Evaluation of the Effect of Display Luminance on the Feature Detection Rates of Masses in Mammograms

Bradley M. Hemminger*, Alan Dillon#, R. Eugene Johnston*,

*Department of Radiology, University of North Carolina,
#Department of Biomedical Engineering, University of North Carolina,
Chapel Hill NC 27599
bmh@rad.unc.edu

ABSTRACT

Purpose To determine the interaction of the luminance range of the display system with the feature detection rate for detecting simulated masses in mammograms.

Methods Simulated masses were embedded in cropped 512x512 portions of mammograms digitized at 50 micron pixels, 12 bits deep. The masses were embedded in one of four quadrants in the image. An observer experiment was conducted where the observer's task was to determine in which quadrant the mass is located. The key variables involved in each trial included the position of the mass, the contrast level of the mass, and the luminance of the display. The contrast of the mass with respect to the background was fixed to one of four selected contrast levels. The digital images were printed to film, and displayed on a mammography lightbox. The display luminance was controlled by the placing neutral density films between the laser printed films of mammographic backgrounds and the lightbox. The resulting luminances examined in this study ranged from a maximum of 10 ftL to 600 ftL. Twenty observers viewed 20 different combinations of the 5 neutral density filters with the 4 contrast levels, for a total of 400 observations per observer, and 8000 observations overall.

Results An ANOVA analysis showed that there was no statistically significant correlation between the luminance range of the display and the feature detection rate of the simulated masses in mammograms. None of the luminance display ranges performed better than any of the others.

Key Words: Image Display, Luminance, Masses, Feature Detection, Display System Characteristics, Mammography, Observer Studies.

SPIE Vol. 3036 • 0277-786X/97/$10.00

2. BACKGROUND AND SIGNIFICANCE

In the past, the medium of film has served as both the storage and the display media for medical imaging. Today, with the advent of digital modalities for most every Radiology examination, and the convenient transmission of digital medical image data via the DICOM communications standard, a decoupling of image storage and image display has occurred. This decoupling is significant, in that images can now be processed prior to their display, and the display of images need not be dependent on limitations of the acquisition and storage systems. As a result, it is important to study the characteristics necessary for medical image display, once the image storage component is separated from the image display component. The specific question addressed in this research is what maximum luminance level is necessary for medical image display.

We chose to evaluate mammography because it has the strictest requirements for luminance range of radiologic medical image display devices. Specifically the ACR recommends 1000ftL luminance lightboxes for the display of analog film-screen mammography films. This requirement is due a number of factors, including the film characteristic curve, limitations inherent in the analog film-screen acquisition techniques, and ambient light of the viewing setting. Now that the image data can be acquired digitally, however, the luminance range of the display device can be determined independently from the acquisition parameters. We would like to determine whether display systems with smaller maximum luminances than the currently proscribed 1000 ftL requirement can perform as well. If they do, then softcopy (video) displays may be satisfactory for mammography image presentation. Additionally, these results should be similar for other, less demanding modalities. This study attempts to determine the effect of the luminance range of display systems on the feature detection rate of masses in mammograms. Masses were chosen because this is similar to many radiology detection tasks (masses in lungs on chest Xray, nodules on chest CT, etc.). The maximum luminances evaluated in the experiment were chosen to match those commercially available for video display systems and mammography lightboxes.

3. MATERIALS AND METHODS

The experimental paradigm used is based on the model we have previously described for evaluating feature detection and contrast enhancement for medical image display.[1,2,3] It allows for the laboratory testing of a range of display parameters (in this case, the luminance range of the display system). Simulated masses were embedded in cropped 512x512 portions of mammograms digitized at 50 micron pixels, 12 bits deep. The masses were embedded in one of four quadrants in the image. An observer experiment was conducted where the observer's task was to determine in which quadrant the mass is located. The key variables involved in each trial included the position of the mass, the contrast level of the mass, and the luminance of the display. The contrast of the mass with respect to the background was fixed to one of four selected contrast levels. The digital images were printed to film, and displayed on a mammography lightbox. The display luminance was controlled by the placing neutral density films between the laser printed films of mammographic backgrounds and the lightbox. The resulting luminances examined in this study ranged from a maximum of 10 ftL to 600 ftL. Twenty observers viewed 20 different combinations of the 5 neutral density filters with the 4 contrast levels, for a total of 400 observations per observer, and 8000 observations overall.

Mammographic Backgrounds

The 80 background images of 512x512 pixels each were taken from clinical mammograms that had been digitized using a Lumiscan digitizer (Lumisys, Inc., Sunnyvale, CA) with a 50

micron sample size and 12 bits of intensity data per sample. The images were selected so as to provide an even distribution of density distributions across density range of breast tissue on clinical mammograms. The mammograms were known to be normal by virtue of 3 years of clinical and mammographic follow-up. They were selected by a radiologist expert in breast imaging from digitized film screen craniocaudal or mediolateral oblique mammograms.

The gray scale values for the mammographic backgrounds are assigned the values recorded by the Lumisys digitizer. The digitizer assigns digital values in the range 495-4095 representing an optical density range of 3.68 - 0.02. The digitizer produces digitized grey values that map one to one with OD values, i.e., the same OD value on film will produce the same grey level when digitized.

Mammographic Mass Stimuli

Mammographic masses were simulated using a locally developed program. A circle of diameter of 90 pixels was generated. When printed on film the mass was 7.2 mm in diameter, and 1' of viewing angle at the average viewing distance of 40cm (about 16"). The circle was gaussian blurred (frequency standard deviation of 0.2) to appear similar to masses presenting on clinical mammograms. Simulated masses were used instead of real features so that we could have precise control over the structure location, and structure to background contrast of the masses. While the simulated masses were not perfectly realistic, our mammographers confirmed that they did possess the same scale and similar spatial characteristics to actual masses seen at mammography.

Contrast

The contrast of the mass to background surround was defined as the luminance ratio ($\Delta L/L$) where ΔL was the luminance of the background surround with the target inserted minus the luminance of the background surround without the target inserted. L is the luminance of the background surround. Several different choices exist for the area under which the mean background surround value could be calculated. Some common choices are depicted in figure 1. While we believe the definition best matched to visual perception would be one that takes into account the structure of the background surround, there are presently no established techniques for this option. We chose for this experiment to use the area just under the inserted target mass feature. We investigated whether choosing a different size area for calculating the mean of the background surround would have effected our calculation of contrast values. Analysis of randomly inserting 1000 target masses into each of the 80 mammographic background images used in the experiment and calculating the resulting mean background surround value, showed that using increasingly larger circles for the background surround area, up to the size of the mammographic background image, did not significantly change the mean digital driving level used for the surround, as compared to the size of the smallest contrast steps used in the experiment. The standard deviation, however, as might be expected, did increase with the larger circles due to the larger inclusion criteria. Thus, using larger diameter circles could possibly reduce sensitivity in measuring the detection rate due to increased variance in calculation of the mean of the background surround. This result supported our decision to use the surround area equal to the area under the target (i.e. a smaller diameter circle).

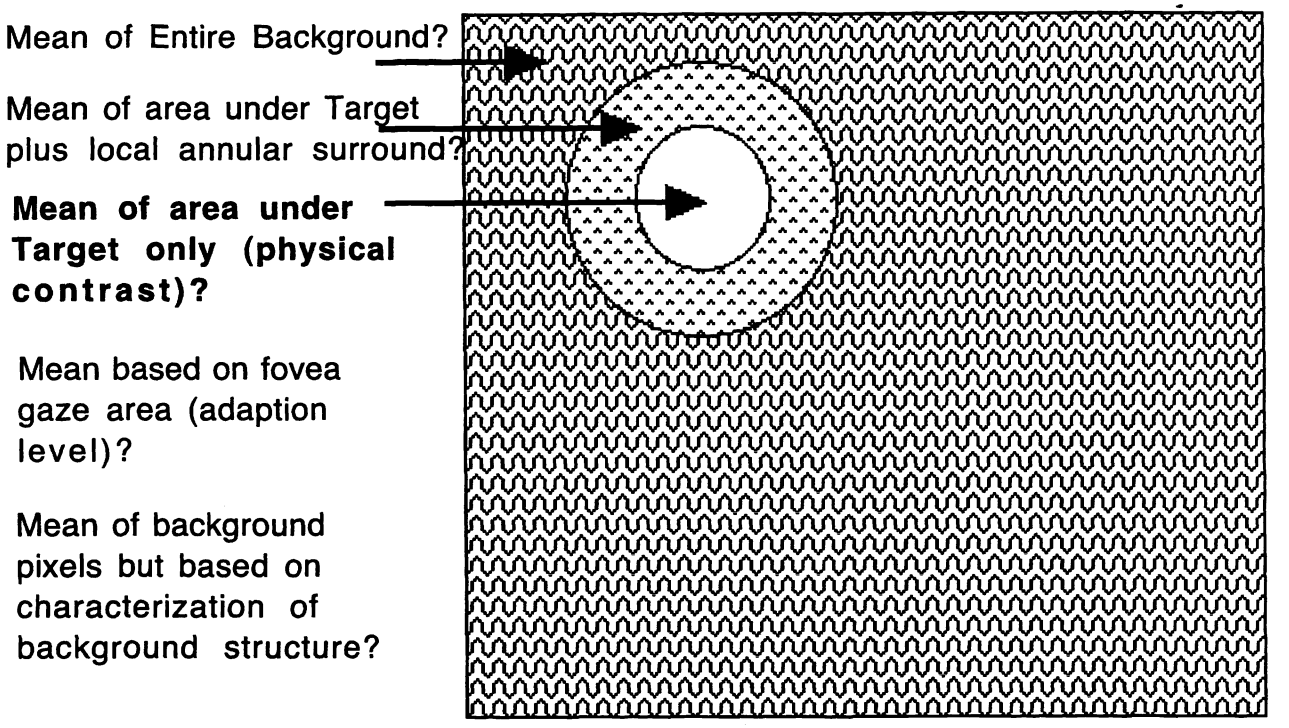

Figure 1. A representation of mammographic background, with white interior circle depicting the mass target insertion area. The surrounding annular ring shows an example of a larger including circle for which the mean could be calculated. Mean values are calculated in digital driving levels of the computer display device, which can be translated into luminance values. Five different methods of calculating the surround value for the contrast definition are given.

The masses were embedded at one of four different contrast levels by pixel-wise addition of the structure and background images. The contrast levels were equally spaced in perceived brightness relative to mean luminance of the background surround area. To calculate the contrast we first calculated the mean DDL in the area of the background where the target would be placed. Then we calculated from this the luminance that would be produced in the experimental setting when this film was placed on our mammography lightbox based on calibration measurements of the printed films on this lightbox. From this surround luminance value, we calculated the luminance value that the target stimulus (mass) should be in order to give us the desired contrast level, and then performed the reverse calculation to determine the DDL values for the mass.

Contrast levels were chosen to provide appropriate calculation of the probit curve. Initial choices of contrast level values were estimated from our prior work. Then we piloted the experiment with 3 observers on a separate set of cropped background images similar to the study ones. Sufficient numbers of trials were used to obtain reasonable estimates of contrast thresholds. The pilot experiments were continued until the chosen contrast levels were appropriately spaced to properly define the probit curve. For this experiment we repeated the pilot three times, each time using 32 or 64 trials repeated with each neutral density film, and with 3 observers. The final contrast levels chosen were contrast values of 4%, 10%, 16% and 22%. This corresponded to percent correct detection rates of 30%, 50%, 80%, 95%, respectively.

Experimental Presentation

The digital images were printed onto standard 14X17 inch single emulsion film (3M HNC Laser Film, 3M, St Paul, MN) using a Lumisys Lumicam film printer (Lumisys Inc, Sunnyvale, CA). Each original 50 micron pixel was printed at a spot size of 80 microns, which produced film images enlarged by a factor of 1.6, approximately 4x4 centimeters in size. Radiologist observers in the previous experiments using this same paradigm reported that they felt this magnification did not make the backgrounds unrealistic.[3] Thirty-two cropped backgrounds were printed per sheet of film. The backgrounds were randomly ordered into an 8X4 grid on each sheet of film. The 8x4 grid was chosen because the mammography lightbox was uniform in luminance only over the central portion of the lightbox, which corresponded to the 32cmx16cm area covered by the 8x4 image grid. The mean luminance of the film test image displayed on the mammography lightbox without any filters was 18 ftL, 26 ftL, and 19 ftL, respectively, for the three film test images in the experiment.

Both the film digitizer and film printer were calibrated, and measurements of the relationship between optical density on film and digital units on the computer were determined in order to generate transfer functions describing the digitizer and film printer.[2] In order to maintain a linear relationship between the optical densities on the original analog film and the digitally printed film, we calculated a standardization function that provided a linear matching between the digitizer and printer transfer function curves, so that, for example, an OD in the 15 percentile on the digitizer curve would map to the OD on the 15 percentile on the film printer curve. This standardization function was applied to the mammographic image backgrounds so that the printed films would maintain a consistent proportional relationship between the original optical densities of the original mammography film and those reproduced on the digitally printed films. The film printer produces films with a constant relationship between an optical density range of 3.62 OD to 0.13 OD, corresponding to a digital input range of 0 to 4095, respectively.

We choose to use neutral density films to control the luminance of the display for consistency, and because of the inherent maximum luminance capability of the lightbox. If we had used a video display system such as a CRT, we would not have be able to reproduce the high luminance levels of lightboxes. Additionally, we would have had to sacrifice contrast resolution (number of grey levels utilized) in order to drive the monitor at reduced luminance ranges in a consistent fashion. Similarly, if we produced films with different luminance ranges, we would have had to decrease the contrast resolution because of only using part of the grey scale range of the display device. It would have additionally caused us to produce multiple films for different luminance values. Since variables in the film printing process could cause differences between films depicting the same contrast levels, producing multiple films for different luminance values might have added a confounding variable to our analysis. For the above reasons, we chose to print a single version of the test images, and use neutral density films instead to modify the luminance of the display. The neutral density films were created using the same Lumicam laser printer used to print the mammographic backgrounds. Uniform flat field films of constant density were produced for the neutral density backgrounds. We also evaluated photographically producing the neutral density filters, but found the variance of OD to be larger for the photographically produced films than for the laser printed neutral density films. We scanned the neutral density films on our Lumisys scanner to check their uniformity. The means and standard deviations of the digitized neutral density films are shown in table 2.

Neutral Density Filter	Mean	STDDEV
10	2230	51
20	2490	27
30	2672	24
200	3539	10
600	3971	6

Table 2. Digital Driving levels of digitized neutral density films. These measurements were used to determine how much noise the neutral density films added to the overall experimental process. Films produced optically had a significantly higher STDDEV.

The resulting luminance levels on the lightbox were exactly controlled by using a voltage regulator inline with the lightbox to adjust the luminance level of the lightbox output up or down. By measuring the luminance output of the lightbox through the neutral density film with a photometer before each experimental session, we could tune the voltage regulator to set the transmitted luminance to exactly the desired output level. This allowed us to consistently maintain the experimental luminance settings for the neutral density films throughout the experiment. The maximum luminance levels in the experiment were chosen to match common commercially available luminance levels for video display systems and mammography lightboxes. The values selected were mammography lightbox (600ftl, see explanation below as to why different from 1000 ftL), high brightness CRT (200 ftL), average workstation monitor (30 ftL), and low end personal computers or hardcopy displays (20 ftL and 10 ftL). The values chosen for the luminance levels are the values as measured by a photometer through one of the five neutral density films combined with either a 0 DDL test film (low end of luminance range) or a 4095 DDL test film (the high end luminance range). These two test films represented the darkest and brightest images possible on the display system with laser printed film. For instance, the high end of the brightest luminance consisted of a clear film undeveloped as the neutral density film, and on top of that a test film produced by the laser printer using the maximum digital driving level of 4095 to produce a uniform flat field. Our mammography lightbox actually produced 790 ftL rather than the expected 1000 ftL. Thus, the highest maximum luminance produced on our mammography display using neutral density films was 600ftL, as measured through the clear neutral density film. Table 3 shows the optical densities of the five neutral density films, the 0 DDL test film, and the 4095 test films, and lists the measured luminances used in this experiment (i.e. what is measured transmitted through the neutral density films and test films on the mammography lightbox).

	0 DDL Test Film OD = 3.62	4095 DDL Test Film OD = 0.13	Range (Max/Min)
ND0 (OD = 1.81)	0.0016 ftL	7.3 ftL	4563
ND1 (OD = 1.56)	0.0031 ftL	14.3 ftL	4613
ND2 (OD = 1.39)	0.0048 ftL	21.8 ftL	4542
ND3 (OD = 0.55)	0.0341 ftL	146.0 ftL	4282
ND4 (OD = 0.13)	0.1120 ftL	457.0 ftL	4080

Table 3. Values show transmitted luminance from lightbox through different neutral density films and min and max test films (DDL 0 and DDL 4095 on laser film printer). Maximum luminance of lightbox without any films is 790 ftL. Rightmost column shows the calculated dynamic range of the display condition (maximum luminance divided by minimum luminance).

The experiment was conducted in our experimental laboratory, which is controlled for light, sound, and other distractions. Room light was 0.043 (day) to 0.0065 (night) lux with no images displayed, and an average of 0.225 lux, 0.376 lux, 0.671 lux, 3.98 lux, 10.63 lux, when experimental films were displayed using the neutral density filters of 10ftL, 20ftL, 30ftL, 200ftL,

600ftL, respectively. Films were displayed on a standard mammography viewbox that was masked to exclude excess light. Observers were free to move, and could use a standard mammography magnifying glass, if desired. Average viewing distance was 16". Observers were dark adapted to the light levels of the experiment for 10 minutes prior to any readings. The neutral density film was placed first on the lightbox. The mammography test film was placed directly on top of the neutral density film.

Observation Task

There were 20 observers for the experiment. They were medical students and graduate students from the University of North Carolina. Performance bonus pay was used to encourage optimal observer performance. Observers selected the quadrant of the image that they thought contained the mass. All images contained a simulated mass, for a 4 Alternative-Forced Choice design. Observers were instructed to make their best guess if they could not tell where the simulated lesion was located in the image.

Prior to beginning the experiment, observers were trained for the task through the use of two films each with 64 images. The first 32 images contained easy (high contrast cases), and the second 32 images contained cases with the contrast matching the levels used in the experiment. An answer sheet overlay provided feedback indicating the correct location of the mass on each image.

The order of presentation of stimuli was counterbalanced so as to eliminate any effects of learning and fatigue. Observers were encouraged to take breaks if needed. Observers were dark adapted to the room upon re-entry. All observers completed the experiment. Each observer examined 80 different images, with the 5 neutral density combinations, for a total 400 images per observer, and a total of 8000 stimuli for all observers for the whole experiment.

Observers took a break at the half way point during the study, and more often if necessary. No time limit was imposed on the observation of the images. Typically, the experiment took 2 hours for each observer, divided into two sessions of 60 minutes each, with a 5 minute break in between sessions.

4. DATA ANALYSIS

Probit models were fit for each subject and display luminance using log10 contrast as the predictor. The probability that a subject gets a correct answer is given by the following equation.

$$\Pr(\text{correct}) = 1/4 + (1 - 1/4)\,\Phi\left[(x - \mu_{ij})\ 1/\sigma_i\right]$$

where i indexes subject and j indexes luminance settings. Here Φ indicates the cumulative Gaussian distribution function. For each subject, this gave a separate location parameter estimate for each luminance setting, and a common spread parameter estimate. A common spread parameter is assumed, since this corresponds with what is known biologically about the human visual system (i.e. it corresponds to an equal change in log contrast producing an equal change in perception throughout the visual response range corresponding to the luminance range of this experiment). The 1/4 arises from the 4 AFC task.

The location parameter, μ_{ij}, is the mean of the corresponding Gaussian distribution for the ith subject and the jth luminance setting. Display luminance conditions that improve detection will cause this parameter to be smaller, and the curve will shift to the left. This occurs because lower contrast levels are required to spot the object. When the display condition makes detection harder, higher contrast levels are needed to locate the mass, and the curve shifts to the right. The values of

σ_i, the spread parameter for the ith subject, correspond to the slope of the curve. Smaller values of σ_i correspond steeper slopes, or greater increases in detection rates per log contrast.

Repeated measures of analysis of variance (ANOVA) allows one to examine the effect of display luminance level, while accounting for the dependence of measurements taken on the same observer. The repeated measures ANOVA model was fitted, with the σj scores as the outcome. The log10 contrast was the predictor for this model.

To compare the processing conditions and to examine the effect of luminance, further analysis was needed. We defined the overall measure to be $\theta ij = \mu_{ij} + \sigma_i$, which corresponds to the log contrast level at which the ith subject viewing the jth luminance condition scored 88% correct. We measured the effect of display luminance condition by calculating the delta (σj) difference between the θ score for the display condition of 600 (reference standard of mammography lightbox) and the θ score for each of the other display luminance conditions, for each subject in this study. A larger positive σj score reflects improved detection, which indicates a more negative θj value. This would indicate better detection with other display luminance conditions than with the standard display luminance condition.

Two analyses were performed using this outcome measure. In order to keep a nominal overall type 1 error rate of 0.05 for experiment, a first repeated measures analysis of variance was done at the 0.04 level, and second set of 4 T-tests was performed at a 0.01 level (0.04 + 0.01 = 0.05). Since there were 4 T-tests, each was performed at 0.01/4 = 0.0025 level. A total of 20 subjects were tested.

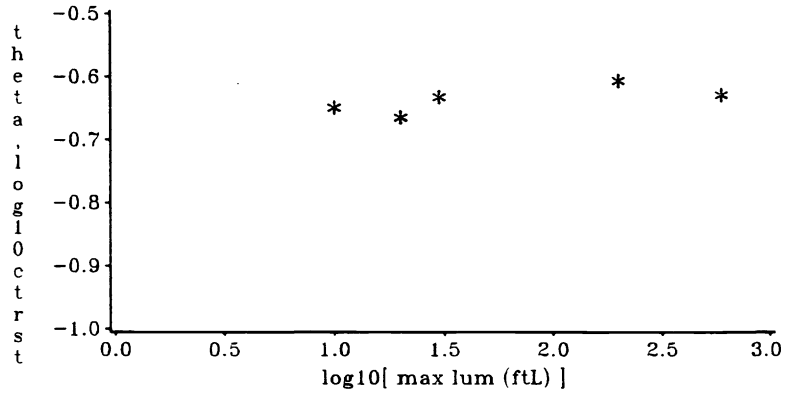

Figure 2. Shows the mean theta values for each level of maximum luminance display condition (luminance is expressed as log10). Rightmost point is 600 ftL condition, and leftmost point is 10 ftL condition. Values closer to the bottom indicate lower contrast thresholds where the observers were more sensitive.

The repeated measures analysis of variance revealed that display luminance condition did not significantly effect the threshold for the detection of masses at the 0.04 level (p-value = 0.0832, Geiser-Greenhouse epilson^ = .6261, df = 4). These results are shown in Figure 2, which depicts the mean θ values for each display luminance condition.

The second analysis, the series of planned step-down tests was implemented at the nominal level of 0.01/4 = 0.0025. The differences between the standard luminance condition and the remaining conditions were examined. None of the P-values were less than 0.0025, and thus none of the display luminance conditions made a significant difference in correctly locating the masses. These results are seen in table 4, which gives the summary statistics for σj at different luminance conditions.

	Mean	Std Deviation	P Value
δ_{10_600}	+ .0203	.1055	0.3998
δ_{20_600}	+ .0358	.0613	0.0173
δ_{30_600}	+ .0038	.0694	0.8094
δ_{200_600}	- .0229	.1034	0.3339

Table 4. Summary statistics for δ at different display luminance level differences, where (δx_y represents difference between scores for display luminance conditions x and y)

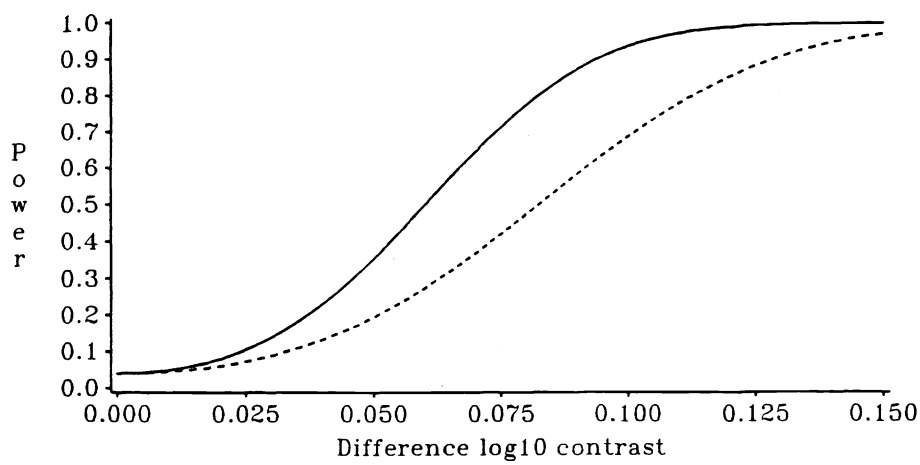

Figure 3. This figure shows the power curve for this experiment. The solid line is estimated power, and the dashed line is the 95% lower bound confidence interval.

Finally a retrospective power analysis was computed. Figure 3 shows the power required to detect a difference in log10 of contrast for a repeated measures of analysis of variance at the 0.04 level The solid line is the estimated power, and the dashed line is the exact 95% one-sided (lower bound) confidence band on power.

5. DISCUSSION

Digital mammography is already beginning to appear in the clinic. It is highly likely that several methods of displaying digital mammograms will be available. It is important to characterize what effects the display system will have on the radiologists' clinical performance. These results suggest the display luminance of the display system is not a significant factor affecting the detection rate of simulated masses inserted in mammographic backgrounds. Vision theory would predict this for uniform backgrounds for this luminance range where Webber's law holds (the value of $\Delta L/L$ is constant). This result validates this for mammographic backgrounds and mass targets. It suggests that lower luminance display systems may function just as well for detection tasks in radiology. Specifically, the option of lower luminance video displays may be a viable option.

The biggest caveat is that the lower luminance levels for which an effect was not found (10ftL to 30ftL) would probably not perform as well under actual clinical conditions. This is because most clinical reading rooms have too much ambient light (overhead fluorescent lights) and glare (from surrounding lightboxes). These light levels are known to cause the contrast thresholds to be larger for the lower luminance display systems. Thus, the result of no significant differences for those display luminance levels may not hold for actual clinical conditions, unless the working environments are changed. Under such clinical conditions these results still suggest that the brighter CRT monitors that are currently commercially available should provide sufficient range for mammographic image presentation, and likely for most other radiological image displays as well, while not being compromised by room lighting conditions.

An important side issue of this talk is the discussion of what contrast definition to use. This is an area requiring further work, especially in the area of background structure and texture based surround luminance measures. Standardization of measures of contrast for non-uniform backgrounds would be of significant help in allowing comparison across different research results.

6. FUTURE WORK

Important future work would be to extend these results to other radiological backgrounds and feature targets, and to test under clinical room lighting conditions. We also plan to conduct similar studies on video displays.

7. ACKNOWLEDGMENTS

This work was supported in part by NIH grants P01-CA47982, R01-CA60193, and R01-CA44060.

8. REFERENCES

1. Puff DT, Pisano ED, Muller KE, Johnston RE, Hemminger BM, Burbeck CA,

McLelland R, Pizer SM, "A Method for Determination of Optimal Image Enhancement for the Detection of Mammographic Abnormalities", Journal of Digital Imaging, Nov 1994.

2. Hemminger BM, Johnston RE, Rolland JR, Muller KE, "Perceptual Linearization of Video Display Monitors for Medical Image Presentation", Proceedings of Medical Imaging 1994: Image Capture, Formatting and Display, SPIE Vol 2164, 1994.

3. Pisano EP, Chandramouli J, Hemminger BM, DeLuca M, Blueck D, Johnston RE, Kuller K, Braeuning MP, Pizer SM, "Does Intensity Windowing Improve the Detectation of Simulated Microcalcifications in Dense Mammograms?", JDI, accepted for May 1997.