THE INFLUENCE OF AMBIENT LIGHT ON THE DETECTABILITY OF LOW-CONTRAST LESIONS IN SIMULATED ULTRASOUND IMAGES

A Thesis Presented to
The Faculty of the
Fritz J and Dolores H Russ
College of Engineering and Technology
Ohio University

In Partial Fulfillment
of the Requirement for the Degree
Master of Science
by
Sharlini Sankaran

March 1999.
# TABLE OF CONTENTS

- TABLE OF CONTENTS ..............................................................................................................i
- LIST OF TABLES ................................................................................................................ii
- LIST OF FIGURES .............................................................................................................. iii
- **1. INTRODUCTION** ...........................................................................................................1
- **2. BACKGROUND** ...........................................................................................................6
  - 2.1 ROC ANALYSIS ...........................................................................................................7
  - 2.2 STATISTICS OF SPECKLE ..........................................................................................16
  - 2.3 RELATIONSHIP BETWEEN SNR AND ROC PARAMETERS ......................................25
  - 2.4 INFLUENCE OF AMBIENT LIGHT ..............................................................................26
- **3. MATERIALS AND METHODS** .....................................................................................33
  - 3.1 IMAGE GENERATION AND DISPLAY .......................................................................33
  - 3.2 VISUAL ANGLE AND LUMINANCE MEASUREMENTS ............................................36
  - 3.3 SUBJECTS AND SESSIONS .......................................................................................39
  - 3.4 SCORING ..................................................................................................................41
  - 3.5 ANALYSIS ................................................................................................................43
- **4. RESULTS AND ANALYSIS** ......................................................................................46
  - 4.1 RESULTS OF ANROC ...............................................................................................46
  - 4.2 CALCULATIONS ..........................................................................................................50
  - 4.3 DISCUSSION OF RESULTS .........................................................................................55
- **5. CONCLUSION** ............................................................................................................58
- REFERENCES ....................................................................................................................63
LIST OF TABLES

Table 1: Measured light levels are shown for the three ambient light conditions: bright, dim, and dark, with three uniform gray scale intensities (0, 64, and 127) displayed on the screen. Values are also shown for measurements taken with the monitor off and with a sample simulated image displayed on the screen. 37

Table 2: Subjects’ responses were scored based on whether or not the simulated image contained a lesion in the center. Points were added to the subjects’ score if they guessed correctly, and subtracted from their score in the same manner if they guessed incorrectly. 42

Table 3: Average, Variance and Standard Deviation values for ROC slopes in the positive-contrast case for all 10 readers, in the three ambient lighting conditions: bright, dim, and dark. 47

Table 4: Average, Variance and Standard Deviation values for ROC slopes in the negative-contrast case for all 10 readers in the three ambient lighting conditions: bright, dim, and dark. 47

Table 5: Positive-contrast accuracy scores ($A_p$) and variances for all 10 readers in all three lighting conditions, bright, dim, and dark. Average values over all readers are also shown. 48

Table 6: Negative-contrast accuracy scores and variances for all readers for the three lighting conditions: bright, dim, and dark. Average values over all readers are also shown. 48

Table 7: Components of standard error for the positive-contrast and negative-contrast cases, for all readers over all lighting conditions. The estimated standard error of the difference, S.E.$_{\text{diff}}$, for both cases is also shown. 52

Table 8: Critical ratios among the three lighting conditions for all 10 readers are shown for the positive-contrast and negative-contrast cases. The subscripts 1, 2, and 3 refer to the bright, dim, and dark lighting conditions respectively. 53

Table 9: Human-ideal efficiency values for the three lighting conditions for all readers are shown for the positive-contrast and negative-contrast cases. The underlying accuracy values used to compute these efficiencies are also shown. 54
LIST OF FIGURES

Figure 1: A sample ROC (receiver-operator characteristic) curve, representing the true-positive (TP) probability against the false-positive (FP) probability for all decision criteria, ranging from strict to lenient. $A_z$ and $d'$ represent the area under the curve and the normal distance from the major diagonal, and are used as bias-free measures of accuracy.

Figure 2: Normal-deviate ROC curve, plotting the normal-deviate true positive values against the normal-deviate false positive values for all decision criteria, ranging from strict to lenient.

Figure 3: Simulated ultrasound image showing speckle, which has a Rayleigh distribution. The image above has a speckle spot size of 8 pixels in the horizontal direction, and 2 pixels in the vertical direction.

Figure 4: Simulated ultrasound image with a positive-contrast lesion. The image has a speckle spot size of 12 and 3 pixels in the horizontal and vertical directions. The lesion has a diameter of 30 pixels, and a contrast factor of +0.4127.

Figure 5: Simulated ultrasound image with a negative-contrast lesion. The image shown has a speckle spot size of 8 and 2 pixels in the horizontal and vertical directions. The lesion has a diameter of 30 pixels, and a contrast factor of -0.2747.

Figure 6: Stimulus response curve showing dark and light adaptation of the visual system. For a given change in intensity $dI$, curve 1 shows a greater change in response, implying better sensitivity, than the light-adapted curve 2. Curve 3, the dark adapted curve, does not elicit a significant change in visual response for the change in intensity shown.

Figure 7: The luminance curve of the monitor used is shown, normalized to the highest gray-scale value (127). The monitor displays a non-linear increase in luminance with respect to the gray-level value of the screen.

Figure 8: Sample session layout for one subject, consisting of three sessions carried out under different lighting conditions: bright, dim, or dark. Each session involves reading two hundred positive-contrast images and two hundred negative-contrast images.

Figure 9: Sample ROC curve formulated from experimental data for one subject reading 200 images in a particular session. The curve shown has an $A_z$ value of 0.79.
Figure 10: Pooled ROC curves for all three lighting conditions, bright, dim, and dark, in the positive-contrast case. The bright and dim curves are relatively close to each other compared to the ROC curve for the dark lighting condition.

Figure 11: Pooled ROC curves for all three lighting conditions, bright, dim, and dark, in the negative-contrast case. The dim and dark curves are relatively close to each other compared to the ROC curve for the bright lighting condition.
1. INTRODUCTION

Lesion detection is a common task in medical imaging, and much research has been conducted on maximizing the reader’s ability to correctly identify the presence/absence of a low-contrast lesion in images produced by computed tomography (CT), single-positron emission computed tomography (SPECT), x-ray, and ultrasound B-mode scans [1]. A lesion refers to an abnormality in the tissue, a region with a different mean density than the rest of the area being imaged. Each imaging modality is best suited for use in specific diagnostic applications, for example, x-ray imaging is used primarily to image areas of high density such as bone. Ultrasound B-mode scans are used to image certain types of soft tissue for various tasks, including lesion detection. Generally, ultrasound imaging is used for differentiation of solid from cystic, or water-filled, structures [1].

Other than the widely known application of ultrasound to scan pregnant uteruses for fetal features, ultrasound imaging is used in gynecological scans to investigate suspected ovarian cysts or malignant tumors. Diagnostic ultrasound is also employed to detect low-contrast lesions in soft tissue such as the liver, the gall bladder, the spleen, and the kidney. Also, ultrasound imaging is used to scan breast tissue and the abdomen for the presence of cysts or tumors [2]. The images produced by the ultrasound scanning are usually read by a radiologist or a physician, who characterizes the tissue as normal or abnormal based on whether or not he/she detects a lesion in the area being imaged. Accuracy in diagnosing the presence or absence of the lesions depends on the
detectability of the lesion, which is governed by factors such as lesion size, lesion contrast, and image resolution. The larger the lesion, the higher the contrast, and the better the resolution, the easier the lesion is to detect.

Ultrasound images exhibit a granular structure known as speckle, which is caused by the scattering of sound by particles in the tissue being imaged [3]. The characteristics of speckle in a particular image are dependent on the characteristics of the transducer used for imaging. A transducer with a smaller resolution will result in a smaller speckle spot size, and vice versa. The presence of speckle in ultrasound images complicates the detection task compared to a monochrome-background detection task, such as a text legibility experiment. When human observers are involved in reading the images, as is usually the case in clinical applications, factors other than the three mentioned above (lesion size, contrast, and resolution) come into play in determining lesion detectability. These factors include a priori knowledge of the case being viewed, the physiological state of the reader, and the environment in which the images are viewed [4]. This study focuses on one particular aspect of the viewing environment: the effect of ambient, or background, light level in the room used for viewing ultrasound images as a factor in determining lesion detectability.

Ambient light is thought to influence lesion detectability in at least three ways: by contributing to glare, being responsible for diffuse and specular reflections, and causing changes in visual sensitivity due to adaptation of the retinal photoreceptors [5]. In clinical settings, the current practice is to view ultrasound images, whether displayed
on film or on a computer screen, with as little ambient light in the room as possible. However, the Illuminating Engineering Society of North America recommends that the ratio of ambient light to light emanating from a visual display terminal be maintained between 1:3 and 3:1 for most office tasks [6]. Other researchers [7,8] also recommend that the lighting level in the room being used for viewing be maintained at approximately the same as the level of light emanated from the visual display unit.

Previous studies addressing this issue [5, 7, 9] have been carried out in various imaging modalities. Alter et al [9] studied the effect of ambient light on the detection of low-contrast targets in a radiograph (x-ray image). They found that lesion detectability was significantly higher if the images were read with the room lights turned off. Baxter et al. [7] studied simulated test patches on both uniform and non-uniform backgrounds, and concluded that all extraneous light (other than that emanated from the image itself) had a detrimental effect on lesion detectability. Rogers et al. [5] conducted a similar study using two test squares on a uniform background. They found opposite trends for positive-contrast (bright stimulus on a dark background) and negative-contrast (dark stimulus on a lighter background) images. A high level of ambient light led to improved discriminability for positive-contrast images, and reduced discriminability for negative-contrast images.

As far as can be discerned, no studies have been carried out attempting to quantify the effect of ambient light on the detectability of low-contrast lesions in ultrasound images. Existing studies for other imaging modalities [5, 7, 9] have found no
quantified justification for choosing one level of ambient light over another. In light of
this, a study was formulated in an attempt to measure the differences, if any, in human
detection performance resulting from changes in ambient lighting. Specifically, the
research conducted seeks to reject the null hypothesis: that the level of ambient, or
background, light has no significant effect on the detectability of low-contrast lesions in
simulated ultrasound images by human observers.

It should be noted that lesions can be of either higher or lower density than the
surrounding tissue. This means that a lesion can be either positive-contrast (lesion
appears brighter than background), or negative-contrast (lesion appears darker than
background). In other words, the polarity of a lesion can be either positive or negative.
Furthermore, ultrasound images can be presented to the observer as either positive or
negative polarity. Thus, this study is conducted for both positive-contrast and negative-
contrast lesions, and a secondary null hypothesis of this study is that ambient lighting
level has the same effect on both positive-contrast and negative-contrast lesions.

Although clinicians and radiologists are trained to be objective when making
decisions [10], decision criteria vary among readers and from case to case. Factors such
as costs and benefits of the decisions made, the physiological state of the observer, and
the environment for the task all influence the detection of lesions [4]. As such, an
analysis method that is independent of the reader’s bias is needed. Since Receiver-
Operator Characteristic (ROC) analysis, discussed in Chapter 2, provides a bias-free
measure of detectability [11], it has been widely used to quantify accuracy in medical
imaging systems, and is the primary method used to analyze data collected in these experiments.

Chapters 2 through 4 of this paper outline the justification for, execution of, and results of the experiments that were carried out with the purpose of rejecting the null hypothesis: that ambient light level does not significantly influence detectability of lesions in ultrasound images. The next chapter reviews existing literature pertinent to lesion detection and ambient lighting. The underlying first and second-order statistics of speckle in ultrasound, factors influencing lesion detectability, mathematical derivation of the ideal observer, and the basis for ROC analysis are explored. The relationship between ROC parameters and the ideal observer is established. The ways in which the level of ambient light influences lesion detectability, namely glare, diffuse and specular reflections, and changes in the sensitivity of the visual system due to light/dark adaptation, are also discussed. Chapter 3, Materials and Methods, outlines the methods used for image generation and data collection, including subject protocol and the layout of the experimental sessions. Physical parameters such as ambient light levels and the visual angle of the screen, are recorded in Chapter 3. Analysis methods for the data obtained in the experiments are outlined. The results, along with accompanying analyses, are presented in Chapter 4. The statistical significance of the results obtained is discussed, and human-ideal efficiency values are computed. The concluding chapter summarizes the findings of the project, discusses probable explanations for the results obtained, and includes recommendations for future work along these lines.
2. BACKGROUND

As mentioned in Chapter One, ROC analysis provides a bias-free method of quantifying the accuracy of an imaging system. In this chapter, ROC analysis methods are outlined, starting from the possible responses of the observer, to the formulation of the ROC curve and the calculation of the various accuracy indices. ROC curves from experiments run under different conditions provide information on the difference in accuracy between conditions. The statistical significance of a difference in accuracy index values for two or more modalities can also be determined from the information obtained by ROC analysis.

However, in order to understand the significance of ROC analysis in terms of lesion detectability, the underlying statistics of speckle and their relationship to lesion detectability must first be outlined. Thus, the presence of speckle in ultrasound images, and the first- and second-order statistics of speckle, are also discussed in this chapter. From this, the detectability of a low-contrast lesion is defined, including the derivation of the optimal signal-to-noise ratio, corresponding to the ideal observer. The relationship between the performance of the ideal observer and the ROC accuracy index is established in the third section of this chapter. Once this relationship is established, the performance of human observers can be related to ideal observer performance in terms of the human-ideal efficiency. Following the postulation that detectability of low-contrast lesions is related to the level of ambient light, the response of the visual system to changes in ambient light levels is discussed in the final subsection of this chapter.
2.1 ROC (RECEIVER-OPERATOR CHARACTERISTIC) ANALYSIS

With a human observer reading images, certain factors can influence or bias the decision criterion adopted. Factors influencing an observer’s decision include a priori knowledge of the probability of abnormality, and the values and costs associated with the decision made [11]. For example, if a missed lesion in a certain case will probably cause the serious illness or death of a patient, the reader will be more likely to adopt a more lenient criteria in detecting lesions, minimizing missed lesions but increasing the probability of a false positive. On the other hand, if a false positive would cause unnecessary anxiety and medical expenditure, and would correspond to a very rare medical condition, the reader would be more conservative, or biased toward a negative decision, in making decisions. In addition, the physiological state of the reader and the viewing environment also may influence the decision made. As such, the decision criterion of the observer (how “strictly” or “leniently” the images are interpreted) is not fixed for any one case, but instead, depends on the factors enumerated above.

As a result of the interplay of those factors mentioned above, lesion detection by human observers is influenced by two aspects of performance: the capacity to discriminate among the alternative states of the object, or accuracy, and the effect of decision factors on this capacity [11]. In many medical imaging detection tasks, the discrimination task consists of identifying the presence or absence of lesions in a uniform background. There are two types of analysis methods for quantifying such a task: bias-free methods, and methods that are dependent on the decision criterion, or bias, employed
by the observer. However, a true measure of the accuracy of an imaging system is independent of the decision criterion employed by the observer. Bias-free methods are generally derived from two types of detection tasks: the two-alternative forced choice (2AFC) task, and the yes-no (ROC) task. The 2AFC task presents the observer with two images displayed side-by-side, with one image containing an abnormality or lesion. The observer then has only to decide in which of the two images the lesion is located. ROC analysis is derived from one of two tasks: a signal-known exactly (SKE) or a signal not-known exactly (SNKE) task. In a signal-known exactly task, also known as a yes-no task, the observer has prior knowledge of the shape, size, and location of the lesion. The task is then to determine whether or not a displayed image contains a lesion. A signal-not-known task offers the observer no prior knowledge of the appearance of a lesion.

The yes-no task is the basis of signal detection theory and the formulation of the Receiver-Operator Characteristic (ROC) curve [12]. The results from the two-alternative forced choice task can be manipulated to form the ROC curve. As a result of this, the underlying distributions for the background and lesion areas are known for the yes-no task, but in the two-alternative forced choice task, the underlying distributions have to be assumed. While the two-alternative forced choice task is easier for the subject than the yes-no experiment, the commonly encountered clinical tasks are more similar to the signal not known task. Unfortunately, it is more difficult to elicit a well-behaved set of data from a signal-not-known task. After weighing the advantages and disadvantages of both types of tasks, it was decided that the yes-no experiment would provide more
information and be more pertinent to the cause than the 2AFC task, and elicit more consistent responses than the signal-not-known task.

In the SKE and SNKE tasks, there are two types of correct responses, and two types of incorrect responses. The correct responses are either true-positives (TP) or true-negatives (TN), also called hits and correct rejections. A TP means that a lesion was present, and the observer identified it; a TN means that there was no lesion in the image, and the observer indicated that it was absent. The two types of incorrect responses are false-positives (FP), and false-negatives (FN), also termed false alarms and misses. In the first instance, the observer indicates the presence of a lesion when none exists, and in the latter case, the observer does not identify a lesion when the image contains one.

These four decision outcomes form the basis of the ROC analysis method. The analysis is based on the observation that there are only two independent decisions possible. This is due to the fact that:

\[ P(A|a) + P(N|a) = 1.0 \]

and \[ P(A|n) + P(N|n) = 1.0 \]

where \( P(A|a) \) is the probability that the observer guessed “abnormal” given that an abnormality, or lesion, is present (corresponding to a “hit”), \( P(N|a) \) is the probability that the observer guessed “normal” given that an abnormality is present (corresponding to a “miss”), \( P(A|n) \) is the probability that the observer guessed “abnormal” when no lesion is present (corresponding to a false alarm), and \( P(N|n) \) corresponds to the probability of a correct rejection.
Since hits and misses are dependent on each other, as are correct rejections and false alarms, ROC analysis represents discrimination and decision performance by only two probabilities, namely the true-positive probability and the false-positive probability. The ROC graph plots the TP probability on the y-axis and the FP probability on the x-axis. A point on the ROC curve represents one possible decision criterion, and the whole curve represents the locus of possible points, ranging from a very strict decision criterion \((P(FP)=0 \text{ and } P(TP)=0)\) to a very lenient criterion \((P(FP) = 1 \text{ and } P(TP)=1)\). The ROC curve is formed by varying decision criteria and recording the TPs and FPs resulting from each decision criterion. The first point of the ROC curve would be at (0,0), and the last point at (1.0, 1.0). Plotting the rest of the points would yield the ROC curve for the imaging system to be studied, valid for all decision criteria. A sample ROC curve is shown in Figure 1.
Figure 1: A sample ROC (Receiver-Operator Characteristic) curve, representing the True-Positive (TP) probability against the False-Positive (FP) probability for all decision criteria, ranging from strict to lenient. \( A_z \) and \( d' \) represent the area under the curve and the normal distance from the major diagonal, and are used as bias-free measures of accuracy.

One of the most commonly used indices of accuracy in medical imaging applications is the area under the ROC curve, \( A_z \). The area under the ROC curve ranges from 0.5, representing random choice, to 1.0, representing perfect discrimination. For the case where \( A_z = 0.5 \), the points of the ROC curve lie along the major diagonal, and \( P(TP) \) is equal to \( P(FP) \) for each decision criterion adopted. For the case where \( A_z = 1.0 \), the ROC curve lies on the line \( P(TP) = 1.0 \), and \( P(FP) = 0 \) regardless of the decision criterion adopted.

The ROC curve can also be represented on a normal-deviate scale instead of a straight probability scale [11]. For the assumption that the background and the lesion both have identical Gaussian distributions, the normal deviate curve will have a slope of
1.0 and a y-intercept of zero. Such a plot is called a binormal ROC, and a sample is shown in Figure 2. The curve shown has a slope of $s=0.67$.

Figure 2: Normal-Deviate ROC curve, plotting the normal-deviate True Positive values against the normal-deviate False Positive values for all decision criteria, ranging from strict to lenient.

$z(A)$ is the normal-deviate value corresponding to a given value of $A_x$. Plotted on the normal-deviate scale, the normal-deviate value corresponding to $z(TP) = 0$ is the y-intercept of the ROC curve, commonly labeled $\Delta m$. For a given set of slope-intercept values, the quantity $z(A)$ can be calculated using the equation [11]:

$$z(A) = \frac{s(\Delta m)}{\left(1 + s^2\right)^{1/2}}$$  \hspace{1cm} (1)

This equation holds only for the binormal assumption, that is, the background and lesion areas are both assumed to have normal (Gaussian distributions). As discussed in the next section, this assumption holds true for the images used in this experiment. Looking up the corresponding value for $z(A)$ in a table of the cumulative standardized normal
distribution gives \( A_x \). The detectability index, denoted by \( d \) or \( d' \), is equal to \( (\sqrt{2})z(A) \), and represents the distance of the ROC curve from the major diagonal. A greater value of \( d \) corresponds to higher detectability.

As mentioned, the accuracy of the subjects' response is given by the area under the ROC curve, which can range from 0.5 to 1.0. In order to compare the performance of two different modalities (or in our case, two different lighting conditions), two separate ROC curves are formed, and corresponding values of \( A_x \) obtained. However, a straightforward comparison of the \( A_x \) values in and of itself is not a sufficient method for determining how significant a difference exists between the two modalities being compared.

The statistical significance of a difference in accuracy scores is quantified by the Critical Ratio (C.R.), or the ratio of the difference in accuracy between the modalities to the standard error of that difference:

\[
C.R. = \left( \frac{\bar{\theta}_1 - \bar{\theta}_2}{S.E.\text{diff}} \right)
\]  

(2)

where \( \bar{\theta}_1 \) and \( \bar{\theta}_2 \) are accuracy indices averaged over \( n \) cases in each modality, read by \( l \) readers on \( m \) independent occasions. The standard error of the difference, S.E.diff, consists of three parts: variance due to case sampling, variance due to reader sampling, and within reader variance. Assuming \( l, m, \) and \( n \) are the same for each modality being studied, and the three variance components are the same for all modalities, S.E.diff can be represented by:
\[ S.E^{(diff)} = \sqrt{2} \left[ S_c^2 (1 - r_c) + \frac{S_{br}^2}{l} (1 - r_{br}) + \frac{S_{wr}^2}{lm} \right]^{1/2} \]  

where

- \( S_c^2 \) is the variability in \( \theta \) due to case sampling, and is proportional to \( 1/n \),
- \( r_c \) is the fraction of \( S_c^2 \) common to both modalities and represents the correlation between cases across modalities,
- \( S_{br}^2 \) is the variability in \( \theta \) due to reader sampling,
- \( r_{br} \) is the correlation between readers across modalities,

and \( S_{wr}^2 \) is the variability in \( \theta \) due to reader inconsistency.

Since the pure variance terms \( S_c^2 \) and \( S_{br}^2 \) cannot be estimated directly [11], a computational formula is used for the calculation of \( S.E^{(diff)} \) in practical cases, where \( m \) is assumed to be equal to 1.

\[ S.E^{(diff)} = \sqrt{2} \left[ S_{c+wr}^2 (1 - r_{c+wr}) + \frac{S_{br+wr}^2}{l} (1 - r_{br+wr}) - S_{wr}^2 \right]^{1/2} \]

where

- \( S_{c+wr}^2 = S_c^2 + S_{wr}^2 \), the observable variance in \( \theta \) that would be found by having one reader read each of a set of different case samples once,
- \( S_{br+wr}^2 = S_{br}^2 + S_{wr}^2 \), the observable variance in \( \theta \) that would be found by having one case sample read once by each of a set of different readers, and
- \( S_{wr}^2 \) is the observable variance of \( \theta \) that would be found by having more than one reader read one case sample on two or more independent occasions.
$r_{br\cdot wr}$ is the observable correlation between the $\theta$s resulting from each of a set of at least 3 different readers reading the same case sample in the two modalities, and $r_{c\cdot wr}$ is the observable correlation between the $\theta$s resulting from a single reader reading each of a set of at least three different case samples in the two modalities.

Commonly, the 95% confidentiality interval for the accuracy of a modality is given by twice the standard error. Alternatively, once the standard error has been found, the critical ratio can then be computed. The critical ratio scales the obtained difference in terms of the normal-deviate $z$. The probability of the observed difference being obtained just by chance can then be determined by looking up the $z$ value in a table of the cumulative standardized normal distribution [11].

In order to see the smallest significant differences, the study needs to be formulated to utilize the maximum statistical power. According to Swets [11], in general, for a $q$-fold decrease in the standard error, (or a $q$-fold increase in the critical ratio), a difference $1/q$ as large can be detected for the same level of power. Statistical power is defined as the probability of correctly rejecting the null hypothesis, or the probability of finding a significant difference among the modalities. Power is dependent on the number of cases being read, case rereading, and matching. In general, the standard error is said to be the smallest if one uses a large number of readers, and only one reading of each sample per reader. However, re-reading is necessary for practical estimations of $S_{wr}$, and to this end, Swets recommends that at least some readers read each case more than once.
The minimum recommendations for these parameters provided a basis for the values chosen in this experiment, as explained in Chapter 3.

2.2 STATISTICS OF SPECKLE

In order to apply the methods of ROC analysis to B-mode scans, the statistical properties of the images must first be understood. Ultrasound images exhibit speckle, a grainy texture, which is caused by interference of sound wavelets scattered by particles in tissue. An image exhibiting speckle is shown in Figure 3.

![Simulated Ultrasound Image showing speckle](image)

Figure 3: Simulated Ultrasound Image showing speckle, which has a Rayleigh distribution. The image above has a speckle spot size of 8 pixels in the horizontal direction, and 2 pixels in the vertical direction.

A lesion in an ultrasound image is an area of different mean density than the background. Lesions in images can be either positive-contrast, with the lesion having a higher mean density than the background, or negative-contrast, where the mean density of the lesion is lower than that of the background. Simulated images containing lesions in the center are shown below, in Figure 4 and 5.
The detectability of lesions in ultrasound images is derived from the statistics of speckle in the images. As such, the first and second-order statistics of speckle need to be examined. The first-order statistics of speckle are Rayleigh, and will be described later in the chapter. To quantify the second-order statistics of speckle in an image, certain assumptions are made: if the number of scatterers within a single resolution cell of the
transducer is large, the scattered waves have uniformly distributed phase, and the
scatterers are randomly located within the resolution cell, the second order statistics of the
speckle depend solely on the characteristics of the transducer used. The result of this is
that the speckle pattern of an object scanned twice under exactly the same conditions
remains unchanged [3]. However, the same object scanned under different conditions
yields a different image. For example, if the frequency of scanning is changed, or the
transducer aperture or orientation is changed, a different image with a different speckle
pattern results.

The first-order statistics of speckle consist of the magnitude and intensity
probability distribution functions, and their means and variances. According to
Burckhardt, the first order statistics of speckle in an image are independent of the
transducer characteristics. The second order statistics involve the covariance and the
correlation functions of speckle, and these are dependent on the resolution, which is
related to the transducer aperture.

The ultrasound signal received at a given time is the sum of the echoes from each
of the individual scatterers, and can be described as a sum of real and imaginary parts [3],
\( V = V_r + V_i \). If each individual reflection is assumed to have random magnitude and
phase, and the number of scatterers is large, the real and imaginary parts of the signal will
be identically distributed, independent, zero-mean Gaussian random variables. The
probability distribution functions of \( V_r \) and \( V_i \) can then be described by:
\[ p(V_r) = p(V_i) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{V_i^2}{2\sigma^2}\right) \]  
(5)

where \( \sigma \) is the standard deviation.

The complex signal \( V \) has the joint distribution of the real and imaginary parts

and can be described as:

\[ p(V_r, V_i) = \frac{1}{2\pi\sigma} \exp\left(-\frac{V_r^2 + V_i^2}{2\sigma^2}\right) \]  
(6)

which is a Rayleigh distribution function [13]. This is conventionally written as:

\[ p(V) = \frac{1}{\psi} \exp\left(-\frac{V^2}{2\psi}\right) \]  
(7)

where \( V \) is the magnitude of the complex signal, and \( 2\psi \) is the mean-square scattering amplitude, or the average backscattered intensity [14].

The mean of a Rayleigh function is given as:

\[ \bar{V} = \sqrt{\frac{\pi}{2\psi}} \]  
(8)

and its variance as:

\[ \sigma^2 = \left(\frac{4 - \frac{\pi}{2}}{2}\right)\psi^2 \]  
(9)

The intensity of the received signal is proportional to the square of the amplitude, and as such, has an exponential distribution function:

\[ p(I) = \frac{1}{2\sigma^2} \exp\left(-\frac{I}{2\sigma^2}\right) \]  
(10)
The detectability of a lesion in an image can be quantified by the contrast-to-speckle ratio (CSR), also known as the signal-to-noise ratio over an area, SNR.<br>
According to Smith [15], a measure of the optimal performance of the observer in detecting low-contrast lesions can be achieved by integrating the average intensity \(<I>\), over the lesion, and comparing it with \(I\) integrated over a background area of the same size. A threshold can then be formed indicating the limit of detectability of the lesion to the ideal observer, and this threshold is the CSR.<br>

For a Rayleigh distribution, the average signal level is \((0.5\pi \psi)^{1/2}\). If the signal strength of each speckle spot is known, the likelihood that the echo strength, or readings of \(\{V_i\}\) has a given level \(\psi_1\) in region 1 is given by the product of \(M\) Rayleigh distributions, since each spot is independent of the others.<br>

\[
L(\{V_i\} / \psi_1) = \prod_{i=1}^{M} (V_i / \psi_1) \exp\left(-\frac{V_i^2}{2\psi_1}\right)
\]

where \(M\) is the number of independent speckle spots. For region 2, the probability of the echo strength having a level \(\psi_2\) is:<br>

\[
L(\{V_i\} / \psi_2) = \prod_{i=1}^{M} (V_i / \psi_2) \exp\left(-\frac{V_i^2}{2\psi_2}\right)
\]

The likelihood ratio for the signals is then:<br>

\[
\gamma_{12} = \frac{L(\{V_i\} / \psi_1)}{L(\{V_i\} / \psi_2)} \prod_{i=1}^{M} (\psi_1 / \psi_2) \exp\left(-\frac{V_i^2}{2}\left(\frac{1}{\psi_1} - \frac{1}{\psi_2}\right)\right)
\]

A function monotonic to the likelihood ratio is used as a "best performance" criterion and is defined as such:
\[ \gamma'_{12} = \log \left( \frac{\psi_1}{\psi_2} \right)^M \gamma''_{12} \]  

Considering the \( \psi \)s as constants, this function reduces to:

\[ \gamma'_{12} = k \sum_{i=1}^{M} V_i^2 \]  

which indicates that the decision criterion is proportional to the average intensity of the region in question. Since \( V \) has a Rayleigh distribution, \( V^2 \), or \( I \), has an exponential distribution.

From equation (15), the mean and variance of the decision criterion are:

\[ \overline{\gamma'_{12}} = M \overline{V_i^2} = 2 M \psi \]  

and \( \sigma^2_{\gamma'_{12}} = 4 M \psi \) for an exponential function.

However, according to the Central Limit Theorem [13], for large values of \( M \) (greater than ten speckle spots), the sum of intensity values tends toward a Gaussian distribution. In most low-contrast medical imaging applications, this assumption holds true for both the lesion areas and the background areas. The performance of an optimal decision rule for differentiating between two Gaussian signals is given by:

\[ SNR_{opt} = \frac{\Delta \mu}{\sigma} \]  

where \( \Delta \mu \) is the difference between the means of the two areas (background and lesion in this case), and \( \sigma \) is the average standard deviation.

This yields:

\[ \Delta \mu = 2 M [\psi_1 - \psi_2] \]  

(18)
and $\sigma = 2M^{1/2}[\psi_1^2 + \psi_2^2]^{1/2}$.

Thus, $SNR_{opt} = M^{1/2} \frac{[\psi_1 - \psi_2]}{[\psi_1^2 + \psi_2^2]^{1/2}}$ \hspace{1cm} (20)

The contrast factor, $C_\psi$, is defined as:

$$C_\psi = \frac{[\psi_1 - \psi_2]}{[\psi_1^2 + \psi_2^2]^{1/2}}$$ \hspace{1cm} (21)

and $M$, the number of independent speckle spots, is:

$$M = \left(\frac{S}{S_c}\right)^{1/2}$$ \hspace{1cm} (22)

where $S$ is the area used for detection, or the area of the lesion, and $S_c$ is the correlation cell size, which is dependent on the speckle spot size. The speckle spot size is measured by the autocovariance of a speckle image. The point-spread function of a transducer is the image produced by a single scatterer. Therefore, for a collection of scatterers, the image consists of the convolution of the point-spread function with the individual scatterers in the tissue. For a B-mode image, the point-spread function is given by a two-dimensional Gaussian function:

$$g(x,z) = \frac{1}{2\pi\sigma_x\sigma_z} \exp\left[-\left(\frac{x^2}{2\sigma_x^2} + \frac{z^2}{2\sigma_z^2}\right)\right]$$ \hspace{1cm} (23)

The complex coherence factor for echo magnitude is the normalized autocovariance function,
For the Gaussian point-spread function given above,

\[ \rho_x(\Delta x, \Delta z) = \frac{C_x(\Delta x, \Delta z)}{C_x(0,0)} \]  

(24)

For the Gaussian point-spread function given above,

\[ \rho_x(\Delta x, \Delta z) = \exp \left[ -\frac{\Delta x^2}{4\sigma_x^2} - \frac{\Delta z^2}{4\sigma_z^2} \right] . \]  

(25)

The correlation cell size is defined as:

\[ S_c = \int_{-\infty}^{\infty} |\rho_x(\Delta x, \Delta z)|^2 \, d\Delta x d\Delta z . \]  

(26)

Smith [15] assumes a circular lesion and an elliptical speckle spot with major and minor axes \( S_{cx} \) and \( S_{cz} \) respectively. \( S_c \) can then be approximated by the area of the ellipse. This definition of \( S_c \) leads to the following equation for \( \text{SNR}_{\text{opt}} \):

\[ \text{SNR}_{\text{opt}} = \left[ \frac{\pi d^2}{4 \pi S_{cx} S_{cz}} \right]^{1/2} \cdot C_w \]

or

\[ \text{SNR}_{\text{opt}} = \frac{d}{\sqrt{S_{cx} S_{cz}}} \cdot C_w \]  

(27)

where \( d \) is the diameter of the lesion, \( S_{cx} \) and \( S_{cz} \) are the speckle spot sizes in the horizontal and vertical directions, and \( C_w \) is the contrast factor, given by Equation (21).

This formula for \( \text{SNR}_{\text{opt}} \) substantiates the claim that detectability of a low-contrast lesion is influenced by contrast, speckle-spot size, and lesion diameter. Since \( \text{SNR}_{\text{opt}} \) is proportional to the diameter \( d \), a larger lesion would mean greater detectability, as one expects. The optimal signal-to-noise ratio is also proportional to \( C_w \), which implies that a lesion with a higher contrast with the background is more easily detectable by the
observer. The detectability of a lesion increases as the speckle spot size decreases, corresponding to an increase in resolution. A smaller speckle spot size implies higher resolution, which would give the observer more detail for a specific area.

Insana’s [16] derivation of the optimal signal-to-noise ratio is similar, with the exception that the correlation cell size is derived by carrying out the integration of equation (26). Performing this integration,

$$S_c = 2\pi \sigma_x \sigma_z$$

(28)

Thus, $M$, the number of independent samples, is:

$$M = \frac{S}{S_c} = \frac{\frac{\pi d^2}{4}}{\frac{2\pi \sigma_x \sigma_z}{8 \sigma_x \sigma_z}} = \frac{d^2}{8 \sigma_x \sigma_z}$$

(29)

The sigmas are related to the speckle spot size by the equations:

$$\sigma_x = \frac{S_{\alpha}}{2\sqrt{\pi}} \quad \text{and} \quad \sigma_z = \frac{S_{\alpha}}{2\sqrt{\pi}}.$$  

(30)

Thus, the equation for $M$ becomes:

$$M = \frac{d^2}{2 \pi S_{\alpha} \sigma_{\alpha}}.$$  

(31)

The equation for $SNR_{opt}$ is then:

$$SNR_{opt} = \frac{\sqrt{\pi}}{2} \cdot \frac{d}{\sqrt{S_{\alpha} \sigma_{\alpha}}} \cdot C_{\psi}.$$  

(32)
2.3 RELATIONSHIP BETWEEN SNR AND ROC PARAMETERS

For a binormal ROC, the detectability index $d'$ is equivalent to the parameter $\Delta m$, which is defined as:

$$\Delta m = \frac{m_s - m_n}{\sigma_n},$$

(33)

where $m_s$ is the mean of the variable with the signal (lesion)

$m_n$ is the mean of the variable without the signal (background)

and $\sigma_n$ is the standard deviation of the variable without the signal $^{11}$.

Referring back to the previous section, for a decision variable of $\gamma_{j2}$ and $M$ speckle spots in the decision region, $m_s$ and $m_n$ can be represented by $2M\psi_s$ and $2M\psi_n$ respectively, and $\sigma_n$ by $2\psi_n\sqrt{M}$.

This implies:

$$\Delta m = \frac{2M\psi_s - 2M\psi_n}{\sqrt{4M\psi_n^2}} = \sqrt{M}\left(\frac{\psi_s - \psi_n}{\psi_n}\right)$$

(34)

The slope of the ROC curve is defined as:

$$s = \frac{\sigma_n}{\sigma_s} = \frac{\sqrt{4M\psi_n^2}}{\sqrt{4M\psi_s^2}} = \frac{\psi_n}{\psi_s},$$

(35)

assuming that the background and lesion areas are both normally distributed. This indicates that the slope of the normal-deviate ROC curve can be expected to be less than one for a positive-contrast (light on dark) lesion, and greater than one for a negative-contrast lesion.
$z(A)$ can be defined by substituting the slope-intercept values into equation (1):

$$z(A) = \sqrt{M} \cdot \frac{\psi_s - \psi_n}{\left(\psi_s^2 + \psi_n^2\right)^{1/2}}$$

or

$$z(A) = \sqrt{M} \cdot C_v^\psi ,$$  \hspace{1cm} (36)

indicating that $z(A)$ is in fact equivalent to the $\text{SNR}_{\text{opt}}$. Therefore, the $A_z$ value (for a yes-no experiment) can be found for the ideal observer by taking the theoretical value of $\text{SNR}_{\text{opt}}$ to a table of the standardized cumulative normal distribution.

$$A_z = \Phi[z(A)]$$  \hspace{1cm} (37)

where $\Phi[]$ refers to the standardized normal probability distribution function.

### 2.4 Influence of Ambient Light

While the performance of the ideal observer can be quantified solely by the ideal signal-to-noise ratio as described in the previous section, the performance of human observers is affected by other factors such as fatigue and viewing environment. Studies [5, 6, 7, 8, 17] show that the level of background, or ambient, light influences an observer's detection threshold for low-contrast lesions. This section discusses the parameters used to measure light, and ways in which the level of light may influence lesion detection.

There are four basic quantities to measure light, corresponding to four stages of light flow: the actual source, the flow of light from the source, the arrival of the light at
the object, and its return from the object [18]. The first of these quantities measures the luminous intensity of a source, in candelas (cd). Historically, one candela is equivalent to the light emitted by a candle. The flow of light from a source, or luminous flux, is measured in lumens (lm), and represents the quantity of light emitted from a source. A point source with a luminous intensity of 1 cd in all directions has a luminous flux of about 12.6 lm [19]. The illuminance of an object, or the amount of light falling on a surface, is measured in lumens per square meter (lux). The pre-metric unit for illuminance is the foot-candle, which is equivalent to 10.76 lux. Finally, the observed brightness, an object’s luminance, is the luminous intensity per unit area, and is measured in candelas per square meter (cd/m²). The luminance of an object is dependent on its reflective properties. Thus, the perceived brightness of an object is its luminance level.

The detectability of a lesion depends, among other things, on the level of luminance of an object. The pre-metric unit of luminance is the foot-Lambert (ft-L), and one foot-Lambert approximately equals 3.4 cd/m² [20].

Light-related factors that may affect the detectability of a lesion include glare, diffuse and specular reflections off the computer screen, and visual adaptation of the retinal photoreceptors [5]. Glare is defined as unwanted light entering the eye from a source placed within the observer’s field of view [21]. The magnitude of sensation of glare is a function of the size, position, and luminance of the extraneous source, the number of sources, and the luminance to which the eyes are adapted to at that time [8]. No actual change in the image occurs due to glare, but the visual system loses sensitivity
towards the target image by trying to adapt toward the extraneous light source. Glare is also a source of visual discomfort, and can be minimized by removing extraneous sources from the field of view of the subject. The consensus among researchers addressing the issue is that glare is debilitating [5, 6, 7, 8, 9, 17, 22]. In view of this, most experiments involving ambient light suggest using indirect light sources, where the light sources are placed outside the observers' field of view.

In addition to glare, another factor affecting detectability is reflection. Reflections can be of two kinds: diffuse or specular. Specular reflection are additions to the image that occur when light emitted by objects or reflected off objects forms images at the front surface of a CRT [5]. Since reflected images are located in front of the displayed images, the visual system alternately focuses on the two sets of images, contributing to possible intermittent blurring of the displayed images. Diffuse reflections are additions to the image involving a fairly uniform increase in luminance, by reflection, at all points across the CRT. The diffuse reflectance characteristic of a particular monitor is based on the percentage of incident light that is reflected off a blank screen. By increasing the luminance of all points, the effective contrast is reduced, thus affecting the detectability of an image. Light reflected from displays is thought to introduce three general categories of problems [21]:

1. reduced contrast of the displayed image
2. competition for the user's attention causes distraction
3. undesirable changes in visual accommodation because the displayed and the reflected images are at different focal distances.

The reduction in contrast caused by light reflected off displays can be quantified as follows. For an image with no incident light and background and lesion illuminance levels $X_i$ and $X_b$ respectively, the contrast is:

$$C_x = \frac{X_b - X_i}{\sqrt{(X_b^2 + X_i^2)}}$$

using equation (22).

With an added ambient light level of $X$, the contrast becomes:

$$C_x = \frac{(X_b + X) - (X_i + X)}{\sqrt{((X_b + X)^2 + (X_i + X)^2)}} = \frac{X_b - X_i}{\sqrt{(X_b^2 + X_i^2 + 2(X^2 + XX_b + XX_i)}}}. \quad (38)$$

However, the effect of the reduced contrast is compensated for by the visual system adapting to the prevailing light condition.

The human visual system can detect a large range of intensities, ranging from $10^{10}$ candelas/m² (the sun at noon: damaging to the eye) to $10^{-6}$ candelas/m². However, the response range of the optic nerve, through which visual information travels, is limited to about 1:100 [23]. As such, a large range of intensities need to be mapped onto a relatively small range of outputs. If the visual response was fixed with respect to the absolute level of illumination incident upon the eye, the system would only be sensitive to large changes in input. To overcome this, the visual system has a mobile response function, which shifts towards the prevailing level of ambient light.
This processes of the visual system adjusting to the prevailing environmental luminance is termed light adaptation or visual adaptation [5]. Visual adaptation occurs in two stages: the instantaneous dilation or contraction of pupil width, and the relatively slow photochemical adaptation of the photoreceptors in the retina. However, the process of retinal adaptation takes less than two minutes for each subject. Studies have shown that the change in visual system sensitivity caused by light adaptation can be attributed largely to chemical changes within the photoreceptor cells themselves [7]. Once the visual system has reached equilibrium, it is said to be in a steady-state adaptation, which corresponds to the shift in the visual operating curve of threshold versus intensity [5]. This shift is called response compression, and enables the eye to detect a much smaller change in illumination than a static response curve would.

At any one time, the light intensities to which the visual system is exposed lie mostly within small ranges [23]. So, it is more efficient to have a flexible mapping of input to output, with the entire operating range of the visual system shifting toward the prevailing ambient light level. The visual system's curve shifts such that a greater range of sensitivity is achieved at an illuminance level close to that of the ambient lighting. Dark-adapted and light-adapted stimulus-response curves for a single photoreceptor cell are shown in Figure 6.
Figure 6: Stimulus response curve showing dark and light adaptation of the visual system. For a given change in intensity $dI$, curve 1 shows a greater change in response, implying better sensitivity, than the light-adapted curve 2. Curve 3, the dark adapted curve, does not elicit a significant change in visual response for the change in intensity shown.

Each of the curves in Figure 6 show a visual system response curve that is adapted toward a different lighting condition. For a small change in luminous intensity, Curve 2 produces a small change in response, implying less sensitivity towards changes in input at that particular operating range. Curve 1 is the visual response curve shifted toward $dI$, and the corresponding change in response is much larger than that of Curve 2, resulting in greater sensitivity to the small change in input.

The visual system is sensitive to contrast, and not to absolute values of illumination, since the stimulus-response curve of the eye’s photoreceptors shifts toward the prevailing ambient light. According to Rogers et al. [5], contrast discrimination works best when the ambient light level and video display terminal have similar
intensities, as the system is then operating. In addition, to minimize undesirable changes due to visual adaptation, the Illuminating Engineering Society of North America [6] recommends that the ratio between the background illuminance and the visual display terminal illuminance be maintained between 1:3 and 3:1. This will ensure that the observer does not endure unnecessary eye strain in attempting to adapt to two very different levels of illuminance. According to Hendee and Wells [8], ergonomic guidelines for lighting a traditional office environment recommend a high level of lighting of about 700 lumens. However, a study was conducted allowing subjects to adjust ambient lighting levels to their comfort, and the results showed median illumination levels of around 125 lumens. In addition, the current practice for viewing ultrasound images in clinical situations is to turn down the background illuminance to the minimum necessary for vision, and view real-time images in the dark.

These two apparently conflicting views on the influence of background illuminance on lesion detection lent further credibility to the necessity of conducting this study. Three ambient lighting levels, differing from each other by at least one order of magnitude, were used as the primary variable in the project. The methods used in formulating the study, setting up the experiments, and collecting data are discussed in the next chapter.
3. MATERIALS AND METHODS

As discussed in Chapter 2, the statistical properties of ultrasound images are well-characterized. This fact was used to produce computer-generated images comparable in appearance and statistical properties to those in clinical applications. This chapter discusses the methods used to formulate the experiments and to analyze the data obtained. Physical measurements of parameters such as the visual angle subtended by the images and the luminance measurements of the different lighting conditions were needed in order to perform the analysis. Subject protocol, encompassing organization and execution of experiments, is also explained. The data collected from these experiments is used to formulate ROC curves, which quantify subjects' accuracy. The chapter concludes with a brief discussion of the analysis done on the ROC curves, the results of which are further explained in Chapter 4.

3.1 IMAGE GENERATION AND DISPLAY

A computer program, ROCGEN, was written to generate these images in a manner similar to that employed by Insana et al. \[16\]. An algorithm that generates random variables (a PASCAL library function) was employed to form two independent, uniformly distributed variables. One of these random variables was then transformed using the Box-Muller transformation \[24\] into a standardized Gaussian variable, with a mean value of zero and a standard deviation of one. This standardized Gaussian was then scaled to have the desired variance, according to the gray-level values for the background area entered by the user. The pixels generated by this process were convolved with a
Gaussian filter in order to smooth the speckle pattern. The generation process was repeated to form a second Gaussian image. Finally, the two Gaussian variables were converted to a Rayleigh distribution. Lesions in images were produced by generating first a speckled background as described above, then a lesion in the center (of a different variance) by the same process.

The program was written such that images numbered 1 through 50 would have a lesion in the center, and images numbered 51 through 100 would not. The image number was later used to keep track of subjects' scores, as explained in the subsection entitled “Scoring”. The speckle spot-size of the images was based on the vertical and horizontal pixel values entered by the user, were related to these by equation (30).

The parameters chosen for the images to be generated were based on typical accuracy scores in medical applications. Based on previous findings [12] that these $A_z$ scores tend to be around 0.83, the difficulty level of the images, controlled by lesion size, contrast, and speckle-spot size, was set by trial and error to yield similar values for $A_z$s in preliminary studies. The diameter of the lesions in both cases was set to be 40 pixels, and the speckle-spot size to be 2 pixels in the vertical direction and 8 horizontally. These values were chosen because ultrasound images typically have poorer resolution in the horizontal direction than in the vertical direction.

Two sets of one hundred images were generated: one set with fifty images containing positive-contrast lesions, and the other consisting of 50 images with negative-contrast lesions. For the positive-contrast case, set B, the background gray-level value,
corresponding to the standard deviation of the underlying normal distribution, was chosen to be 729, and the lesion shade was chosen to be 900. These values were simply reversed for the negative contrast case (set C), i.e. the background shade was 900, and the lesion shade 729. In this way, the absolute values of contrast for the positive-contrast case and the negative-contrast case, as given by equation 22, are the same.

A computer program, TRAINER, was written to introduce subjects to the nature of the experiments to be conducted. Ten images, half of which contained positive-contrast lesions at a higher contrast than the experimental images, were generated for use in this program. A corresponding negative-contrast set of ten images was also generated. These images, generated by ROCGEN, were displayed at the center of a uniform, speckled, background with matching variance. The speckled image and background covered about 2/3 of the screen. The rest of the screen was taken up by a menu bar presenting the subject with possible choices, which are explained in the “Scoring” section of this chapter. Instructions to the subject were given on screen and tasks were performed with a mouse. The images presented to the subject in this program were not scored, and subjects were given feedback as to whether they had guessed correctly after each image in order to aid in the training process.

The images used for data collection during the experimental sessions, sets B and C as described above, were displayed by yet another computer program called RDROC. This program first presented the subject with a set of 20 images to train on, with feedback. The subject was then given the option to train on another set of twenty images,
to a maximum of 5 sets of 20, or stop when he/she felt comfortable enough to start the actual experiment. The initial training images were taken from the 100 images generated by ROCGEN, and displayed in random order. After the training portion was concluded, the program displayed the same 100 images in random order, without feedback. A mandatory one-minute break followed this, after which the 100 images were again displayed, in a different order, without feedback. The subject’s responses were written to a file for easy retrieval during analysis.

3.2 VISUAL ANGLE AND LUMINANCE MEASUREMENTS

Since subjects’ responses to the images presented are dependent on the viewing environment, the physical parameters of this environment were measured. Given that the subjects sat at an average distance of 20 inches from the screen, and the screen measures 8 by 11 inches, the horizontal visual angle subtended by the screen during the experiments was about 22.6 °. The visual angle of each pixel also depends on the distance of the observer from the screen. The images displayed on screen during the experiments were 480 pixels high and 480 pixels wide. For these images, each pixel occupied about 0.002 minutes of arc both horizontally and visually.

Ambient light measurements for all three lighting conditions, bright, dim, and dark, were made with a Pritchard photometer, both with the monitor turned off and the monitor on. The photometer was situated at approximately the same location that the subject would be sitting in during the experiment. With the monitor on, measurements were made with the screen displaying a gray-scale level of 0, 64 (half of maximum), and
127 (the maximum), and finally, with a sample experimental image displayed on the screen. The results are shown in Table 1.

Table 1: Measured light levels are shown for the three ambient light conditions: bright, dim, and dark, with three uniform gray scale intensities (0, 64, and 127) displayed on the screen. Values are also shown for measurements taken with the monitor off and with a sample simulated image displayed on the screen.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Luminance (ft-L)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bright</td>
</tr>
<tr>
<td>Monitor off</td>
<td>1.081</td>
</tr>
<tr>
<td>Gray Scale 0</td>
<td>1.02</td>
</tr>
<tr>
<td>Gray Scale 64</td>
<td>8.29</td>
</tr>
<tr>
<td>Gray Scale 127</td>
<td>14.95</td>
</tr>
<tr>
<td>Sample image</td>
<td>4.01</td>
</tr>
</tbody>
</table>

As can be seen from the table, the three lighting levels differed from each other by at least one order of magnitude. With the sample image on the monitor, lighting levels for the second and third conditions were below the precision of the measurements made. The IESNA recommendation that the ratio between the ambient light level and the display unit light level be kept between 1:3 and 3:1 was fulfilled only by the brightest of the three lighting conditions.

While taking measurements, it was noted that the luminance scale of the monitor is not linear with respect to the gray-level scale. In light of this, luminance measurements were made for all the gray-levels using the Optical photometer. This instrument was chosen because measurements can be made directly off the screen, with the probe attached to a suction cup pressed onto the screen to ensure that no extraneous light is
measured. The luminance curve of the monitor, normalized to the luminance of the highest gray-level (127), is shown in Figure 7.

![Luminance Curve](image)

**Figure 7:** The luminance curve of the monitor used is shown, normalized to the highest gray-scale value (127). The monitor displays a non-linear increase in luminance with respect to the gray-level value of the screen.

The luminance curve above was formed using the measurements taken from the center of the screen, corresponding to where the lesion would be displayed during experiments. While taking measurements, it was noticed that the luminance distribution of the monitor itself is not uniform. Overall, the luminance level of the bottom of the screen was found to be about 11% higher than that of the center of the screen. The top of the screen had about the same luminance levels as the center.

Before beginning the experiments, several rudimentary steps were taken to control the parameters of the viewing environment. Subjects were cautioned not to adjust the contrast knob of the monitor, and to remain in approximately the same position.
throughout the course of the experiment. Light sources were placed outside the subjects’ field of view, but not directly behind the subject, to minimize glare and specular reflection. All windows in the room were covered with black paper so as to block out external light.

3.3 SUBJECTS AND SESSIONS

Ten subjects, six female and four male, participated in the research. All were university students with no prior experience in reading ultrasound images. These subjects were chosen as opposed to trained clinicians because previous studies [25] have shown no significant difference in the ability of subjects with and without radiology training to detect low-contrast signals in structured background noise, and furthermore, all subjects would then have the same level of training in reading ultrasound images.

Subjects received monetary compensation for their participation. This consisted of a flat per hour fee, and an added incentive rate based on their performance, which will be explained in the “Scoring” section of this chapter. To ensure confidentiality and guard against any possible biases, subjects were identified by number only (3 through 12) for data recording and computational purposes. At the beginning of the first session, subjects were required to sign a consent form stating that all procedures had been satisfactorily explained to them. They then participated in an introductory session, which lasted between ten and fifteen minutes, during which TRAINER, the computer program described previously, was run. The introductory session was designed to familiarize the
subject with the appearance of speckle and introduce them to the appearance of low-contrast lesions in an ultrasound image.

Each subject attended three sessions: one for each lighting condition. Each session was divided into two parts: one dealing with only positive-contrast tasks, and the other with only negative-contrast cases. Each of the six parts started with a set of 20 mandatory training images, taken at random from the actual experimental images, as described in the Image Generation and Display portion of this chapter. The purpose of the training images was to allow the subject to reach a stable error rate, and thus minimize the training effect. Also, going through training gave the subjects’ visual systems time to adjust to the prevailing ambient light conditions.

A sample session for one subject, disregarding the training portions, is shown in Figure 8:

**Sample Session**

![Diagram](Image)

**Figure 8:** Sample session layout for one subject, consisting of three sessions carried out under different lighting conditions: bright, dim, or dark. Each session involves reading two hundred positive-contrast images and two hundred negative-contrast images.
A short break was allowed between each set of 100 images. Subjects read 200 positive-contrast (100 images read twice, as two separate sets) and 200 negative-contrast images (also 100 images read twice) at each session for a total of 400 images. The light conditions were presented in a different order for different subjects, with the purpose of assuring that any difference in accuracy scores was not due to learning effects. The positive- and negative-contrast image sets were also alternated, and the images were presented in random order.

During the experiments, the experimenter was in the room to answer any questions but did not otherwise attempt to direct the subjects’ responses one way or another. No verbal communication took place once subjects had started reading experimental images, and nobody was permitted to enter the room once testing had begun. No time restrictions were placed, but after the initial training, it was found that subjects spent an average of 4 to 6 seconds on each image.

3.4 SCORING

Subjects were asked to rate their certainty of the presence/absence of a lesion in a particular image on a confidence scale of 1 through 6, as such:

1: YES!!! - absolutely sure that lesion is present
2: YES  - lesion is probably present
3: YES?  - lesion might be present
4: NO?  - lesion might be absent
5: NO   - lesion is probably absent
These choices were presented to them as a menu bar alongside the images displayed. In the training sessions, the subjects were given feedback as to whether they had guessed correctly. No feedback was given for the actual experimental images.

The responses of the subjects for each category were recorded into a file. If the subject guessed correctly, points were given based on the category chosen, and if the subjects guessed incorrectly, points were subtracted from their total score in the same manner, as shown in Table 2.

Table 2: Subjects' responses were scored based on whether or not the simulated image contained a lesion in the center. Points were added to the subjects' score if they guessed correctly, and subtracted from their score in the same manner if they guessed incorrectly.

<table>
<thead>
<tr>
<th>Lesion Present</th>
<th>Response</th>
<th>Score</th>
<th>Lesion Absent</th>
<th>Response</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>+4</td>
<td></td>
<td>1</td>
<td>-4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>+2</td>
<td></td>
<td>2</td>
<td>-2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>+1</td>
<td></td>
<td>3</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-1</td>
<td></td>
<td>4</td>
<td>+1</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-2</td>
<td></td>
<td>5</td>
<td>+2</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>-4</td>
<td></td>
<td>6</td>
<td>+4</td>
</tr>
</tbody>
</table>

Subjects were encouraged to perform the discrimination task to the best of their ability, since a higher score would mean higher incentive pay. The compensation scheme was formulated such that subjects would receive a penny for every point scored in addition to the fixed per hour rate previously agreed upon. This was designed to discourage students from random guessing, and to use the 6 decision choices wisely.
3.5 ANALYSIS

ROC (Receiver-Operator Characteristic) Analysis was carried out to determine the accuracy of the subjects’ response in each lighting condition. The data recorded during the experiments was fed into a computer program based on that of Swets [11], called ANROC. The ROC curve was formulated using five different decision criteria, ranging from strict to lenient. The first decision criterion considered only answers in category 1 (for images with a lesion) as a true positive, and all the others as false positives. Similarly, the second decision criterion encompasses categories 1 and 2, and so on. These decision criteria are used to form five points on the ROC curve.

The ROC curve plots the probability of correctly identifying a lesion (“true positive” or “TP”) on the vertical axis and the probability of indicating the presence of a lesion when one does not exist (“false positive” or “FP”) on the horizontal axis. The program ANROC computes the x-y coordinates of these five points, as well as the accuracy, or area under the ROC curve resulting from these points. ANROC also calculates slope-intercept values and detectability indices for ROC curves plotted on a normal-deviate scale. A sample ROC curve, based on data taken from one session, is shown in Figure 9. The true-positive and false-positive values given by ANROC were plotted to form the curve. According to the results of the computer program, the Az value corresponding to this particular curve is 0.7909.
Figure 9: Sample ROC curve formulated from experimental data for one subject reading 200 images in a particular session. The curve shown has an $A_z$ value of 0.79.

The variance of the $A_z$ value is also calculated, and this variance is one of the parameters used to calculate the standard error between modalities. The different variance components, between-reader, within reader, and case variance, were calculated, and these values were used to find the standard error between lighting modalities. The critical ratio of the difference between modalities was computed for each case, and this was used to predict the statistical significance of the difference in accuracy scores.

The experimental results were compared to the ideal observer in two ways: theoretical and computational. The theoretical $\text{SNR}_{\text{opt}}$ was calculated using equation (33). To verify that the theoretical ideal corresponded to the computational ideal, a computer program was written to simulate the ideal observer. The images used in the actual experiments were read by this program, which calculates the average pixel value over the
lesion area and the average value over the background area. The ratio of these two values is taken, and a decision made on whether or not the lesion is present, depending on the decision criterion. The program was run for both sets of images: negative and positive-contrast, with varying decision criteria. For example, with a decision criterion of 1, for the positive-contrast case, a ratio greater than 1.0 would result in the system identifying the presence of a lesion. In the negative-contrast case, the converse is true: a ratio less than 1.0 would result in a positive decision. The (TP, FP) pairs resulting from each decision criterion were used to form a ROC curve, and the area under the curve, $A_z$ was computed. Finally, the efficiency of human observers compared to the theoretical ideal was computed for all lighting conditions, for both positive- and negative-contrast images using the equation below [16]:

$$\eta_{HI} = \frac{SNR_{opt_H}}{SNR_{opt_I}},$$

(39)

where the subscripts $H$ and $I$ refer to human and ideal parameters respectively.
4. RESULTS AND ANALYSIS

The data from the experiments was analyzed using ANROC in the manner described in the previous chapter, and a summary of these results is presented in this chapter. In order to compare the difference in accuracy over the three lighting conditions, the statistical significance of the difference in $A_z$ values must first be computed. This is achieved by calculating the critical ratio between each pair of lighting conditions, and between polarities. In order to find the critical ratio, the various components of standard error are first calculated, and the average standard error of the difference in accuracies is estimated. Following this, the critical ratio of the difference between modalities is computed for each case. The last step in the analysis of data consists of comparing the performance of the human observers to the theoretical observer in terms of human-ideal efficiencies.

4.1 RESULTS OF ANROC

Data from both the positive-contrast case and the negative-contrast case was analyzed in the same manner, but kept separate. The $A_z$ and slope-intercept values of the ROC curves were computed both from the averages and the pooled data. Average and variance values of slope for all readers were computed for the positive-contrast and negative-contrast tasks, and are tabulated in Table 3 and 4 respectively.
Table 3: Average, variance and standard deviation values for ROC slopes in the positive-contrast case for 81110 readers, in the three ambient lighting conditions: bright, dim, and dark.

<table>
<thead>
<tr>
<th></th>
<th>bright</th>
<th>dim</th>
<th>dark</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>slope(ave)</td>
<td>0.96126</td>
<td>0.95396</td>
<td>0.94723</td>
<td>0.95415</td>
</tr>
<tr>
<td>variance</td>
<td>0.04071</td>
<td>0.07107</td>
<td>0.06676</td>
<td>0.05951</td>
</tr>
<tr>
<td>stdev</td>
<td>0.20178</td>
<td>0.26659</td>
<td>0.25838</td>
<td>0.24225</td>
</tr>
</tbody>
</table>

Table 4: Average, variance and standard deviation values for ROC slopes in the negative-contrast case for all 10 readers in the three ambient lighting conditions: bright, dim, and dark.

<table>
<thead>
<tr>
<th></th>
<th>bright</th>
<th>dim</th>
<th>dark</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>slope(ave)</td>
<td>0.84592</td>
<td>1.09938</td>
<td>0.90775</td>
<td>0.95102</td>
</tr>
<tr>
<td>variance</td>
<td>0.08561</td>
<td>0.12736</td>
<td>0.02541</td>
<td>0.07946</td>
</tr>
<tr>
<td>stdev</td>
<td>0.29259</td>
<td>0.35687</td>
<td>0.1594</td>
<td>0.26962</td>
</tr>
</tbody>
</table>

The average ROC curve slope values were 0.954 and 0.951 for the positive- and negative-tasks.

A similar analysis was carried out for the accuracy index \( A_z \). The average \( A_z \) values for the bright, dim, and dark lighting conditions were 0.8143, 0.8134, and 0.7991 in the positive-contrast case. The average variance for the \( A_z \) values, corresponding to the case-and-within-reader variance, \( S^2_{cw\text{,}z} \), for this case was 0.001. For the negative-contrast task, the average \( A_z \) values were 0.8893, 0.9008, and 0.8981 respectively, and the average variance 0.0006. The \( A_z \) values and variances for all 10 readers are presented in Table 5 and Table 6 for the positive-contrast and the negative-contrast cases respectively. Both sets of readings (100 images read twice) were used in finding these values.
Table 5: Positive-contrast accuracy scores (Az) and variances for all 10 readers in all three lighting conditions, bright, dim, and dark. Average values over all readers are also shown.

<table>
<thead>
<tr>
<th>Reader</th>
<th>Bright</th>
<th>Dim</th>
<th>Dark</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Az</td>
<td>var Az</td>
<td>Az</td>
</tr>
<tr>
<td>3</td>
<td>0.8247</td>
<td>0.0009</td>
<td>0.7982</td>
</tr>
<tr>
<td>4</td>
<td>0.9019</td>
<td>0.0005</td>
<td>0.8727</td>
</tr>
<tr>
<td>5</td>
<td>0.8459</td>
<td>0.0008</td>
<td>0.8297</td>
</tr>
<tr>
<td>6</td>
<td>0.8043</td>
<td>0.001</td>
<td>0.7731</td>
</tr>
<tr>
<td>7</td>
<td>0.7731</td>
<td>0.0011</td>
<td>0.7734</td>
</tr>
<tr>
<td>8</td>
<td>0.8097</td>
<td>0.001</td>
<td>0.8199</td>
</tr>
<tr>
<td>9</td>
<td>0.8076</td>
<td>0.001</td>
<td>0.8427</td>
</tr>
<tr>
<td>10</td>
<td>0.8413</td>
<td>0.0009</td>
<td>0.7909</td>
</tr>
<tr>
<td>11</td>
<td>0.8354</td>
<td>0.0009</td>
<td>0.8347</td>
</tr>
<tr>
<td>12</td>
<td>0.6995</td>
<td>0.0017</td>
<td>0.7991</td>
</tr>
<tr>
<td>average</td>
<td>0.8143</td>
<td>0.00098</td>
<td>0.81344</td>
</tr>
</tbody>
</table>

Table 6: Negative-contrast accuracy scores and variances for all readers for the three lighting conditions: bright, dim, and dark. Average values over all readers are also shown.

<table>
<thead>
<tr>
<th>Reader</th>
<th>Bright</th>
<th>Dim</th>
<th>Dark</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Az</td>
<td>var Az</td>
<td>Az</td>
</tr>
<tr>
<td>3</td>
<td>0.9271</td>
<td>0.0008</td>
<td>0.8942</td>
</tr>
<tr>
<td>4</td>
<td>0.8892</td>
<td>0.0006</td>
<td>0.9156</td>
</tr>
<tr>
<td>5</td>
<td>0.8845</td>
<td>0.0006</td>
<td>0.9051</td>
</tr>
<tr>
<td>6</td>
<td>0.9163</td>
<td>0.0004</td>
<td>0.9590</td>
</tr>
<tr>
<td>7</td>
<td>0.8559</td>
<td>0.0007</td>
<td>0.8862</td>
</tr>
<tr>
<td>8</td>
<td>0.947</td>
<td>0.0002</td>
<td>0.8915</td>
</tr>
<tr>
<td>9</td>
<td>0.8188</td>
<td>0.0015</td>
<td>0.8684</td>
</tr>
<tr>
<td>10</td>
<td>0.9325</td>
<td>0.0003</td>
<td>0.9175</td>
</tr>
<tr>
<td>11</td>
<td>0.8711</td>
<td>0.0007</td>
<td>0.8750</td>
</tr>
<tr>
<td>12</td>
<td>0.8506</td>
<td>0.0008</td>
<td>0.8956</td>
</tr>
<tr>
<td>average</td>
<td>0.8893</td>
<td>0.00066</td>
<td>0.9008</td>
</tr>
</tbody>
</table>

The data from all readers was pooled, and the ROC curves resulting from the pooled data for all three lighting conditions are shown below. The ROC curves for the
positive-contrast case are shown in Figure 10 and the negative-contrast ROC curves are shown in Figure 11.

Figure 10: Pooled ROC curves for all three lighting conditions, bright, dim, and dark, in the positive-contrast case. The bright and dim curves are relatively close to each other compared to the ROC curve for the dark lighting condition.
Figure 11: Pooled ROC curves for all three lighting conditions, bright, dim, and dark, in the negative-contrast case. The dim and dark curves are relatively close to each other compared to the ROC curve for the bright lighting condition.

Although the pooled curves have a smaller $A_x$ variance than the averaged curves, the general trend of the change in accuracy values caused by different lighting conditions remains the same. As such, the pooled ROC curves provide an adequate visual measure of the relationship between the averaged ROC curves for the three ambient light levels, which were used in calculations.

4.2 CALCULATIONS

The values obtained from the computer analysis were compared with the corresponding theoretical values. Theoretically, the slope of the ROC curve is related to the underlying Rayleigh variances by equation (35). Thus, the theoretical slope of the positive contrast cases is 729/900, or 0.81, and the theoretical slope of the negative-contrast cases is the inverse of that, or 1.23. As will be discussed in the next section, the
theoretical slopes were not in agreement with the observed slopes. In view of this, the accuracy index $A_z$ is a better predictor of the difference between modalities. The significance of this difference is given by the critical ratio between each pair of lighting conditions, which necessitates first estimating the standard error.

The five components of the standard error, given in equation (4), were computed based on the data in Table 5 and Table 6. The case and within-reader variance, $S^2_{c+wr}$, is just the average variance in $A_z$, as given in these tables. The observed correlation when a single reader reads each of a set of at least three case samples, $r_{c-wr}$, was estimated by writing a computer program to separate a case sample of 100 images into 3 subsets. This program was run using the results of three readers, and the correlation in $A_z$ scores was computed.

$S^2_{br+wr}$, the between- and within-reader variance, was calculated for each of the lighting conditions from the data in Table 5 and Table 6 using the formula:

$$S^2_{br+wr} = \frac{1}{10} \sum_{i=1}^{10} (\bar{A}_z - A_z)^2.$$

(40)

The average $S^2_{br+wr}$ value for all three lighting conditions was used, and this value was calculated to be 0.00208 in the positive-contrast case and 0.00088 in the negative-contrast case. $r_{b-wr}$, the correlation observed when ten different readers read the same case sample in two modalities, was calculated using the formula:

$$r_{b-wr} = \frac{\frac{1}{10} \sum_{i=1}^{10} (A_{z1}A_{z2} - \bar{A}_{z1}\bar{A}_{z2})}{\sqrt{\text{var} A_{z1} \text{var} A_{z2}}}$$

(41)
where $A_{z1}$ and $A_{z2}$ refer to the accuracy indices of the two modalities being compared. As with $S^2_{b-wr}$, the correlation coefficients were computed for all three modalities, and the average value was used in determining the standard error.

The within-reader variance, $S_{wr}$, was calculated using the formula:

$$S_{wr}^2 = \frac{1}{10} \sum_{i=1}^{10} \left( \frac{A_{z1i} - A_{z2i}}{2} \right)^2$$

(42)

This variance was found to be 0.0016 and 0.0019 for the positive- and negative-contrast cases. The components of standard error for both cases, along with the estimated $S.E._{diff}$, are shown in Table 7.

Table 7: Components of standard error for the positive-contrast and negative-contrast cases, for all readers over all lighting conditions. The estimated standard error of the difference, $S.E._{diff}$, for both cases is also shown.

<table>
<thead>
<tr>
<th></th>
<th>$S^2_{b+wr}$</th>
<th>$R_{br-wr}$</th>
<th>$S^2_{c+wr}$</th>
<th>$R_{cr+wr}$</th>
<th>$S^2_{wr}$</th>
<th>$S.E.\text{diff}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>positive-contrast</td>
<td>0.00208</td>
<td>0.6675</td>
<td>0.00103</td>
<td>0.1077</td>
<td>0.00159</td>
<td>0.00636</td>
</tr>
<tr>
<td>negative-contrast</td>
<td>0.00088</td>
<td>0.6255</td>
<td>0.0006</td>
<td>0.3805</td>
<td>0.0019</td>
<td>0.00754</td>
</tr>
</tbody>
</table>

The average standard error, over both positive-contrast and negative-contrast cases, is then 0.007. The critical ratios for each pair of lighting conditions: bright and dim, dim and dark, and bright and dark, were then calculated using equation (2). These results are presented below, in Table 8. The subscripts 1, 2, and 3 in the table refer to the bright, dim, and dark lighting conditions.
Table 8: Critical Ratios among the three lighting conditions for all 10 readers are shown for the positive-contrast and negative-contrast cases. The subscripts 1, 2, and 3 refer to the bright, dim, and dark lighting conditions respectively.

<table>
<thead>
<tr>
<th></th>
<th>CR$_{12}$</th>
<th>CR$_{23}$</th>
<th>CR$_{13}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>positive-contrast</td>
<td>0.1377</td>
<td>2.1924</td>
<td>2.3300</td>
</tr>
<tr>
<td>negative-contrast</td>
<td>1.5257</td>
<td>0.3566</td>
<td>1.1691</td>
</tr>
</tbody>
</table>

Taking the critical ratio values to a table of the Q-function, the probabilities of the difference in accuracy scores obtained being purely due to chance were found to be 89%, 2.8%, and 2.0% for the positive-contrast case. For the negative-contrast task, these probabilities were 12%, 78%, and 24% respectively. Between the positive-contrast and negative contrast cases, the critical ratio was computed to be 12.49, using the average values for $A_z$ and $S.E.\text{diff}$. This corresponds to a less than $1^{-12}$ probability that the difference was due to chance alone.

In order to compare the subjects' performance with that of the ideal observer, the theoretical value of $\text{SNR}_{\text{opt}}$ was calculated using equation (32). For a 40-pixel diameter lesion, and a vertical and horizontal speckle spot size of 2 and 8, the number of speckle spots in the lesion, $M$, was found to be 157 using equation (31):

$$M = \frac{40^2}{\frac{2}{\pi} (8 \times 2)} \approx 157$$

The contrast factor, given by equation (21), was calculated to be:

$$C_v = \frac{\psi_n^2 + \psi_s^2}{\sqrt{\psi_n^4 + \psi_s^4}} = 0.1476$$

and thus, $\text{SNR}_{\text{opt}}$, corresponding to $z(A)$, is given as:
\[ z(A) = \sqrt{M \cdot C_\nu} = 1.850 \]

This theoretical SNR\textsubscript{opt} value of 1.850 corresponds to an \( A_z \) of 0.9678 for the ideal observer. A computer program was written to read the images used in the experimental sessions and act as a computational ideal observer, as described in Section 3.5. The values returned by this program were between 0.97 and 0.98 for both positive-contrast and negative-contrast case samples. Since the area under the ROC curve is slightly overestimated in the computer program, it was concluded that the theoretical and computational ideal values are in agreement. The theoretical ideal was used as a measure of observer efficiency. Human-ideal efficiencies were calculated using equation (39) for each of the lighting conditions used. These values are given in the table below.

Table 9: Human-ideal efficiency values for the three lighting conditions for all readers are shown for the positive-contrast and negative-contrast cases. The underlying accuracy values used to compute these efficiencies are also shown.

<table>
<thead>
<tr>
<th></th>
<th>Positive-Contrast</th>
<th>Negativ Contrast</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( A_z )</td>
<td>( z(A) )</td>
<td>efficiency</td>
<td>( A_z )</td>
</tr>
<tr>
<td>Bright</td>
<td>0.8143</td>
<td>0.895</td>
<td>0.234</td>
<td>0.8893</td>
</tr>
<tr>
<td>Dim</td>
<td>0.8134</td>
<td>0.89</td>
<td>0.2314</td>
<td>0.9008</td>
</tr>
<tr>
<td>Dark</td>
<td>0.7991</td>
<td>0.84</td>
<td>0.1866</td>
<td>0.8984</td>
</tr>
</tbody>
</table>

These calculations indicate that the subjects were about twice as efficient in detecting darker lesions on a lighter background, than vice-versa. However, the darker lesions had the same absolute value of contrast as the positive-contrast lesions used in the experiment. The significance of the values obtained is discussed next.
4.3 DISCUSSION OF RESULTS

The results obtained indicate opposite trends in detectability for the positive-contrast and negative-contrast cases. For positive-contrast detection tasks, subject performance was significantly better in the bright and dim lighting conditions compared to that of the darkest ambient lighting condition. No significant difference in accuracy was observed between the bright and dim lighting conditions. The converse was true for the negative-contrast case: subjects tended to perform better in the dark and dim conditions as compared to the darkened room. This suggests that, in order to maximize the possibility of detection of both positive-contrast and negative-contrast lesions, the optimal environment when viewing ultrasound images should be dim ambient light.

However, although the difference in $A_2$ values obtained is statistically significant, the corresponding change in resolution is not apparent in terms of clinical applications. For example, in the bright-contrast case, the maximum difference in $\text{SNR}_{\text{opt}}$, corresponding to a change in resolution, is 6.5%. This translates to the observer being able to detect a lesion that is 6.5% smaller in bright light than in the dark. In the negative-contrast case, the observer is able to detect a 4.9% smaller lesion in the dark than in bright light.

The results obtained in this study are consistent with the conclusions of Rogers et. al. [5], who detected a slight trend towards higher detectability in lower ambient light levels for the case of positive-contrast targets, and the opposite trend for the case of negative-contrast tasks. Rogers' study was done with monochromatic stimuli, using the
just-noticeable difference to analyze data. Since this study found that ROC analysis shows the same trend for low-contrast lesions in ultrasound images, it can be concluded that negative-contrast and positive-contrast stimuli are processed in different ways by the brain, and this causes the level of ambient light to have opposing effects for opposing lesion polarities.

A significant difference was found when lesion polarity was involved, that is, the detectability of negative-contrast lesions was consistently higher than that of positive-contrast lesions. This held true for all lighting conditions, and the average increase in $A_z$ was 30% for the negative-contrast case. Although Insana [15] found no noticeable difference in detectabilities of equivalent positive-contrast and negative-contrast stimuli, Bernecker et al. [22] found that for a text legibility experiment, darker stimuli on a bright background are significantly more legible. Blackwell [17] found that for negative-contrast stimuli, the detection threshold was about 20% lower than that of bright stimuli on a dark background in the case of relatively large stimuli and low ambient brightnesses. Blackwell’s experiments were conducted using simple monochromatic stimuli and backgrounds, and Bernecker’s used text as stimuli on a plain background, but the results from the simulated ultrasound images used in this experiment seem to echo both their findings. The range of ambient light in Blackwell’s experiment was between 0 and 100 foot-Lamberts, and the ambient light conditions used in this study correspond to Blackwell’s lower range of brightnesses. Therefore, the 30% decrease in detection threshold for the negative-contrast case seems to be in agreement with Blackwell’s
findings. The explanation proffered by Bernecker et al. is that the visual adaptation luminance is higher with a bright background screen due to the much larger area of pixels being stimulated. However, for a low-contrast lesion, the difference between the mean background luminance level and the lesion luminance level is only 1.5 cd/m². Although the results obtained by Bernecker et al. are in agreement with those obtained in this experiment, the detection task in a specular image is more complex than that in a monochrome image. Thus, the speculations for one type of may not be directly applicable to the other, especially when cognitive processes are involved.

In terms of efficiency, observers were about twice as efficient on the negative-contrast cases as the positive-contrast ones. Insana's values for human-ideal efficiency were about 0.6 for both the negative-contrast and the positive-contrast cases. The results from the experiments conducted were well below that, especially in the positive-contrast case, when the average efficiency was 21.8%. For the negative-contrast case, the efficiency was 46.4%. Possible reasons for the discrepancy include reader expertise and the choice of detection task used. However, Rackow [25] contends that experienced readers do not produce significantly better results than untrained readers. In terms of the choice of detection task, Insana's study utilized a two-alternative forced choice task. It seems likely that in this type of task, the detectability of lesions would be higher than in the yes-no task, since the observer has only to identify which of the two images presented contains a lesion.
5. CONCLUSION

The project was formulated to investigate the influence of ambient light on the detectability of low-contrast lesions in simulated ultrasound images displayed on a computer screen. Two sets of a hundred images were generated to simulate actual ultrasound B-mode scans. The first set consisted of fifty images with no stimulus, and fifty images with a positive-contrast lesion located in the center of the image. The second set was similar, with the exception that the stimuli were negative-contrast. In generating these images, the speckled background was first generated, followed by a lesion in the center of half the images, 40 pixels in diameter. A random number generator was called for each subject, and this generated three sets of random numbers ranging from 1 to 100.

The subjects participated in an introductory session, followed by three sessions each, conducted under three lighting conditions: bright, dim, or dark. Each session was separated into two parts: one dealing with two hundred positive-contrast images (one set of a hundred, read twice), and the other with two hundred negative-contrast images (also one set of a hundred, read twice). The images were presented in random order to the subjects, but the negative-contrast sessions were kept distinct from the positive-contrast ones. Subjects were told ahead of time which polarity they would be looking for.

The responses were recorded into a file, and this data was then fed into a computer program to perform the ROC (Receiver-Operator Characteristic) analysis. Subjects' accuracy in detecting lesions was determined based on the ROC curves obtained. The variances of the accuracy indices were also obtained from the computer
program. In order to determine the significance of the observed differences in accuracy, the standard error, $S.E_{\text{diff}}$, for both the negative-contrast cases and the positive-contrast cases were estimated. As expected, the $S.E_{\text{diff}}$ values for both cases were about the same. The average standard error value was then used to compute the critical ratio between lighting conditions and between the positive-contrast and negative-contrast cases.

To compare the observed values with the ideal, the theoretical value for the ideal observer, $\text{SNR}_{\text{opt}}$, was calculated. A computer program was also written to predict the performance of the ideal observer to verify that the theoretical value held true for the images used in the experiment. The two values were found to be in agreement with each other. The theoretical value of $\text{SNR}_{\text{opt}}$ was then used as a basis for calculating the performance efficiency of human observers in all lighting modalities for both the positive-contrast and the negative-contrast cases.

Data analysis showed that the detectability of positive-contrast lesions was significantly higher in the bright and dim lighting conditions than in the dark. Conversely, the detectability of negative-contrast lesions was found to be lower in the dark than in bright and dim light. This effect was also observed by Rogers et al. [5], who found a trend towards better detectability for negative-contrast lesions in higher ambient light, and vice-versa. The overall detectability of negative-contrast lesions was found to be much higher than that of positive-contrast lesions. Although this difference was not reported by Insana [16], Blackwell [17] and Bernecker et al. [22] reported that the
detection threshold for negative-contrast stimuli was lower than that of equivalent positive-contrast stimuli, implying better detectability in the negative-contrast case.

The few studies that have been done on the influence of ambient light on low-contrast detectability have produced varying conclusions. Baxter et al. [7] acknowledged that extraneous light from adjacent viewboxes contributed to veiling glare, and concluded that a darkened room corresponded with the highest level of accuracy. Rogers et al.[5], although noticing a trend towards improved detectability with lower ambient light in the positive-contrast case and vice-versa in the negative-contrast case, did not find a significant difference caused by ambient lighting levels. However, the conclusion of this study, that dim light provides the optimal lighting condition to permit better detectability of both positive- and negative-contrast stimuli in the case of simulated ultrasound images, is not contradictory to the findings of other studies conducted for other imaging modalities. Although the differences in accuracy scores are statistically significant, these differences correspond to only a 6.5% change in the resolution perceived by the observer in the positive-contrast case, and a 4.9% change for the negative-contrast task. Clinically, the differences in lesion detectability caused by different ambient lighting levels are not apparent, and therefore, the level of ambient light, although statistically significant, has no appreciable effect on image resolution in a clinical context.

The opposite trends in the change in lesion detectability caused by a change in ambient light seem to point to a difference in the way low-contrast positive and negative stimuli are processed by the brain. This speculation is further reinforced by the fact that
negative-contrast stimuli are an average of 44% more detectable than positive-contrast stimuli of equal contrast factor and size. That the difference is due to cognition and not to a flaw in the image generation process was verified by the fact that the computational observer returned similar accuracy values for the positive-contrast and negative-contrast cases. However, although the subjects were given no information to bias them towards either positive-contrast or negative-contrast lesions, it seems likely that the observed difference is due to the psychological mindset, or bias, of the subject, as speculated by Bernecker et al. [22].

Exactly how and why this difference in cognition occurs has not been quantified yet, to the best of the researcher’s knowledge. This is an area that needs to be explored for further study. Another area of this project that needs to be clarified in the future is the calculation of ROC slopes for the positive- and negative-contrast cases. As noted in the Results section of this paper, the ROC slopes did not conform to the theoretical value, and were noisy. A study involving a larger case sample would reduce the variance of the ROC curve slope. The difference of the slope values from the ideal could then used to provide some insight on the cognitive processing of positive- and negative-contrast images.

This paper documents the first known systematic study of the effect of ambient light on the detectability of low-contrast lesions in ultrasound images. This study is also the first to quantify such differences in terms of ROC analysis and resolution losses for any imaging modality. In summary, the null hypothesis, that ambient light level has no
significant effect on the detectability of low-contrast ultrasound lesions, has been successfully rejected. The secondary null hypothesis, that positive-contrast and negative-contrast lesions are influenced in the same manner, has also been successfully rejected. With the suggestions for future work outlined earlier, this paper lays out the groundwork for further studies involving the effect of ambient light on lesion detectability.
REFERENCES


