides while the other side

displayed a homogenous area with the same luminance as the pattern baseline luminance, as can be seen in Figure 27. A small random horizontal shift was used for every pattern in order to reduce the influence of fixed pattern noise in the monitor and a small delay of approximately one second was used between patterns to minimize the possibility of detection by temporal change. All patterns were displayed in random order and the observers were free to decide the length of each viewing session.
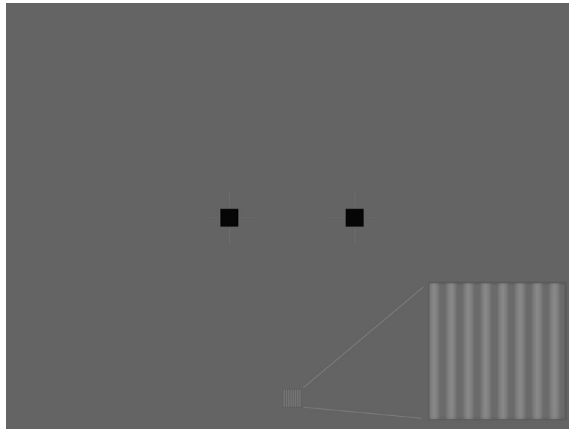

Figure 27: Example of test patterns used in the 2AFC studies in Papers II-IV. A high-contrast test pattern (enlarged in the lower right part of the figure) was always visible at the bottom of the screen during observations. The actual low-contrast pattern to be detected was either present to the left or to the right.

4.7 Study designs

An overview of the used patterns and observer studies are summarized in Table 4.

Table 4: Overview of the used patterns and observer studies. Luminance settings are different combinations of pattern/background/ambient luminance. The contrast levels are the number of pattern contrast levels used to determine the contrast threshold. Repetitions are the number of repeated observations for each combination of luminance setting and contrast level. PC is the percentage of correctly identified patterns used for determination of the contrast threshold.

	Sinusoidal patterns				2AFC observer studies		
	Pattern generation technology	Luminance settings	Contrast levels	Repetitions	Observations per observer	Observers	PC
Paper II	8 bit subpixel	6	4	30	720	15	75
Paper III	10 bit subpixel	11	4	25	1100	10	80
Paper IV	10 bit subpixel	9	10	50	4500	11	75

4.7.1 Paper I

The effects obtained when combining the GSDF (derived at variable adaptation), with Barten's f-factor (derived at fixed adaptation) were studied for a large number of luminance ranges

using computer simulations. An increasing luminance range increases the total contrast in terms of JNDs, while the f-factor predicts a decrease in contrast for objects with a luminance far from the adaptation luminance. The purpose of the study was to quantify the obtained contrast for all DDLs when using displays with different luminance ranges. A quantity called effective JND (JND_{eff}) was derived by combining the GSDF with the f-factor and, due to quantization effects, converted to an average probability (P_{ave}) of detecting the difference in luminance between neighboring DDLs. P_{ave} was used as a figure-of-merit (FOM) together with the maximum (P_{max}) and minimum (P_{min}) probabilities for detection. The probability for detecting a difference in luminance between two neighboring DDLs was determined for three eye adaptation models. Variable adaptation is assumed for GSDF where an observer is always adapted to every possible luminance level in the displayed image. Two other models with fixed adaptation were also studied with their respective adaptation luminance levels calculated by the linear average and the logarithmic average of the minimum and maximum luminances.

4.7.2 Paper II

Paper II aimed at generating sinusoidal test patterns suitable for vision research on computer displays and also to obtain a value for the contrast threshold at the adaptation luminance level and two other levels, one lower and one higher than the adaptation luminance. Patterns were generated with 8-bit DDLs using subpixels. Three pattern luminance levels (1.8, 20 and 350 cd/m^2) were used and for each luminance level, patterns at four contrast levels were created. The study was conducted with two different backgrounds: one with the background area set to 20 cd/m^2 homogenously distributed, and the other with bright white stripes on a black background, still with an average luminance of 20 cd/m^2. The purpose was to study if the background luminance distribution could influence the level of the contrast threshold. 15 observers viewed every combination of pattern luminance, contrast and background 30 times in a 2AFC study. The contrast threshold was determined at 75% correct responses.

4.7.3 Paper III

In Paper III, patterns were generated with 10-bit DDLs using subpixels with the purpose of determining contrast thresholds at three luminance levels under both variable adaptation and fixed adaptation. Also, the effects of using low and high ambient luminance levels were studied in order to quantify the influence of light originating from outside the display. In total, 11 luminance area combinations (Figure 28) were used together with 4 contrast levels for each combination. Ten observers viewed every combination of luminance area and pattern contrast 25 times in a 2AFC study. Due to generally high detection rates, a level of 80% correct responses was used as the contrast threshold instead of 75%. In this way, extrapolation could be avoided.
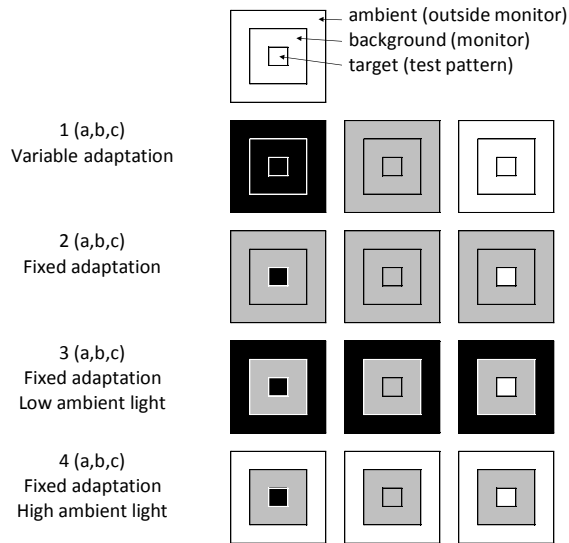
Figure 28: Illustration of luminance settings used in Paper III. Black, gray and white correspond to 2 cd/m^2, 26.5 cd/m^2 and 350 cd/m^2, respectively. Note that settings 1b and 2b were identical and the corresponding part of the observer study was only performed once.

4.7.4 Paper IV

A new calibration method compensating for the effects of fixed adaptation was developed and evaluated in Paper IV. Since the contrast thresholds obtained in Paper II and Paper III differed from those described by the f-factor, and also because other studies[33] showed non-consistent results, the contrast thresholds for fixed adaptation were determined at nine luminance levels. In the papers by Barten[63] and Flynn et al[33] described previously, the contrast sensitivities were almost symmetric when expressed as a function of the logarithmic ratio between adaptation luminance and pattern luminance. In order to obtain evenly distributed sampling points of the contrast sensitivity function, low-contrast patterns in the luminance range 2 cd/m^2 to 600 cd/m^2 were constructed at nine different luminance levels distributed evenly on a logarithmic scale, i.e. 2, 4, 8, 17, 35, 71, 144, 294 and 600 cd/m^2. Test patterns at 10 contrast levels close to the detection threshold were used for each luminance level in a 2AFC observer study. Eleven observers viewed each combination of pattern contrast and luminance 50 times. The PC was converted to the detectability index d' and the contrast thresholds were determined at a level corresponding to 75% correct responses. A Gaussian function was fitted to the obtained data and was used to calculate a new calibration function based on the GSDF, but fixed-adaptation compensated (GSDF$_{FAC}$), see section 4.9. The developed method was used for determining the GSDF$_{FAC}$ for a display with a luminance range of 2 to 600 cd/m^2. In order to study the impact of the choice of adaptation level, the GSDF$_{FAC}$ was determined at three different adaptation levels: (A) the logarithmic average of the maximum and minimum luminance, 35 cd/m^2; (B) the linear average of the luminance levels

corresponding to all DDLs, 173 cd/m^2; (C) 7 cd/m^2 which corresponds to a factor 5 lower than (A), since (B) is a factor 5 higher than (A).

The two calibration methods were used in a VGC study where the visibility of low-contrast (but clearly visible) circular test objects was evaluated using the same equipment as when determining the contrast sensitivity function. In this study, a comparison of the visibility of test objects at different pixel value levels was made between calibrations according to the GSDF and according to the GSDF$_{FAC}$. As described above, the contrast sensitivity function is expected to be almost symmetric when expressed as a function of the logarithmic ratio between adaptation luminance and pattern luminance. In order to approximately center the contrast sensitivity function with regard to the luminance range of the display, an adaptation level of 30 cd/m^2 (in turn approximately corresponding to the logarithmic average of the minimum (1.06 cd/m^2) and maximum luminance (673 cd/m^2) of the display) was used for L$_{adapt}$ when calibrating the display according to GSDF$_{FAC}$. For both calibration methods, a pixel value of 100 approximately corresponded to a luminance of 30 cd/m^2. In order to achieve an adaptation level of 30 cd/m^2 in the evaluation study, relatively small test patterns were placed in a homogenous background of pixel value 100. Test patterns with a diameter of 40 pixels were placed in another homogenous circular area with a diameter of 140 pixels. An example of a test pattern is found in Figure 29. The difference in luminance between the two circular areas corresponded to 2 pixel values (on an 8-bit scale). In order to evenly sample the pixel value range, these patterns were displayed at pixel values of 8, 68, 128,188 and 248. 10 observers viewed every combination of calibration method and luminance level 50 times in random order. The perceived contrast for each displayed pattern was rated on a 5-grade ordinal scale from 1 (worst) to 5 (best).
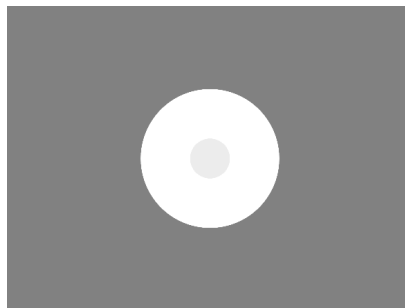

Figure 29: Example of test pattern used in the evaluation study in Paper IV. The contrast between the two circular areas has been increased for demonstration purposes.

4.8 Determination of the contrast thresholds

The contrast thresholds were determined by the obtained relationships between pattern contrast and the proportion of correctly identified patterns (PC). In Papers II and III, the contrast thresholds were determined at 75% and 80% correct responses, respectively. The reason for using a higher detection threshold in Paper III was that the obtained results were

Calibration of medical displays

generally higher than 75%, and a detection threshold corresponding to a PC of 80% did not require extrapolation for any combination of observer and luminance settings. In Paper IV, the detectability index (d') was calculated using Eq. 9 for all obtained PC. For each observer and luminance level, a linear fit between d' and contrast was determined. Since the human visual system contains internal noise even when no signal is present, the linear fit was forced to intercept the y-axis (contrast=0) at d'=-0.4.[122] A 100% detection rate returns a d' of infinity and all such values were therefore excluded from the analysis. The detection threshold of 75% correct responses corresponds to a d' of 0.95, see Eq. 9.

4.9 Derivation of a new calibration method

The obtained Gaussian function in Paper IV, describing the contrast sensitivity at fixed adaptation, was used in order to derive a new calibration method that compensates GSDF for fixed adaptation (GSDF$_{FAC}$). Although this Gaussian function is denoted with the letter f, it is not identical to Barten's f-factor. Both functions describe the contrast sensitivity at fixed adaptation but were derived using different methods and take different values. The Gaussian function (f) in Paper IV is described by

$$f = A_0 \times e^{-\frac{z^2}{2}},$$
(10)

where z is obtained from

$$z = \frac{\log\left(\frac{L_{adapt}}{L}\right) - A_1}{A_2},$$
(11)

where L_{adapt} is the adaptation luminance, L is the pattern luminance and A_0, A_1, A_2 are the height, center and width of the Gaussian, respectively.

For a display, the relationship between input values (DDLs) and luminance output can be divided into two steps where the first step is a function between DDLs and JND index values. For a GSDF calibrated display, this is a linear relationship where the available JND index values are evenly distributed over all DDLs. The calibration method proposed in this thesis will instead exhibit a non-linear relationship. In the second step, the JND index values are transformed to luminance values where both calibration procedures use the same method. Correction factors for each DDL (DDL$_i$ where $0 \leq i \leq N$) can be calculated from the Gaussian function if

a) the adaptation luminance is known, and

b) the luminance corresponding to that DDL is known.

Since the new calibration method is intended for clinical images with an unknown luminance distribution displayed on devices with varying luminance ranges, the adaptation luminance has to be estimated. However, it is outside the scope of the present thesis to define a general adaptation luminance level.

The luminance corresponding to a particular DDL can only be determined if the display calibration is known. Since the display calibration is initially not known, an iterative method is required. The first set of luminance values associated with the DDLs was obtained by a normal calibration to the GSDF and used as input values, i.e. the first associated luminance value for DDL_i was $L_{GSDF,i}$. The corresponding JND index value is denoted $JND_{GSDF,i}$.

The correction factor (c_i) for DDL_i ($1 \leq i \leq N$) was calculated from

$$c_i = \frac{1}{f_{L_i}},$$
(12)

where f_{L_i} is the contrast sensitivity for the luminance level (L_i) associated with DDL_i as determined by Eq. 10.

To keep the corrected JND index values within the available JND index range, the correction factors needed to be normalized to an average of one:

$$c_{norm,i} = \frac{c_i}{c_{average}},$$
(13)

where $c_{norm,i}$ is the normalized correction factor for DDL_i, c_i is the correction factor for DDL_i and $c_{average}$ is the average of all correction factors.

For every DDL, a new JND index was calculated by multiplying the increment in JND_{GSDF}, which is constant for all DDLs, with the corresponding correction factor and adding the new JND index from the previous DDL. The first new JND index was the same as for the GSDF and, since the correction factors were normalized, the last JND index also ended up the same as for the GSDF:

$$JND_{new,0} = JND_{GSDF,0}$$
(14)

and

$$JND_{new,i} = c_{norm,i} \times (JND_{GSDF,i} - JND_{GSDF,i-1}) + JND_{new,i-1},$$
(15)

where $JND_{GSDF,i}$ is the JND index associated with the GSDF luminance value ($L_{GSDF,i}$) used as the first input luminance value. For every new JND index, a corresponding luminance value was calculated by using the same relationship between JND index and luminance as for the GSDF. The new luminance values were used as input to a repeated calculation until the difference between any obtained luminance value and the corresponding input luminance value became negligible.

4.10   Statistical analysis

In Paper II, the software IDL (Exelis Visual Information Solutions Inc., Boulder, CO, USA) was used in order to obtain the best possible linear fit between pattern contrast and the proportion of correctly identified patterns. The standard errors of the proportions of correctly identified patterns were also used as input in order to obtain the slope and intercept errors.

The errors of the contrast thresholds and f-factors were calculated using normal error propagation rules.

Bootstrapping is a statistical method that can be used to estimate the confidence intervals from a single data set with an unknown probability distribution.[123] A new data set of the same size as the original data set is obtained by sampling (with replacement) from the original data set. The new data set differs from the original and the obtained mean value will therefore also be different. By obtaining a large number of bootstrapped data sets, a distribution of mean values is obtained, and the 95% confidence interval is obtained by removing the 2.5% lowest means as well as the 2.5% highest means.

In Paper IV, a total of 2000 bootstrapped data sets were obtained, where each data set included the data from 11 randomly sampled (with replacement) observers. For each observer, luminance level and contrast level within the data set, 50 randomly sampled (with replacement) observations were selected. For each of these bootstrapped data sets, the contrast threshold at each luminance level and a Gaussian fit of the contrast sensitivity were determined. The asymmetric 95% confidence intervals were determined from the 2.5 and 97.5 percentiles of each bootstrap distribution.

In the VGC analysis[77, 78] in Paper IV, the visibility of each combination of luminance level and calibration was compared against the average visibility of all combinations. For each analysis, the evaluated class thus consisted of the ratings for a given luminance level/calibration combination, whereas the reference class consisted of all ratings. The VGC analysis was performed using the software VGC Analyzer (In-house developed computer software, University of Gothenburg). This software determines the AUC_{VGC} averaged over observers and a bootstrapping technique is used to determine the asymmetric 95% confidence interval of the AUC_{VGC}. The statistical analysis of the AUC_{VGC} was based on the trapezoidal VGC curve and the resampling was done on both patterns and observers in order for the results to be generalizable to the population of patterns and observers. A 95% confidence interval not covering 0.5 was interpreted as indicating that the visibility for that luminance/calibration combination significantly differed from the average combination.

5 RESULTS

5.1 Theoretical determination of optimal luminance settings (Paper I)

Some of the results from the theoretical study that combined GSDF with the f-factor, Paper I, is shown in Figure 30. The minimum luminance was always 1 cd/m^2 and the maximum luminance varied between 1 and 1000 cd/m^2.
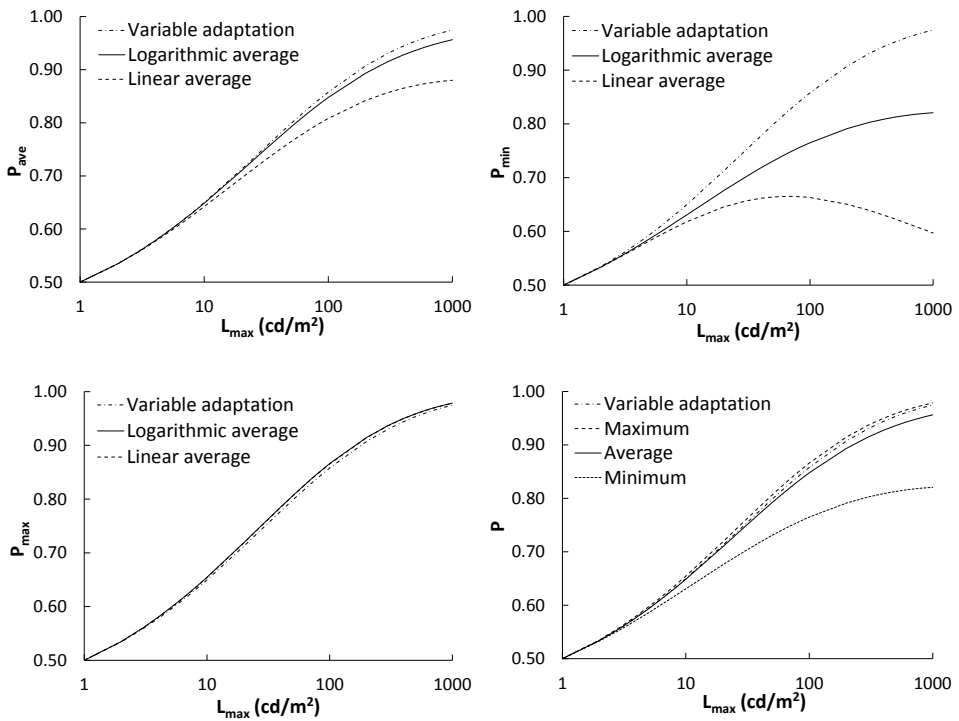

Figure 30: P_{ave} (upper left), P_{min} (upper right) and P_{max} (lower left) as a function of L_{max} when $L_{min} = 1$ cd/m^2 for an adaptation luminance given both by the logarithmic and the linear average of L_{max} and L_{min}. Lower right: P_{ave}, P_{min} and P_{max} for the logarithmic adaptation presented together. As a comparison, the value of P for variable adaptation is presented.

The difference in P_{ave} between variable adaptation and logarithmic average adaptation was small for all luminance ranges, and the corresponding difference for P_{max} was negligible. In other words, fixed adaptation caused only a minor contrast reduction for most DDLs compared to variable adaptation. However, the difference in P_{min} between the two types of adaptation was substantial, especially for larger luminance ranges. For some of the DDLs, the contrast was reduced considerably. The reduction increased with increasing luminance range and, due to the Gaussian shape of the f-factor, was caused by a reduction in contrast for the

lowest and highest DDLs, i.e. the darkest and brightest parts of an image. For the linear average adaptation model, an increasing luminance range caused an increase in contrast in all parts of the image up to a certain luminance range. When the luminance range was increased even further, contrast in some parts of the image started do decrease (P_{min} reached a maximum). Even though the luminance range increased, low-contrast objects in some parts of the image were more difficult to detect.

5.2 Effects of non-uniform luminance distributions (Papers II and III)

By using sinusoidal test patterns in a homogenous background, a contrast threshold of 0.75 ± 0.02 JND (95% CI) was obtained in Paper II when both the target and background average luminance were close to 20 cd/m^2. When a striped pattern with the same average luminance was used as background, the contrast threshold was 0.74 ± 0.03 JND. At least for this particular pattern, no significant difference was found when compared to the homogenous background.

Table 5 shows the obtained contrast thresholds from Paper III for the luminance conditions described in Figure 28. Luminance conditions 1 (a,b,c) corresponded to variable adaptation where the pattern luminance equaled the background luminance. Luminance conditions 2 (a,b,c) corresponded to fixed adaptation where the pattern luminance varied but the background luminance was constant. Luminance conditions 3 (a,b,c) and 4 (a,b,c) were similar to fixed adaptation except for the ambient luminance that was lower and higher, respectively, than in 2 (a,b,c). Luminance conditions 1b and 2b were identical, but patterns with a luminance far from the adaptation level (2a and 2c) were more difficult to detect than patterns with the same luminance under variable adaptation conditions (1a and 1c). When the luminance from the surrounding area was increased from a very low level (3 a,b,c) to the same level as the display luminance (2 a,b,c), the contrast thresholds were unaffected, but when the luminance from the surrounding area was increased considerably (4 a,b,c), patterns in bright areas (4c) were detected more easily and patterns in in dark areas (4a) were more difficult to detect.

Table 5: Contrast thresholds in JND for all luminance settings in Paper III (cf. Figure 28). Luminance settings 1b and 2b were identical and the results were therefore obtained from the same data in the observer study.

1 (a,b,c)	0.63	0.64	0.5
2 (a,b,c)	1.06	0.64	0.77
3 (a,b,c)	1.08	0.65	0.74
4 (a,b,c)	1.53	0.62	0.67

5.3 Determination of contrast thresholds and comparisons with the f-factor (Papers II-IV)

The GSDF is based on variable adaptation and the luminance change required for detection of the DICOM standard target is, per definition, one JND. Some of the patterns in Paper II-IV were studied using luminance conditions equal to variable adaptation, i.e. the average luminance of the pattern was equal to the background luminance, and the obtained contrast thresholds in the different papers are presented in Table 6. The obtained contrast thresholds were in the range 0.47 to 0.75 JND and were determined at luminance levels between 2 and 350 cd/m^2. The detection threshold level was 75% correct responses in Papers II and IV and 80% correct responses in Paper III.

Table 6: Obtained contrast thresholds for variable adaptation in the different papers.

	Contrast threshold (JND)	Luminance (cd/m^2)	Detection threshold level (%)
Paper II	0.75	20	75
Paper III	0.63	2	80
Paper III	0.64	26.5	80
Paper III	0.50	350	80
Paper IV	0.47	35	75

The relative contrast sensitivities for patterns with a luminance different from the adaptation luminance were determined in Papers II-IV. The results are shown in Figure 31 together with the corresponding relationships by Flynn et al.[33] and Barten.[63]
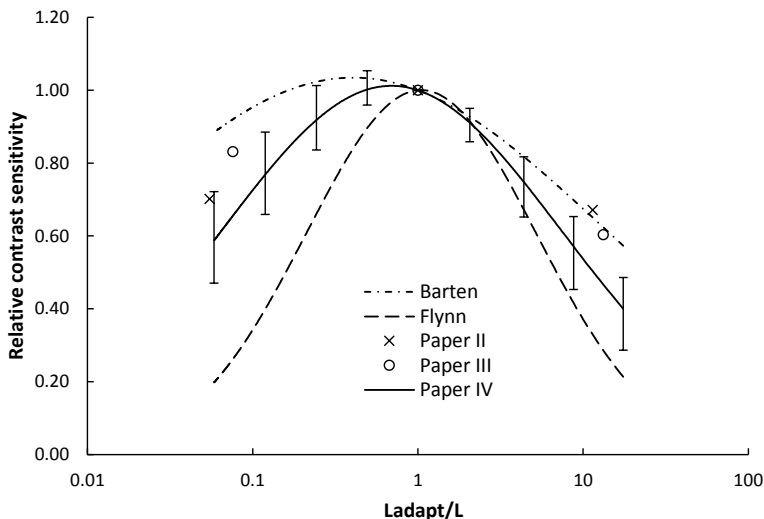

Figure 31: A Gaussian fit to the relative contrast sensitivity values obtained in Paper IV and the results from Paper II and III together with the corresponding relationships by Flynn et al.[33] and Barten.[63] L is the pattern luminance and L$_{adapt}$ is the adaptation level of the observer. Error bars correspond to the 95% CI.

In Papers II and III, only three pattern luminance levels (including the adaptation level) were studied. In Paper IV, patterns at nine luminance levels were studied, and shown in the figure is a Gaussian fit to the data together with the bootstrapped 95 % confidence interval for the Gaussian fit. The experimentally determined values for A_1 and A_2 (Eq. 11) with 95% confidence intervals were -0.16 [-0.33, -0.02] and 1.03 [0.94, 1.13], respectively.

Most obtained contrast thresholds from Papers II-IV fall between the results of Flynn et al.[33] and Barten.[63] While the contrast sensitivity according to Flynn peaks at the adaptation luminance, the maximum contrast sensitivity according to Barten occurs for patterns having a luminance slightly higher than the adaptation luminance. This shift in peak contrast sensitivity was also noticed in the results from Papers II-IV, although not as pronounced.

5.4 Generation and evaluation of a new calibration method (Paper IV)

The obtained contrast sensitivities from Paper IV (Figure 31) were used to derive a new calibration method (GSDF$_{FAC}$) that compensates the GSDF for fixed adaptation. The luminance output for the two calibration methods (GSDF and GSDF$_{FAC}$) is shown in Figure 32 for a display calibrated between 2 and 600 cd/m^2.
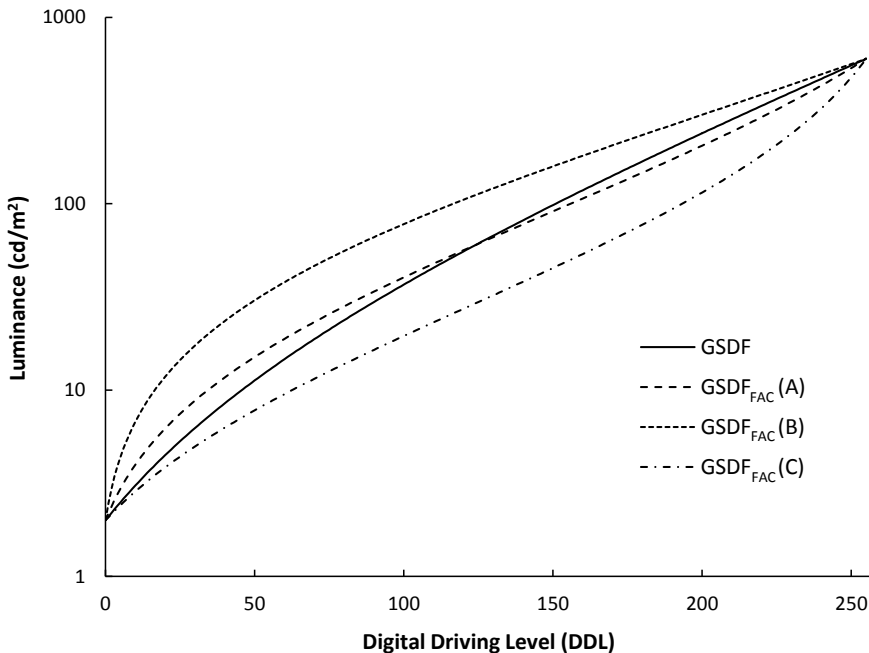

Figure 32: Luminance output as a function of DDL, for a display calibrated between 2 and 600 cd/m^2. The GSDF$_{FAC}$ is shown for three possible adaptation levels: (A) the logarithmic average of the maximum and minimum luminance, 35 cd/m^2; (B) the linear average of the luminance levels corresponding to all DDLs, 173 cd/m^2; (C) 7 cd/m^2.

The GSDF$_{FAC}$ is shown for three choices of adaptation level: (A) the logarithmic average of the maximum and minimum luminance, 35 cd/m^2; (B) the linear average of the luminance levels corresponding to all DDLs, 173 cd/m^2; (C) 7 cd/m^2. Compared to the GSDF, the slope of the GSDF$_{FAC}$ (A) was higher at low and high luminance levels while it was lower at mid-luminance levels. When the adaptation level increased, as for GSDF$_{FAC}$ (B), the contrast increased for low luminance levels and decreased for high luminance levels, while the opposite was true when the adaptation level decreased, as for GSDF$_{FAC}$ (C).

Results from the evaluation study are shown in Figure 33. As can be seen, for a display calibrated according to the GSDF, the perceived contrast was highest for the two objects closest to the background luminance (pixel value 100). For objects far from the background luminance, the perceived contrast was lower. For this calibration, the AUC$_{VGC}$ ranged from 0.18 to 0.71. When the display was calibrated according to the GSDF$_{FAC}$ (assuming an adaptation level equal to the background luminance which was set to the logarithmic average of the minimum and maximum luminance), the score ranged from 0.44 to 0.68 and the available contrast was thus distributed more evenly than for the GSDF. The two lowest scores for the GSDF were obtained for the darkest and brightest patterns. These scores were significantly lower than the lowest score obtained for the GSDF$_{FAC}$. For the GSDF, the scores for all luminance levels were statistically separated from the average value; three were lower and two were higher. For the GSDF$_{FAC}$, three of the scores could not be separated from the average value.
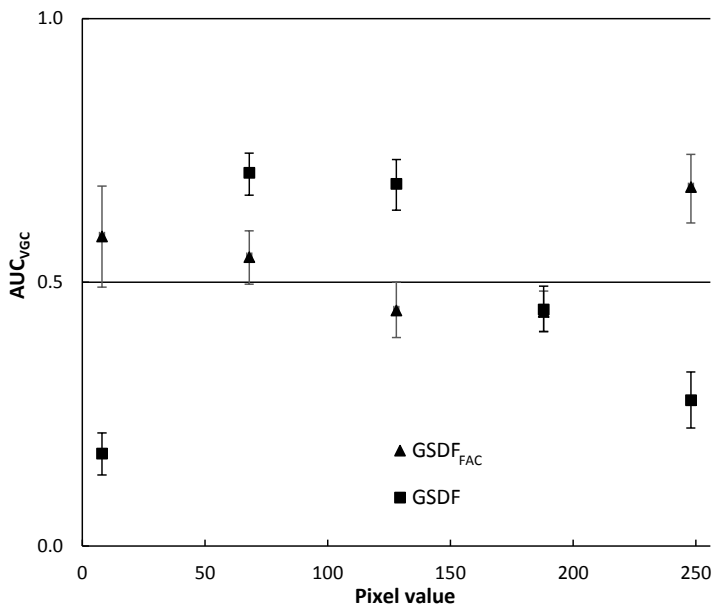

Figure 33: AUC$_{VGC}$ scores from the evaluation study of the two calibration methods in Paper IV. Error bars correspond to the 95% CI.

6 DISCUSSION

Eye adaptation and the limited dynamic range of the adapted human visual system is something that must be considered when viewing medical images – especially on high brightness displays. In Paper I, the theoretical effects of combining the GSDF with a model for fixed adaptation were studied. The results showed that the perceived contrast was distributed unevenly throughout the grayscale with more perceived contrast in the mid-gray region than in the dark and bright regions. The effects were stronger when using a high luminance range and for some eye adaptation models. In the literature there are at least two models for fixed adaptation[33, 63] – and they differ. In order to obtain a model that was realistic when viewing medical displays, observer studies of low-contrast patterns had to be conducted under realistic conditions. The methodology for generating low-contrast patterns on computer displays was developed and tested in Paper II. Since the used luminance resolution turned out to be insufficient, the methodology was further refined in Paper III. Finally, in Paper IV, a model for fixed adaptation was obtained and used to derive a version of the GSDF that is compensated for the effects of fixed adaptation. The obtained model for fixed adaptation showed a more predominant effect than the model used in Paper I, further strengthening the conclusion that fixed adaptation must be considered when viewing medical images. In an evaluation study, the newly developed compensated version of the GSDF distributed the perceived contrast more evenly than the GSDF.

The following sections discusses important aspects of the methods used and the obtained results.

6.1 Theoretical determination of optimal luminance settings (Paper I)

As the luminance range of a display increases, each change in DDL corresponds to an increasing number of JNDs per DDL. On the other hand, an increasing luminance range also causes the contrast sensitivity to decrease for objects with a luminance far from the adaptation luminance, as described by the f-factor. Since the f-factor is always close to unity in the vicinity of the adaptation luminance, it is not surprising that the obtained P_{max} was almost the same for all adaptation models. Larger discrepancies will only occur under the unlikely circumstance that the adaptation luminance is outside the luminance range of the display. The P_{ave}, on the other hand, is the average probability of detection for all DDLs and it is clear that the linear average model suffered from having the adaptation level far from the majority of the other luminance levels, especially for large luminance ranges. Since the distribution of f-factor values on a logarithmic scale is more symmetric for low adaptation levels than for high adaptation levels within the luminance range, the low logarithmic average adaptation level will avoid very low f-factor values. The P_{min} is determined by the largest distances between the adaptation luminance and the luminance of an observed object, and for the linear average model in combination with an increasing luminance range, the increase

in JNDs per DDL no longer compensates for the reduction in contrast sensitivity due to the f-factor. By using a large luminance range, some of the displayed details are unlikely to be detected. The logarithmic average model did not reach the point where perceived contrast is reduced, even for the maximum luminance range included in this study, although P_{min} appeared to be close to the maximum possible value. However, the results indicate than an even larger luminance range or an f-factor model that predicts a faster decrease in contrast sensitivity than the f-factor described by Barten[63] will result in a loss of perceived contrast even for the logarithmic model.

In the lower right part of Figure 30, P_{max}, P_{ave} and P_{min} for the logarithmic average model are compared together with the variable adaptation model. Even though perceived contrast increases in all parts of the image (P_{min} is increasing), even for objects with a luminance as far as possible from the adaptation luminance, it is obvious that displays with a large luminance range will not be perceptually linear. As the luminance range increases, the distance between P_{max} and P_{min} also increases, and objects close to the adaptation luminance will be displayed with a higher contrast than objects with a luminance far from the adaptation luminance. Although the maximum luminance range used in this study is large by clinical standards today, it is not technically impossible to achieve by using state-of-the-art displays.[124, 125] Also, even for the logarithmic adaptation model, which is close to ideal with regards to the symmetry of the f-factor, P_{min} is close to its peak value.

In Paper I, it was concluded that the results indicate that the full luminance range of a display calibrated according to the GSDF should be used in order to best visualize the image information. Under the assumption that Barten's f-factor is valid and that the logarithmic average adaptation model is likely, the conclusion is correct. By increasing the luminance range, the perceived contrast increases in all parts of the image, although the distribution of perceived contrast becomes increasingly less uniform. However, the results from Paper IV indicate that the dynamic range of the HVS is smaller than that described by Barten.[63] By using a contrast model with a smaller dynamic range (discussed further in section 6.3) or luminance adaptation models (discussed further in section 6.5) than in Paper I, reductions in perceived contrast is likely even for some of the studied luminance ranges. Therefore, the conclusion from Paper I has to be modified; when calibrating a display according to the GSDF, there is a maximum luminance ratio beyond which perceived contrast will start to decrease in parts of the image.

6.2 Effects of non-uniform luminance distributions (Papers II and III)

In Paper II, the spatial luminance distribution of the background did not affect the contrast threshold. The obtained thresholds for patterns displayed on a homogenous background and a non-homogenous background with the same average luminance, but with bright lines on a low-luminance background, could not be separated statistically. However, it should be noted that studying static patterns on a static background is relatively far from the clinical conditions

when radiologists rapidly study different parts of clinical images. When the luminance distribution on the retina varies temporally, the adaptation level is determined not only by the average luminance but also by spatial correlations within the scene, a process one hundred times slower than the immediate light response.[126]

Paper III concluded that ambient light originating from outside the display itself did have an effect on the ability of the observer to detect low-contrast objects in dark areas but this effect seemed to be moderate unless the luminance from the surrounding area was higher than the display average luminance (see Figure 28 and Table 5). A high luminance from outside the display did affect the ability to detect low luminance patterns negatively. The viewing angle contributing to eye adaptation has been suggested to correspond to 11 degrees,[64] which corresponds to a circular area of 23 cm when a viewing distance of 60 cm is used, which is well inside the borders of a display. The decrease in contrast sensitivity is probably caused by veiling glare, i.e. the scattering of light inside the eye that causes a reduction in perceived contrast. Similar effects have been demonstrated in other studies.[107, 109]

6.3 Determination of contrast thresholds and comparisons with the f-factor (Papers II-IV)

Most of the studies that the GSDF is based on were performed using the method of adjustment in which an observer adjusts the contrast of a pattern until it is barely visible. Since the adjusted pattern ends up visible, the obtained contrast threshold is higher than that obtained in a 2AFC study, where detection performance rather than visibility confidence is measured. The ratio between contrast thresholds obtained with 2AFC and the method of adjustment has been proposed to be between 0.64 and 0.68.[127]

The differences in contrast thresholds obtained in the different studies, as shown in Table 6 and Figure 31, may be attributed to three different effects: the shape of the patterns were not ideal in Paper II (discussed further in section 6.6), the observer studies in Papers II and III were relatively small, and the composition of observers varied between all three observer studies. The observer studies are discussed further in section 6.7. Another reason for the higher contrast threshold values obtained in Paper III compared to Paper IV is the choice of detection level. In paper III, the contrast thresholds were determined at 80% correct responses compared to 75% in Paper IV. However, the conclusions in Papers II-IV are all based on relative contrast sensitivities, where the obtained contrast sensitivities for all luminance conditions are compared to a reference condition. The absolute detection threshold and observer composition are therefore of minor importance.

The results from Paper IV are probably more accurate than the results from Papers II and III. In Paper IV, contrast thresholds at nine luminance levels were determined using ten levels of pattern contrast for each luminance level. In Paper II and III, four pattern contrast levels were used for each of the three luminance levels. The number of observations for each contrast

level was also considerably larger in Paper IV than in Paper II and III. A brief summary of the study setups is shown in Table 4 in section 4.7.

In Paper IV, the obtained contrast sensitivities for patterns with a luminance different from the adaptation luminance were lower than the values obtained by Barten's[63] f-factor. The results of Paper IV thus indicate that the effect of fixed adaptation is more pronounced than expected from the f-factor. Since the f-factor values were used in Paper I, the results from this paper are likely underestimated. If Paper I were to be recalculated using the results from Paper IV, the differences between the variable adaptation model and the two models for fixed adaptation (logarithmic average and linear average) would, in most cases, be more pronounced. P_{max} would not be affected since the peak of the Gaussian reaches unity even though the width is narrower. However, due to the narrower width of the Gaussian curve that describes the effect of fixed adaptation, P_{min} would be clearly reduced. An increasing luminance range would thus increase the contrast more slowly at the extreme luminance levels than in the mid-gray area. Also, the peak for P_{min} would occur at a lower luminance range than predicted when using Barten's f-factor.

6.4 Generation and evaluation of a new calibration method (Paper IV)

Image consistency is essential in medical imaging. When a new set of patient images is compared to images taken at a previous occasion, any changes should ideally be caused by changes in the patient itself. There are however several factors other than patient anatomy that can influence image appearance, such as imaging equipment, used technique factors at image acquisition, image processing and image display. As imaging technology advances, old equipment is replaced with new technology and new image processing algorithms are developed, causing the obtained images to be better than before, but also different. It is important to introduce such changes in a controlled manner in order for the radiologists to interpret the new images correctly. Any sudden unannounced changes in image quality will leave the radiologists wondering if the patient has changed or if the change is caused by something else. Although consistency in image appearance is desired over time, it is not always possible to achieve since patients are examined in different hospitals using different equipment. Another large source of possible image inconsistency is displays. Images presented on uncalibrated displays will not be perceived as similar if there is a difference in luminance range, characteristic curve, ambient light or reflection properties. For example, the same digital image will take a different appearance in the radiology reading room than when image processing was optimized in the examination room. An image presented in a bright operating room will be different from when the surgeon first saw it at a review station. Displays next to each other on the same workstation will render images differently. The GSDF is a huge improvement compared to uncalibrated displays and a large degree of image similarity is obtained across a large range of luminance conditions. However, since the GSDF is based on variable adaptation, large luminance ranges are not accounted for. As display technology advances, the luminance ranges are increasing and modern displays are brighter

than ever before. A higher maximum luminance and lower reflection properties make the luminance range increasingly wider, which is beneficial since the overall contrast increases and it is possible to use displays in even brighter locations with a decreased sensitivity for variations in ambient lighting. However, as described above, when a display with a high luminance range is calibrated according to the GSDF, perceived contrast in the darkest and brightest parts of an image will be reduced compared to a display with a lower luminance range.

The $GSDF_{FAC}$ aims at providing an equally distributed contrast over the full luminance range of the display under the assumption of fixed adaptation. For a low luminance range, the viewing conditions may be approximated as variable adaptation and GSDF will thereby provide a satisfactory result. For a large luminance range, perceived contrast decreases for objects far from the adaptation level and in order to calibrate the display according to the $GSDF_{FAC}$, the adaptation level must be estimated. However, the choice of adaptation level has a large influence on the obtained display function, as can be seen in Figure 32 where the $GSDF_{FAC}$ is shown for three adaptation levels. Since an adaptation level corresponding to the logarithmic average of the minimum and the maximum luminance of the display causes the contrast sensitivity function to be almost symmetric over the luminance range of the display, the contrast in the darkest and the brightest regions of the luminance range are equally affected, and the $GSDF_{FAC}$ is thereby relatively similar to the GSDF. An adaptation level far from the logarithmic average will cause the contrast sensitivity function to be asymmetric, and the $GSDF_{FAC}$ thereby increases the contrast for the luminance levels where the contrast sensitivity is low and decreases the contrast for the luminance levels where the contrast sensitivity is high.

The influence of adaptation level also affects the GSDF, although the effect is not as obvious since only one single display function is possible. However, the GSDF is based on variable adaptation and is therefore only valid at one single point – the adaptation level. At all other points on the GSDF, the validity decreases as the distance between object luminance and adaptation luminance increases. Due to the symmetry of the contrast sensitivity function obtained at the logarithmic average of the minimum and maximum luminance of the display, very low contrast sensitivities at the extreme luminance levels are avoided. If, on the other hand, the adaptation level is close to one of the extreme luminance levels, the GSDF provides a high contrast at the adaptation luminance but, due to very low contrast sensitivity values at the other extreme luminance level, the perceived contrast will there be very low. For example, a GSDF-calibrated display that displays an image that is mostly white will require a much higher contrast in the low-luminance region than stated by the GSDF in order for details to be visible.

Although the evaluation study was performed with an adaptation level corresponding to the logarithmic average of the two extreme luminance levels of the display, which is the most beneficial operating point of the GSDF due to the symmetry of the contrast sensitivity function, the $GSDF_{FAC}$ clearly distributed the perceived contrast more evenly than the GSDF. Other choices of adaptation levels would likely increase the difference between the two calibration methods. In the evaluation study (Figure 33), the patterns displayed with the

GSDF$_{FAC}$ still had variations in perceived contrast – objects in dark and bright parts received higher scores than objects in the mid gray. Subjective contrast is difficult to quantify and although the observers had been instructed to rate the contrast between the small circle and the larger surrounding circle, it is possible that they were also influenced by the contrast between the larger circle and the background. Another explanation might be that the pattern size was larger than the sinusoidal pattern used when deriving the new calibration method, and possibly large enough to influence the adaptation level of the observer and thereby pushing it slightly in the direction of the studied luminance level.

Image processing and display characteristics are always closely connected to each other. When optimizing the image processing for an image, the image only looks as intended on the used display, or on a display where the distribution of perceived contrast throughout the luminance range is similar. If uncalibrated displays are used, image processing has to be adjusted manually for every display, a process that is not only time consuming, but also highly dependent on the operator. The loss of contrast in dark and bright regions in GSDF-calibrated high-luminance-range displays can of course be compensated for by increasing the image processing contrast accordingly, but this does not solve the basic problem with inconsistencies between displays with different luminance ranges. Since the GSDF$_{FAC}$ distributes the perceived contrast more evenly throughout the grayscale than the GSDF does, the increase in contrast at one luminance level unavoidably causes a decrease in contrast for another luminance level. The main benefit of using a display calibration that is perceptually uniform across displays with different luminance ranges is that image processing can be left unaltered while maintaining the intended appearance. However, if the display calibration is changed from GSDF to GSDF$_{FAC}$, image processing must be changed accordingly in order not to lose contrast at luminance levels where actual lesions are likely to occur. Once the image processing is optimized for the new calibration method, the need for successive alterations of the image processing is reduced compared to using the GSDF.

6.5 Eye adaptation

Eye adaptation is a complex biological process and the most visible mechanism is the change in pupil size with varying light intensity. However, since the pupil size only varies between 2 and 8 mm in diameter,[128] the pupil only accounts for a minor modulation of the incoming light, and the remaining part of the 14 orders of magnitude under which the HVS operates is regulated by up to eight distinct molecular mechanisms,[129] making eye adaptation difficult to quantify.

When determining the contrast thresholds using observer studies in Papers II-IV, the adaptation luminance level was expressed as the luminance of a homogenous display surface. As described above, the adaptation luminance for clinical images where the luminance distribution is far from homogenous, may be approximated by the average luminance incident on the eyes within a viewing angle of 11 degrees,[64] and the choice of fixation point within the image therefore probably affects the adaptation level as well.[64] It is also possible that the luminance distribution within the viewed area plays a role, although results from Paper II

indicate that the adaptation level is determined mainly by the average luminance. A first approximation of the adaptation level when viewing medical images is to use the average of the maximum and minimum luminance. Since the GSDF is close to exponential, a linear average will probably overestimate the average luminance and a logarithmic average will therefore probably be closer to the correct value. By using the extreme luminance values, only the luminance range is considered and not the shape of the characteristic curve. A better approximation is to calculate the average value of all possible luminance levels. Under the assumption of a flat p-value histogram, i.e. an equal number of all possible p-values, the luminance distribution is determined only by the characteristic curve of the display. For example, consider a GSDF calibrated display with a minimum luminance of 1 cd/m^2 and a maximum luminance of 600 cd/m^2. The linear average of the maximum and minimum luminance is 300.5 cd/m^2, the logarithmic average of the maximum and minimum luminance is 24.5 cd/m^2 and the linear average of all luminance levels is 128.1 cd/m^2. More research is required in order to determine an adaptation level that best represents the viewing of medical images. This adaptation level can then be used in order to calibrate displays according to GSDF$_{FAC}$.

Medical images do not have flat p-value histograms. The histogram varies with imaging technique and examined body part. Also, the distribution of p-values depends on image processing such as window/level, and the distribution of displayed p-values also changes when only a part of the image is displayed by using zoom/pan functions. Since a typical display is considerably larger than the area contributing to the adaptation level, the adaptation level is determined by the local luminance distribution close to the fixation point in the viewed image. If a large homogenous part of an image is studied, the viewing conditions might be described as close to variable adaptation, but most clinical images have the full range of luminance levels present even in relatively small areas, causing the viewing conditions to be closer to fixed adaptation. Based on the assumption that viewing conditions for clinical images are best described by fixed adaptation with an adaptation luminance between the linear and the logarithmic average of the maximum and minimum luminance of the display, the results from Paper I that best describes real viewing conditions are probably somewhere between the logarithmic and the linear average adaptation models.

6.6 Pattern generation (Papers II-IV)

Low-contrast patterns generated on an 8-bit computer display will most likely always be above the detection threshold. Unless the display has a poor luminance range, is placed in a high brightness environment or is poorly calibrated, a difference of one DDL will be visible. In order to generate complex patterns with a low contrast the number of luminance levels must be increased, either by increasing the bit-depth of the display and/or by using sub-pixels. An example of an obtained pattern using 8-bit DDLs with subpixel modulation from Paper II is shown in Figure 34. At the indicated luminance level, and for the indicated pattern contrast, the available luminance levels were too few, resulting in a pattern with only three luminance levels. The shape of the obtained pattern is closer to a bar pattern than a sinusoidal. In Paper

II, patterns with a shape far from the ideal sinusoidal were detected more easily than expected. Campbell et al. concluded that the contrast threshold of a pattern is determined only by the amplitude of the fundamental Fourier component of its wave form.[7, 130] A square wave would thus have a contrast sensitivity $4/\pi$ times greater than a sinusoidal pattern of the same spatial frequency.
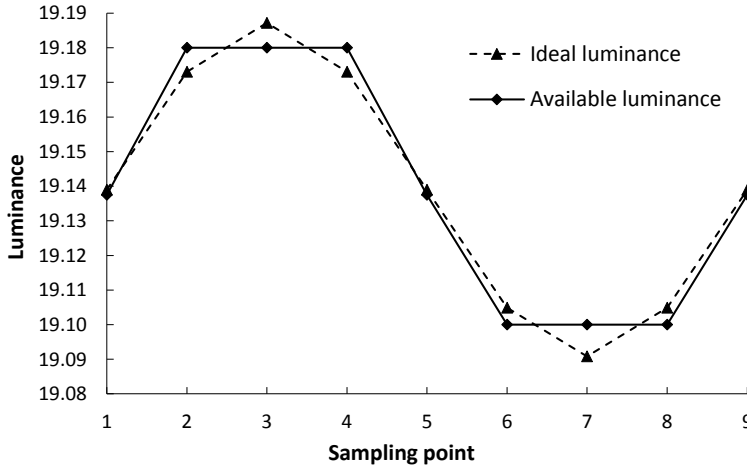

Figure 34: Sinusoidal fits for one combination of luminance level and contrast. The triangles indicate the ideal luminance levels while the squares are the available luminance levels. The pattern is from Paper II and was generated with 8-bit DDLs using subpixels.

In order to obtain more available luminance levels, the luminance resolution has to be increased. One possible alternative is to allow the individual subpixels to deviate by two values, instead of one, from the other two subpixels. However, it is then possible for two neighboring subpixels to differ by up to four values. For example, two neighboring pixels with the subpixel values [75,75,77] and [73,75,75] will be close in luminance, but the luminance difference between individual subpixels will possibly be high enough to be noticeable by an observer with high spatial vision.

10 bit DDLs in combination with subpixel modulation was sufficient in Papers III and IV in order to create sinusoidal patterns that only marginally differed from the ideal. By using subpixel modulation, the number of available luminance levels is increased by a factor of seven, similar to increasing the luminance resolution by three bits (a factor eight). Even though the number of DDLs of most displays is limited to 8 or 10 bits, it is possible that the luminance resolution of the display is higher. When generating sinusoidal patterns, it is the output luminance resolution that is critical, not the number of DDLs. The number of DDLs limits the number of simultaneously displayed grayscale levels, and for a sinusoidal test pattern it would be sufficient for only five grayscale levels to be displayed simultaneously. Theoretically, Papers III and IV could have been performed using 8-bit DDLs together with 10-bit luminance resolution. Subpixel modulation would still have been required since this method increases the luminance resolution. However, by using 10-bit DDLs, the characteristic curve can be measured with

higher resolution and thereby avoiding interpolations. Nevertheless, it would actually be possible to also measure the same amount of luminance levels using 8-bit DDLs by repeating the measurement of the characteristic curve four times with the display LUT values increased by one between each measurement.

6.7 Observer studies (Papers II-IV)

As always in observer studies, the outcome is heavily dependent on the observers. However, as previously mentioned, the conclusions in Papers II-IV are all based on relative contrast sensitivities, where the obtained contrast sensitivities for all luminance conditions are compared to a reference condition. The absolute detection threshold and observer composition are therefore of minor importance. The average age and age range of the observers are shown in Table 7.

Table 7: Age statistics of the observers in the observer studies.

Study	Paper II	Paper III	Paper IV, Part I	Paper IV, Part III
Number of observers	15	10	11	10
Average age	31	32	36	33
Median age	28	31	33	28
Minimum age	22	27	23	23
Maximum age	59	43	65	63

There was no need for previous experience in viewing images since the tasks did not require any kind of advanced decision making. The task was usually a simple matter of deciding whether or not a pattern was present to the left or to the right. As radiologists were not needed, the process of finding observers was simplified. Co-workers and medical physics students had the, somewhat doubtful, honor of viewing thousands of test patterns. Although the task of deciding whether a pattern was present to the left or to the right might seem like an easy task, it most definitely was not. If it were easy, the outcome would be 100% correctly identified patterns and thus of no use. The patterns had to be close to the detection threshold, which in practice meant that they were experienced as not visible at all. The observers had to base their decisions on the fact that the area on either side was slightly more inhomogeneous than on the other. The main problem in these types of studies is to use test patterns with just the right contrast. If the contrast is too low, the detection rate will be close to 50%, i.e. nothing but chance in a 2AFC study. If the contrast is too high then all patterns will be correctly identified. The useful contrast span for a particular observer is small and on top of that, the contrast sensitivity differs among observers. Some patterns may be easily detected by one observer and almost impossible to see for another. The range of pattern contrast levels could of course be expanded to compensate for this effect, but that would also increase the workload considerably. The observers would have had to put a lot of effort into patterns that are either clearly below or clearly above the detection threshold. Due to the statistical nature of this detection task, there had to be many observations of the same test pattern, and to increase a large study with even more patterns where a large part of the observations would

be meaningless did not seem like the right way to go. A better alternative for future studies would probably be to generate a large number of patterns over a relatively large contrast range and use some variant of the staircase method. In this case, the range of pattern contrasts would adjust to each individual observer, resulting in faster studies.

7 CONCLUSIONS

The main conclusions drawn from the studies presented in this thesis are

- that perceived contrast is possible to quantify on medical displays using vision research close to the detection threshold, even though the available luminance levels are discrete and irregularly distributed across the luminance range;

- that a new calibration method that compensates the GSDF for fixed adaptation distributes the perceived contrast more evenly than the GSDF across a high luminance range.

The specific conclusions drawn from the studies presented in this thesis are

I. that displays with a relatively high luminance range calibrated to the GSDF are far from perceptually uniform and that a further increase in luminance range may even cause a reduction in perceived contrast in the darkest and the brightest parts of an image;

II. that vision research close to the detection threshold using sinusoidal patterns is possible to conduct on medical displays, but a luminance resolution of 11 bits (8 bit DDLs using subpixel modulation) is not enough in order to achieve a sufficient number of test patterns;

III. that a luminance resolution of 13 bits (10 bit DDLs using subpixel modulation) is enough to produce sinusoidal patterns across the entire luminance range, and that luminance originating from outside the display may have an effect on the low-contrast detection ability if the luminance level is higher than the adaptation level;

IV. that the calibration method developed provides a more uniform distribution of perceived contrast across the entire range of a high luminance range display, compared to the GSDF.

8 ACKNOWLEDGEMENTS

Sist men inte minst vill jag även tacka min familj. Min mamma **Bodil** med sambon **Sven-Olof** och min pappa **Göran** med sambon **Gittan** hjälper alltid till när det behövs. Er hjälp med barnpassning har varit fantastisk när jag har kombinerat skrivandet med föräldraledighet. När forskningsarbetet gått trögt har mina underbara barn **Erik, Viktor, Sigrid och Freija** påmint mig om att en tur till stranden eller en stund i sandlådan egentligen är mycket viktigare än att sitta inne och arbeta. Min älskade fru **Hillevi** har även dragit ett tungt lass genom att sköta mycket av det praktiska på hemmaplan, särskilt när jag under en period tillbringade många långa kvällar på jobbet. Men det viktigaste är trots allt ditt stöd och att du alltid finns i min närhet. Tack för att jag får leva med dig. Jag älskar er alla.

9 REFERENCES

1 *NEMA PS3 / ISO 12052: Digital Imaging and Communications in Medicine (DICOM) Standard* (Rosslyn, VA, USA (available free at http://medical.nema.org/), 2014).

2 *NEMA PS3.14 / ISO 12052: Digital Imaging and Communications in Medicine (DICOM) Standard - Grayscale Standard Display Function* (Rosslyn, VA, USA (available free at http://medical.nema.org/), 2014).

3 J.J. DePalma and E.M. Lowry, "Sine-wave response of the visual system. II. Sine-wave and square-wave sensitivity," Journal of the Optical Society of America **52**(3), 328–335 (1962).

4 A.S. Patel, "Spatial resolution by the human visual system. The effect of mean retinal illuminance," Journal of the Optical Society of America **56**(5), 689–694 (1966).

5 J. Robson, "Spatial and temporal contrast sensitivity functions of the visual system," Journal of the Optical Society of America **56**(8), 1141–1142 (1966).

6 F.L. van Nes and M.A. Bouman, "Spatial modulation transfer in the human eye," Journal of the Optical Society of America **57**(3), 401–406 (1967).

7 F.W. Campbell and J.G. Robson, "Application of Fourier analysis to the visibility of gratings," The Journal of Physiology **197**(3), 551–566 (1968).

8 A. Watanabe, T. Mori, S. Nagata, and K. Hiwatashi, "Spatial sine-wave responses of the human visual system," Vision Research **8**(9), 1245–1263 (1968).

9 E. Howell and R. Hess, "The functional area for summation to threshold for sinusoidal gratings," Vision Research **18**(4), 369–374 (1978).

10 C. Carlson, "Sine-wave threshold contrast-sensitivity function: dependence on display size," Rca Review **43**, 675–683 (1982).

11 V. Virsu and J. Rovamo, "Visual resolution, contrast sensitivity, and the cortical magnification factor," Experimental Brain Research **37**(3), 475–494 (1979).

12 J. Rovamo, R. Franssila, and R. Näsänen, "Contrast sensitivity as a function of spatial frequency, viewing distance and eccentricity with and without spatial noise," Vision Research **32**(4), 631–637 (1992).

13 J. Rovamo, H. Kukkonen, K. Tiippana, and R. Näsänen, "Effects of luminance and exposure time on contrast sensitivity in spatial noise," Vision Research **33**(8), 1123–1129 (1993).

14   J. Rovamo, O. Luntinen, and R. Näsänen, "Modelling the dependence of contrast sensitivity on grating area and spatial frequency," Vision Research **33**(18), 2773–2788 (1993).

15   A. Van Meeteren and J. Vos, "Resolution and contrast sensitivity at low luminances," Vision Research **12**(5), 825–833 (1972).

16   M. Sachs, J. Nachmias, and J. Robson, "Spatial-frequency channels in human vision," Journal of the Optical Society of America **61**(9), 1176–1186 (1971).

17   P.G.J. Barten, "Physical model for the contrast sensitivity of the human eye," Proceedings of SPIE **1666**, 57–72 (1992).

18   P.G.J. Barten, "Spatio-temporal model for the contrast sensitivity of the human eye and its temporal aspects," Proceedings of SPIE **1913**, 2–14 (1993).

19   E. Muka and B.R. Whiting, "On the human visual system intrascene luminance dynamic range," Proceedings of SPIE **4686**, 169–176 (2002).

20   E. Samei, A. Badano, D. Chakraborty, K. Compton, C. Cornelius, K. Corrigan, M.J. Flynn, B. Hemminger, N. Hangiandreou, J. Johnson, M. Moxley, W. Pavlicek, H. Roehrig, L. Rutz, J. Shepard, R. Uzenoff, J. Wang, and C. Willis (eds.), *Assessment of Display Performance for Medical Imaging Systems, Report of the American Association of Physicists in Medicine (AAPM) Task Group 18* (Madison, WI, USA, 2005).

21   *CIE:15 Technical report Colorimetry*, 3rd ed. (Vienna, 2004).

22   P. Ledda, L.P. Santos, and A. Chalmers, "A local model of eye adaptation for high dynamic range images," Proceedings of the 3rd international conference on Computer graphics, virtual reality, visualisation and interaction in Africa - AFRIGRAPH '04 (2004).

23   G. Wyszecki and W.S. Stiles, *Color science: Concepts and methods, quantitative data and formulae. 1982*, 2nd ed. (John Wiley & Sons, New York, 1982).

24   G. Wyszecki, "Color appearance," in *Handbook of perception and human performance. Sensory processes and perception (vol.1)*, edited by K.R. Boff, L. Kaufman and J.B. Thomas (John Wiley & Sons, New York, 1986).

25   L.E. Arend, "Mesopic lightness, brightness, and brightness contrast," Perception & Psychophysics **54**(4), 469–476 (1993).

26   R.G. Kuehni, "The early development of the Munsell system," Color Research & Application **27**(1), 20–27 (2002).

27   M. Heidelberger, *Nature from Within: Gustav Theodor Fechner and His Psychophysical Worldview* (University of Pittsburgh Press, Pittsburgh, 2004).

28   S. Hecht, "The visual discrimination of intensity and the Weber-Fechner law," The Journal of General Physiology **7**(2), 235–267 (1924).

29   R.E. Fancher and A. Rutherford, *Pioneers of psychology: A history*, 4th ed. (W. W. Norton & Company, New York, 2011).

30   R.W.G. Hunt, "Light energy and brightness sensation," Nature **179**(4568), 1026–1026 (1957).

31   H. Pauli, "Proposed extension of the CIE recommendation on 'Uniform color spaces, color difference equations, and metric color terms,'" Journal of the Optical Society of America **66**(8), 866–867 (1976).

32   B. Anderson and J. Winawer, "Image segmentation and lightness perception," Nature **434**(3), 79–83 (2005).

33   M.J. Flynn, J. Kanicki, A. Badano, and W.R. Eyler, "High-fidelity electronic display of digital radiographs," Radiographics **19**(6), 1653–1669 (1999).

34   C.C. Chen, C.Y. Wu, and T.F. Wu, "LED back-light driving system for LCD panels," Twenty-First Annual IEEE Applied Power Electronics Conference and Exposition. 381–385 (2006).

35   E.A. Krupinski, K. Siddiqui, E. Siegel, R. Shrestha, E. Grant, H. Roehrig, and J. Fan, "Influence of 8-bit versus 11-bit digital displays on observer performance and visual search: a multi-center evaluation," Proceedings of SPIE **6515**, 65150L (2007).

36   S. Bender, K. Lederle, C. Weiss, S.O. Schoenberg, and G. Weisser, "8-bit or 11-bit monochrome displays--which image is preferred by the radiologist?," European Radiology **21**(5), 1088–1096 (2011).

37   F.D. Luna, *Introduction to 3D Game Programming with DirectX 11* (Mercury Learning & Information, Dulles, VA, USA, 2012).

38   G. Sellers, R.S. Wright, and N. Haemel, *OpenGL SuperBible: Comprehensive Tutorial and Reference*, 6th ed. (Addison-Wesley Professional, Upper Saddle River, NJ, USA, 2013).

39   A. Ryer, *Light measurement handbook*, 2nd ed. (Technical Publications Dept., International Light, Inc., Newburyport, MA, USA, 1997).

40   P. Mildenberger, M. Eichelberg, and E. Martin, "Introduction to the DICOM standard," European Radiology **12**(4), 920–927 (2002).

41   J. Morishita, K. Dogomori, S. Hatanaka, T. Hiwasa, Y. Nakamura, N. Hashimoto, Y. Higashida, F. Toyofuku, and M. Ohki, "Effect of test patterns on measurement of the luminance of LCD devices by use of a telescopic-type luminance meter," Radiological Physics and Technology **1**(1), 95–99 (2008).

42   T. Shiiba, N. Tanoue, S. Tateoka, M. Maeda, F. Toyofuku, and J. Morishita, "Effects of ambient-light correction in luminance measurements of liquid-crystal display monitors by use of a telescopic-type luminance meter," Radiological Physics and Technology **3**(1), 65–69 (2010).

43   H.R. Blume, A.M.K. Ho, F. Stevens, and P.M. Steven, "Practical aspects of grayscale calibration of display systems," Proceedings of SPIE **4323**, 28–41 (2001).

44   T. Gillespy and A.H. Rowberg, "Displaying radiologic images on personal computers," Journal of Digital Imaging **6**(3), 151–163 (1993).

45   K.A. Fetterly, H.R. Blume, M.J. Flynn, and E. Samei, "Introduction to grayscale calibration and related aspects of medical imaging grade liquid crystal displays," Journal of Digital Imaging **21**(2), 193–207 (2008).

46   M. Matsuyama, K. Takahashi, H. Akamine, S. Awamoto, Y. Nakamura, N. Hashimoto, and J. Morishita, "SU-GG-I-50: When should we recalibrate the grayscale standard display function in different ambient lighting conditions?," Medical Physics **37**(6), 3113–3113 (2010).

47   M. Matsuyama, J. Morishita, H. Akamine, S. Tokurei, H. Yabuuchi, Y. Nakamura, N. Hashimoto, and M. Ohki, "Effect of ambient lighting on liquid-crystal displays with different types of surface treatment," Radiological Physics and Technology **6**(1), 121–129 (2013).

48   E. Samei, A. Badano, D. Chakraborty, K. Compton, C. Cornelius, K. Corrigan, M.J. Flynn, B. Hemminger, N. Hangiandreou, J. Johnson, D.M. Moxley-Stevens, W. Pavlicek, H. Roehrig, L. Rutz, J. Shepard, R.A. Uzenoff, J. Wang, C.E. Willis, and T.G. 1. 8. AAPM, "Assessment of display performance for medical imaging systems: executive summary of AAPM TG18 report," Medical Physics **32**(4), 1205–1225 (2005).

49   *DIN V 6868-57: Image quality assurance in diagnostic X-ray departments - Part 57: Acceptance testing for image display devices* (Münich, 2001).

50   N. Yoshikazu (ed.), *JESRA X-0093: Quality Assurance (QA) Guideline for Medical Imaging Display Systems* (Tokyo, Japan, 2005).

51   *IPEM Report 91: Recommended Standards for the Routine Performance Testing of Diagnostic X-Ray Imaging Systems* (2005).

52   *Qualitätssicherungs- Richtlinie (QS-RL)* (2014).

53   N. Perry, M. Broeders, C. de Wolf, S. Törnberg, R. Holland, and L. von Karsa (eds.), *European guidelines for quality assurance in breast cancer screening and diagnosis*, 4th ed. (Luxembourg, 2006).

54   N. Perry, M. Broeders, C. de Wolf, S. Törnberg, R. Holland, and L. von Karsa (eds.), *European guidelines for quality assurance in breast cancer screening and diagnosis - Supplements*, 4th ed. (Luxembourg, 2013).

55   *ACR–AAPM–SIIM Practice parameter for determinants of image quality in digital mammography* (2014).

56   *PAS 1054: Requirements and testing of digital mammographic X-ray equipment* (2005).

57   *IAEA human health series No.17: Quality assurance programme for digital mammography* (Vienna, Austria, 2011).

58   *Guidelines for Quality Assurance in Mammography Screening*, 3rd ed. (Dublin, Ireland, 2008).

59   *Guidelines for quality control testing for digital (CR DR) mammography*, 3rd ed. (Sydney, Australia, 2012).

60   C. Poynton and B. Funt, "Perceptual uniformity in digital image representation and display," Color Research & Application **39**(1), 6–15 (2014).

61   N. Belaïd, Perceptual linearization of soft-copy displays for achromatic images (Eindhoven University of Technology, 1999).

62   S.M. Pizer, "Intensity mappings to linearize display devices," Computer Graphics and Image Processing **17**(3), 262–268 (1981).

63   P. Barten, "Formula for the contrast sensitivity of the human eye," Proceedings of SPIE **5294**, 231–238 (2004).

64   W.L. Makous, "Fourier models and the loci of adaptation," Journal of the Optical Society of America **14**, 2323–2345 (1997).

65   J.G. Rogers and W.L. Carel, *Development of design criteria for sensor displays. Report HAC Ref. No. C6619* (Hughes Airport Company, Culver City, CA, 1973).

66   S. Hecht and Y. Hsia, "Relation between visual acuity and illumination," The Journal of General Physiology **11**(3), 255–281 (1928).

67   R.M. Glantz, "Visual adaptation: a case of nonlinear summation," Vision Research **12**(1), 103–109 (1972).

68   D.A. Baylor and M.G. Fuortes, "Electrical responses of single cones in the retina of the turtle," The Journal of Physiology **207**(1), 77–92 (1970).

69 R.A. Normann and I. Perlman, "The effects of background illumination on the photoresponses of red and green cones," The Journal of Physiology **286**, 491–507 (1979).

70 B. Farell and D. Pelli, "Psychophysical methods, or how to measure a threshold and why," in *Vision research: A practical guide to laboratory methods*, edited by C. R.H.S. and R. J.G. (Oxford University Press, New York, 1999).

71 G.A. Gescheider, *Psychophysics: The Fundamentals*, 3rd ed. (Lawrence Erlbaum Associates, Hillsdale, NJ, 1997).

72 F.A. Wichmann and N.J. Hill, "The psychometric function: I. Fitting, sampling, and goodness of fit," Perception & Psychophysics **63**(8), 1293–1313 (2001).

73 B. Treutwein, "Adaptive psychophysical procedures," Vision Research **35**(17), 2503–2522 (1995).

74 A.E. Burgess, "Comparison of receiver operating characteristic and forced choice observer performance measurement methods," Medical Physics **22**, 643–655 (1995).

75 H. Strasburger, "Converting between measures of slope of the psychometric function," Perception & Psychophysics **63**(8), 1348–1355 (2001).

76 P.B. Elliott, *Signal detection and recognition by human observers* (Wiley, New York, 1964).

77 M. Båth and L.G. Månsson, "Visual grading characteristics (VGC) analysis: a non-parametric rank-invariant statistical method for image quality evaluation," British Journal of Radiology **80**(951), 169–176 (2007).

78 M. Båth, "Evaluating imaging systems: practical applications," Radiation Protection Dosimetry **139**(1-3), 26–36 (2010).

79 A. Carlander, J. Hansson, J. Söderberg, K. Steneryd, and M. Båth, "Clinical evaluation of a dual-side readout technique computed radiography system in chest radiography of premature neonates," Acta Radiologica **49**(4), 468–474 (2008).

80 A. Carlander, J. Hansson, J. Söderberg, K. Steneryd, and M. Båth, "The effect of radiation dose reduction on clinical image quality in chest radiography of premature neonates using a dual-side readout technique computed radiography system," Radiation Protection Dosimetry **139**(1-3), 275–280 (2010).

81 S. Zachrisson, J. Hansson, A. Cederblad, K. Geterud, and M. Båth, "Optimisation of tube voltage for conventional urography using a Gd2O2S:Tb flat panel detector," Radiation Protection Dosimetry **139**(1-3), 86–91 (2010).

82   L. Larsson, M. Båth, E.L. Engman, and L.G. Månsson, "Harmonisation of the appearance of digital radiographs from different vendors by means of common external image processing," Radiation Protection Dosimetry **139**(1-3), 92–97 (2010).

83   J.M. Goo, J.Y. Choi, J.G. Im, H.J. Lee, M.J. Chung, D. Han, S.H. Park, J.H. Kim, and S.H. Nam, "Effect of monitor luminance and ambient light on observer performance in soft-copy reading of digital chest radiographs," Radiology **232**(3), 762–766 (2004).

84   C. Wade and P.C. Brennan, "Assessment of monitor conditions for the display of radiological diagnostic images and ambient lighting," British Journal of Radiology **77**(918), 465–471 (2004).

85   J.E. Sheedy, R. Smith, and J. Hayes, "Visual effects of the luminance surrounding a computer display," Ergonomics **48**(9), 1114–1128 (2005).

86   M.F. McEntee, P. Brennan, M. Evanoff, P. Phillips, W.T. O`Connor, and D. Manning, "Optimum ambient lighting conditions for the viewing of softcopy radiological images," Proceedings of SPIE **6146**, 61460W (2006).

87   A.S. Chawla and E. Samei, "Ambient illumination revisited: A new adaptation-based approach for optimizing medical imaging reading environments," Medical Physics **34**(1), 81–90 (2007).

88   P.C. Brennan, M. McEntee, M. Evanoff, P. Phillips, W.T. O'Connor, and D.J. Manning, "Ambient lighting: effect of illumination on soft-copy viewing of radiographs of the wrist," American Journal of Roentgenology **188**(2), W177–W180 (2007).

89   M.F. McEntee, J. Ryan, M.G. Evanoff, A. Keeling, D. Chakraborty, D. Manning, and P.C. Brennan, "Ambient lighting: setting international standards for the viewing of softcopy chest images," Proceedings of SPIE **6515**, 65150M (2007).

90   B.J. Pollard, A.S. Chawla, D.M. Delong, N. Hashimoto, and E. Samei, "Object detectability at increased ambient lighting conditions," Medical Physics **35**(6), 2204–2213 (2008).

91   N.M. O'Connell, R.J. Toomey, M. McEntee, J. Ryan, J. Stowe, A. Adams, and P.C. Brennan, "Optimization of region of interest luminances may enhance radiologists' light adaptation," Academic Radiology **15**(4), 488–493 (2008).

92   M.F. McEntee and B. Martin, "The varying effects of ambient lighting on low contrast detection tasks," Proceedings of SPIE **7627**, 76270N (2010).

93   B. Pollard, E. Samei, A. Chawla, C. Beam, L. Heyneman, L.H. Koweek, S. Martinez-Jimenez, L. Washington, N. Hashimoto, and H.P. McAdams, "The Effects of Ambient Lighting in Chest Radiology Reading Rooms," Journal of Digital Imaging **25**(4), 520–526 (2012).

94   T. Kimpe and T. Tuytschaever, "Increasing the number of gray shades in medical display systems--how much is enough?," Journal of Digital Imaging **20**(4), 422–432 (2007).

95   D.L. Leong, L. Rainford, T.M. Haygood, G.J. Whitman, P.M. Tchou, W.R. Geiser, S. Carkaci, and P.C. Brennan, "Verification of DICOM GSDF in complex backgrounds," Journal of Digital Imaging **25**(5), 662–669 (2012).

96   D.L. Leong, L. Rainford, T.M. Haygood, G.J. Whitman, W.R. Geiser, B.E. Adrada, L. Santiago, and P.C. Brennan, "Trend of contrast detection threshold with and without localization," Journal of Digital Imaging **26**(6), 1099–1106 (2013).

97   J. Fan, H. Roehrig, W. Dallas, and E.A. Krupinski, "Achieving consistent color and grayscale presentation on medial color displays," Proceedings of SPIE **6917**, 69170P (2008).

98   J. Fan, H. Roehrig, W. Dallas, and E. Krupinski, "Medical color displays and their calibration," Proceedings of SPIE **7450**, 74500F (2009).

99   S.F. Hashmi, H. Roehrig, and M. Arthur, "Calibration and verification of DICOM software at the U-of-A," Proceedings of SPIE **8143**, 81430N (2011).

100   H. Roehrig and S.F. Hashmi, "Verification of the QUBYX perfectlum calibration software using a PR-670 spectro radiometer and associated verification facility," Proceedings of SPIE **7966**, 79660W (2011).

101   L.D. Silverstein, S.F. Hashmi, K. Lang, and E.A. Krupinski, "Paradigm for achieving color-reproduction accuracy in LCDs for medical imaging," Journal of the Society for Information Display **20**(1), 53–62 (2012).

102   E.A. Krupinski, L.D. Silverstein, S.F. Hashmi, A.R. Graham, R.S. Weinstein, and H. Roehrig, "Influence of LCD color reproduction accuracy on observer performance using virtual pathology slides," Proceedings of SPIE **8318**, 83180P (2012).

103   A. Badano, C. Revie, A. Casertano, W.C. Cheng, P. Green, T. Kimpe, E. Krupinski, C. Sisson, S. Skrøvseth, D. Treanor, P. Boynton, D. Clunie, M.J. Flynn, T. Heki, S. Hewitt, H. Homma, A. Masia, T. Matsui, B. Nagy, M. Nishibori, J. Penczek, T. Schopf, Y. Yagi, and H. Yokoi, "Consistency and standardization of color in medical imaging: a consensus report," Journal of Digital Imaging **28**(1), 41–52 (2015).

104   T. Kimpe, J. Rostang, A. Avanaki, K. Espig, A. Xthona, I. Cocuranu, A. V. Parwani, and L. Pantanowitz, "Does the choice of display system influence perception and visibility of clinically relevant features in digital pathology images?," Proceedings of SPIE **9041**, 904109 (2014).

105   A. Avanaki, K. Espig, T. Kimpe, A. Xthona, C. Marchessoux, J. Rostang, and B. Piepers, "Perceptual uniformity of commonly used color spaces," Proceedings of SPIE **9041**, 90410V (2014).

106   E.A. Krupinski, H. Roehrig, and J. Fan, "Do aging displays impact observer performance and visual search efficiency?," Proceedings of SPIE **6515**, 65150Q (2007).

107   L. Paulsson and J. Sjostrand, "Contrast sensitivity in the presence of a glare light. Theoretical concepts and preliminary clinical studies," Investigative Ophthalmology & Visual Science **19**(4), 401–406 (1980).

108   I. Lie, "Visual detection and resolution as a function of adaptation and glare," Vision Research **21**(12), 1793–1797 (1981).

109   M. Abrahamsson and J. Sjostrand, "Impairment of contrast sensitivity function (CSF) as a measure of disability glare," Investigative Ophthalmology & Visual Science **27**(7), 1131–1136 (1986).

110   E. Colombo, S.A. Comastri, L. Issolio, and R. Echarri, "Pupil Light Reflex Produced by Glare under Mesopic Adaptation," Journal of Light & Visual Environment **31**(2), 70–79 (2007).

111   M. Choi, D. Sharma, F. Zafar, W.C. Cheng, L. Albani, and A. Badano, "Does veiling glare in the human eye hinder detection in high-dynamic-range displays?," Journal of Display Technology **8**(5), 273–282 (2012).

112   M. Choi, J. Wang, L. Albani, and A. Badano, "60.2: Minimizing Veiling Glare in the High-Luminance-Range Visualization of Medical Images," SID Symposium Digest of Technical Papers **43**(1), 816–819 (2012).

113   M. Choi, L. Albani, and A. Badano, "An image-dependent model of veiling glare effects on detection performance in large-luminance-range displays," Proceedings of SPIE **8318**, 831804 (2012).

114   M.J. Murdoch and I.E.J. Heynderickx, "Veiling glare and perceived black in high dynamic range displays," Journal of the Optical Society of America **29**(4), 559–566 (2012).

115   M. Choi, J. Wang, W.C. Cheng, G. Ramponi, L. Albani, and A. Badano, "Effect of veiling glare on detectability in high-dynamic-range medical images," Journal of Display Technology **10**(5), 420–428 (2014).

116   A. Badano, M.J. Flynn, M. Sandrine, and J. Kanicki, "Angular dependence of the luminance and contrast in medical monochrome liquid crystal displays," Medical Physics **30**(10), 2602–2613 (2003).

117   A. Badano, S. Schneider, and E. Samei, "Visual assessment of angular response in medical liquid crystal displays," Journal of Digital Imaging **19**(3), 240–248 (2006).

118   E. Samei and S.L. Wright, "Viewing angle performance of medical liquid crystal displays," Medical Physics **33**(3), 645–654 (2006).

119 T. Kimpe and G. Van Hoey, "Increasing display performance over a wide range of viewing angles by means of improved calibration algorithms," Proceedings of SPIE **6808**, 680810 (2008).

120 S. Hatanaka, J. Morishita, T. Hiwasa, K. Dogomori, F. Toyofuku, M. Ohki, and Y. Higashida, "Comparison of viewing angle and observer performances in different types of liquid-crystal display monitors," Radiological Physics and Technology **2**(2), 166–174 (2009).

121 Y. Ikushima, J. Morishita, H. Akamine, Y. Nakamura, and N. Hashimoto, "Visual assessment method of angular performance in medical liquid-crystal displays by use of the ANG test pattern: effect of ambient illuminance and effectiveness of modified scoring," Radiological Physics and Technology **7**(1), 51–56 (2014).

122 W. Huda, K.M. Ogden, E.M. Scalzetti, D.R. Dance, and E.A. Bertrand, "How do lesion size and random noise affect detection performance in digital mammography?," Academic Radiology **13**(11), 1355–1366 (2006).

123 B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans* (Society for Industrial and Applied Mathematics, Philadelphia, 1982).

124 S. Takarabe, H. Akamine, H. Yabuuchi, Y. Nakamura, N. Hashimoto, and J. Morishita, "SU-E-I-44: Investigation of the contrast ratio required for medical-grade liquid-crystal displays," Medical Physics **38**(6), 3405 (2011).

125 L. Albani, L. De Paepe, S. Bonfiglio, G. Guarnieri, and G. Ramponi, "HDR medical display based on dual layer LCD," 8th International Symposium on Image and Signal Processing and Analysis (ISPA) 511–515 (2013).

126 S.M. Smirnakis, M.J. Berry, D.K. Warland, W. Bialek, and M. Meister, "Adaptation of retinal processing to image contrast and spatial scale," Nature **386**(6), 69–73 (1997).

127 G. Legge, "A power law for contrast discrimination," Vision Research **21**(4), 457–467 (1980).

128 A.B. Watson and J.I. Yellott, "A unified formula for light-adapted pupil size," Journal of Vision **12**(10), 1–16 (2012).

129 E.N. Pugh, S. Nikonov, and T.D. Lamb, "Molecular mechanisms of vertebrate photoreceptor light adaptation," Current Opinion in Neurobiology **9**(4), 410–8 (1999).

130 F.W. Campbell, R.H. Carpenter, and J.Z. Levinson, "Visibility of aperiodic patterns compared with that of sinusoidal gratings," The Journal of Physiology **204**(2), 283–298 (1969).