

# Visual methods for determining ambient illumination conditions when viewing medical images in mobile display devices

Fahad Zafar<sup>a,b</sup>, Mina Choi<sup>a,c</sup>, Joel Wang<sup>a</sup>, Peter Liu<sup>a</sup>, Aldo Badano<sup>a</sup>

<sup>a</sup>*Division of Imaging and Applied Mathematics, Office of Science and Engineering Laboratories, Center for Devices and Radiological Health, FDA, Silver Spring, MD.*

<sup>b</sup>*Department of Electrical and Computer Engineering, University of Maryland Baltimore County, MD.*

<sup>c</sup>*Department of Electrical and Computer Engineering, George Washington University, DC.*

---

## Abstract

The quality of the displayed image in mobile devices is affected by the varying ambient illumination conditions. Determining appropriate viewing conditions for particular visual tasks requires time and appropriate instrumentation. To this end, we explored the usefulness of more practical, visual tests for use in clinical environments. We conducted experiments to determine the limitations of mobile displays in terms of the visibility of subtle targets for different background luminance, ambient illumination with two mobile devices. We compare a noise-embedded text detection task and a threshold estimation staircase technique for a range of illuminance between 1 and 80,000 lux encompassing conditions found in dark reading rooms, office spaces, and outdoor scenarios. We found that the text detection method holds promise as a surrogate for more complicated tests in the framework of a clinically practical implementation.

*Keywords:* mobile displays, medical imaging, ambient illumination.

---

## 1. Introduction

Mobile devices possess extensive portability incentives. Any application ported to such devices can bring maximum availability to users. Physicians could make use of these devices in emergency situations and remote environments. However, the use of these devices to interpret medical images raises questions that need to be answered before their use becomes widespread. It has not yet been demonstrated that mobile displays provide accurate and quality information to the user for correct decision making based on imaging data(1). Many concerns including security, cost, data confidentiality, network speed and battery life must be taken into account besides the display requirements. Some studies have been conducted with CT brain images displayed on personal digital assistants with limited success(2, 3). These experiments did not account for all the factors that contribute to medical image viewing for non standard environments. Even if a mobile device provides promising results compared to a medical display device in standard conditions, changing ambient environments can dramatically

affect user performance. While desktop display systems have been extensively analyzed(4, 5), mobile displays have not. One study(6) analyzed mobile displays under different viewing conditions to show how the CIECAM02 model can be adopted under different ambient lighting conditions. CIECAM02 is the most recent color appearance model ratified by the International Commission on Illumination (CIE). The model is used to predict the color that occurs in certain illumination conditions considering factors such as color constancy effect, adapting backgrounds, stimulus characteristics and surrounding luminance information. The authors showed that the screen luminance differs under different ambient conditions and proposed an enhancement model to compensate for that distortion. Another study(7) discussed factors affecting subjective image quality measures including clearness, naturalness, sharpness and colorfulness for liquid-crystal displays. However, to date, objective assessment of how ambient illumination conditions affect visual task performance on mobile displays has not been reported with respect to solutions solving this problem.

Method	Subjects	Ambient	Trials	Classes	Views
Staircase	5	3	35	5	2625
DENOTE	5	3	10	5	750

Table 1: Number of observations for all experiments. Ambient represents the number of ambient illumination conditions (average, bright, dark) used. Classes indicates the mean value of noise backgrounds with 5 different luminance subsets (0-51, 51-102, 102-153, 153-204, 204-255). Views indicate the number of decisions.

In this paper, we study and discuss the effect of ambient conditions on detection performance. We introduce a novel methodology based on noise-embedded text detection as a practical test which can be applied to multiple devices and illumination conditions. We compare results obtained with an established methodology based on staircase techniques for threshold sensitivity experiments to results obtained using the text detection methodology.

## 2. Methods

In all our studies a human subject is presented with a detection or search task. The two mobile device used for experiments include the Google Nexus One and the Samsung Galaxy S with 3.7 inch and 4 inch screens respectively. Both these devices are active-matrix organic light-emitting displays with a 480 by 800 pixel resolution. The observer experiments were conducted for three lighting conditions: bright (60,000 - 80,000 lx), average (400 - 500 lx), and dark ( $< 1$  lx), and five background luminance levels. Each of the five luminance classes represent a subset of the gray levels with a class interval of about 51 gray levels each (0 - 51, 51 - 102, 102 - 153, 153 - 204, 204 - 255). All observers were seated in the particular ambient condition for 10 minutes before conducting the experiment. The set of observers included two females, ages 23 to 25 with perfect vision, and three males, age around 25, two with corrected vision. The observers were trained and fully disclosed on the purpose of the experiment, and were allowed test trials before conducting the actual data gathering experiments under each illumination condition.

Table 1 presents the number of total observations conducted for the main experiments presented in this paper. Three observers performed the experiments in 3 illumination conditions simulating dark, office (“average”) and outside (“bright”) environment with 5 background luminance classes.

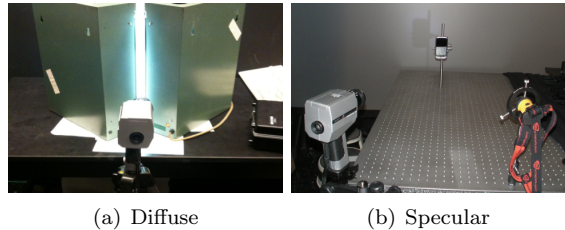


Figure 1: (a) Experimental setup for measuring the reflected luminance from the two diffuse light sources using the photometer. (b) Experimental setup for measuring the luminance reflected off the mobile display using a flash light and a photometer where the angle of reflection is 15 degrees.

The techniques were implemented as an application using Java and the Android SDK 2.2 that can be ported to any tablet/mobile/device running the Android OS. No attempt was made to correct the look up table used in converting gray level values to luminance.

### 2.1. Physical Characterization

We measured display reflections for commonly used mobile devices as well as for two medical displays. The reflectance of display devices can be separately characterized involving specular and diffuse light scattering. Both components have markedly distinct effects on image quality and require different experimental measurement methods. Our methodology for calculating the coefficients is adopted from previously reported techniques(8, 9).

For calculating the diffuse coefficient, first, three screws were placed on the back of a white Styrofoam box to hold up the mobile device. Subsequently, two fluorescent lamps with light diffusers were positioned at the opening of the box with a small gap between them, and white paper/Styrofoam was placed to cover the openings. A T-10 illuminance meter (Konica Minolta, New Jersey, NJ) was placed into the Styrofoam box to measure the illuminance inside the box (see Fig. 1), which at the face was measured at 6580 lux of quasi-isotropic illumination. Lines were marked on the white paper that was taped beneath the box to maintain the exact opening (in terms of angle) for each measurement and the box was closed off with the fluorescent lights and the luminance response from the diffuse reflection off the screen was measured using a Minolta CS-100 spot photometer through the gap between the fluorescent lights. Finally, the diffuse reflectance coefficient was calculated as

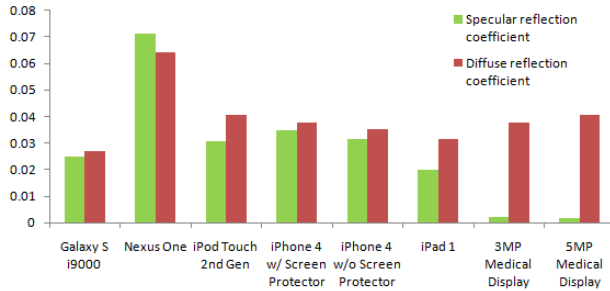


Figure 2: Specular and diffuse coefficients for a set of display including medical and mobile devices.

$R_d = L_{ref}/I_{box}$ , where  $L_{ref}$  is the reflected luminance measured with the spot photometer and  $I_{box}$  is the illuminance at the face of the devices when the illuminating box is on. The units of  $R_d$  are  $\text{sr}^{-1}$ .

For the specular coefficient, the device screen was consistently placed perpendicular to the table. Next, an LED flashlight was fixed 31 inches away from the device holder at an angle of 15 degrees from the normal relative to the center of the screen, pointing at the middle of the mobile display (see Fig. 1). The CS100 photometer was similarly fixed 31 inches away at 15 degree from the normal relative to the screen and focused at the center of the mobile display. The photometer was then used to measure the luminance from the direct reflection from the mobile display. Next, the photometer was placed 62 inches away from the LED light and a direct luminance value from the LED lights was measured. Finally, the specular reflectance coefficient was calculated as  $R_s = L_{ref}/L_{dir}$  where  $L_{ref}$  and  $L_{dir}$  are the measured luminance values from the reflection and from direct view respectively. It should be noted that in our setup, the margin of uncertainty is between 5 and 15 percent. In the specular reflection test, 1 degree off in alignment can change the reflected luminance response by approximately  $30 \text{ cd/m}^2$ . In the diffuse reflection test, the luminance response only changed by around  $15 \text{ cd/m}^2$ .

Fig. 2 shows a comparison of devices in terms of reflectivity. Nexus One has unusually higher specular reflectivity which causes problems in viewing, as consistently reported by our observers. Our findings show that the medical displays have a lower

specular reflectance coefficient, most likely because the medical display screens have a matte finish overcoat rather than the glossy-style screens commonly found in mobile displays. For the diffuse reflectance coefficient, the coefficient values are at the higher, in the range  $0.005\text{-}0.010 \text{ sr}^{-1}$ , above the highest mobile reflectance display measurement. This is also most likely due to the medical display matte topcoat which is acceptable in controlled indoor environments. It should be noted that medical displays are used under controlled environments thus eliminating some hindrance caused by diffuse reflectivity. However, when deploying mobile displays, both the higher specular and diffuse reflectivity contribution might affect performance.

We also measured the luminance response of the mobile devices used in our experiments. Our measurements were performed with the PR730 (Photo Research Inc. Chatsworth/CA). A software application was written to synch the photometer and transmitted images from a computer to the mobile device. The transmission was done over a wireless network. The photometer was positioned at the normal to the screen at less than 1 cm away. The program would then send an image to the mobile device, the mobile device would display the image, and then the photometer would measure and record the luminance response of the mobile display. This was repeated for however many images/color levels needed. We did not attempt to control the lookup tables for each device and left the mapping to the intrinsic performance that the device carries from the manufacturer. The luminance range measured was from around  $7.051 \times 10^{-3} \text{ lx}$  to a maximum luminance of  $296.2 \text{ lx}$ . The luminance range was found to be similar in both devices.

## 2.2. Perceptual Tests

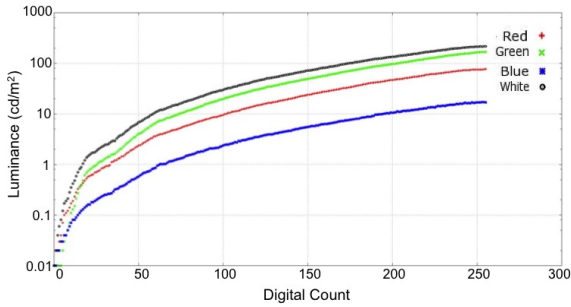
We compare results obtained using two methodologies: a well-established, staircase technique that involves several minutes of trials, and a novel methodology introduced in this paper that relies on a more practical and less time consuming technique of text identification.

### 2.2.1. Noise-embedded text detection (DENOTE)

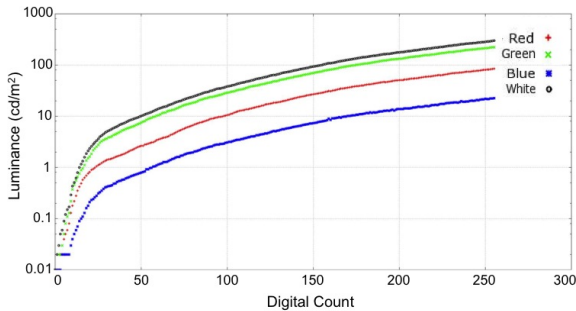
The methodology we introduce in this paper is the DEtection of NOisy TExt (DENOTE), a practical experiment to judge the current device and ambient conditions for image reading tasks. This method is inspired by CAPTCHA and the Ishihara

	3 MP		5 MP		Nexus	Galaxy	iPhone	iPad	iPod
Model	Eizo Radi-	force R31	Eizo Radi-	force G51	1 <sup>st</sup> Gen	i9000	4 <sup>th</sup> Gen	1 <sup>st</sup> Gen	2 <sup>nd</sup> Gen
Display	LCD		LCD		AMOLED	AMOLED	LCD	LCD	LCD
Resolution	1536x2048		2048x2560		480x800	480x800	960x640	1024x768	320x480
In-plane Switching	Yes		Yes		No	No	Yes	Yes	Yes
Luminance	250 cd/m <sup>2</sup>		500 cd/m <sup>2</sup>		229 cd/m <sup>2</sup>	296 cd/m <sup>2</sup>	500 cd/m <sup>2</sup>	-	-

Table 2: Display specifications. Specifications measured (maximum luminance) or taken from product documentation.



(a) Nexus One



(b) Galaxy S

Figure 3: Luminance response for the mobile display devices used in the study.

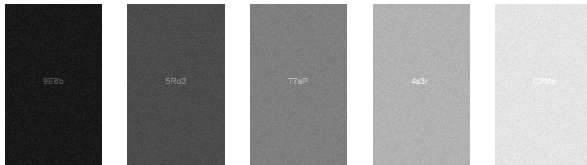


Figure 4: Image samples for the DENOTE method for each of the five luminance classes.

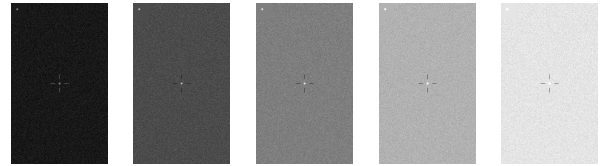


Figure 5: Image samples for the staircase experiment from each of the five luminance classes. There is a reference signal on the top left while the signal in the image center is embedded within the background noise.

vision test (10). CAPTCHA is a visual test used in computer security while the Ishihara test is used to test an observer for color blindness. In this method, a word is distorted by immersing the characters in a noisy background and adjusting the contrast appropriately (see Fig. 4). The observer is asked to identify four alphanumeric characters displayed in the middle of the screen with no time limitations. A percent correct measure on a binary scale is used for the analysis. If the observer gets all the four characters correct, he gets a correct answer otherwise it is considered incorrect. Initial trial runs were conducted to figure out an acceptable gray level for the display text. This proposed method is easy to set up since no other equipment or standardization procedures are required. It provides the observer with a simple task with an execution time of less than 3 minutes. The entire task includes testing the subject on 5 different background levels, with 10 trials per level. This helps establish a good estimate for the user performance for each background class. We believe this methodology is practical for a clinical reading condition.

The results from this type of experiment cannot be used for statistical modeling with much confidence but will offer an approximate assessment for any mobile device in different ambient lighting conditions. For this experiment, the mobile device was not stationary; the user was allowed to move the de-

vice at will and identify the text on the screen. Four characters were displayed randomly on the screen from a dataset including the characters A-Z, a-z, 1-9. Certain characters that might cause confusion for the observer and present unnatural variability in the results, such as '0' and 'O' or 'l' and '1' were excluded. The character amplitude chosen for the embedded text was fixed at 15 gray levels added on top of the noise. The user response was recorded and compared to the displayed word. A success ratio was generated to gauge user performance. The results are assessed on a binary scale. The observer achieves success only if all four characters in the image are correctly identified. In our analysis, all the other combinations of correct and incorrect answers are considered failures.

### 2.2.2. Threshold estimates

The staircase method is a well understood methodology frequently used for estimating thresholds in sensitivity experiments and detection tasks(11, 12, 13). Using a staircase approach, we determined contrast thresholds for the same experimental conditions.

The target was a Gaussian spot 20 pixels in diameter using a single, ascending staircase paradigm. Noise gray level values clamped to the luminance class limits were displayed using an equal probability occurrence function. Each gray level value in the respective luminance class was equally likely to appear in the background of the signal. The Gaussian signal has the highest amplitude in the center and a variance of 1.5 pixels for the amplitude drop off. The size of the signal is 20 by 20 pixels. Cross hair lines around the signal and a reference signal is provided at the top left of the screen to aid in detection (Fig. 5). The signal amplitude increases or decreases depending on each response and a threshold value is calculated from the data using the methodology from Ref. (13).

The mobile device was fixed in front of the observer eye for perpendicular viewing to the screen at a distance of about 30 cm. The observer was provided with a chin rest support for fixed eye position in front of the mobile device.

### 2.2.3. Search Task

Finally, we conducted a search task observer study using a Gaussian spot signal. The user had to identify the number of signals on the screen picking either one of 0, 1, 2, 3 or 4 options. Since the test was conducted on a smart phone with a touch

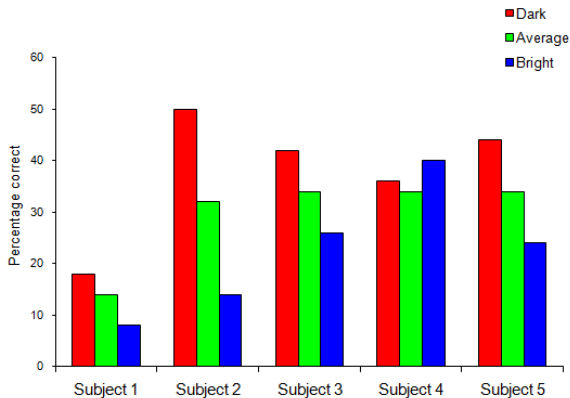
screen display, the user was supposed to identify the spot by tapping on it. Observers had trouble touching the spots due to visual parallax. Satisfaction of search (14) was also a problem since most observers were annoyed by the fact that they cannot locate tap on the spot even when they see it.

Increasing the tap dimensions around the spot to a  $\pm 20$  pixels did not provide satisfactory results as many observers were still not able to touch the signal when holding the device at an angle. Further increasing the window would have annulled the purpose of the experiment. Continuously tapping on the screen caused smudges on the screen and required wiping the screen periodically before use. Display degradation is a relevant factor for image viewing on smart phones, but our current focus did not cover that aspect. Thus, the test was modified to fixed options in the menu for the user, making it easier to count the number of visual spots. This experiment did suffer from false positives and false negatives outcomes and the results had a high variance.

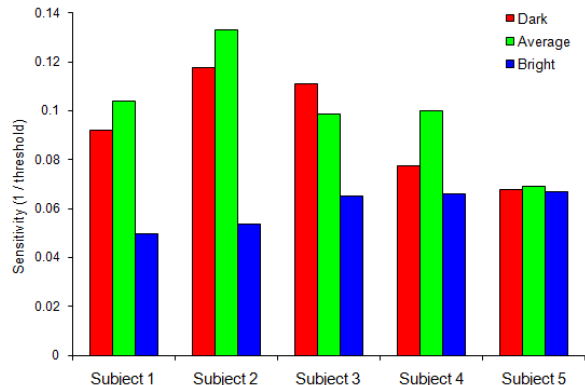
## 3. Results

Fig. 6 presents results for a 5 observer study in which we compare the DENOTE method with the staircase technique. The observers were asked to identify the four characters randomly placed in the middle of the screen in different ambient illumination conditions. Each bar in the DENOTE section represents the sum of performance across five background luminance classes with 10 trials each. The staircase method estimates the contrast sensitivity in the same ambient conditions. It can be noted that user performance generally deteriorates as ambient illumination increases from dark to bright. It appears fairly intuitive as more light reflected off the screen will increase the difficulty of detection tasks. Users of the mobile device reported that the reflectivity of the screen was a major issue for the performance drop during the experiments. The observer was allowed to hold the device while performing the DENOTE experiment. Observers got the best results in dark conditions. Bright conditions which represent an outside environment, presented the highest level of difficulty. A relative performance drop exists for each observer.

The staircase results in Fig. 6 suggest that the signal detection threshold increases in a highly lit ambient environment. The mobile screen reflects



(a) DENOTE



(b) Staircase

Figure 6: User performance for embedded text detection and average threshold estimates conducted for 3 different ambient lighting conditions using the Nexus One.

(mostly specularly for this technology) the ambient lighting into the user’s eye which makes it difficult to spot the signal. The subject settles on a higher threshold compared to when he/she performs the same experiment in the dark. The staircase method is time consuming and requires test trials to determine starting thresholds for different conditions. The observers were unable to locate the signal within a reasonable gray level distance for the first and second luminance class in bright conditions. The results presented in Fig. 6 are an aggregate of performance across all the five luminance class images viewed by the observer, however, Fig. 6 presents data for only the first luminance class for two ambient conditions showing prominently similar trends for all subjects. We analyzed data for each luminance class for identifying similar patterns in the data. It is our conclusion that though some background images are harder to view in one condition compared to the other, no consistent trends were detected for all observers for luminance classes except for the first one.

### 3.1. DENOTE vs Staircase

Our analysis shows that DENOTE can be a better substitute as a practical test in ambient illumination conditions for mobile displays before clinical practices. The staircase technique used to estimate the visibility threshold is analyzed against out DENOTE technique used to estimate suitability of ambient conditions for medical imaging on mobile display device. Fig. 6 shows that the results are coher-

ent for observer response when using both of these methods. Performance for all observers decreases when ambient conditions get bright. It should be noted that the staircase methodology is a simpler task of detection compared to DENOTE. DENOTE relies on character recognition where an observer can fail a trial even if he observes some presence of a signal, but cannot correctly identify it. Fig. 6 presents results for all the luminance classes combined and the DENOTE data for dark and bright environment follows similar trends with that of the staircase results.

### 3.2. DENOTE for exploring other issues with mobile displays

We explored some additional issues related to mobile image viewing such as the use of multiple devices as well as free movement versus a fixed positioned mobile device. Fig. 8 presents results for the two devices used in this study. It was found that the two male subjects; one with perfect vision and the other with corrected vision performed better using the Galaxy S compared to the Nexus One. We express the results in terms of  $\Delta P$ , the difference between percent correct achieved on one device and that obtained with the other device. Fig. 8 represents a difference graph where bars above the x-axis indicate better observer results when the Nexus One was used while the bars below indicate otherwise for particular conditions.

The experimental results identify that some mobile devices might be more usable under a par-

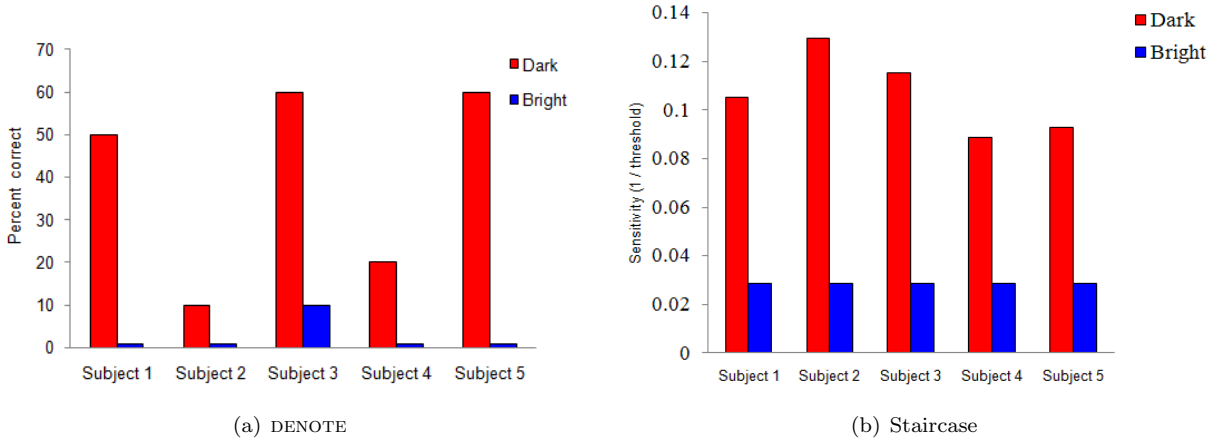


Figure 7: User performance for embedded text detection and average threshold estimates analyzed only for the first luminance class.

ticular ambient environment while others would not. It should be noted that this experiment is not a brand/manufacture comparison but it only highlights the fact that signal visibility in ambient conditions varies from device to device. This trend was present when the device was fixed in a static/stationary position in front of the observers as well as when they were handed the device to use it at will in order to successfully perform the task. Though a general trend for both the subjects agrees with the fact that a device with higher reflectivity hinders viewing, subject 1 in the bright conditions has an abnormal occurrence of being more successful with the Nexus One (the device with higher diffuse and specular). One reason for this occurrence is the fact that the results presented in the graph are an aggregate for all the five luminance classes, and subject 1 was able to perform better only for luminance class 5 for the Nexus One which lead to a peak in the aggregate result analysis. Secondly, since experiments conducted in the bright ambient conditions were actually performed outside, the illuminance conditions were not consistently constant. Some factors (moving clouds) might have caused a reduced illuminance on the device display which helped the observer achieve better results. This form of noise can be removed through consistent monitoring of the ambient conditions synced with the DENOTE experiment running in parallel. We plan to include this technique in our future experiments.

This pilot study also aimed to identify if non-static device movement and use of different devices

has an affect on observer performance. The device being static or non-static did not effect observer performance. This attribute is very user specific as an observer moving the device can find a sweet spot to identify characters while the other could not. Moving the device can dramatically change the impact of light incident on the display, thus how an observer is moving the device is a complex variable to analyze and record. We intend to observe the effect of changing ambient environment by recording it in real time while the observer is conducting the experiment, but that research activity remains as part of our future work.

### 3.3. Search tasks complicates the test

We conducted an additional set of experiments following a Signal Known Exactly with Search (SKEWS) paradigm. For these experiments the a reference signal is displayed to the observer at a known position, while the observer has to search the display for such a signal occurrence. The position of the detection signal is unknown unlike the previous experiments mention in this study (no cross-hair).

The results suffered from high variability perhaps due to satisfaction of search(14). If the attention level of an observer decreases as he or she scans the whole image, high variability is added to the experimental results. This experimental layout also suffers from a false positive and a false negative uncertainty. The user might get the number of signals right but there is no guarantee he saw the right signals to make that decision. He might be considering noise clumps as signals while actually missing the actual signal.

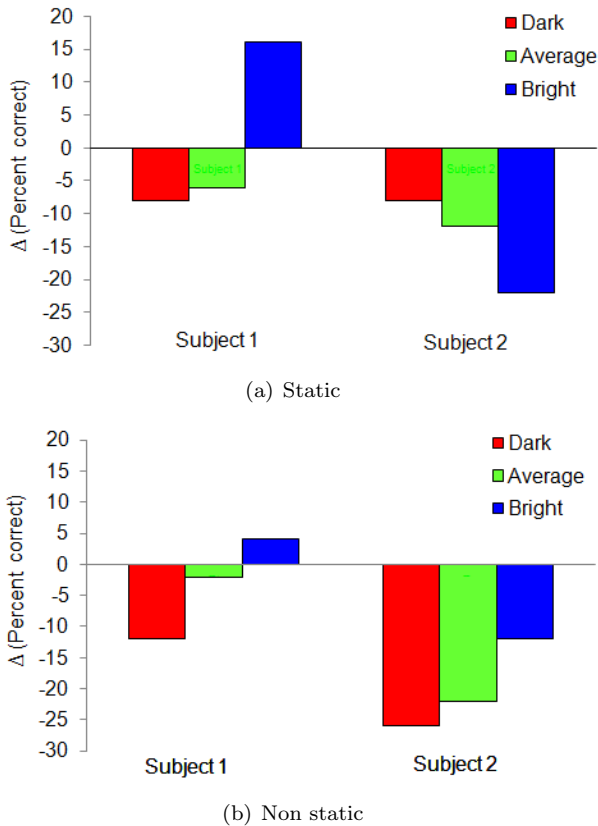


Figure 8: Results for  $\Delta P$ , the difference between observer performance for two mobile devices used in the experiment.

In one of the variations of the search experiments, we allowed the user to touch the Gaussian signal on the screen for identification. This test sounds intuitive as most mobile devices have touch sensitive screens. This method added two additional variables to the experiment. First, the observers were unable to click a 10x10 signal on the screen due to parallax error while holding the device. Multiple attempts were needed for some observers who identified the spot early in the trial but took 10 seconds to touch the screen for identification affecting observer attention. The second problem was related to the highly specular screen present in the device. One touch can leave marks on the screen that need to be wiped repeatedly for every trial. We noticed that the condition of the screen has a significant role in the observer performance. Scratches and finger prints present added difficulty to the user especially when environmental lighting is present. This is one of the areas of this study that needs further research. Fig. 9 shows that user performance

for all conditions was lower when compared to each other. The worst performance occurs when using the mobile device in average conditions specifically with the fifth background luminance class. The results show a substantial uncertainty (though 10 trials were conducted for each single data point). This added variability in the results might present a challenge for search tasks on mobile devices.

The SKEWS experiment was also limited by the number of touch events you can perform, to halt the observer from touching the whole screen and getting a definite positive result. The DENOTE test outputs results based on a binary format. If the observer gets the entire string of characters correct, the answer is correct, otherwise it is false. SKEWS was designed around the idea that observers can have partial correctness. Thus the variability amongst successive trials is higher if the observer sees all the signals in 1 trial and does not see any in the other. This problem was the resultant of all the issues discussed earlier in this section. When observers keep pressing the wrong position, they are forced to the next trial due to touch restrictions of the trial. Thus if the observer was able to see and touch most of the signals successfully in the earlier trial, the current trial had swayed the results by a considerable factor resulting in a higher variability in the output results.

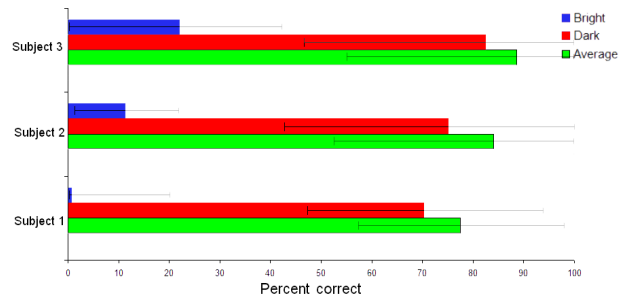


Figure 9: Results from the SKEWS experiment.

#### 4. Discussion and summary

We have explored some factors that must be taken into account for medical imaging on mobile displays. We also introduce a practical test that can be used to detect the perceptual limit of a human observer under varying ambient conditions. The results unequivocally show that detection tasks are hindered to some extent by the ambient lighting

conditions in mobile display devices. The experiments conducted indicate that recognizing letters is a harder task compared to identifying the presence of a Gaussian signal on the display thus making DENOTE a more robust and effective testing technique. Suitability of devices for use in clinical practice is highly dependent upon characteristics of the mobile displays' response to environmental lighting. Our results indicate that for devices (smart phones) that though offer the same feature set, might still have very different display characteristics. This presents the problem of being able to use one device in certain illumination conditions and refraining from the use of another.

We have presented the DENOTE method as a surrogate for a more accurate but time consuming methodology. In average ambient conditions however, sensitivity is higher compared to the percent correct for DENOTE. This occurrence can be explained by the fact that DENOTE is a harder test, asking observers to characterise the string on the screen correctly, not merely detect its presence (as in the case for staircase). We analyzed the DENOTE results on the basis of a binary confidence threshold. If any 1 of the four string characters was identified incorrect, the observer scored a 0 percent success for that trial. This indicates DENOTE being a more robust test that correctly tests observer response and its characterization part giving a better idea of their visual perception.

The luminance values for average and dark conditions might be too similar to extract clear trends and relationships between contributing factors in the experiments. Results for bright conditions stand out indicating that the effect of an increase in luminance to 60,000 - 80,000 lx causes the performance to significantly decrease. We identify that there is some effect of background luminance on the observer performance as some noisy patterns are easier to view in particular ambient conditions. Our study did not focus on which gray level ranges are difficult or easier to view for which ambient illumination but we identify this as a factor. Our conclusions about this subject are drawn from the fact that all observers recorded a 0 percent success for luminance class 1 and 2 in the bright conditions and their visual threshold in the staircase experiment was above 35<sup>th</sup> gray level. They were however able to gain some percentage correctness viewing images in the next 3 classes and coherently their visual thresholds were below 35<sup>th</sup> gray level in the staircase experiments. This case only occurred

for the bright ambient condition.

When conducting the additional experiment exploring issues such as static and non-static handling of the device, we conclude that data from static and non static movement of respective devices did not provide sufficient trends to be classified as a factor. Observers adopted to each condition with personal variability thus the results were not clear. However, we believe that devices with higher reflectivity (specular and diffuse) are an issue for consideration for medical image viewing and that not all devices might be used with in certain ambient condition. The DENOTE test helps identify the observer detectability limit, which might allow one device to be used in a certain conditions while deeming another device unusable due to its display characteristics.

The current experiments were aimed at comparing DENOTE with the staircase technique in order to verify trends. These two methods in our experiments have advantages and shortcomings. In the staircase method for threshold estimation, the signal location and amplitude are known to the observer through the cross hair indicator and the reference image. This technique does not test the observer for a search task with out added complexities such as the *nby*m AFC (Alternate Forced Choice) approach. The screen needs to be divided into tiles where each tile has a signal in the center. This process also complicates the output analysis. However, the embedded text experiment is a simple task to assess the use of any mobile display which come in vastly different specifications (resolution, luminance). The test can be performed quickly with out any equipment (chin rest, luminance meter) and results from such a study can help identify the limitations of a certain mobile display in any condition. DENOTE can also be easily extended to a search task methodology. Some alterations to DENOTE can allow us to test observers response for search tasks. Placing a character at a random location on the screen can help identify if the observer successfully identifies the target. This test provides a higher merit for testing since most problems discussed in section 3.3 will be dealt with. The occurrence of false positives or false negatives tends to zero if characters are displayed on the screen with no repetition. When using a Gaussian spot target, the observer can identify the signal with a lesser degree of confidence. In our DENOTE search experiment, we improve the degree of confidence when the observer identifies a target on screen by checking if it was the correct character or not. Search

experiments do not require a tiled split of the display screen. This testing technique remains as future work which will help model a better and more improved DENOTE.

Testing methodologies for mobile displays have some added complexities. If you restrict the viewing angle of the mobile display in front of the user, it denies the user the advantage of adjusting the viewing and makes the experiment inconsistent with its traditional use. On the other hand, if free movement of the device is allowed for an experiment, some observers might move the device into a position where experimental factors might begin to change. Using one hand to shield light or moving the device in a position where the ambient illumination changes significantly are factors hard to control during experimentation. We did however notice that fixing the mobile device restricts many viewing angles which the observer can use to their advantage but some observers might be good at locating good viewing angles than others.

## 5. Conclusion

Mobile displays bring added complexity along with their portability benefits to medical imaging. Porting medical image viewing to mobile devices is not a straightforward task. Users must be aware of certain limitations and characteristics of their mobile devices. Here, we show that mobile devices have considerably higher specular reflectivity than medical displays. This factor might account for the observed increase of the perceptual limits of mobile displays in high illumination environment. Our DENOTE method, a noise-embedded text detection task, shows promise as a practical tool for determining the ambient illumination conditions for viewing medical images in mobile devices. Our results suggest that some mobile displays might be less suitable than others in certain ambient lighting conditions. Our study did not account fully for other factors affecting performance on mobile devices including adjusting viewing direction which is known to also affect target visibility.

## Acknowledgements

The authors acknowledge support from the Research Participation Program administered by ORISE through an interagency agreement between DOE and FDA. The mention of commercial products herein is not to be construed as either an actual

or implied endorsement of such products by the Department of Health and Human Services. This is a contribution of the Food and Drug Administration and is not subject to copyright.

## References

- [1] R. Toomey, Handheld devices for radiologists: as good as monitors, *Imaging Med* (2010).
- [2] R. Jarmo, I. Eila, J. L. et. al., Initial experience with a wireless personal digital assistant as a teleradiology terminal for reporting emergency computerized tomography scans, *J. Telemed.* (2000) *Telecare* 6, 45-49.
- [3] Y. LG, W. DR, A demonstration of instant pocket wireless ct teleradiology to facilitate stat neurpsurgical consultation and future telemedicine implications, *Am. J. Emerg. Med* (2000), No.18, 423-426.
- [4] A. Badano, D. Sharma, G. Guarnieri, G. Ramponi, L. Albani, Effect of case variability on the measurement of reader performance when comparing medical display technologies: application to the study of high-luminance-range displays, *Proceedings of SID* (2009).
- [5] M. Flynn, A. Badano, Image quality degradation by light scattering in display devices, *Journal of Digital Imaging* 12 (1999) 50–59.
- [6] Y. Park, C. Li, Y. Kwak, D. Park, C. Kim, Applying CIECAM02 for mobile display viewing conditions, 15th Color Imaging Conference Final Program and Proceedings (2007).
- [7] Y. J. Kim, M. R. Luo, W. Choe, H. S. Kim, S. O. Park, Y. Baek, P. Rhodes, S. Lee, C. Y. Kim, Factors affecting the psychophysical image quality evaluation of mobile phone displays: the case of transmissive liquid-crystal displays, *J. Opt. Soc. A.* 25 (2008) 2215–2222.
- [8] A. Badano, M. J. Flynn, J. Kanicki, High-Fidelity Medical Imaging Displays, 2010.
- [9] E. Samei, A. B. A, D. Chakraborty, K. Compton, C. Cornelius, K. Corrigan, M. Flynn, B. Hemminger, N. Hangiandreou, J. Johnson, M. Moxley, W. Pavlicek, H. Roehrig, L. Rutz, J. Shepard, R. Uzenoff, J. W. J, C. Willis, Assessment of display performance for medical imaging systems, Report of the American Association of Physicists in Medicine (AAPM) Task Group 18, Medical Physics Publishing, Madison, WI, AAPM On-Line Report No. 03 (2005).
- [10] S. Ishihara, Tests for colour-blindness, Handaya, Tokyo, Hongo Harukicho (1917).
- [11] T. Cornsweet, The staircase-method in psychophysics, *The American journal of psychology* 75 (1962) 485–491.
- [12] J. Nachmias, R. Steinman, An experimental comparison of the method of limits and the double staircase-method, *The American journal of psychology* 78 (1965) 112–115.
- [13] M. García-Perez, Yes-no staircases with fixed step sizes: Psychometric properties and optimal setup, *Optometry and Vision Science* 78 (2001) 1.
- [14] S. Samuel, H. L. Kundel, C. Nodine, L. C. Toto, Mechanism of satisfaction of search: eye position recordings in the reading of chest radiographs, *RadioGraphics* (1995) 895–902.