Trend of Contrast Detection Threshold with and without Localization

David L. Leong · Louise Rainford ·
Tamara Miner Haygood · Gary J. Whitman ·
William R. Geiser · Beatriz E. Adrada ·
Lumarie Santiago · Patrick C. Brennan

Published online: 16 March 2013
© Society for Imaging Informatics in Medicine 2013

Abstract Published information on contrast detection threshold is based primarily on research using a location-known methodology. In previous work on testing the Digital Imaging and Communications in Medicine (DICOM) Grayscale Standard Display Function (GSDF) for perceptual linearity, this research group used a location-unknown methodology to more closely reflect clinical practice. A high false-positive rate resulted in a high variance leading to the conclusion that the impact on results of employing a location-known methodology needed to be explored. Fourteen readers reviewed two sets of simulated mammographic background images, one with the location-unknown and one with the location-known methodology. The results of the reader study were analyzed

using Reader Operating Characteristic (ROC) methodology and a paired *t* test. Contrast detection threshold was analyzed using contingency tables. No statistically significant difference was found in GSDF testing, but a highly statistical significant difference (*p* value <0.0001) was seen in the ROC (AUC) curve between the location-unknown and the location-known methodologies. Location-known methodology not only improved the power of the GSDF test but also affected the contrast detection threshold which changed from +3 when the location was unknown to +2 gray levels for the location-known images. The selection of location known versus unknown in experimental design must be carefully considered to ensure that the conclusions of the experiment reflect the study's objectives.

D. L. Leong (✉)
Analogic Corporation, 8 Centennial Drive,
Peabody, MA 01960, USA
e-mail: dleong@analogic.com

D. L. Leong · L. Rainford
Diagnostic Imaging, UCD School of Medicine
and Medical Sciences, University College Dublin,
Belfield, Dublin 4, Ireland

T. M. Haygood · G. J. Whitman · B. E. Adrada · L. Santiago
Department of Diagnostic Radiology,
The University of Texas MD Anderson Cancer
Center, Houston, TX 77030, USA

W. R. Geiser
Department of Imaging Physics,
The University of Texas MD Anderson Cancer
Center, Houston, TX 77030, USA

P. C. Brennan
Brain and Mind Research Institute,
Faculty of Health Science, University of Sydney,
Sydney, New South Wales 2006, Australia

Key words Image perception · contrast threshold · GSDF · ROC · SKE · LKE

Introduction

Detection of objects is an important part of a radiologist's interpretation task, and the contrast of the object plays an important role in object detection. The Digital Imaging and Communications in Medicine (DICOM) standard has been used for many years to perceptually linearize image display, which in theory reduces the effects of image luminance on contrast detection [1]. The American College of Radiology [2], the American Association of Physicists in Medicine [3], and Integrating the Healthcare Enterprise [4] all recommend the use of DICOM calibrated monitors in diagnostic interpretation. Knowledge of the impact of the display luminance and perceived brightness on contrast detection is critically important to the design of schemes to optimize image quality. These

optimization procedures are commonly formulated to optimize image contrast and patient dose to produce the best clinical images [5]; thus, detailed knowledge of image display contrast is important to researchers attempting to improve the diagnostic efficacy of radiographic images. Barten [6] developed what is currently the most comprehensive model of the human visual system for monochrome images used in radiology. However, the publishing of this seminal work has not ended research into contrast detection [7–10].

The method of choice in contrast detection threshold experiments has been, and continues to be, the two-alternative forced choice (2AFC) or 4AFC methodology. AFC methodology employs a signal-known-exactly (SKE) and location-known-exactly (LKE) task. The 2AFC method allows for quick assessment of a large number of images and has helped to clearly show the relationship between object size and contrast [9]. However, this method eliminates search from the task, which is not a true representation of many of the detection tasks in radiology [11].

In previous work using a search task (location unknown) [12], a very high inter-observer variability with a high false-positive rate was seen when readers were asked to detect simple contrast disks on mammographic-like backgrounds. The goal of the original work was to test the hypothesis that the DICOM Grayscale Standard Display Function (GSDF) [13] perceptually linearized image display as median image luminance changed in complex images. However, with the observed high variability and false-positive rate, it was difficult to achieve sufficient power to state conclusively that the GSDF did perceptually linearize display monitors. Since the previous work used a location-unknown methodology, the study was repeated using the same readers but with a specified location given to eliminate search from the task.

Using location-known and location-unknown methodologies, the current work attempted to establish with both approaches the contrast detection threshold in a noisy background that simulated a representative radiographic image; confirm within each method that, with GSDF calibrated monitors, changes in median background intensity will not affect the contrast detection threshold; and compare the data produced with each of the methodological approaches.

Materials and Methods

One hundred and fifty mammographic-like images created for a previous investigation were used in this study [12]. The images were based on 30 background images of 512×512 pixels in size and 8 bits deep using Bochud et al.'s [14] synthetic clustered lumpy background method. All image creation was performed using IDL 7.0 (ITT Visual Solution, Boulder, CO, USA), and background images had a total pixel range from the median of approximately ± 60 gray

levels. The background intensity was varied by subtracting or adding an offset to the base image to produce five images with different median levels while maintaining the original image distribution. The five median backgrounds produced were 65, 100, 135, 175, and 205 gray level values. One hundred images had one randomly placed uniform contrast disk with a 20-pixel radius, inserted with a value ranging from +1 to +5 gray levels from the background pixel value, thus maintaining the background variation at the location of insertion. This constituted the set of location-unknown images (Fig. 1a). For the LKE images, the same images with same lesion locations were used, but this time a ring of white or black dots (based on the brightness of the image) with a radius of 40 pixels was added around each disk, and for the images with no disk, a ring of dots was placed in a random location (Fig. 1b). For all images, the readers knew the size and the shape of the object, making this a SKE test. Fifty images had no disks applied upon them.

The reader study was performed using ViewDEX [15] 1.0 software (Sahlgrenska University Hospital, Göteborg, Sweden) to run an ROC reader study with a five-point confidence scale. A score of five indicated high confidence that a disk was present, while a score of one was used to represent that a reader was very confident that a disk was not present. Images were viewed in full resolution on standard personal computers with off-the-shelf color LCD monitors all calibrated to be compliant with DICOM Part 14 (typical luminance range of 0.7 to 100 cd/m^2). VeriLUM software (Image Smiths, Bethesda, MD, USA) with a single photometer was used for all calibrations. Each reader performed two reading sessions, a location-unknown session [12] and a second session with the location-known images with an average of 4 months between the two reading sessions. There was no time limit for viewing the images and for both readings, indirect lighting was used to control the ambient lighting to below 20 lx. All image manipulation tools were turned off, and the readers were required to view the images as displayed. In order to become accustomed to the ViewDEX software, all readers practiced with a training set of images just prior to each of the scored reading sessions, which were not part of the study data set. Readers were aware that contrast disks were randomly located on approximately two out of three images.

Fourteen readers participated in both reading sessions required for this study, producing 14 paired readings. The 14 readers were from three institutions and had diverse levels of clinical experience. Seven readers were attending radiologists, and seven were naïve readers from the medical imaging field. The radiologists had, on average, 13.2 years of post-residency experience with a range of 3 to 35 years. The naïve readers included two engineers, three academic medical physicists, and two technology academic lecturers.

Analysis of reader data was performed by ROC analysis using the DBM-MRMC [16–22] 2.2 software (University of

Fig. 1 **a** A sample image showing a contrast disk on a clustered lumpy background when the location is unknown. The *white arrow* points to the upper-right edge of the disk. **b** A sample of the same image as it was used in the location known readings. The *ring of dots* was changed

from *white to black* as the median background level went to 175 and 205 to allow readers to find the location of the disk quickly. Notice that the background texture is visible through the disk in both the location-known and location-unknown formats

Chicago, Kurt Rossmann Laboratories for Radiologic Image Research, Chicago, IL, USA). Since the scores were non-parametric, the Trapezoidal/Wilcoxon method was used, and ANOVA tests compared the area under the ROC curves (AUC) between the different median background levels to determine the impact of luminance change on reader performance. JMP®, Version 9 statistical software (SAS Institute Inc., Cary, NC, USA), was used to perform the paired *t* test to compare the data produced with the location-unknown and location-known methods. JMP was also used to generate the contingency tables, to establish the contrast threshold, and to generate the reading time histograms. The contrast threshold was defined as the gray level where detection performance exceeded the 50 % mark. Performance estimates, using a power of 0.8 and an alpha of 0.05, were computed using scripts created at the University of Iowa's Medical Image Perception Laboratory [19, 23] with SAS 9.1.3 SP4 (SAS Institute Inc., Cary, NC, USA) software.

Results

The ROC analysis produces one area under the ROC curve (AUC) for each reader and at each background level. Figure 2 is a compilation of all of the AUC scores. Table 1 displays a summary the AUC scores with the minimum and maximum values, mean, and the standard deviations. The results of ANOVA test determined that for both locations unknown and known, there were no statistical significant differences between median background levels, with *p* values of 0.55 and 0.76, respectively. Using a power of 0.80 and an alpha of 0.05, this study was able to detect a change in AUC for the location-unknown and location-known methodologies of a 0.095 and 0.066, respectively. The mean AUC score for all readers across all median

background levels was 0.755 for location unknown and 0.900 for location known, these values being significantly different ($p < 0.0001$).

The scores for all readers and all median background levels are represented in Fig. 3.

The paper employs a graphical representation of the contingency table, called a mosaic plot (Figs. 3, 4, 5) [24]. The main plot in a mosaic plot appears as a stacked bar chart, with the width of the bars being proportional to the number of counts in each column. Since the dataset contained 150 images with 100 of the images having contrast disks divided across five possible values and 50 images having no disks, the width of each of the contrast disk columns is 13.3 % of the entire width and the *no disk*, 33.3 % of the total width. The scale on the left is a cumulative percent scale. On the right side of the graph is a summary bar

Fig. 2 The graph shows the area under the curve (AUC) for each reader at each median background level as box whisker plot, which shows the distribution of the scores. *Each box whisker graphic* shows the mean score at the center line of the central box with the *ends of the box* being the 25 and 75 % quantiles. The *length of the lines* shows the maximum and minimum values with the *short lines* near the ends showing the 5 and 95 % quantiles

Table 1 Summary of AUC values for all readers

Reader	Unknown location				Known location			
	Max	Min	Mean	Std dev	Max	Min	Mean	Std dev
1	0.81	0.62	0.72	0.087	0.97	0.81	0.90	0.064
2	0.86	0.70	0.79	0.059	0.91	0.84	0.87	0.031
3	0.83	0.65	0.76	0.072	0.88	0.83	0.85	0.018
4	0.88	0.64	0.78	0.095	1.00	0.86	0.93	0.050
5	0.86	0.61	0.76	0.098	0.98	0.93	0.94	0.023
6	0.80	0.72	0.75	0.033	0.93	0.86	0.88	0.025
7	0.78	0.68	0.73	0.037	0.94	0.86	0.90	0.029
8	0.78	0.64	0.71	0.062	0.93	0.88	0.90	0.020
9	0.87	0.73	0.80	0.055	0.94	0.88	0.91	0.023
10	0.79	0.65	0.74	0.064	0.95	0.92	0.93	0.012
11	0.72	0.62	0.69	0.040	0.95	0.86	0.91	0.036
12	0.79	0.67	0.73	0.047	0.94	0.85	0.89	0.038
13	0.85	0.70	0.79	0.058	0.91	0.87	0.90	0.019
14	0.88	0.78	0.83	0.038	0.91	0.88	0.90	0.015

that gives the overall distribution as proportional lengths of the values listed just to the left of the bar and also serves as a color key to those values. The size of each section is proportional to the overall distribution of the rows. In these graphs, the values of 1 to 5 are the numeric scores given by the readers. As an example, in Fig. 3b in the +2 column, the white section at the top represents the percent of score of 5, which was a reader rating of *very confident object is present*. This represents 40 % of the total scores received for this contrast level.

Summary findings of the location-unknown images and the location-known images are shown in Fig. 3a and b, respectively. For the location-unknown case, readers scored a disk as present 24, 33, and 67 % of the time for the +1, +2, and +3 contrast disks, respectively. For the *no disk* images, readers scored a disk as present 24 % of the time. In the location-known cases, readers scored a disk as present 23,

88, and 100 % for the +1, +2, and +3 contrast disks, respectively. In this case, the *no disk* images only received 8 % of the overall scores as a disk was present.

Figure 4 takes the readers' scores and combines the 4 and 5 score to be *yes* a disk is present. Scores 1 and 2 were combined into *no* a disk is not present. Score 3, while described as *somewhat confident object present*, which is a positive detection, for this plot is resigned the value of *maybe* to signify the uncertainty of the readers.

Figure 5 shows mosaic plots broken down by median background level. The location-unknown images are on the left and the location-known images on the right. Across the same row are images of the same median background level.

The distribution of reading times per image is shown in Fig. 6 for all readers for the unknown-location readings (6a) and known-location reading (6b). The median reading time

Fig. 3 Mosaic plot showing the reader-confidence score as the contrast in the disk changed. Graph **a** shows the case when location was unknown, and graph **b** when the location was known. The scores listed on the right side of the graph are related to the 5-point confidence scale.

The varying shades of gray indicate different scores starting with black as a 1 (very confident that an object is not present) progressing to 5 (very confident that an object is present) in white. The y-axis is the cumulative percent distribution of the score

Fig. 4 Mosaic plot shows the reader-confidence score as the contrast in the disk changed. Graph **a** shows the case when location was unknown, and graph **b** when the location was known. The scores have been combined with 4 and 5 as *Yes* for a disk present and, 1 and 2

was 9 s with a range of 2 to 171 s for the unknown-location cases. For the known-location cases, the median reading time was 4 s with a range of 2 to 62 s.

Discussion

For both location unknown and location known, there was no statistically significant difference (p values of 0.55 and 0.76, respectively) in detection as the median background gray level changed in this study, supporting the theory that the DICOM GSDF does perceptually linearize monitor contrast detection for images of complex backgrounds. The lower variances seen with the location-known methodology show an improved ability to detect a change in AUC with the same number of readers, as seen in Fig. 2. Overall, the standard deviation for readers was also reduced by using the location-known methodology, as is seen in Table 1. For a statistical test of significance, more readers would be required to reduce the detection threshold.

The improvement in AUC scores with the location known over location unknown is expected as removing scanning errors would increase scores as described by Kundel et al. [25]. This highlights the potential value of accurate and detailed radiologic request information, from the referring physician, focusing wherever possible on lesion location, so that the radiologist's attention can be directed to the appropriate regions of interest. Upon review of the literature, a paucity of research in this area highlights the need for further research into radiologic requests improving radiologist's performance. A decrease in the inter-observer variability of 40 % was observed when the location was known. As discussed further below, a large part (but not all) of the improvement in AUC scores is a decrease in the false-positive rate.

Analysis of contrast detection threshold findings identified a difference between the location-unknown and

combined to *No* for no disk present. The three scores were reassigned to maybe. *White* indicates a score of *Yes* (an object is present) and *black* a score of *No* (an object is not present). The y -axis on the left is the percent distribution of the score

location-known images. Figure 3a shows the data for the location-unknown images and indicates that the contrast detection threshold, as defined as the contrast level where the majority of observations were 4 or 5 when a disk was present, occurred at +3. For the location-known images, as seen in Fig. 3b, the detection appears at +2. To better visualize the contrast detection threshold, Fig. 4 shows combined values for the score, and the change from +3 (location known) to +2 (location unknown) is clearly seen. This reduction in the contrast detection threshold identifies that removing search from detection tasks improves readers' ability to detect low contrast objects. In the case of the GSDF test, a location-known methodology gives the most power to the study, but it overestimates the contrast detection threshold that would be experienced in a clinical environment by reducing the detection threshold. To determine clinically relevant contrast perception, search should be included in the study design, and since a large proportion of previous contrast detection threshold published work is based on *location-known*, the amount of contrast required for detection is most likely to have been underestimated when taking into account that search is an important part of radiology interpretation.

The improvement in the false-positive rate with the *location-known* technique can also be seen in Fig. 3 and the scores given for the *no disk* images. In the location-unknown study, 24 % of images that did not have a disk were scored as positive for disk inclusion, which was calculated by taking the sum of scores from *very confident*, *confident*, and *maybe* present ratings. This dropped to 8 % with the location-known approach, which indicates that readers were distracted to a small extent by the clustered lumpy background. When comparing the +1 *disk* to the *no disk* for the location-unknown case, readers scored a *disk present* at 24 % for both. Since the distribution of 5, 4, and 3 scores is statistically equivalent between the +1 *contrast disk* and the *no disk* images, it is reasonable to conclude

Fig. 5 Mosaic plots showing the reader-confidence score as the contrast in the disk changed for each of the five median background levels. The *left side column* of graphs **a**, **c**, **e**, **g**, and **i** are for the location-unknown case, and on the *right*, graphs **b**, **d**, **f**, **h**, and **j** for the location-known case. Thus, side by side are the same median background levels. The lowest median background starts at the *top a* and **b**, working to the brightest images at the *bottom i* and **j**

Fig. 6 Distribution of reading time per image of all scores for all readers given in seconds. Graph **a** shows the results for the location-unknown images, and **b** shows the results for the location-known images

that false localization contributes substantially to these “correct” scores (for the $+1$ disks). Performing the same comparison in the location-known case, there is an increase to 23 % of disk being present in the $+1$ contrast disk images as compared to the 8 % in the *no disk* images, which cannot be just false localizations as in the location-unknown case. The contribution of false positives to apparently “correct” answers in the location unknown paradigm, highlights the need for localization ROC techniques such as LROC or JAFROC, when using search methodologies.

To ensure that there was no effect by median background level, the mosaic plots were broken down by median background level, as seen in Fig. 5. The left columns of plots are all *location unknown* and the right side *location known*. On the same row is the same median background level. There is uniformity in both the *location-unknown* and the *location-known* columns for each of the median backgrounds, as seen in the overall plots in Fig. 5. This is also true for the comparison of *location unknown* versus *known* at each median background level, suggesting that the overall pattern of the results is largely unaffected by the methodology employed.

The median reading time per image dropped from 9 down to 4 s when the location was known, while the maximum time to read an image dropped from 171 to 62 s (Fig. 6). The drop in median time to score the image accounted for an overall reduction of 30 % in the total time required to perform the readings, indicating an important reduction when planning a reader study.

Conclusion

This study highlights the importance of choosing a methodology that matches up to the aims of the work. In particular, this work has determined that for low contrast SKE experiments, the contrast detection threshold drops when search was taken out of the task, and this response impacts upon establishing how much object contrast is necessary for detecting relevant regions of interest in a clinical image. If absolute contrast values are required, then location known methodologies may offer excessive values that do not reflect the influences of radiologic search; on the other hand, if trends are being observed and it is paramount to minimize

observer number and time, then the location-known option presents important advantages. Nonetheless, the overall conclusion that GSDF perceptually linearizes object contrast is consistent with both methodologies.

Acknowledgments The authors would like to thank the participants who devoted two reading sessions in support of this research.

References

1. Hemminger BM, Johnston RE, Rolland JP, Muller KE: Introduction to perceptual linearization of video display systems for medical image presentation. *J Digit Imaging* 8:21–34, 1995
2. American College of Radiology. Practice Guideline for Digital Radiography, Reston, VA, 2007
3. Samei E, Badano A, Chakraborty D, Compton K, Cornelius C, Corrigan K, Flynn MJ, Hemminger B, Hangiandreou N, Johnson J, Moxley-Stevens DM, Pavlicek W, Roehrig H, Rutz L, Shepard J, Uzenoff RA, Wang J, Willis CE: Assessment of display performance for medical imaging systems: executive summary of AAPM TG18 report. *Med Phys* 32:1205–1225, 2005
4. IHE Technical Framework Volume I Integration Profiles, Chicago, IL, 2007
5. Kroon H: Overall x-ray system simulation model developed for system design and image quality versus patient dose optimization. SPIE, San Diego, California, USA, 2003
6. Barten PGJ: Contrast sensitivity of the human eye and its effects on image quality. SPIE Optical Engineering Press, Bellingham, WA, 1999
7. Burgess AE, Jacobson FL, Judy PF: Human observer detection experiments with mammograms and power-law noise. *Med Phys* 28:419–437, 2001
8. Wang J, Xu J, Baladandayuthapani V: Contrast sensitivity of digital imaging display systems: contrast threshold dependency on object type and implications for monitor quality assurance and quality control in PACS. *Med Phys* 36:3682–3692, 2009
9. Burgess AE, Jacobson F, Judy P: Mass discrimination in mammography: experiments using hybrid images. *Acad Radiol* 10:1247–1256, 2003
10. Tchou P, Flynn MJ, Peterson E: 2AFC assessment of contrast threshold for a standardized target using a monochrome LCD monitor. *Proc. Medical Imaging 2004: Image Perception, Observer Performance, and Technology Assessment*: San Diego, CA, USA, 344–352
11. Kundel HL: Reader error, object recognition, and visual search San Diego. SPIE, California, USA, 2004
12. Leong DL, Haygood TM, Whitman GJ, Tchou PM, Geiser WR, Carkaci S, Rainford L, Brennan PC: Verification of DICOM GSDF in Complex Backgrounds. *J Digit Imaging* 25:662–669, 2012

13. National Electrical Manufacturers Association: Digital Imaging and Communications in Medicine (DICOM) Part 14: Grayscale Standard Display Function. National Electrical Manufacturers Association, Rosslyn, VA, 2008
14. Bochud F, Abbey C, Eckstein M: Statistical texture synthesis of mammographic images with super-blob lumpy backgrounds. *Opt Express* 4:33–42, 1999
15. Borjesson S, Hakansson M, Bath M, Kheddache S, Svensson S, Tingberg A, Grahn A, Ruschin M, Hemdal B, Mattsson S, Mansson LG: A software tool for increased efficiency in observer performance studies in radiology. *Radiat Prot Dosimetry* 114:45–52, 2005
16. Dorfman DD, Berbaum KS, Lenth RV, Chen YF, Donaghy BA: Monte Carlo validation of a multireader method for receiver operating characteristic discrete rating data: factorial experimental design. *Acad Radiol* 5:591–602, 1998
17. Dorfman DD, Berbaum KS, Metz CE: Receiver operating characteristic rating analysis. Generalization to the population of readers and patients with the jackknife method. *Invest Radiol* 27:723–731, 1992
18. Hillis SL: A comparison of denominator degrees of freedom methods for multiple observer ROC analysis. *Stat Med* 26:596–619, 2007
19. Hillis SL, Berbaum KS: Power estimation for the Dorfman-Berbaum-Metz method. *Acad Radiol* 11:1260–1273, 2004
20. Hillis SL, Berbaum KS: Monte Carlo validation of the Dorfman-Berbaum-Metz method using normalized pseudovalues and less data-based model simplification. *Acad Radiol* 12:1534–1541, 2005
21. Hillis SL, Berbaum KS, Metz CE: Recent developments in the Dorfman-Berbaum-Metz procedure for multireader ROC study analysis. *Acad Radiol* 15:647–661, 2008
22. Hillis SL, Obuchowski NA, Schartz KM, Berbaum KS: A comparison of the Dorfman-Berbaum-Metz and Obuchowski-Rockette methods for receiver operating characteristic (ROC) data. *Stat Med* 24:1579–1607, 2005
23. Medical Image Perception Laboratory [Internet]. Available at <http://perception.radiology.uiowa.edu/>. Accessed July 27, 2010
24. Friendly M: Mosaic Displays for Multi-Way Contingency Tables. *J Amer Statist Assoc* 89:190–200, 1994
25. Kundel HL, Nodine CF, Carmody D: Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. *Invest Radiol* 13:175–181, 1978