16

Effect of display luminance on the feature detection rates of masses in mammograms

Bradley M. Hemminger[a)]
Department of Radiology, University of North Carolina 27599-7515

Alan W. Dillon
Department of Biomedical Engineering, University of North Carolina 27599-7575

R. Eugene Johnston
Department of Radiology, University of North Carolina 27599-7515

Keith E. Muller, Marla C. Deluca, and Christopher S. Coffey
Department of Biostatistics, University of North Carolina 27599-7400

Etta D. Pisano
Department of Radiology, University of North Carolina 27599-7515

(Received 4 February 1999; accepted for publication 18 August 1999)

Our purpose in this study was to determine the importance of the luminance range of the display system for the detection of simulated masses in mammograms. Simulated masses were embedded in selected portions (512×512 pixels) of mammograms digitized at 50 μ pixels, 12 bits deep. The masses were embedded in one of four quadrants in the image. An observer experiment was conducted in which 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 placing neutral density films between the laser printed films of mammographic backgrounds and the lightbox. The resulting maximum luminances examined in this study ranged from 34 cd/m^2 to 2056 cd/m^2. Twenty observers viewed 80 different images (20 observations at each of 4 different mass contrast levels) under each of the 5 luminance conditions for a total of 800 independent observations per observer. An analysis of variance yielded no statistically significant correlation between the luminance range of the display and the feature detection rate of the simulated masses in mammograms. However, the performance of the lower luminance display systems (less than 300 cd/m^2), may be reduced due to the high levels of ambient light found in many reading environments. © *1999 American Association of Physicists in Medicine.* [S0094-2405(99)01811-8]

Key words: image display, display systems, digital imaging, mammography, image perception

I. INTRODUCTION

In the past, film has served as both the detector 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 detection 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 independently study the characteristics necessary for medical image display. The general question addressed in this research is what effect the maximum luminance level of the display system has on medical image feature detection.

We chose to evaluate mammography because it has the most demanding luminance range requirements of the radiology modalities. Specifically, the ACR recommends 3426 cd/m^2 (1000 ft L) luminance lightboxes for the display of analog film-screen mammography films. This requirement is due to 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. The purpose of this research is to investigate the effect of the maximum luminance of which the mammography display system is capable. If display systems with smaller maximum luminance values, but comparable luminance ranges, perform as well as the currently proscribed 3426 cd/m^2 requirement for film on lightboxes, then softcopy (video) displays may be satisfactory for mammography image presentation. Additionally, these results should hold for other, less demanding medical imaging modalities such as computed tomography (CT), magnetic resonance (MR), computed radiography (CR), and digital radiography (DR). In this study we attempt to determine the effect of the luminance range of display systems on the feature detection rate of simulated masses in mammograms. Masses were chosen because this is similar to many radiology detection tasks (masses in lungs

on chest x-ray, nodules on chest CT, etc.). The luminance ranges evaluated in the experiment were chosen to match those of commercially available video display systems and mammography lightboxes.

While it may seem counter-intuitive to expect observer performance on a feature detection task to remain the same if the maximum luminance of the display device is reduced, vision theory predicts that as long as the dynamic range is the same, then observer performance should not change.[1–4] The dynamic range is the number of perceivable shades of gray in the luminance range (i.e., from minimum luminance to maximum luminance of the display system). Thus, maximum luminance should not be a factor as long as the maximum luminance does not become so low that it forces a reduction in the perceivable dynamic range, and the room lighting conditions allow the full dynamic range to be utilized (i.e., large amounts of ambient light will significantly reduce the effective dynamic range of displays with lower maximum luminances by washing out perception of the darkest shades of gray). When ambient light is kept to a minimum, so that the full dynamic range of the display device can be utilized, there should be no change in feature detection due to maximum luminance of the display device. Similar results, that found no difference in the detection of masses and microcalcifications with display luminance conditions of 274 cd/m^2 vs 480 cd/m^2 when reading mammograms on video CRT monitors, have been reported by Krupinski *et al.*[5]

II. 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.[6–8] 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 selected subparts of mammograms (the subparts are referred to as mammographic backgrounds). The masses were embedded in one of four quadrants in the background. 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 range was controlled by placing the neutral density films between the laser printed films of mammographic backgrounds and the lightbox. The resulting luminance ranges examined in this study had maximum luminance values ranging from 34 cd/m^2 to 2056 cd/m^2. Twenty observers viewed 80 different images (20 observations at each of 4 different mass contrast levels) under each of the 5 luminance conditions for a total of 800 independent observations per observer.

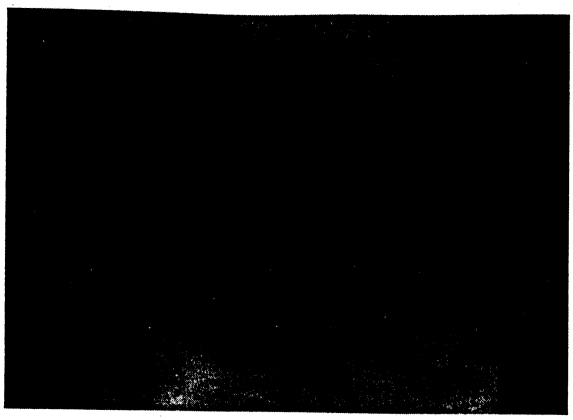

FIG. 1. Example from the experiment of background with a mass stimulus in the lower right hand corner (arrow points to the mass). This is an example of an easily discernable mass (contrast of mass to background was 0.22).

A. Mammographic backgrounds

The background images used in the experiment were subparts of digitized film screen mammograms. There were a total of eighty different background images. Each background image was 512×512 pixels, 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 by a radiologist expert in breast imaging, from digitized film screen craniocaudal or mediolateral oblique mammograms. They were selected so as to provide an even distribution of density distributions across the 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.

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 gray values that map one to one with OD values; i.e., the same OD value on film will produce the same gray level value when digitized.

B. Mammographic mass stimuli

Mammographic masses were simulated using a locally developed program. A circle 4.5 mm in diameter was first generated. Then 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 on mammograms. An example background with inserted simulated mass is shown in Fig. 1.

C. Contrast

The contrast of the mass to background surround was defined as the luminance ratio $\Delta L/L$, where ΔL was the mean luminance of the background surround with the target inserted minus the mean luminance of the background surround without the target inserted. L is the mean luminance of the background surround. Several different choices exist for the area under which the mean background surround value could be calculated. 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. Thus, we limited our consideration to mean luminances of the area surrounding the inserted target. We investigated whether choosing a different size area for calculating the mean of the background surround would have affected our calculation of contrast values, and found no substantial difference between using areas ranging from slightly smaller than the target to using the full 512×512 background surround area.[9] However, using the largest areas increased the standard deviations, which would reduce our sensitivity in measuring the feature detection of the masses given our contrast definition. As a result, we chose for this experiment to use the area of the background under the inserted target mass feature.

The masses were embedded at one of four different contrast levels by the pixel-wise addition of the target mass 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 pixel digital value 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 the 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 digital pixel 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. The experiment was piloted with three observers on a separate set of background images similar to the study 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 0.04, 0.10, 0.16, and 0.22. This corresponded to approximate percent correct detection rates of 30%, 50%, 80%, 95%, respectively.

D. Experimental presentation

The digital images were printed onto standard 14×17 inch single emulsion film (3M HNC Laser Film, 3M, St. Paul, MN) using a Lumisys Lumicam film printer (Lumisys Inc., Sunnyvale, CA). Each original digitized 50 micron pixel was printed at the film printer's smallest spot size of 80 microns, which produced film images enlarged by a factor of 1.6. This resulted in each 512×512 pixel background being reproduced at approximately 4×4 centimeters in size, and the 4.5 mm masses being reproduced at 7.2 mm. Radiologist observers in the previous experiments using this same paradigm reported that they felt this magnification did not make the backgrounds unrealistic.[4] Thirty-two backgrounds were printed per sheet of film. The experiment consisted of three test films, with the third film having only 16 trials since there were 80 total trial backgrounds. The backgrounds were randomly ordered into an 8×4 grid on each sheet of film. The 8×4 grid was chosen because the mammography lightbox was uniform in luminance only over the central portion of the lightbox, which corresponded to the 32 cm×16 cm area covered by the 8×4 image grid.

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.[10] 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 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 chose to use neutral density films to control the luminance of the display for consistency, and to keep the inherent maximum luminance capability of the lightbox. If we had used a video display system such as a CRT, we would not have been able to reproduce the high luminance levels of lightboxes. Additionally, we would have had to sacrifice contrast resolution (the number of gray 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 gray scale range of the display device. Further, it would have caused us to produce multiple films for different luminance values, which may have caused differences between films depicting the same contrast levels due to variables in the film printing. For the above reasons, we chose to print a single version of the test images, and use neutral density films 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.[9]

The resulting luminance levels on the lightbox were exactly controlled by using a voltage regulator in line with the lightbox to adjust the luminance level of the lightbox output. 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 represent common commercially available luminance levels for video display systems and mammography lightboxes. The values selected were mammography lightbox (2056 cd/m^2), high brightness CRT (685 cd/m^2), average workstation monitor (102 cd/m^2), and low end personal computers or hardcopy displays (69 cd/m^2 and 34 cd/m^2). To be consistent across our display luminance condition, we used a neutral density film for each luminance condition, which meant that even the maximum luminance range had a neutral density film behind the test film. The reason the maximum display luminance (simulating a mammography lightbox) was 2056 cd/m^2 is because the 2707 cd/m^2 mammography lightbox when measured with a transparent (unexposed) film in place produced 2056 cd/m^2. While 2056 cd/m^2 may seem substantially less than the 3426 cd/m^2 recommended by ACR, the actual difference is small for two reasons. First, the lightbox used in the experiment is within the range of most commercial mammography lightboxes, which produce a maximum luminance in the range of 2500 cd/m^2–3500 cd/m^2. Second, the effective difference due to having the neutral density filter in place is small compared to the effect of the mammogram itself. Further, the data analysis for the experiment included examining the interaction of feature detection with luminance to determine if any trends, such as improved detection at higher luminances, were suggested.

In order to test the effect of luminance ranges, the dynamic range of each luminance condition was held constant as well. The dynamic range is defined as the maximum luminance of a display condition divided by the minimum luminance of that display condition. Table I shows the optical densities of the five neutral density films, and the measured dynamic ranges for each of the five display luminance conditions in the experiment.

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 34 cd/m^2, 69 cd/m^2, 103 cd/m^2, 2056 cd/m^2 maximum luminance display conditions. 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. The average viewing distance was 40 cm. 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.

TABLE I. Column 1 shows the Optical Densities of the neutral density films used for the five display luminance conditions. Column 2 shows the measured dynamic range of each display luminance condition (maximum luminance divided by minimum luminance). Maximum luminance was measured with neutral density film in combination with the smallest OD produced on the laser printed film. Minimum density was measured with neutral density in combination with largest OD produced by the laser printed film.

Optical density of five neutral density (ND) films	Dynamic range (Max÷Min)	Display condition (cd/m^2)
1.81 (ND_0)	5331	34
1.56 (ND_1)	5382	69
1.39 (ND_2)	5293	103
0.55 (ND_3)	5426	685
0.13 (ND_4)	5444	2056

E. Observation task

There were 20 observers for the experiment. They were medical students and graduate students from the University of North Carolina. A previous study using this same paradigm established that the students' performance parallels that of radiologists for this feature detection task.[6] Performance bonus pay was used to encourage optimal observer performance. Observers selected the quadrant of the image that they thought contained the mass, and indicated the location on a paper form, which matched the film format. 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 with two films, with each film containing 64 images. The first 32 images contained easy cases (higher contrast than used in the experiment), and the second 32 images contained cases with contrast values 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 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. The observers were dark adapted to the room each time they

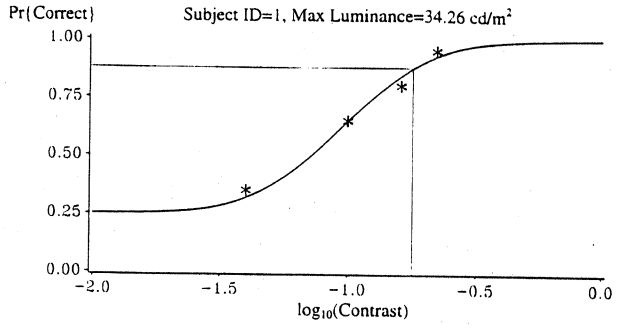

FIG. 2. A probit curve from the experiment for the 34 cd/m^2 display luminance condition for observer 1. The probit curve is fit to the four measured percent correct response points at each of the four contrast levels. A horizontal line shows the location of the 88% percent correct location on the probit curve for observer 1 and display condition 1 ($\theta_{11}=\mu_{11}+\sigma_1$).

entered or re-entered the room. All observers completed the experiment. Each observer examined 80 different images (20 independent observations at each of the four contrast levels), with each of the 5 neutral density combinations, for a total of 400 images per observer, and 8000 stimuli for all observers for the whole experiment.

III. DATA ANALYSIS

Probit models were fit for each subject and display luminance using $\log_{10}$ (contrast), hereafter referred to as log (contrast), as the predictor.[11] The probability that a subject gets a correct answer is given by the following equation:

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

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).[2,4] The 1/4 arises from the 4 AFC task. An example probit curve for observer 1 is shown in Fig. 2.

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 i^t subject, correspond to the slope of the curve. Smaller values of σ_i correspond to steeper slopes or greater increases in detection rates per log(contrast).

To compare the processing conditions and to examine the effect of luminance, we defined the overall measure of performance 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 difference (δ) between the θ score for the display condition of 2056 cd/m^2 (the reference standard of the mammography lightbox) and the θ score for each of the other display luminance conditions. This was done over all subjects in this study. A larger positive δ_j score reflects improved detection, which indicates a more negative θ_j value. This would imply better detection with other display luminance conditions than with the standard display luminance condition.

Two statistical analyses were planned: a repeated measures ANOVA model fit with the δ_j scores as the outcome, and planned series of step-down tests. The log(contrast) was the predictor for this model. In order to keep a nominal overall type I error rate of 0.05 for the experiment, a repeated measures analysis of variance was done at the 0.04 level. with a set of 4 T-tests at a 0.0025 level for each. A total of 20 subjects were tested.

TABLE II. Percent correct totals for observer feature detection rate across all observers for each display maximum luminance condition.

Display luminance (cd/m^2)	Percent correct (%)
34	66.5
69	66.9
103	64.6
685	63.3
2056	64.9

IV. RESULTS

The percent correct totals across all observers and all contrast levels are shown for each of the maximum luminance conditions in Table II. There is minimal difference between the overall percent correct and the display luminance conditions.

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, G-G ϵ=0.6261, df=4). These results are shown in Fig. 3, which depicts the mean θ values for each

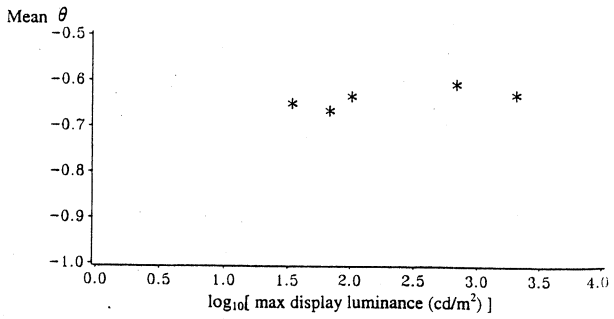

FIG. 3. Mean θ values for each level of the maximum display luminance condition (luminance is expressed as $\log_{10}$). The rightmost point is the 2056 cd/m^2 condition, and the leftmost point is the 23 cd/m^2 condition. Values closer to the bottom indicate lower contrast thresholds where the observers were more sensitive to perceiving contrast thresholds.

TABLE III. Summary statistics for display luminance level differences, where δ_j represents the difference in θ scores between the jth luminance level and the reference mammography lightbox display luminance level, 2056 cd/m^2.

	Mean	Standard deviation	p-value
δ_{34}	0.0203	0.1055	0.3998
δ_{69}	0.0358	0.0613	0.0173
δ_{103}	0.0038	0.0694	0.8094
δ_{685}	−0.0229	0.1034	0.3339

display luminance condition. The rightmost point is the 2056 cd/m^2 condition and the leftmost point is the 34 cd/m^2 condition. Values closer to the bottom indicate lower contrast thresholds where the observers were more sensitive.

The second analysis was the series of planned step-down tests 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 display luminance conditions made a significant difference (p-values <0.0025) in correctly locating the masses. These results are seen in Table III, where the summary statistics for δ_j are given at different luminance conditions.

Finally, a retrospective power analysis was computed.[12] Figure 4 shows the power required to detect a difference in log(contrast) for a repeated measures Analysis of Variance at the 0.04 level. The solid line is the estimated power and the short and long dashed line is the exact 95% one-sided lower bound confidence band on power. The fainter dashed lines in Fig. 4 show we would have an estimated power of 0.872 for the clinically relevant task of detecting the presence of a simulated mass one contrast threshold above the background, which is calculated to be 0.09 in log(contrast) for this experiment (Appendix A). As a result, we believe the experimental data strongly supports this experiment having adequate power to detect a contrast threshold difference in contrast. However, the power may be reduced somewhat if the contrast thresholds for the detection of masses in mammograms turns out to be less than measured in this experiment.

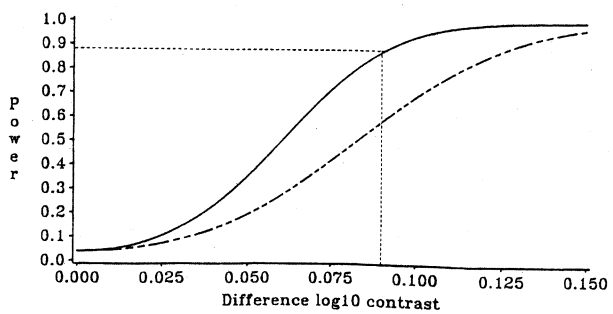

FIG. 4. Power curve for this experiment. Solid line is the estimated power and the short and long dashed line is the exact 95% one-sided lower bound confidence band on power. Fainter dashed line is the specific power estimate for the clinical mass detection task of this experiment.

V. DISCUSSION

Digital mammography is becoming available clinically. 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 maximum luminance of the display system is not a significant factor affecting the detection rate of simulated masses inserted in mammographic backgrounds. Vision theory predicts this for the detection of mammographic features on uniform backgrounds for these luminance ranges when the dynamic range is the same across display conditions.[1,2,4] This result for mammographic backgrounds and mass targets is consistent with the prediction for uniform backgrounds. 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 regarding displays operating at the lower luminance levels for which an effect was not found (34 cd/m^2 to 103 cd/m^2). These displays would probably not perform as well under actual clinical conditions because of higher levels of ambient light and the lack of observer dark adaptation for low luminance levels. While most mammography reading rooms are more carefully controlled, this is not the case for general reading rooms where CR images are often displayed under high ambient light levels with significant amounts of glare present. Under these conditions higher luminance display systems, or changes in room lighting conditions would be required. Under average clinical conditions these results still suggest that commercially available high brightness CRT monitors (300 cd/m^2 and brighter) would not be significantly compromised by moderate room lighting conditions, and should provide a sufficient luminance range for mammographic and other radiological image presentations.

ACKNOWLEDGMENTS

This work was supported in part by NIH Grants No. P01-CA47982, No. R01-CA60193, and No. R01-CA44060. A preliminary version of this work was reported in the proceedings of the SPIE 1997 Medical Imaging conference, and is used with the permission of SPIE.[9] The authors would also like to thank the Medical Physics reviewers for their helpful comments.

APPENDIX: CALCULATION OF CLINICALLY RELEVANT THRESHOLD FOR POWER CURVE

To determine whether this experiment had sufficient power to detect the clinically relevant change of seeing the mass versus not seeing the mass for different display luminance conditions, we calculated the contrast of probable detection for this experimental paradigm. The estimated power[13] for detecting the difference in contrast between seeing a mass and not seeing a mass is 0.872 (the intersection of the fainter dashed lines on the power curve in Fig. 4). Our definition of a contrast threshold for this experimental para-

digm, the contrast of probable detection, is the shoulder point on the probit curve (Fig. 2), that is, the mean plus one standard deviation as described in Puff,[6] which for this experiment works out to be the 88% percent correct level. While contrast thresholds for simple targets on uniform backgrounds are well described in experimental vision literature, there is not published data for contrast thresholds for mammography backgrounds and lesions. Thus, the best estimate we have for the contrast threshold for this task is the experimental data from this work. We averaged the probit curves across all observers and all luminance conditions, and calculated the contrast threshold to be 23% for the 88% percent correct level. This equates to a difference of 0.09 in log contrast. Plotting 0.09 on the retrospective power graph Fig. 4, shows we would have an estimated power of 0.872 for detecting a difference of 0.09 in log(contrast).

[a]Electronic mail for corresponding author: bmh@rad.unc.edu

[1] T. N. Cornsweet, *Visual Perception* (Academic, New York, 1970).
[2] C. Murchison, in *A Handbook of General Experimental Psychology* (Clark University Press, Worcester MA, 1934), p. 769.
[3] S. M. Pizer, "Intensity mapping to linearize display devices," Comput. Graph. Image Process. **17**, 262–2268 (1981).
[4] G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed. (Wiley, New York, 1982), pp. 567–570.
[5] E. Krupinski, H. Roehrig, T. Furukawa, and C. Tang, "Influence of monitor luminance and tone scale on observer detection performance," Proc. SPIE **3340**, 99–104 (1998).
[6] D. T. Puff, E. D. Pisano, K. E. Muller, R. E. Johnston, B. M. Hemminger, C. A. Burbeck, R. McLelland, and S. M. Pizer, "A method for determination of optimal image enhancement for the detection of mammographic abnormalities," J. Digital Imaging **7**, 161–171 (1994).
[7] E. D. Pisano, J. Chandramouli, B. M. Hemminger, M. DeLuca, D. Blueck, R. E. Johnston, K. Muller, M. P. Braeuning, and S. M. Pizer, "Does intensity windowing improve the detection of simulated microcalcifications in dense mammograms?" J. Digital Imaging **10**, 79–84 (1997).
[8] E. D. Pisano, J. Chandramouli, B. M. Hemminger, M. DeLuca, D. Blueck, R. E. Johnston, K. Muller, M. P. Braeuning, and S. M. Pizer, "Does intensity windowing improve the detection of simulated masses in dense mammograms?," J. Digital Imaging **10**, 174–182 (1997).
[9] B. M. Hemminger, A. W. Dillon, and R. E. Johnston, "Evaluation of the effect of display luminance on the feature detection of simulated masses in mammograms," Proc. SPIE **3036** (1997).
[10] B. M. Hemminger, R. E. Johnston, J. R. Rolland, K. E. Muller, "Perceptual Linearization of Video Display Monitors for Medical Image Presentation," *Proceedings of Medical Imaging 1994: Image Capture, Formatting and Display*, Vol. 2164 (1994).
[11] D. J. Finney, *Probit Analysis*, 3rd ed. (Cambridge University Press, London, 1971).
[12] K. E. Muller and V. A. Benignus, "Increasing scientific power with statistical power," Neurotoxicol. Teratol. **14**, 211–219 (1992).
[13] K. E. Muller and V. B. Pasour, "Bias in linear model power and sample size due to estimating variance," Commun. Stat. Theory Methods **26**, 839–851 (1997).