Comparative Performance Evaluation of Contrast-Detail in Full Field Digital Mammography (FFDM) Systems Using Ideal (Hotelling) Observer versus Automated CDMAM Analysis

Ioannis Delakis[1,2], Robert Wise[2], Lauren Morris[2], and Eugenia Kulama[2]

[1] Sidra Medical and Research Centre, Doha, Qatar
idelakis@sidra.org
[2] Radiological Sciences Unit, Charing Cross Hospital, Fulham Palace Road, London, United Kingdom
{robert.wise2,lauren.morris,eugenia.kulama}@imperial.nhs.uk

Abstract. The purpose of our work was to evaluate contrast-detail performance for a range of full field digital mammography systems using Hotelling observer SNR analysis and ascertain whether it can be considered an alternative to CDMAM evaluation. Five FFDM systems were evaluated, which differed in generation (age), Automatic Exposure Control (AEC) behaviour, tube/target combination and detector type. Contrast-detail performance was first analysed using CDMAM phantom analysis and then using the Hotelling observer SNR methodology. The Hotelling observer SNR was calculated for input signal originating from gold discs of varying thicknesses and diameters and then used to estimate the threshold gold thicknesses for each diameter as per CDMAM analysis. There were small differences between the two techniques, especially in small diameter details, which can be attributed to structural characteristics of the CDMAM phantom. The Hotelling observer SNR technique showed lower variability than results from CDMAM analysis. Overall, the Hotelling observer SNR methodology showed variations in the FFDM systems performance consistent with previous findings, demonstrating its value as a performance assessment metric.

Keywords: CDMAM, Hotelling Observer, Ideal Observer, image quality.

1 Introduction

The main methodology used to evaluate mammographic image quality in European quality control programmes is based on the analysis of threshold detectability characteristics, using the contrast-detail phantom for mammography (CDMAM) [1-3]. Images acquired using the CDMAM phantom can be analysed either with observer or automated readings. Observer-based readings are affected by intra-observer error, which can compromise the reliability and confidence of the results. In addition, reading CDMAM images can be time consuming, and therefore often not practical for routine assessment of image quality. Although recent work has tried to link automated

H. Fujita, T. Hara, and C. Muramatsu (Eds.): IWDM 2014, LNCS 8539, pp. 166–173, 2014.

readings with human-observer performance, results can be further dependent on structural differences between CDMAM phantoms [3-4].

As an alternative to CDMAM analysis, the ideal-observer methodology can be used to evaluate threshold detectability characteristics of FFDM systems. An ideal observer is a hypothetical device that performs a given task at the optimal level possible, given the available information and any specified constraints. The correlation between ideal observer results and human performance is dependent on the type of ideal observer used. In this study we used the Hotelling ideal observer as it takes into account both first- and second-order statistics of image data to incorporate some of the human observer limitations [5-6]. Previous work has generalised the definition of Hotelling observer to include the effects of focal spot unsharpness, magnification and scattering, and successfully applied the methodology in the evaluation of FFDM systems [7-12]. The purpose of our work was to study the threshold detectability performance for a range of FFDM systems, using both CDMAM and ideal (Hotelling) observer analysis and ascertain whether the ideal observer methodology can offer some advantages over CDMAM evaluation.

2 Materials and Methods

The FFDM systems included in our study are shown in Table 1 and the properties of their detectors are shown in Table 2.

Table 1. List of FFDM systems included in the study

Manufacturer	System Type	Installation	AEC setting
GE Healthcare	Senographe DS	July 2007	Contrast
GE Healthcare	Senograhe Essential	August 2010	Standard
Hologic	Selenia (Mo target)	June 2007	Autofilter
Hologic	Selenia (W target)	March 2011	Autofilter
Hologic	Dimensions	November 2010	Autofilter

Table 2. Detector properties of FFDM systems included in the study

Manufacturer	System Type	Detector Type	Size (cm)	Pixel size (μm)
GE Healthcare	Senographe DS	CsI (indirect)	19x23	100
GE Healthcare	Senographe Essential	CsI (indirect)	24x29	100
Hologic	Selenia (Mo target)	Selenium (direct)	24x29	70
Hologic	Selenia (W target)	Selenium (direct)	24x29	70
Hologic	Dimensions	Selenium (direct)	24x29	70

The FFDM systems differ with respect to generation, detector, tube technology, as well as the behaviour of their Automatic Exposure control (AEC) and the subsequent choice of target/filter combination, tube potential (kVp) and mAs for clinical exposures. The AEC setting used in this work is the same as what is currently applied on each system for breast screening studies of 60mm breast thickness, as prescribed by European and UK evaluation protocols [1], [13]. All FFDM systems had a nominal focal spot of 0.3mm.

2.1 CDMAM Methodology

FFDM systems were evaluated with the CDMAM phantom following the methodology described by EUREF, the European Reference Organisation for Quality Assured Breast Screening and Diagnostic References [1] and the NHS Breast Screening programme [13]. For each of the FFDM systems listed in Table 1, sixteen "for processing" CDMAM images were evaluated using the CDMAM Analyser (version 1.5.5) and CDCOM (version 1.6), software provided by EUREF.

Threshold detectability was also evaluated with five different CDMAM phantoms, currently in use by mammography physics services in London, as shown in Table 3. CDMAM data acquisition and analysis was repeated on the same FFDM system (GE Healthcare, Senograph DS) and results were compared to identify variability across CDMAM phantoms.

Table 3. CDMAM phantoms used in cross-phantom evaluation

CDMAM #	Serial #	Clinical site
1	1013	The Royal Marsden NHS Trust
2	1036	Bart's Health NHS Trust
3	1683	Imperial College Healthcare NHS Trust
4	1227	Mount Vernon Hospital - Hillingdon Hospital NHS Trust
5	1512	Royal Free London NHS Foundation Trust

2.2 Hotelling Observer SNR Methodology

GNNPS and GMTF Data Acquisition. Estimating the Hotelling observer SNR requires calculation of the Generalised Modulation Transfer Function (GMTF), Generalised Normalised Noise Power Spectrum (GNNPS) and their respective Generalised Noise Equivalent Quanta (GNEQ). The steps involved in the calculation of the Modulation Transfer Function (MTF), Normalised Noise Power Spectrum (NNPS) and Noise Equivalent Quanta (NEQ) have been described extensively in previous work, for example by Marshall (2006) [14]. The Generalised definition of MTF includes the effect of detector blur, focal spot unsharpness, magnification and scatter properties of the system [11]. In our work, the GMTF was measured by calculating the MTF on an image acquired by placing a thin (0.2mm), sharp edge of tungsten foil between two slabs of 25mm PMMA, at a slightly oblique angle (1-2°). The Generalised definition of the NNPS includes the effect of scatter on the input signal and was measured by calculating the NNPS of four "for processing" images of 50mm of PMMA, using the same exposure settings as for the CDMAM methodology. By using the GMTF and GNNPS we then calculated the GNEQ as:

$$GNEQ(f_x,f_y)=\frac{GMTF(f_x,f_y)^2}{GNNPS(f_x,f_y)} \quad (1)$$

where f_x and f_y is spatial frequency in x and y directions, respectively.

Hotelling Observer SNR Calculation. The Hotelling observer SNR can be calculated for any given signal. In order to compare results with CDMAM methodology, the signal (ΔS) used in this study is the Fourier transform of golden discs with varying thickness (h) and diameter (r):

$$\Delta S(f,h,r)=\alpha(h).\frac{\sqrt{3}r}{4f}.J_1.(2\pi fr) \quad (2)$$

where f is the vectorial sum of f_x and f_y, J_1 is the Bessel function of the first kind and $\alpha(h)$ is the radiographic contrast of gold at thickness h. The attenuation characteristics of gold for each kVp and filter/target combination were based on work published by the National Health Service Breast Screening Programme (NHSBSP) and are different for each kV and spectrum (target/filter) [13]. The SNR for this signal was calculated as an integral of the GNEQ over all spatial frequencies, weighted by the spectrum of the signal ΔS [15]:

$$SNR^2=\iint_{f_x,f_y} GNEQ\left(f_x,f_y\right)\ \Delta S(f,h,r)\ df_x\ df_y \quad (3)$$

The SNR calculated for a disc of set diameter and thickness was then compared to a threshold SNR value to determine whether the disc can be considered detected. In order to have a detection task comparable to CDMAM analysis, the threshold SNR was determined on the basis of multiple alternative forced choice (MAFC) analysis, as the CDMAM methodology requires the observer to make a decision on which of four corners the signal is present. Work performed by Burgess (1995) [16] outlines the probability of detection in a MAFC experiment for a given SNR. As typical for 4AFC tests, CDMAM details were considered detected when the probability of detection, based on their Hotelling observer SNR, was equal or higher than 0.625, which is the midway point between 0.25 (random guessing) and 1.00 (perfect response).

Intra-System Reproducibility. In order to compare the short-term reproducibility of CDMAM and Hotelling observer SNR methodologies, the GE Healthcare, Senographe Essential, FFDM system was evaluated at the same time (noon) over a period of five consecutive working days. The mean and standard deviation values of the threshold detectability values were then calculated for results from each methodology.

3 Results and Discussion

Fig. 1 shows the magnitude difference between CDMAM results from each phantom and the overall mean value. The dotted lines in the figure show the average error margin (2sem) expected from CDMAM results at each disc diameter, and indicate that the magnitude difference at small details is equal or greater than the error margin. It is worth noting that CDMAM #3, which is used in subsequent parts of this study, appears to be overestimating threshold detectability, especially for small details, suggesting under-performance of the FFDM system across the range of detail diameters.

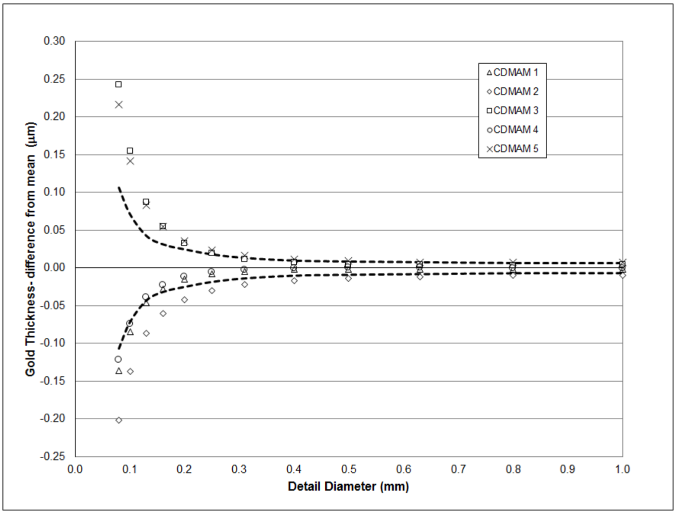

Fig. 1. Magnitude difference of threshold detectability gold thickness results from mean

As shown in Fig. 2, both methodologies indicate that threshold detectability is lower for Selenium compared to CsI detectors, which is consistent with previous work [17] and reflects differences in detector type (direct vs. indirect) and pixel size.

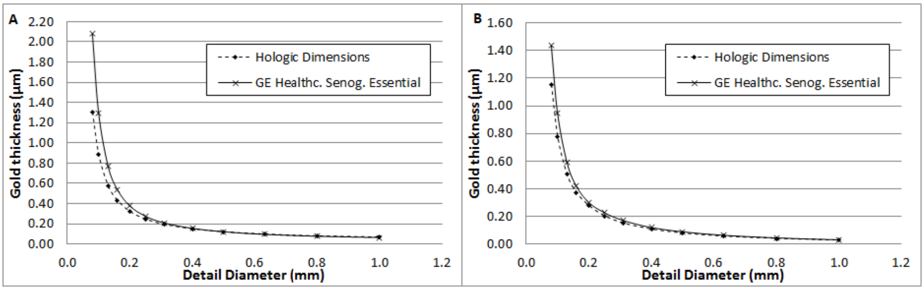

Fig. 2. Threshold detectability results using CDMAM (**A**) and Hotelling observer SNR (**B**) methodologies, for FFDM systems of different detector type

Both methodologies also show an improvement in the performance of new generation FFDM systems, which may be attributed to technological improvements but could also indicate the impact of wear and tear on account of age and workload. As shown in Table 4, the average ratio of threshold detectability for New-to-Old generation FFDM systems tends to be less than unity, indicating the superiority in performance of new compared to older generation FFDM systems. Overall, results show that the Hotelling observer SNR methodology can be a more sensitive performance differentiator.

Table 4. Average ratio of threshold detectability results (New-to-Old FFDM system)

	GE Sen. Essential/ GE Senog. DS	Hologic Selenia (W)/ Hologic Selenia (Mo)	Hologic Dimensions/ Hologic Selenia (Mo)
CDMAM fit to predicted	0.83	0.89	0.99
Hotelling observer	0.89	0.90	0.89

Fig. 3 shows the mean and standard deviation of threshold detectability results using both methodologies on the same FFDM system for five consecutive days. The standard deviation, as indicated by the error bars in Fig. 3, is a measure of the intra-system variability of results from each methodology. On average, there was 7% variation in CDMAM results, reaching 10% for small details (0.08mm diameter). Results using the Hotelling observer SNR methodology were more consistent, with average variation of approximately 1%.

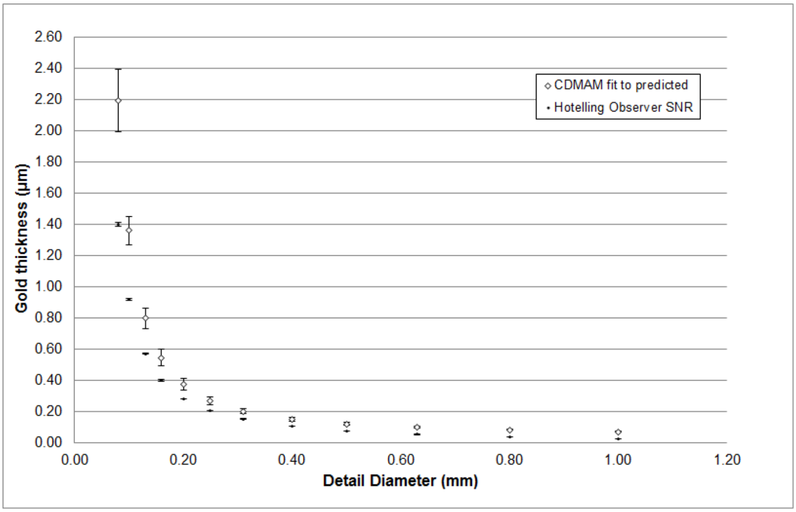

Fig. 3. Mean and standard deviation (error bars) of threshold detectability results using both methodologies on the same FFDM system for five consecutive days

Threshold detectability results using the Hotelling observer SNR methodology were consistently lower than CDMAM results, across all detail diameters and for all FFDM systems, as shown in Table 5. This may reflect the structural characteristics of the CDMAM phantom #3 used in this study, which appears to be overestimating threshold detectability, as previously discussed. The difference between threshold detectability results is higher at small diameter details, which is also consistent with the behaviour of the CDMAM phantom #3, as shown in Fig. 1.

Table 5. Average ratio of threshold detectability results (Hotelling observer-to-CDMAM methodology) for each FFDM system

Manufacturer	System Type	Ratio
GE Healthcare	Senographe DS	0.67
GE Healthcare	Senographe Essential	0.71
Hologic	Selenia (Mo target)	0.83
Hologic	Selenia (W target)	0.85
Hologic	Dimensions	0.74

4 Conclusions

Threshold contrast detectability was evaluated on a number of FFDM systems using both CDMAM and Hotelling observer SNR methodologies. Results showed that the Hotelling observer SNR methodology can be used as a performance metric for FFDM systems, displaying differences with respect to system generation and detector type, in the same way as CDMAM analysis. In addition, Hotelling observer SNR results showed lower variability than CDMAM analysis results in intra-system evaluation.

Our work also identified differences in threshold contrast detectability results when different CDMAM phantoms were used, indicating a potential dependence on structural characteristics of phantoms. The CDMAM phantom used for inter-system comparison in our study appeared to overestimate threshold contrast detectability across all diameter details. CDMAM results were also consistently higher than Hotelling observer SNR results, reflecting a similar behavior.

In conclusion, we have shown that the ideal observer methodology could provide a more reproducible and performance-representative alternative to CDMAM analysis, as it is shows lower variability and is not phantom-specific. The ideal observer methodology also requires fewer exposures than CDMAM methodology, which can be of practical benefit for regular quality control of a large number of clinical FFDM systems, as is the case for large-scale breast screening programmes.

Future work will extend the range, type and number of FFDM systems evaluated using both methodologies, and will perform further comparisons between different CDMAM phantoms.

References

1. Van Engen, R., Young, K., Bosmans, H., Thijssen, H.: The European protocol for the quality control of the physical and technical aspects of mammography screening. Euref, Luxembourg (2006)
2. Young, K., Johnson, B., Bosmans, H., Van Engen, R.: Development of minimum standards for image quality and dose in digital mammography. In: Digital Mammography IWDM 2004, Durham NC, USA (2005)

3. Young, K., Alsager, A., Oduko, J., Bosmans, H., Verbrugge, B., Geerste, T., et al.: Evaluation of software for reading images of the CDMAM test object to assess digital mammography systems. In: Proc. SPIE, vol. 6913, p. 69131C (2008)
4. Young, K.C., Cook, J.J.H., Oduko, J.M., Bosmans, H.: Comparison of software and human observers in reading images of the CDMAM test object to assess digital mammography systems. In: Proc. SPIE, vol. 6142, p. 614206 (2006)
5. Sandrik, J., Wagner, R.: Absolute measures of physical image quality: measurement and application to radiographic magnification 9(4), 540-9 (1982)
6. Barrett, H., Yao, J., Rolland, J., Myers, K.: Model observers for assessment of image quality. Proc. Natl. Acad. Sci. 90(21), 9758–9765 (1993)
7. Kyprianou, I.S.: A method for total x-ray imaging system evaluation. PhD Thesis. University of New York, Buffalo, NY (2004)
8. Kyprianou, I.S., Rudin, S., Bednarek, D., Hoffman, K.: Generalizing the MTF and DQE to include x-ray scatter and focal spot unsharpness: application to a new microangiographic system. Med. Phys. 32(2), 613–626 (2005)
9. Kyprianou, I., Ganguly, A., Rudin, S., Bednarek, D., Gallas, B., Myers, K.: Efficiency of the Human Observer Compared to an Ideal Observer Based on a Generalized NEQ Which Incorporates Scatter and Geometric Unsharpness: Evaluation with a 2AFC Experiment. In: Proc. Soc. Photo. Opt. Instrum. Eng., vol. 5749, pp. 251–262 (2005)
10. Liu, H., Kyprianou, I.S., Badano, A., Myers, K.J., Jennings, R.J., Park, S., et al.: SKE/BKE Task-based methodology for calculating Hotelling observer SNR in mammography. In: Proc. of the SPIE, Medical Imaging, vol. 7258 (2009)
11. Liu, H., Badano, A., Chakrabarti, K., Kaczmarek, R., Kyprianou, I.: Task specific evaluation of clinical full field digital mammography systems using the Fourier definition of the Hotelling observer SNR. In: Proc. of the SPIE, Medical Imaging, vol. 7622 (2010)
12. Liu, H.: Task specific evaluation methodology for clinical Full Field Digital Mammography. PhD Thesis. University of Maryland, College Park (2012)
13. Commissioning and routine testing of Full Field Digital Mammography Systems. NHSBSP Equipment Report 0604. NHS Cancer Screening Programmes, Sheffield. Report No.: version 4 (in press)
14. Marshall, N.: A comparison between objective and subjective image quality measurements for a full field digital mammography system 51(10), 2441–63 (2006)
15. Wagner, R.F., Brown, D.: Unified SNR analysis of medical imaging systems. Phys. Med. Biol. 30, 489–518 (1985)
16. Burgess, A.: Comparison of receiver operating characteristic and forced choice observer performance measurement methods. Medical Physics 22(5), 643 (1995)
17. Review of measurements on full field digital mammography systems. NHSBSP Equipment Report 0901. NHS Cancer Screening Programme (2009)