

The Effect of Ambient Illumination on Handheld Display Image Quality

Peter Liu · Fahad Zafar · Aldo Badano

Published online: 10 October 2013
© Society for Imaging Informatics in Medicine 2013

Abstract Handheld devices such as smartphones and tablets are becoming useful in the medical field, as they allow physicians, radiologists, and researchers to analyze images with the benefit of mobile accessibility. However, for handheld devices to be effective, the display must be able to perform well in a wide range of ambient illumination conditions. We conducted visual experiments to quantify user performance for testing the image quality of two current-generation devices in different ambient illumination conditions while measuring ambient light levels with a real-time illuminance meter. We found and quantified that due to the high reflectivity of handheld devices, performance deteriorates as the user moves from dark areas into environments of greater ambient illumination. The quantitative analysis suggests that differences in display reflection coefficients do not affect the low illumination performance of the device but rather the performance at higher levels of illumination.

Keywords Clinical image viewing · Data display · Diagnostic display monitors · Diagnostic image quality · Digital display · Display device · Human–computer interaction · Human visual system · Image display · Liquid crystal display (LCD) · Medical displays · Personal digital assistant (PDA) · Visual perception

P. Liu · F. Zafar · A. Badano (✉)
Division of Imaging and Applied Mathematics, Office of Science and Engineering Laboratories, Center for Devices and Radiological Health, FDA, 10903 New Hampshire Ave, Silver Spring, MD 20993, USA
e-mail: aldo.badano@fda.hhs.gov

P. Liu
Department of Computer Science, University of Maryland, College Park, MD, USA

F. Zafar
Department of Electrical and Computer Engineering, University of Maryland Baltimore County, Baltimore, MD, USA

Introduction

The use of handheld devices has become prominent. Many advanced applications developed for mobile platforms provide critical availability of tools for physicians, radiologists, and scientists. With the advancement of display technology, current generation handheld displays have higher number of pixels, greater pixel density, and wider color gamut. However, in order for a handheld device to serve as a reliable tool for medical image analysis, it is critical that the display performance be good and consistent in different environments. While workstation medical displays have been extensively studied and analyzed [1, 2], mobile displays have not yet been proven to have the ability to accurately provide quality information to the user for correct diagnoses based on imaging data for all modalities [3]. In fact, even if a handheld device provides promising results when compared to a medical display device in standard illumination conditions, changing environments can dramatically affect image quality [4].

How ambient illumination levels affect visual task performance on handheld displays has yet to be studied [5]. In a recent report by Lin and Kuo [6], the authors studied the image quality of a handheld display device under different ambient illumination levels. Considering the physical characteristics of color temperature, luminance, contrast, and display resolution, Lin and Kuo concluded that outdoor environments around 7,000 lx are not suitable for handheld displays. In a critique of Lin and Kuo's study, Badano expressed that there exists a difference between the metrics of preference and image quality [7]. Lin and Kuo's study relied on preference as their metric for determining image quality, which can be highly subjective. On the other hand, preference can depend on several other factors (unrelated to performance) suggesting that analyzing image quality through a task-based approach such as the one developed by Vogel et al. [8] is better suited and provides useful quantitative information.

In this paper, we discuss the effect of ambient illumination conditions on visual task performance. Using the DENOTE technique developed by Zafar [5], a methodology based on noise-embedded text detection, we tested handheld display devices and user performance in several illumination conditions ranging from 0.01 to 31,622 lx.

Methods

Our methodology is based on the methods described in Ref. [5]. A human subject is presented with a task to detect and distinguish four characters embedded in a white noise background. We tested two current-generation smartphone devices: A and B, with 3.7 and 4-in. screens measured diagonally, respectively. Both of these devices have active-matrix organic light-emitting displays (AMOLED) with a 480 by 800 pixel array and a PenTile (RGBG-Matrix) pixel arrangement. The AMOLED display on device A is categorized as Normal TSP (Touch Sensor Panel). The screen on device B is labeled Super AMOLED and advertised as having a brighter screen with less overall reflection and reduced power consumption, resulting from the capacitive touch screen being directly integrated onto the display panel (On-Cell TSP).

The setup for observer experiments included illumination conditions in the super bright (3,000–30,000 lx), medium bright (1,000–3,000 lx), average (300–1,000 lx), medium dark (10–300 lx), and super dark (0–10 lx) illuminance levels binned in $100.5n$ lx intervals (see Fig. 2), with n in the range of 0–9.

Three observers performed the experiments in five illumination conditions simulating dark room (“super dark” and “medium dark”), office (“average”), and outdoors (“medium bright” and “super bright”) environments. The technique was implemented using a Java application and the Android SDK 2.2 that can be ported to any compatible tablet/handheld/device running the Android OS. No attempt was made to correct the look up table used in converting gray level values to luminance.

Physical Characterization

The reflectance of display devices can be separately characterized by two components: specular and diffuse scattering. These components distinctly affect image quality and require different experimental measurement methods for their characterization. Our methodology for measuring the coefficients was adopted from Ref. [9] (see Fig. 1.)

The diffuse coefficient was measured by the following procedure. First, three screws were placed on the back of a white Styrofoam box (area of size 16×16 in.) 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

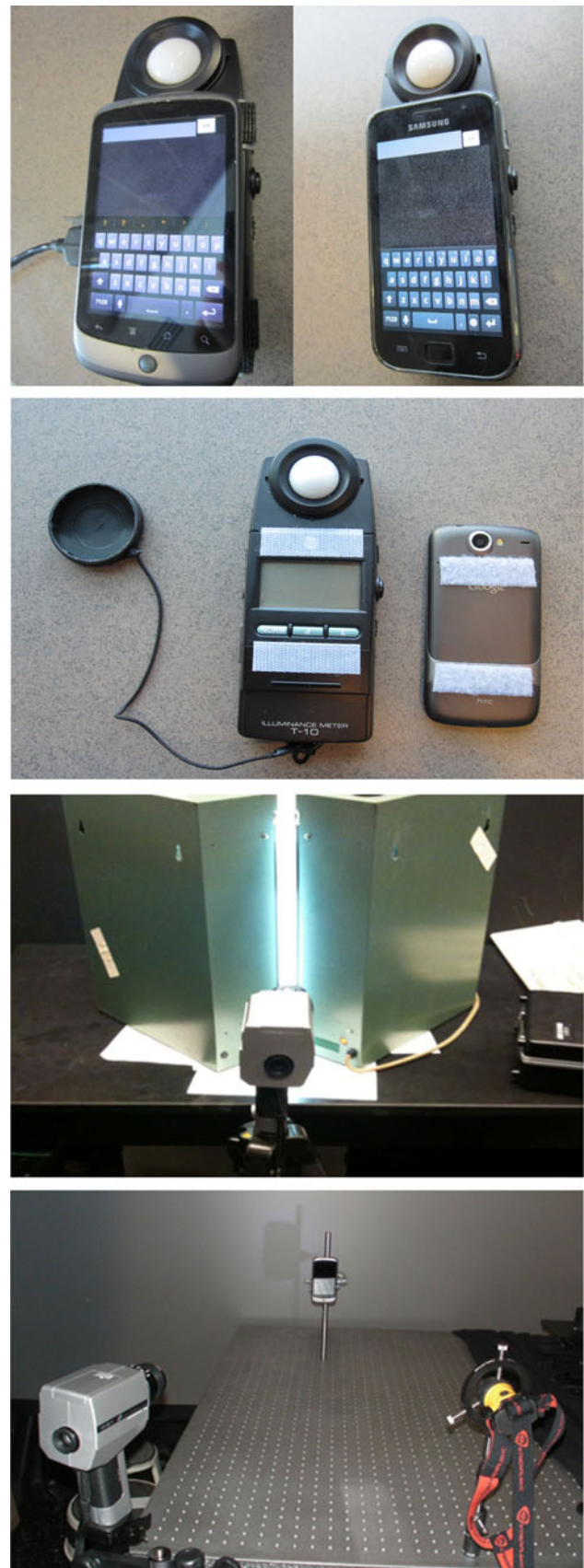


Fig. 1 Experimental setup

cover the openings. The T-10 illuminance meter (Konica Minolta, New Jersey) was placed into the Styrofoam box to measure the illuminance inside the box (see Fig. 1), which measured at approximately 6,580 lx 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. For each measurement, a different mobile device was placed at the back of the box. The box was closed off with the fluorescent lights and the luminance response of the diffuse reflection off the screen was measured using the Minolta CS-100 spot photometer through the gap between the fluorescent lights. Finally, the coefficient of diffuse reflectance was calculated as $R_d=L_r/I_b$, where L_r is the reflected luminance measured with the spot photometer and I_b is the illuminance at the face of the devices when the illuminating box is on. The units of R_d are $\text{cd}/\text{m}^2 \cdot \text{lx}$ (Fig. 2).

The specular coefficient was measured by the following procedure. First, a device holder was set up and the mobile device was placed in the clamp. The device screen was consistently placed perpendicular (90°) to the table. Next, a 7-LED flashlight was fixed 78.7 cm away from the device holder at an angle of 15° from the normal relative to the center of the screen, pointing at the center of the mobile display. The CS-100 photometer was similarly fixed 78.7 cm away at 15° from the normal relative to the screen, and then focused at the center of the mobile display. Next, the CS-100 photometer was placed 157.5 away from the LED light and a direct luminance response from the LED lights was measured. Finally, the specular reflectance coefficient was calculated as $R_s=L_r/L_d$ where L_r and L_d 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 %. In the specular reflection test, 1° off in alignment can change the reflected luminance response by approximately $30 \text{ cd}/\text{m}^2$. In the diffuse reflection test, the luminance response can change by approximately $15 \text{ cd}/\text{m}^2$ with 1° off in alignment.

Perceptual Testing

Through our testing, we determined at which illuminance level the user performed using each handheld device. The DENOTE method is practical and most appropriate for testing in dynamic reading conditions. The observer was asked to identify four alphanumeric characters displayed in the middle of the screen with no time limitations. The mobile application DENOTE was preloaded onto each handheld device before experimentation. The DENOTE method implementation was similar to its initial build, with four random characters being displayed on the screen from a dataset that included the characters A–Z, a–z, 1–9. In order to more accurately assess user performance, we modified the technique so that all characters that may cause confusion for the observer ([Z, 2], [1, l, i, l], [P, R], [p, q], [Q, O, 0, o], [B, 8], [2, 7]) were omitted. We used 0–51 subset of gray levels for the background noise levels for each image tested and repeated this subset 10 times for each illuminance level at 50 trials each, for a total of 500 trials per experiment. Each subject ran the experiment on both devices. All observers were placed in the particular environment for 5 min prior to experimentation. The set of observers used included three males, ages 18–25, one with corrected vision. The observers were knowledgeable of the tasks they needed to perform. The device was also set at maximum brightness with the automatic brightness adjustment turned off.

For each experiment, the subject ran the DENOTE application on a given handheld device. The observers were given the task to identify the four randomly generated alphanumeric characters placed in the middle of the screen under different ambient illumination conditions. The subject was allotted an unrestricted amount of time to complete the task. An overseer was present to verify that the task was executed properly. We believe that display degradation due to fingerprints, oily smudges, and dust was not a significant factor that contributed to user performance after users were provided with screen wipers to clean the screen.

One issue that hindered user performance during experimentation, however, was glare or direct reflection of light sources in bright and very bright viewing conditions, which we defined as greater than 1,000 lx. Users noted that it was nearly impossible to discern any characters on the screen. While this glare is due to the specular reflectivity of the devices, with device A being the more reflective of the two tested devices, users attempted to alleviate this issue and reduce glare by tilting the screen in varying positions and

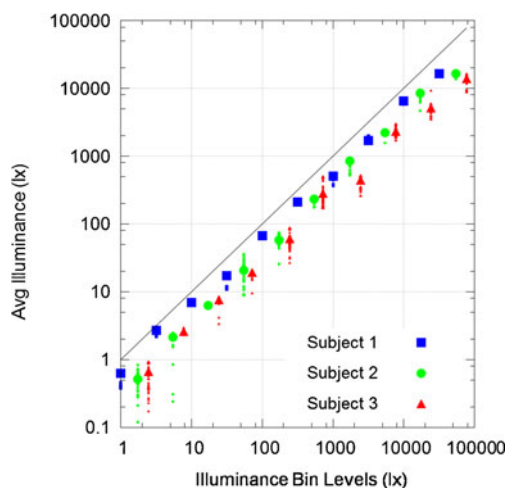


Fig. 2 Average illuminance (lx) for each test subject under each bin for both devices. Each experiment took place in a controlled environment (either in a dark lab, office, or outside space) with increasing illuminance beginning in the dark environment. The illuminance distribution was binned into ten illuminance levels (x axis). Each bin consists of *three points* representing the average illuminance for each of the three subjects in that particular illuminance range. Each point was displayed along all individual illuminance measurement data points



Fig. 3 Experimental setup. The handheld device was attached to the face of the T-10 such that the sensor was able to capture illuminance readings of the background. The T-10 was connected to a laptop computer through a proprietary RS232 cable because due to Android OS limitations, the handheld devices were not able to directly interface with the T-10. The T-

10 reading and writing program was run on the laptop so that it was able to read data from the T-10 and write measurements into a text file which was used for data analysis. A time stamp synchronized each user response from the DENOTE application with the exact illuminance measurement

clearing the screen of any fingerprints, oils, and dust that might have become very noticeable under high illumination.

Illuminance Measurement

In order to measure ambient light levels, we used an illuminance meter (Konica Minolta T-10). Prior to each experiment, the illuminance meter was calibrated appropriately. Because the meter does not automatically store the illuminance measurements on the instrument itself, a Java program was developed to read each set of measurements and save them in a file on a separate computer. The program was designed to read four measurements per second, with a time stamp to accurately synchronize each user response to its corresponding illuminance measurement. Because of OS firmware limitations on the Android platform, we were not able to install this automated reading program directly into the device and have the device communicate directly with the illuminance meter. The data sent from the illuminance meter thus needed to be stored on a separate laptop computer. The illuminance meter needed to be tethered to the laptop computer in order to read and send

data between them. This, to some degree, hindered the mobility of the user when conducting the experiments. To maximize the efficiency of use of the T-10, we firmly attached the handheld device onto the meter (see Fig. 1). This allowed for increased accuracy of measurements and adjustability for the user. For each experiment, the handheld device was not stationary; the subject was allowed to change the viewing angle and direction and move the device freely (Fig. 3).

Statistical Analysis

We collected user performance data (results of the DENOTE test) for each subject, each device, and each illuminance bin level (Fig. 4). Because the responses can be either correct or incorrect, we can estimate the uncertainty in user performance per subject with a binomial variance model: $2P(1-P)/Nc$ [10], where r is the reader, d is the device, P is the performance, and Nc is the number of observations, in this case 50. The variance was used to calculate the uncertainty seen in Fig. 4.

We then calculated the difference in user performance ΔP_{rd} for each reader r between the responses recorded with device

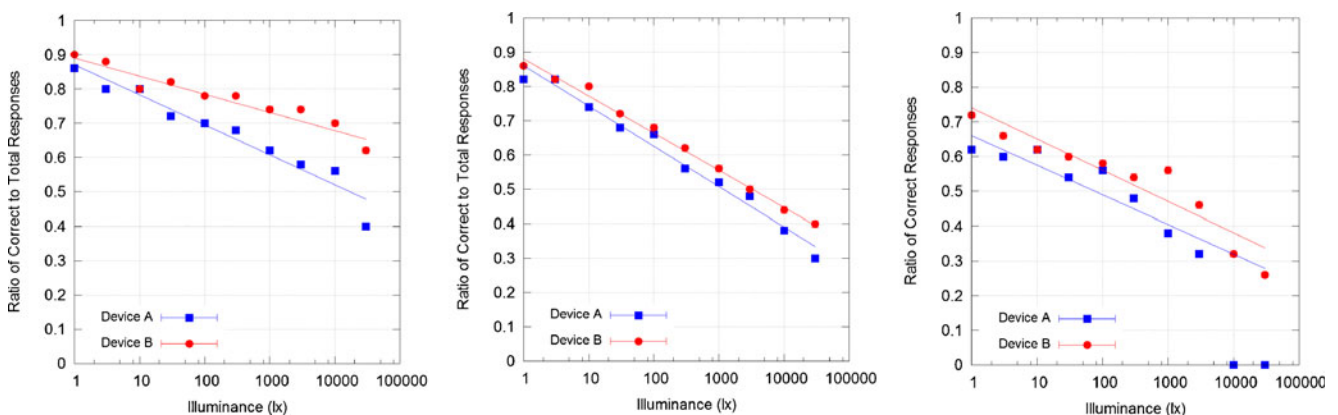


Fig. 4 Ratio of correct to total responses comparison of the two devices

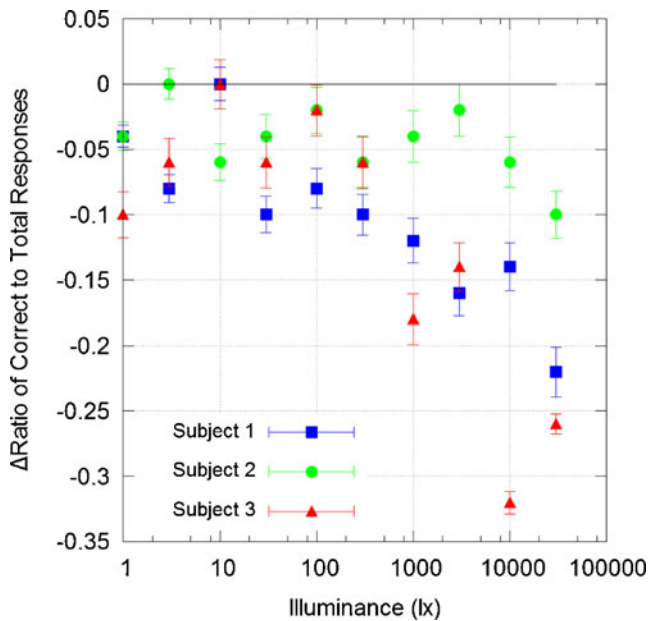


Fig. 5 Difference of ratio of correct to total responses of device A and device B. The illuminance distribution was binned into 10 illuminance levels (*x* axis). The *negative points* show that in almost all cases, the user performed better when using device B than when using device A

A and the responses recorded with device B (Fig. 5). A ΔP_{rd} value of zero indicates no effect while negative values indicate a degradation of performance. The error in ΔP_{rd} was computed as the sum of the variances corresponding to the data for device A and device B (Table 1).

This binomial variance model was used to calculate the sample standard deviations in Fig. 5 and Table 3.

Results

Table 2 shows the reflectance coefficient measurements performed in our laboratory. We can observe that device A has over 3× higher specular and diffuse reflection coefficients. This leads to increased reflectivity and glare in all ambient conditions. The specular and diffuse reflection coefficients of

device A are among the highest of any handheld device display we have measured.

Figure 4 shows results for the 3-observer study comparing the ratio of correct to total responses (the number of correct responses out of 50) of each level for each device. Each data point in Fig. 4 plots represents 50 readings. It can be noted that user performance generally deteriorates as ambient illumination increases from dark to bright. This finding is intuitive as more light reflected off the screen will increase the difficulty of detection tasks by decreasing target contrast. The data shows that observers achieved the best detection performance in dark conditions while bright conditions, which represent outside environments, presented the highest level of difficulty. This can be seen in a relative performance decrease that exists for each observer. In addition, users also performed better when using the improved screen technology of device B over the screen of device A. Figure 5 shows the difference of the ratio of correct to total responses by subtracting the ratio of correct to total responses of device B from device A. In Fig. 5, all points are less than or equal to zero, which indicates that in all circumstances, observers performed the same or better when using device B. Additionally, we note a trend among the data which shows that in environments of greater illuminance, the difference in performance is higher for all of the subjects tested. Each data point in Fig. 5 plot represents 100 readings.

Based on the data we collected, we calculated the correlation coefficient for a log-linear fit. A log-linear model was used because it was simple and appropriately matched the trend of our data. Table 3 presents the correlation coefficients for each experiment for each subject. The data confirms that there is a negative correlation between illuminance and user performance for each device. A larger sample of observers would be needed in order for this trend to be more definitive.

Discussion

We have explored how the DENOTE method can be used to evaluate screen quality performance on handheld devices

Table 1 Display specifications measured (*) or taken from official product documentation

	Device A	Device B
Manufacturer	Google/HTC	Samsung
Model	1st Gen	GT-I9000
Display	Normal TSP (AMOLED)	On-Cell TSP (SAMOLED)
Resolution	480×800	480×800
Pixel arrangement	PenTile	PenTile
Pixel density	252 ppi	233 ppi
Luminance range*	36,379	51,155
Minimum luminance*	0.0058 cd/m2	0.0058 cd/m2
Maximum luminance*	229 cd/m2	296 cd/m2

Table 2 Specular and diffuse reflection coefficients for device A and device B. The uncertainty was calculated by taking the average of ten measurements

	Device A	Device B
R_s	0.071±0.002	0.025±0.001
R_d (cd/m ² ·lx)	0.064±0.001	0.027±0.001

Because device A has higher specular and diffuse reflection coefficients, its display has higher reflectivity, and therefore increased glare

under varying ambient conditions. This study allowed us to demonstrate and quantify the effect of illuminance levels on the ability of human observers to correctly perform a detection task on a handheld device. We analyzed the results based on a binary confidence threshold [5] wherein all four alphanumeric characters must be correctly identified in order to achieve a correct response, making the detection task difficult and more effective for testing.

We compared our results with the previous work presented by Vogel et al. and found our results to be somewhat conflicting with their findings. In the previous study, the authors concluded that illuminance had a minor impact on mobile-display performance [8]. Our study, however, suggests the opposite. Screen reflectance is highly dependent on display technology. It is interesting to note that even though devices A and B have displays with the same pixel array size, user performance on device B was overall greater. This apparent discrepancy can be explained by pointing out that Vogel et al. defined “bright” conditions to be approximately 126 lx, which may not be representative of many outdoor environments. In this illuminance range, their results do not show a drastic change in correct observation ratio with increased illuminance. In our experiments, we expanded the range of illuminance to over 30,000 lx to better test the devices under more varied and representative environments. Because of the expanded illuminance range, we are able to more accurately determine the perceptual limits on mobile displays. Lin and Kuo determined that environments around 7,000 lx

are not suitable for using handheld displays [6]. In addition to supporting this claim, our data suggests that in environments over 1,000 lx, user performance begins to decline at a faster pace, with most users only achieving near 50 % correct in our text detection test and noticeably worse as the environment illuminance increases. From this, our results suggest that it might not be optimal for radiologists to read images in environments with an illumination greater than 1,000 lx.

Subjects reported that the reflectivity of the screen was a major issue for the decrease in performance during the experiments, and this is reflected in our measurements. For smartphones in particular, display characteristics improve with each device generation. Screen technology plays an important role in image quality and screen reflectivity. We showed that since device A has higher reflectivity in both specular and diffuse modes than device B, subjects generally performed poorer when using device A. This also indicates that in general, devices with higher reflectivity may be neither optimal nor appropriate for medical image viewing. The DENOTE test can help identify an observer detectability limit, which may allow one device to be appropriate for use in certain conditions while deeming another device unusable due to its reflection characteristics [5].

In the analysis of our measurement results, we saw that devices A and B have different reflectance coefficients (see Table 2). In Fig. 4, we noticed that for all subjects, performance was higher when using the device with the lower reflectance coefficient. However, we also noticed that the performance trend was relatively similar for both devices. This suggests that the reflectivity of the screen has a similar effect on image quality degradation and decreased user performance as ambient illuminance levels increase. In other words, due to other display characteristics, the extrapolated dark-environment performance is higher for device A. As illuminance increases, because device A has lower reflectance coefficients, user performance is greater when using device A. This user performance trend is maintained across the illuminance range. However, because the slopes of the log-linear fit

Table 3 First bin, intercept point, slope of the log-linear fit model (see Fig. 4), and correlation coefficients for each subject for each device, taking into account the ratio of correct to total responses and the illuminance. The first bin and intercept information shows how the user performed under the optimal condition for viewing on mobile displays: very dark environments. We determine the slopes of the fitted log lines for

each subject to be relatively similar, which show that the effect of ambient illumination on user performance degradation is similar even though the devices have differing reflectance coefficients. Because each coefficient is close to -1, we can infer that the ratio of correct to total responses and illuminance are strongly inversely correlated

Device	Subject 1		Subject 2		Subject 3	
	A	B	A	B	A	B
Performance first bin	0.860±0.005	0.900±0.004	0.820±0.005	0.860±0.005	0.620±0.009	0.720±0.008
Intercept of fit	0.870±0.014	0.890±0.009	0.860±0.015	0.880±0.008	0.660±0.010	0.740±0.005
y Slope of fit	-0.038±0.003	-0.023±0.001	-0.051±0.003	-0.047±0.014	-0.037±0.010	-0.039±0.005
Correlation coeff.	-0.82	-0.79	-0.75	-0.69	-0.81	-0.83

lines are relatively similar, we cannot conclude that performance degradation for device A is worse than for device B with increasing illuminance. We can see in Fig. 5 that as the illuminance increases, the difference of performance between devices A and B also increases. This shows an increased decline in performance of device A that is not evident in the log-linear fit curves. We plan on running more tests to study this trend.

Using the DENOTE method to test screen quality has limitations. Motion of the device allows the subject to move the device into a position where experimental factors may be altered. We attempted to measure any changes by attaching the device directly onto the illuminance meter. It should be also noted that some users were more adept at using full-screen, touch-based devices and discerning characters that are imbedded in noise. All observers did not use the same viewing angles throughout the trial due to personal preference. This change in viewing angles and viewing distance may affect the reflectance. This was an uncontrolled parameter of our experiment, since we allowed users to dynamically alter the position and angle of the devices during testing.

Another challenge of this study was controlling the illuminance levels. This was prominent in the medium bright and super bright illuminance levels (greater than 1,000 lx). While it may be advantageous to run the experiments in an outdoor environment, we found that it was very difficult to maintain within the level, as sunlight is often and rapidly changing. To address this issue, we used high-power Tungsten lights (Lowel Tota-Light, Lowel-Light Manufacturing, Inc.) to simulate bright conditions. While this may be effective in simulating the illuminance conditions, we found that the effect from the light was not as severe as in outdoor conditions, in part due to the 3D distribution of the light sources.

In this study, we did not analyze the effect of display characteristics on user performance but rather the influence of illuminance using our DENOTE visual test. We attempted to address any variable factors by turning the display to maximum brightness for all trials with the automatic brightness adjustment turned off. In addition, we used the illuminance data gathered during the experiments to determine the exact illuminance for each user response generated by the use of time stamps. We also controlled the experiment in which the subjects were stationed in a certain illumination environment so that the illuminance stayed within the range allotted for each trial. In future testing, we plan on allowing the test subject to move around in several illuminance levels, encompassing dark, average, and bright environments, while running DENOTE tests. Ideally, the experiment would be run in, for example, a moving vehicle in normal use conditions where the illuminance is constantly changing. The illuminance meter would similarly record data in real time. We will then be able to further confirm our results regarding the effect of ambient illuminance on user performance and its validity in a real-world environment.

Conclusion

Handheld display devices for medical imaging exhibit characteristics that may in some circumstances significantly affect image quality. Users must be aware of the limitations when the screen is viewed under different ambient illuminations. We show that handheld displays have relatively high reflectivity that varies with screen technology and causes glare in high-illumination environments. We also show that this high reflectivity leads to decreased detection performance using the DENOTE technique, a noise-embedded text detection task, for analyzing which ambient illuminations are suitable for viewing medical images on handheld devices. We found that as illuminance increases, user performance for the detection task considered greatly decreases.

Acknowledgments 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. We thank the volunteers who participated in the experiments.

References

1. Badano A, Sharma D, Guarnieri G, Ramponi G, Albani L: Effect of case variability on the measurement of reader performance when comparing medical display technologies: application to the study of high-luminance- range displays. *Proc SID* 2009
2. Flynn M, Badano A: Image quality degradation by light scattering in display devices. *J Digit Imaging* 12:50–59, 1999
3. Toomey R: Handheld devices for radiologists: as good as monitors. *Imaging Med* 2010
4. Xu X, Kerofsky L: Improving content visibility for high-ambient-illumination viewable display and energy- saving display. *JSID* 19, 2011
5. Zafar F, Choi M, Wang J, Liu P, Badano A: Visual methods for determining ambient illumination conditions when viewing medical images in mobile display devices. *J SID* 2011
6. Lin P-H, Kuo W-H: Image quality of a mobile display under different illuminations. *Percept Mot Skills* 113:215–228, 2011
7. Badano A: Predicting perceived image quality: a critique of Lin and Kuo. *Percept Mot Skills* 114:236–238, 2012
8. Vogel R, Saha A, Chakrabarti K, Badano A: Evaluation of high-resolution and mobile display systems for digital radiology in dark and bright environments using human and computational observers. *J SID* 15(6):357–365, 2007
9. Samei E, Chakraborty ABAD, Compton K, Cornelius C, Corrigan K, Flynn M, Hemminger B, Hangiandreou N, Johnson J, Moxley M, Pavlicek W, Roehrig H, Rutz L, Shepard J, Uzenoff R, Willis JWJC: 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, 2005. AAPM On-Line Report No. 03
10. Badano A: Effect of slow display on detectability when browsing large image datasets. *J SID* 17/11