RESEARCH

Could standardizing “commercial off-the-shelf” (COTS) monitors to the DICOM part 14: GSDF improve the presentation of dental images? A visual grading characteristics analysis

D J McIlgorm*,1, C Lawinski2, S Ng3 and J P McNulty4

1Dublin Dental University Hospital, Trinity College Dublin, Dublin, Ireland; 2King’s Centre for the Assessment of Radiological Equipment, King’s College Hospital, London, UK; 3Department of Dental and Maxillofacial Radiology, Dental Institute, King’s College London, London, UK 4Diagnostic Imaging, School of Medicine and Medical Science, Health Sciences Centre, University College Dublin, Belfield, Dublin, Ireland

Objectives: To investigate whether standardizing commercial off-the-shelf (COTS) display devices to the digital imaging and communications in medicine part 14: greyscale standard display function (DICOM part 14: GSDF) would affect the presentation of dental images.

Methods: Two COTS display devices from the radiology department of a dental teaching hospital and a laptop computer monitor for reference were calibrated to conform to DICOM part 14: GSDF. Four dental surgeons and two final-year students undertook a relative visual grading analysis of the two devices before and after calibration, under control of the viewing environment.

Results: Calibrating COTS display devices to conform to the DICOM part 14: GSDF and viewing under reduced ambient light result in a consistent, perceived visual sensation for the presented radiological image. The area under the visual grading characteristics curve (AUCVGC) before calibration is 0.62 CI (0.56, 0.68) and AUCVGC after calibration is 0.51 CI (0.45, 0.57).

Conclusions: Standardizing COTS display devices to the DICOM part 14: GSDF can improve image presentation.

Dentomaxillofacial Radiology (2013) 42, 20130121. doi: 10.1259/dmfr.20130121

Cite this article as: McIlgorm DJ, Lawinski C, Ng S, McNulty JP. Could standardizing “commercial off-the-shelf” (COTS) monitors to the DICOM part 14: GSDF improve the presentation of dental images? A visual grading characteristics analysis. Dentomaxillofac Radiol 2013; 42: 20130121.

Keywords: calibration; greyscale; luminance

Introduction

The digital imaging and communications in medicine (DICOM) part 14: greyscale standard display function (GSDF) is a standard of greyscale uniformity that has been specified by the American College of Radiologists and the National Electrical Manufacturers Association for display systems.1 The luminance requirements that meet this standard for radiological applications are based on complex psychovisual experiments that centre on luminance measurements and their relationship to perceived brightness.2 The adaptive phenomenon of the human visual system means that one’s sensitivity is increased to small brightness variations when the area of interest is surrounded by bright elements. Barten2 investigated these brightness variations and the DICOM part 14: GSDF was defined, with the distance between two luminances that the human eye could just about detect called just noticeable differences (JNDs).

For most display devices when the pixel values contained within the processed image file are presented as digital driving levels (DDLs) and then subsequently converted to luminance levels by the display controller, the displayed image can take on a “uniqueness” for that particular display device. If the digital driving levels are
achieved the display device is said to be “perceptually linearised”.3

The American Association of Physicists in Medicine Task Group 18 (AAPM TG18)4 issued standard guidelines that addresses the issue of performance evaluation for medical display devices that are used for diagnostic and clinical reading purposes, and this includes a recommendation that such display devices should be standardized to the DICOM part 14: GSDF.1 AAPM TG184 also recommends assessing and scoring anatomical images for the visibility and quality of detail presented on the display device.

This investigation uses a visual grading characteristics (VGC) analysis to test the validity of this recommendation.5 VGC analysis draws upon the methods associated with preference studies, image criteria studies, visual grading analysis (VGA) and receiver operating characteristic (ROC) analyses.6,7 VGC analysis is characterized by its simplicity and discriminating power and allows subjective opinions to be quantified and analysed. Comparing the images produced by two test modalities, observers grade the visibility of the structure on the test image relative to that of a reference image using an arbitrary step scale (much worse to much better), with the middle value of the scale meaning a visibility equal to the reference image.5,6 The VGA results can then be used to characterize the difference between the two modalities in the same way as the difference in the response of the observers to the signal and noise distributions are used to characterize the observers in an ROC study.5 It is an observer performance method that is quick and straightforward to conduct, and relevant radiographic anatomy or pathology may meritoriously be used for the image analysis.5 The objective of this investigation was to use a VGC analysis to determine whether standardizing the luminance response for a display device would have an effect on the presentation of clinical images.

Materials and methods

Ethical approval (BDM/11/12-16), as well as informed consent from the participating observers, was obtained prior to this investigation.

Selection of monitors
Six commercial off-the-shelf (COTS) liquid crystal display (LCD) monitors in the radiology department of the Dublin Dental University Hospital, Dublin, Ireland, with 8-bit capability display controllers and set to a resolution of 1024 x 768 pixels had their characteristic luminance response curves transformed by Verilum® v. 5.02 (Image Smiths Inc., Bethesda, MD) software to conform to the DICOM part 14: GSDF luminance response.8 It was then possible to switch between the characteristic luminance response and the DICOM part 14: GSDF luminance response for each monitor. The software is also used as a quality assurance tool and provides information on how well a monitor conformed to the DICOM part 14: GSDF luminance response transformation.8 The monitor that had undergone the best transformation to the DICOM part 14: GSDF luminance response was selected, IBM ThinkVision L171 (IBM, Armonk, NY), as well as the monitor that had undergone the poorest transformation, NEC Multisync 1550V (NEC, Tokyo, Japan). Table 1 presents technical information and conformance information for all six monitors. The null hypothesis was then tested for the following four situations.

- There is no difference in the overall perceived visual sensation between the characteristic luminance response of the NEC monitor and when it is transformed to the DICOM part 14: GSDF luminance response.
- There is no difference in the overall perceived visual sensation between the characteristic luminance response of the IBM monitor and when it is transformed to the DICOM part 14: GSDF luminance response.
- There is no difference in the overall perceived visual sensation between the IBM monitor and the NEC monitor when both are using their own characteristic luminance responses.
- There is no difference in the overall perceived visual sensation between the IBM monitor and the NEC monitor when both are transformed to the DICOM part 14: GSDF luminance response.

To perform the tests, the method described by Geijer et al9 was used, whereby the visibility of a target structure in a presented image on each of the two selected monitors was compared with the corresponding structure presented on a monitor displaying a reference image, and the observer would give a score for each of the selected monitors. The laptop monitor that was used to present the reference image was an ACER emachines e525s-902G25Mi (ACER Inc., New Taipei City, Taiwan), which was set to a resolution of 1024 x 768 pixels and also had its characteristic luminance response curve transformed to the DICOM part 14: GSDF by the Verilum software application. This monitor had a maximum luminance that would be classed as primary (reporting or diagnostic) by AAPM TG18.4,10 Table 1 also presents the technical information and the conformance information for this laptop monitor.

Evaluation of images
4 clinical supervisors (registered dentists with a mean of 22 years of experience interpreting dental radiographs) and 2 final year dental students (each with 3 years’ experience interpreting dental radiographs) carried out the VGA.
Table 1  Technical specifications and digital imaging and communications in medicine part 14: greyscale standard display function (DICOM part 14: GSDF) conformance data provided by the Verilum® software application (Image Smiths, Inc., Bethesda, MD) for the six commercial off-the-shelf (COTS) liquid crystal display (LCD) monitors in the radiology department and for the laptop monitor used to display the reference image.

| Monitor                        | Product type | Year of manufacture | Max luminance (cd/m²) | Screen size (inches) | JND per luminance intervala | Minimuma | Maximuma | Root mean square errorb |
|--------------------------------|--------------|---------------------|-----------------------|----------------------|----------------------------|----------|----------|------------------------|
| IBM Thinkvision L171 (Armonk, NY) | LCD          | 2005                | 67.85                 | 17                   | 1.27                       | 1.18     | 1.36     | 0.051                  |
| HP 1740 (Palo Alto, CA)        | LCD          | 2006                | 74.35                 | 17                   | 1.44                       | 1.33     | 1.55     | 0.062                  |
| Philips Brilliance 150P2 (Amsterdam, Netherlands) | LCD          | 2001                | 127.11                | 15                   | 1.64                       | 1.51     | 1.82     | 0.077                  |
| HP 1740 (Palo Alto, CA)        | LCD          | 2006                | 51.91                 | 17                   | 1.55                       | 1.39     | 1.72     | 0.081                  |
| Philips Brilliance 150P2 (Amsterdam, Netherlands) | LCD          | 2001                | 128.10                | 15                   | 1.66                       | 1.47     | 1.87     | 0.119                  |
| NEC Multisync 1550 V (Minato, Tokyo) | LCD          | 2002                | 63.16                 | 15                   | 1.04                       | 0.68     | 1.20     | 0.124                  |
| ACER emachines E525 (New Taipei City, Taiwan) | LCD          | 2009                | 192.20                | 15                   | 1.79                       | 1.65     | 1.90     | 0.063                  |

aThese measurements are an indicator of how well the just noticeable differences (JNDs) of the DICOM part 14: GSDF have been linearly mapped to the digital driving levels of the monitor. Large differences between the maximum and minimum values would suggest inconsistency.

bThe smaller the root mean square error the more closely the monitor approximates the DICOM part 14: GSDF and would give a more consistent perceptual response. It shows how well the monitor has been perceptually linearized.

At the first period of VGA, each monitor including the laptop monitor to display the reference image was using its own characteristic luminance response (before calibration), and at the second period of VGA, after an interval of 6 weeks, all three monitors were using the DICOM part 14: GSDF luminance response (after calibration). 15 anonymous dental images for teaching dental radiology were used for the VGA. The 8-bit JPEG images with horizontal and vertical resolution of 96 dots per inch were displayed on a Windows® (Microsoft Corporation, Redmond, WA) viewing application in slide show mode. The image files were displayed as they were saved, having been optimized to facilitate their interpretation during the teaching module. They consisted of nine intraoral images, four panoramic images, one upper standard occlusal image and one lateral cephalometric image. The participants were asked not to alter the display of the images, which were scaled to fit the screen of the monitors. They were asked to view the images at 90° to the screen at a distance of between 30 cm (close inspection) and 60 cm (normal viewing). The ambient lighting was set to between 25 lux and 40 lux and was measured with a calibrated Luxi photometer (Unfors Instruments, Billdal, Sweden) set to illuminance mode. The participants were given no information about any of the parameters of the monitors, except that the three devices were going to be technically adjusted after the first period of VGA. At each period of VGA, the participants were given information on two predefined criteria relating to target structures within each image, which were designed by the investigators.

The participants were instructed to score the display quality of the pre-defined criteria of the target structure on the test image relative to the corresponding landmark on the reference image using the following scoring system.

1. Test image is clearly superior to the reference image.
2. Test image is somewhat superior to the reference image.
3. Test image is equal to the reference image.
4. Test image is somewhat inferior to the reference image.
5. Test image is clearly inferior to the reference image.

Figures 1 and 2 are examples of images that were used.

Criteria for Figure 1. It is a periapical radiograph of the 36.

(i) A low-density radio-opaque area is visible between the roots of 36 above the clearly defined radiolucent area.
(ii) A low-density radio-opaque band (soft tissue) is well demarcated from the higher density radio-opacity (mandibular bone) in the 37 region.

Criteria for Figure 2. It is a panoramic radiograph of an 8-year-old child.

(i) The position of the radiolucency of the right mental foramen can be accurately located.
(ii) The complete radio-opaque outline of both 38 and 48 tooth crypts are visible.

Analysis

The VGA scores were analysed in a manner similar to that used in ROC analysis. A constructed VGC curve described the relationship between the proportions of fulfilled image criteria for the two compared monitors, taking into consideration all the possible thresholds for a fulfilled criterion. The area under the VGC curve (AUCVG) was used as a single measure of the difference in image quality between the two compared monitors. An AUCVG value greater than 0.5 would...
indicate that one of the compared monitors offered a preferred image quality over the other and an AUC\textsubscript{VGC} value of 0.5 would indicate that the image quality offered by the two monitors was similar in preference.

Results

The collected data consisted of 180 observations for each of the two monitors (IBM and NEC) when using their own characteristic luminance response (before calibration), and 180 observations for each of the two monitors after they had been transformed to the DICOM part 14: GSDF luminance response (after calibration). Based on the specific criteria and observers used in this investigation, the NEC monitor offered a preferred perceived visual sensation for the displayed image after it was calibrated to the DICOM part 14: GSDF (Figure 3), AUC\textsubscript{VGC} = 0.73 confidence interval (CI) (0.67, 0.79). The IBM monitor also offered a preferred perceived visual sensation for the displayed image after it was calibrated to the DICOM part 14: GSDF (Figure 4), AUC\textsubscript{VGC} = 0.63 CI (0.56, 0.69).

The IBM monitor offered a preferred perceived visual sensation for the displayed image over that of the NEC monitor before either of them was calibrated to the DICOM part 14: GSDF (Figure 5), AUC\textsubscript{VGC} = 0.62 CI (0.56, 0.68). One possible explanation may be that the display controller system for the IBM monitor is more flexible at using its DDLs or repeating DDL values to give a characteristic luminance response that is preferred by the observer, and it seems this flexibility might be further improved by the calibration process (Figure 4). Another explanation may simply be the age of the NEC monitor compared to the IBM monitor.

However, both the NEC monitor and the IBM monitor showed a similar preferred perceived visual sensation for the displayed image after both had been calibrated to the DICOM part 14: GSDF (Figure 6), AUC\textsubscript{VGC} = 0.51 CI (0.45, 0.57). These results suggest that the modified luminance values after calibration for both these COTS monitors have resulted in an overall preferred perceived visual sensation for the observers.

Discussion

For a COTS monitor using its own characteristic luminance response, there is a tendency that the shades
are emphasized in the low luminance range and compressed in the middle and high luminance ranges. With the DICOM part 14: GSDF calibration process, a DDL is selected to produce a luminance value nearest to the desired luminance on the DICOM part 14: GSDF curve, so that for a DICOM part 14: GSDF transformed monitor, differences between any two shades should be perceived as the same degree of change in luminance in the greyscale displayed. The concept of “perceptual linearization” attributed to the DICOM part 14: GSDF calibration process makes sure that the mapping from the image data space of the grey-scale processed image to the human observers visual sensory space accurately transmits changes in intensities in the image to the human observer. The practical significance is that distortions in the transferred clinical information between the presented processed image data and the perceived visual sensation are minimized. This investigation has shown that an image presented on a DICOM part 14: GSDF calibrated display is preferred by the observers for the tested monitors, even if the contrast sensitivity of the human visual system were pre- 

dentification could assist practitioners in choosing display 

Figure 6 The 95% confidence interval of the VGC curve (AUCVGC = 0.51) does cover 0.5. The IBM monitor (IBM ThinkVision L171; IBM, Armonk, NY) and the NEC monitor (NEC Multisync® 1550 V; NEC, Tokyo, Japan) present an equally preferred image quality. AUCVGC, area under the VGC curve; CI, confidence interval; Std Dev, standard deviation; VGA, visual grading analysis; VGC, visual grading characteristics
Although it is widely accepted that ROC studies using a “gold standard” are generally considered the most reliable method for evaluating the diagnostic value of medical imaging modalities and techniques, visual grading studies are relatively straightforward to conduct and clinical images are often readily available and could provide a useful additional image quality evaluation method in the dental radiology assessment toolkit. The concept behind a VQA is that if the practitioner knows the anatomical landmarks or appearance of a pathological lesion, and is presented with a high-quality image, then he/she would be confident to make a diagnosis. The benefit of visual grading is that it is not limited to a specific diagnosis, while the disadvantage is that it is a subjective method of evaluating the quality of an image, and not a measure of the ability to make the correct diagnosis. VGC analysis is used as a measure of the difference in image quality between modalities and is not an absolute value of subjective image quality. In addition, the measurement of other factors that could potentially influence the grading may not be possible with this type of assessment. In the absence of any well-defined criteria for evaluating dental image quality, it would be necessary to design one’s own criteria for these type of studies based on diagnostically relevant anatomical structures and features, and one should not underestimate the ability of any clinician to recognize anatomical features in order that a diagnosis could be reached.22

In conclusion, display systems can influence the contrast resolution requirements for clinical applications, standardizing COTS display devices to the DICOM part 14: GSDF can improve the presentation of dental radiological images.

References

1. National Electrical Manufacturers Association (NEMA). Digital imaging and communications in medicine (DICOM) part 14: grayscale standard display function. Rosslyn, VA: NEMA Publishing; 2006.
2. Barten PGJ. Contrast sensitivity of the human eye and its effects on image quality. Bellingham, WA: SPIE Optical Engineering Press; 1999.
3. Bradley M, Hemminger R, Johnston RE, Rolland JP, Muller KE. Introduction to perceptual linearization of video display systems for medical image presentation. J Digit Imaging 1995; 8: 21–34.
4. Samei E, Badano A, Chakraborty D, Compton K, Cornelius C, Corrigan K, et al. Assessment of display performance for medical imaging systems, report of the American Association of Physicists in Medicine (AAPM) Task Group 18. AAPM online report no. 03. Madison, WI: Medical Physics Publishing; 2005.
5. Båth M, Månsson LG. Visual grading characteristics (VGC) analysis: a non-parametric rank-invariant statistical method for image quality evaluation. Br J Radiol 2007; 80: 169–176. doi: 10.1259/bjr/35012658
6. Månsson LG. Methods for the evaluation of image quality: a review. Radiat Prot Dos 2000; 90: 89–99.
7. van Erkel AR, Pattynama PM. Receiver operating characteristic (ROC) analysis: basic principles and applications in radiology. Eur J Radiol 1998; 27: 88–94.
8. Verilum. 5.0:2006. Users guide. Bethesda, MD: Image Smiths Inc, 1999.
9. Geijer H, Geijer M, Forsberg L, Khedache S, Sund P. Comparison of color LCD and medical-grade monochrome LCD displays in diagnostic radiology. J Digit Imaging 2007; 20: 114–121. doi: 10.1007/s10278-007-9028-5
10. Bextell DS. Display considerations for hospital-wide viewing of soft copy images. Br J Radiol 2007; 80: 503–507.
11. Samei E. AAPM/RSNA physics tutorial for residents: technological and psychophysical considerations for digital mammographic displays. Radiographics 2005; 25: 491–501. doi: 10.1148/rg.252045185
12. Brennan P, McEntee M, Evanoff M, Philips P, O’Connor WT, Manning DJ. Ambient lighting: effect of illumination on soft-copy viewing of radiographs of the wrist. AJR 2007; 188: W177–W180. doi: 10.2214/AJR.05.2048
13. Haak R, Wicht MJ, Nowak G, Hellmich M. Influence of displayed image size on radiographic detection of approximal caries. Dentomaxillofac Radiol 2003; 32: 242–246.
14. Hellen-Halme K, Nilsson M, Petersson A. Effect of monitors on approximal caries detection in digital radiographs—standard versus precalibrated DICOM part 14 displays: an in vitro study. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2009; 107: 716–720. doi: 10.1016/j.tripleo.2008.12.011
15. Haak R, Wicht MJ, Hellmich M, Nowak G, Noack MJ. Influence of room lighting on grey-scale perception with a CRT-and a TFT monitor display. Dentomaxillofac Radiol 2002; 31: 193–197. doi: 10.1038/sj/dmfr/4600668
16. Robson KJ, Kotre CJ. Pilot study into optimisation of viewing conditions for electronically displayed images. Radiat Prot Dosimetry 2005; 117: 298–303. doi: 10.1093/rpd/nci744
17. Lowe JM, Brennan PC, Evanoff MG, McEntee MF. Variations in performance of LCDs are still evident after DICOM gray-scale standard display calibration. AJR Am J Roentgenol 2010; 195: 181–187. doi: 10.2214/AJR.09.2851
18. MacDonald-Jankowski DS, Orpe EC. Some current legal issues that may affect oral and maxillofacial radiology. Part 2: digital monitors and cone-beam computed tomography. J Can Dent Assoc 2007; 73: 507–511.
19. Hellen-Halme K, Petersson A, Warfvinge G, Nilsson M. Effect of the detection of approximal caries in digital radiographs: an in vitro study. Dentomaxillofac Radiol 2008; 37: 380–384. doi: 10.1259/dmfr/26038913
20. Institute of Physics and Engineering in Medicine. Recommended standards for the routine performance testing of diagnostic X-ray imaging equipment. Report 91. York, UK: IPEM; 2006.
21. Hemminger BM, Blume HR. Are medical image display systems perceptually optimal? Measurements before and after perceptual linearization. Proceedings of medical imaging 1996: image capture, formatting and display. Vol. 2707–58. Bellingham, WA: SPIE Press; 1996.
22. EUR 16260. European guidelines on quality criteria for diagnostic radiographic images. Luxembourg, Belgium: Office for official publications of the European Communities; 1996.