1040-5488/16/9412-0000/0 VOL. 94, NO. 12, PP. 00–00
OPTOMETRY AND VISION SCIENCE

ORIGINAL ARTICLE

Inter-Display Reproducibility of Contrast Sensitivity Measurement with iPad

Manuel Rodríguez-Vallejo*, Juan A. Monsoriu†, and Walter D. Furlan†

ABSTRACT

Purpose. To evaluate the reliability of measuring CS with uncalibrated iPads.

Methods. Six random iPads with retina display were calibrated with a colorimeter and the correlation between Luminance (*L*) and pixel level (*y*) was computed according to an exponential function. The mean and confidence interval (±2SD) obtained from the six iPads were calculated and the bit-stealing technique was applied for expanding *y* from 256 to 2540 possible values. The *L* of the optotype was computed for the selected contrast values (log*C*) represented in log units, using 0.1 log and 0.05 log steps. At each particular *y*, the contrast was considered reliable when the mean *L* plus 2SD was less than half the difference of luminance between two consecutive levels of contrast. Differences between the iPads for the *Experimental logC* were evaluated with the Friedman test.

Results. Luminance properties vary between devices, which were reflected in the computed *Experimental logC* ($p < 0.0005$). The contrast was found to be reliable for 0.1 log steps in the range from 0 to −2.2 log. On the other hand, for steps of 0.05 log, the contrast was only reliable for values ranging from 0 to −1.7 log.

Discussion. Both luminance and contrast steps differed between iPads with the same retina display, making it necessary to calibrate each display to achieve accurate luminance and contrast steps of 0.05 log units or less. However, for screening purposes utilizing contrast steps of 0.1 log unit or greater for a validated psychophysical test, calibration is not required to achieve accurate results across the displays described herein.

(Optom Vis Sci 2016;94:00–00)

Key Words: contrast sensitivity, iPad, reliability, calibration

Computer-based contrast sensitivity (CS) testing requires accurate calibration of the display. Contrast is computed as the difference between the stimulus and the background luminance (usually in cd/m^2) relative to the background for optotypes or to the mean level of luminance for sinusoidal gratings. CS is the reciprocal of this value; hence, the lower the contrast, the higher the CS.[1] The main reason to use an iPad for testing CS instead of another Android tablet is due to its expected less variation between units. First is because all iPad retina displays come from the same manufacturer (LG displays), and second is because iPad 3rd generation have not apparently changed their retina displays (LP097QX1) in the following models: 4th generation, Air, and Air 2. Authors who have previously characterized the iPads used in their studies have raised the assumption that uniformity in screen luminance properties might be assumed even though it has not been demonstrated yet.[2] The main aim of this study is to evaluate the need of calibrating the luminance response of an iPad before using apps devoted to the measurement of CS. This paper is, to the best of our knowledge, the first study that analyzes the CS values that can be reliably measured with an iPad without a previous calibration of its luminance response.

METHODS

Devices and Calibration

Six iPads with retina display, models A1458 (three units), A1430 (two units), and A1416 (one unit), were measured with a Syper4Elite colorimeter after setting *auto-lock* as *never*, fixing the brightness at 50%, and waiting 15 minutes to ensure the stabilization of the screen luminance.[3] Each iPad was in use for an unknown number of hours except a new one which was measured after receiving its first complete charge; this new iPad was re-calibrated 1 month after the first calibration to assess the variability attributable to the colorimeter. The characterization of

*MS
†PhD
Qvision, Unidad de Oftalmología, Vithas Hospital Virgen del Mar, Almería, Spain (MR-V); Centro de Tecnologías Físicas, Universitat Politècnica de València, Valencia, Spain (JAM); and Departamento de Óptica, Universitat de València, Burjassot, Spain (WDF).

the screens was conducted with an app that allowed to manipulate the y from 0 to 255 for each RGB isolated channel. The L was then measured for 52 evenly spaced values of y per channel. The mean L and the standard deviation (SD) obtained from the measurement of the six iPads were computed for each L level for the three RGB channels. The nonlinearity of the mean L value was estimated by fitting the scatterplot of the measurements to an exponential function as follows[1]:

$$L_R(y) = \left(\frac{y}{y_{max}}\right)^{\gamma} \quad (1)$$

where y is the digital level on the display, $y_{\max}$ is the maximum value ($y_{\max}$ = 255), $L_R(y) \in [0,1]$ is the corresponding relative luminance, and γ is the gamma value. With the data obtained from calibration, the relative R:G:B luminances were calculated by dividing $L(y)$ for each RGB channel by the sum of the three and approximating the result to the first decimal. These relative luminances were required for selecting the best matrix (δ_R, δ_G, δ_B) to apply the bit-stealing[4] technique and for later computing the Look up Table (LuT) as it has been previously described by To et al.[5]

The L corresponding to each contrast level commonly used for testing CS with optotypes were calculated considering Weber's law:

$$C = \frac{L_b - L_f}{L_b} \quad (2)$$

where L_b is the background luminance and L_f the luminance of the optotype or foreground. L_f was then computed for the logarithm of the contrast ($\log C$) ranging from $\log C$ = 0 to $\log C$ = −2.2 considering that L_b is 1 after normalization:

$$L_f = 1 - 10^{\log C} \quad (3)$$

Statistical Analysis

Fig. 1 shows the criteria we adopted to assess the reliability of presenting a contrast in a non-calibrated display. It was computed among the selected $\log C$ values ranging from $\log C$ = 0 to $\log C$ = −2.2 in steps of −0.1 (23 levels) and −0.05 (45 levels). The horizontal dashed lines represent the luminances (L_i) corresponding to three consecutive $\log C$ steps along the mean luminance curve (L_m) obtained from all the iPads. Therefore, the uniformity between devices for presenting a stimulus of certain contrast was considered as reliable when $[(L_i - L_{i-1})/2]$, *a* in Fig. 1, was higher than 2 standard deviations (SD) from the mean, *b* in Fig. 1. In this case, if we present a contrast with the same y value in different devices, the luminance offered by the device for this y will be closer to the required luminance for a $\log C$ value than to the luminance required for the next consecutive $\log C$ value. MATLAB software (R2013a; MathWorks, Natick, MA) was used for processing data. The fitting to the exponential functions for obtaining the gammas was completed with the Curve Fitting Toolbox.

The y corresponding to the selected *Theoretical* contrast values from $\log C$ = −0.05 to $\log C$ = −2.2 in −0.05 steps and −0.1 steps were calculated for the *LuT* table obtained from the mean gamma of the six iPads. The equivalent contrasts for these y in each iPad were computed, resulting in the *Experimental logC* values. Therefore, the *Theoretical logC* are the contrasts that we want to reliably reproduce and the *Experimental logC* are the true contrasts that each iPad displays for these *Theoretical* contrasts. Hence, the higher the difference between the *Theoretical* and *Experimental*

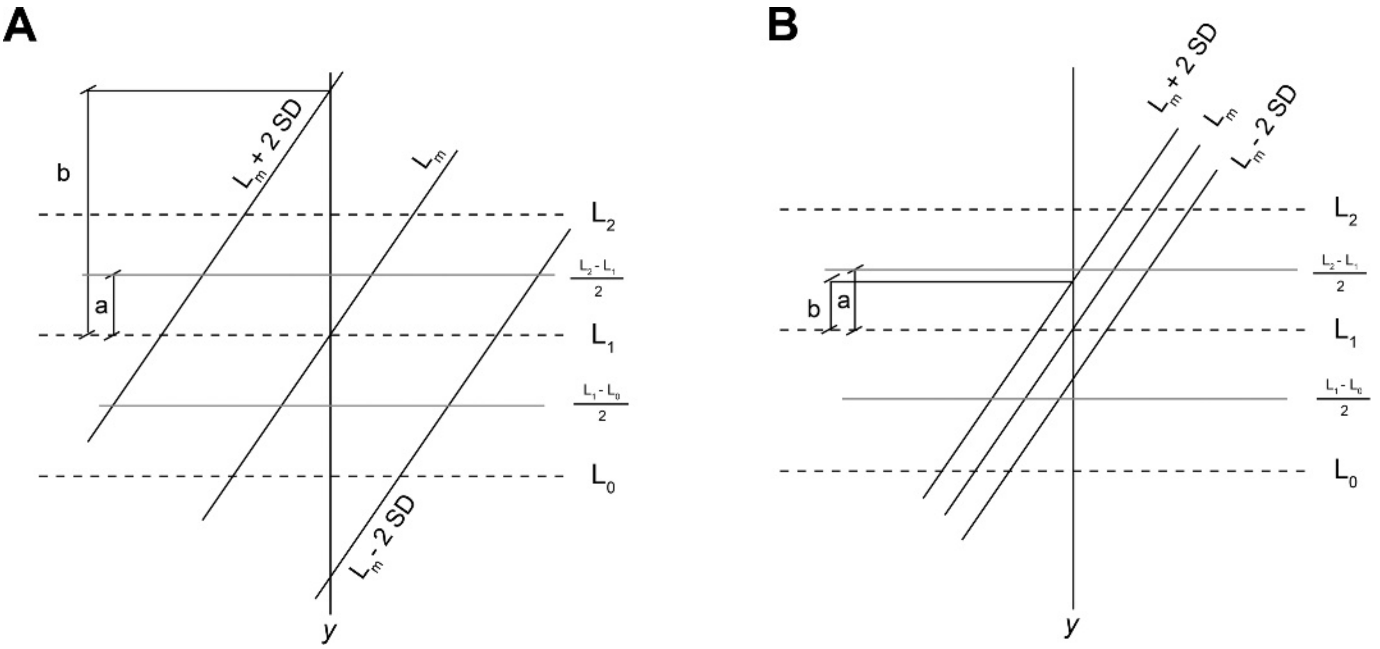

FIGURE 1.

(A) Two standard deviation from the mean (b) was greater than half the difference of luminance between two consecutive log*C* steps (a), not reliable. (B) Two standard deviation from the mean (b) was less than half the difference of luminance between two consecutive log*C* steps (a), reliable.

TABLE 1.
Gamma (γ) and maximum luminance (L_{max}) for the *RGB* channels of each iPad

Id	Model	γ_R (L_{max})	γ_G (L_{max})	γ_B (L_{max})	γ_W (L_{max})
1 (1st)	A1458	2.195 (27.35)	2.158 (91.61)	2.024 (9.15)	2.156 (128.11)
(2nd)*		2.197 (27.64)	2.157 (92.07)	2.027 (9.27)	2.156 (128.98)
2	A1416	2.219 (25.98)	2.248 (84.26)	2.096 (9.01)	2.230 (119.25)
3	A1458	2.373 (25.17)	2.351 (79.18)	2.204 (8.24)	2.345 (112.59)
4	A1430	2.074 (24.86)	2.052 (81.45)	1.955 (8.21)	2.049 (114.52)
5	A1430	2.035 (24.95)	2.029 (82.21)	1.921 (8.47)	2.022 (115.63)
6	A1458	2.139 (22.53)	2.124 (79.24)	2.031 (7.07)	2.121 (108.84)
+2SD		2.101 (28.03)	2.051 (91.45)	1.986 (9.71)	2.056 (129.20)
Mean		2.168 (25.14)	2.154 (82.99)	2.034 (8.36)	2.148 (116.48)
−2SD		2.253 (22.25)	2.291 (74.53)	2.103 (7.00)	2.270 (103.77)

γ_W values were computed by the combination of measured *R*, *G*, and *B* values. Mean values and standard deviations (SD) are presented at the bottom of the table.

L_{max} = value in cd/m^2 for y = 255 at each channel.

*A second measure was taken 1 month after the first for checking variability due to Spyder4Elite. This 2nd measure has not been considered for computing mean and standard deviation (SD).

logC, the poorer the reliability for contrast displaying in each non-calibrated iPad. The Kolmogorov-Smirnov test resulted in non-normal distributions of the *Experimental logC*. Therefore, a Friedman test was run with SPSS version 20 (SPSS Inc., Chicago, IL) to determine if there were differences in the *Experimental logC* between the six iPads.

RESULTS

Table 1 shows L_{max} and gamma values for *R*, *G*, *B* for each display and mean ± 2SD.

Fig. 2A represents the mean values of *L* and ±2SD from the mean (dashed lines) for the six devices (solid line). The relative *R: G: B* luminance, approximated to the first decimal, was constant, 0.2: 0.7: 0.1, for digital levels over y = 50. This means that $L(y)$ for *R* channel is two times higher than for the *B* channel whereas for *G* channel is seven times higher than for *B* channel. These relative luminances ensured that the bit-stealing technique could be obtained by adding nine levels more of luminance between each one of the 8-bit gray levels according to the next matrix:

$$(\delta_R, \delta_G, \delta_B) \in \{(0,0,1),(1,0,0),(1,0,1),(1,0,2),(2,0,1),(2,0,2),(0,1,0),(0,1,1),(1,1,0)\}$$

Fig. 2B represents the result obtained after applying the bit-stealing technique to the mean of the six devices (a total of 2540 pixel gray levels). The L_{bg} was set as y_{max} = 254 at the three *RGB* channels and the bit-stealing curve was normalized to L_{bg} = 1 by computing the relative luminance $L_R(y)$ described in equation 1.

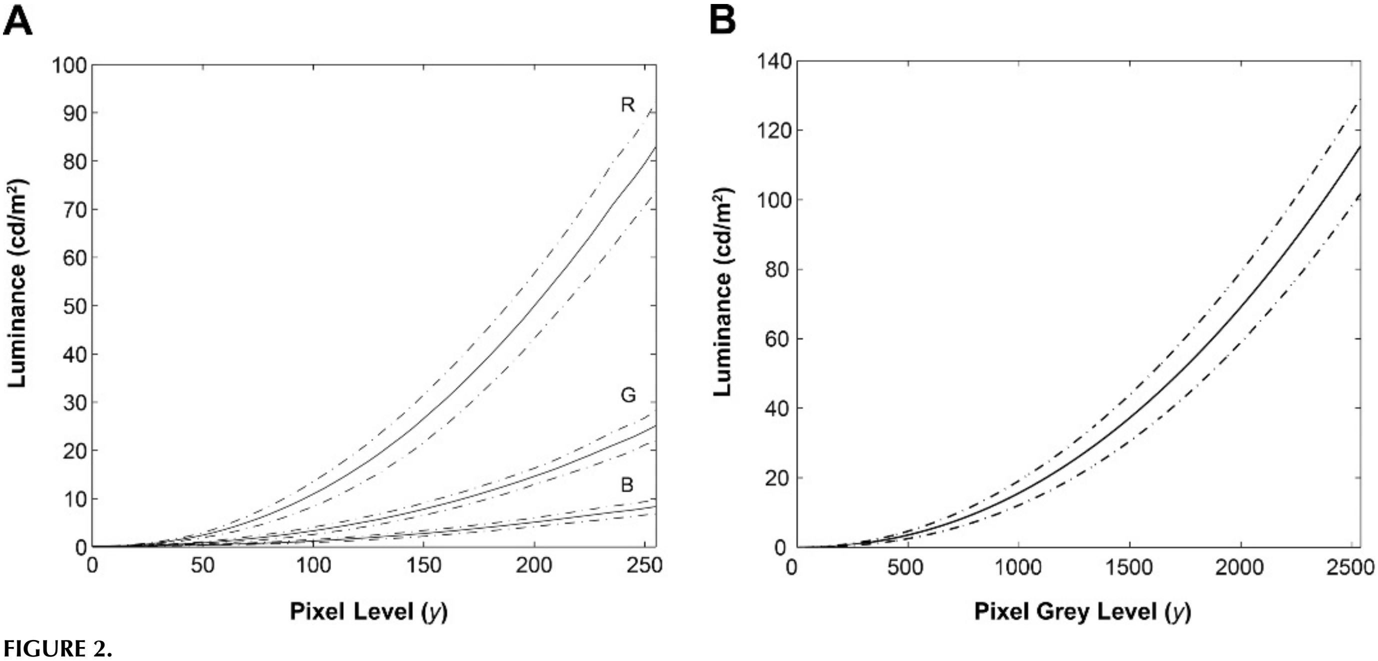

FIGURE 2.
Correlation between luminance and pixel level. (A) Mean (solid lines) and two standard deviations (dashed lines) of the six iPads for each RGB channel. (B) Mean and two standard deviations after bit-stealing application considering y_{max} = 254 (a total of 2540 pixel gray levels).

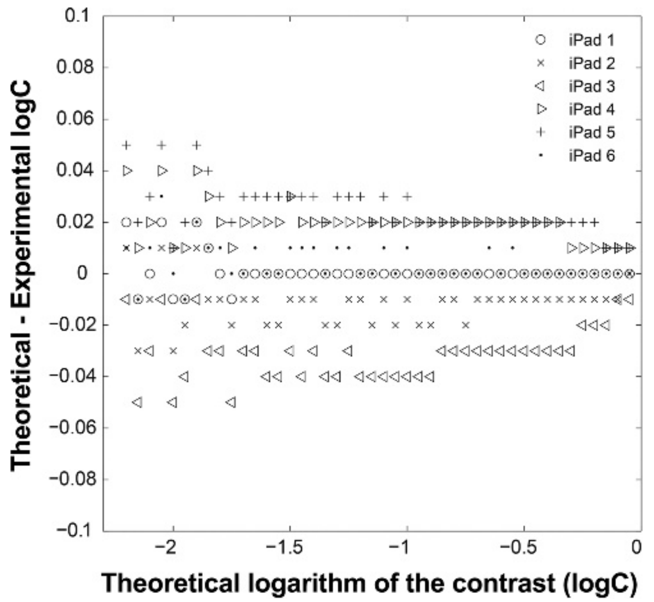

FIGURE 3.

Difference between *Theoretical logC* and *Experimental logC* against the selected contrast values obtained for comparison of the six iPads.

The difference between the *Theoretical logC* and the *Experimental logC* for all the iPads against the *Theoretical logC* values is shown in Fig. 3. The *Experimental logC* was statistically significantly different between the iPads for the −0.1 steps $\chi^2(6) = 126.189$, $p < 0.0005$ and for the −0.05 steps $\chi^2(6) = 257.485$, $p < 0.0005$. We found that for the iPad 1, the *Experimental logC* obtained 1 month apart were exactly the same for all levels except for −1.5 log$C$ (with a variation of 0.01 log$C$), whereas the variation was considerably greater with the other iPads, especially for lower contrasts (see Fig. 3). In fact, the iPad 1 was the closest to the *Theoretical logC* and a deviation from the *Theoretical logC* up to 0.05 log units was obtained for finest values of contrast for iPads 3 and 5.

In all, as was previously mentioned, the contrast variability among devices was considered as reliable when $[(L_i - L_{i\text{-}1})/2] > 2SD$ (see Fig. 1). Bearing in mind this condition, Fig. 4A shows that this condition (a > b, in Fig. 1) was satisfied for contrasts from log$C$ = 0 to log$C$ = −2.2 in both 0.1 log and 0.05 log steps. However, as can be seen in Fig. 4B for contrasts lower than log$C$ = −1.7, it is not adequate to use even 0.05 log steps because the L steps required for these finest contrasts are very small with regard to the L variability of the screen.

DISCUSSION

The proliferation of apps for testing vision alerted the research community because many of them are not correctly designed and do not follow the current international standards for testing vision.[6,7] Moreover, although the gamma function for a given iPad retina was considered in several papers,[3,8,9] the variability among the same iPad displays has not been previously reported. For this reason, in this paper, we have studied the variability among different iPads that share the same model of retina display. We demonstrated that there exist statistically significant differences between the luminances of iPads. This evidence, presumed previously by Lin,[10] confirms that for a precise measure of the CS, a previous calibration is necessary. Despite this, in this work, we show that for screening purposes, on which contrast patches vary in log unit steps, the variability among different units of the same iPad is small enough to avoid the previous calibration of each particular device. Particularly for iPad retina displays working at 50% of screen brightness, the calibration should be required only for contrasts decreasing in −0.05 log units if the contrast of the optotype is less than −1.7 log units. On the other hand, if the optotype contrast decreases in −0.1 log units, the results would be reliable for any iPad without a previous calibration. Our results suggest that for normal users who perceive very finest levels of CS (finer than −1.7 log units), the test should be designed in −0.1 log steps for being reliable. However, in low vision patients, who

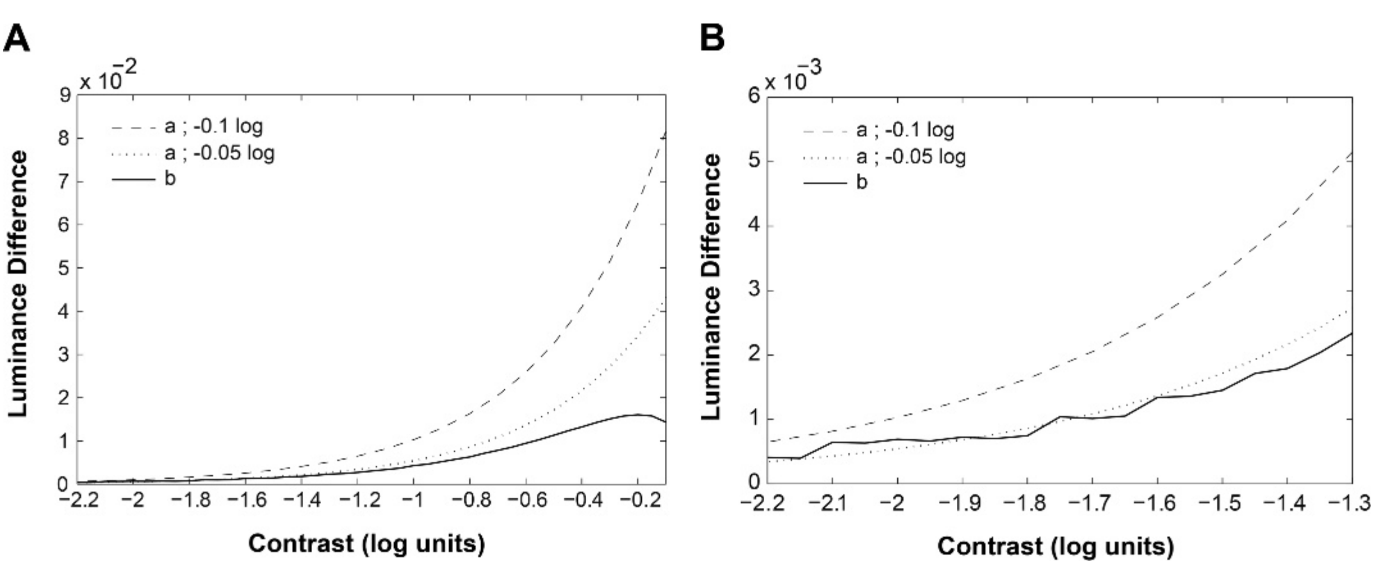

FIGURE 4.

Graphs representing the reliability of using an iPad considering a test developed for the gammas obtained from the mean of the six iPads. The *a* and *b* variables are explained in Fig. 1 and (A) represents the contrasts from −0.0 to −2.2 and (B) from −1.3 to −2.2 for a clearer view.

normally just perceive contrasts greater than −1.7 log units, the number of contrasts can be expanded by means of designing the test in −0.05 log steps. For instance, Kollbaum et al.[11] reported that the mean of CS ($-\log C$) in low vision patients measured with iPad retina was 1.43 ± 0.42 log units, which means that a test designed in −0.05 steps might be used for screenings with this population without having to worry about previous calibration.

One negative factor regarding our methodology is that we have conducted the study by means of computing $L_R(y)$. This means that even though a reliable contrast value can be extrapolated between the same devices, the luminance may differ between iPads because at the same y_{max}, the L could be different and thus better CS could be achieved with higher background luminances.[12] Furthermore, we found that the new iPad had the highest screen brightness, which suggests that it is likely that, for a given screen, L decreases with the time of use.

We are aware that our calibration data may differ from previous reports,[3,8,9] not just for using different instruments for calibration, a spectroradiometer instead of a colorimeter, but also for performing the calibration at different brightness of the screen. We decided to select the 50% of brightness instead of the 100% to reduce the glare of the screen as it has been previously reported.[11] On the other hand, the Spyder4Elite Colorimeter has been previously used in vision research.[13,14] Furthermore, a colorimeter is considerably cheaper than a photospectroradiometer, making it accessible for users who desire to calibrate their own tablets for personalizing the calibration data of any app that supports this option. However, it is important to note that brightness of the screen can be underestimated with the Spyder4Elite. This is the reason our mean brightness was 116.48 cd/m^2 versus the 150 cd/m^2 reported in other studies.[11] Therefore, increasing screen brightness to a percentage slightly over 50% would be acceptable mainly for screening elderly/low vision patients who often need high luminance displays for optimal performance. Our results are only applicable for iPad retina displays and for measuring achromatic CS. For other purposes, such as the design of tests based in chromatic thresholds or in any color discrimination task, the previous calibration should be complemented with a chromatic characterization of the device.[9]

In conclusion, differences in L properties between iPads suggest that a previous calibration is always recommended before conducting precise CS measurements; however, we have demonstrated that for screening purposes, reliable measures of CS can be obtained without a previous calibration by using an app that expands the range of y by bit-stealing and that uses a gamma correction of γ_R = 2.168, γ_G = 2.154, and γ_B = 2.034.

ACKNOWLEDGMENTS

This study was supported by the Ministerio de Economía y Competitividad and FEDER (Grant DPI2015-71256-R), and by the Generalitat Valenciana (Grant PROMETEOII-2014-072), Spain.

MR-V has designed and programmed apps that are currently distributed by the Apple Store with his own developer account. The other authors report no conflicts of interest and have no proprietary interest in any of the materials mentioned in this article.

Received October 14, 2015; accepted April 28, 2016.

REFERENCES

1. To L, Woods RL, Goldstein RB, Peli E. Psychophysical contrast calibration. Vision Res 2013;90:15–24.
2. Rodríguez-Vallejo M, Remón L, Monsoriu JA, Furlan WD. Designing a new test for contrast sensitivity function measurement with iPad. J Optom 2015;8:101–8.
3. Aslam TM, Murray IJ, Lai MY, Linton E, Tahir HJ, Parry NR. An assessment of a modern touch-screen tablet computer with reference to core physical characteristics necessary for clinical vision testing. J R Soc Interface 2013;10:20130239.
4. Tyler CW. Colour bit-stealing to enhance the luminance resolution of digital displays on a single pixel basis. Spat Vis 1997;10:369–77.
5. To L, Woods RL, Peli E. 17.3: Visual calibration of displays for accurate contrast reproduction. SID Int Symp Dig Tech Pap 2009;40:216–9.
6. Perera C, Chakrabarti R, Islam FMA, Crowston J. The Eye Phone Study: reliability and accuracy of assessing Snellen visual acuity using smartphone technology. Eye 2015;29:888–94.
7. Rodríguez-Vallejo M. Comment on 'The Eye Phone Study: reliability and accuracy of assessing Snellen visual acuity using smartphone technology'. Eye (Lond) 2015;29:1627.
8. Tahir HJ, Murray IJ, Parry NR, Aslam TM. Optimisation and assessment of three modern touch screen tablet computers for clinical vision testing. PLoS One 2014;9:e95074.
9. De Fez D, Luque MJ, García-Domene MC, Camps V, Piñero D. Colorimetric characterization of mobile devices. Optom Vis Sci 2016;93:1–9.
10. Lin TP, Rigby H, Adler JS, Hentz JG, Balcer LJ, Galetta SL, Devick S, Cronin R, Adler CH. Abnormal visual contrast acuity in Parkinson's disease. J Park Dis 2015;5:125–30.
11. Kollbaum PS, Jansen ME, Kollbaum EJ, Bullimore MA. Validation of an iPad test of letter contrast sensitivity. Optom Vis Sci 2014;91:291–6.
12. Bühren J, Terzi E, Bach M, Wesemann W, Kohnen T. Measuring contrast sensitivity under different lighting conditions: comparison of three tests. Optom Vis Sci 2006;83:290–8.
13. Gu XJ, Hu M, Li B, Hu XT. The role of contrast adaptation in saccadic suppression in humans. PLoS One 2014;9:e86542.
14. Bogfjellmo L-G, Bex PJ, Falkenberg HK. The development of global motion discrimination in school aged children. J Vis 2014;14:19.

Manuel Rodríguez-Vallejo
Qvision, Unidad de Oftalmología
Vithas Hospital Virgen del Mar
04120 Almería
Spain
e-mail: manuelrodriguezid@qvision.es