Image quality assessment in digital mammography: part II. NPWE as a validated alternative for contrast detail analysis

IOP PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **56** (2011) 4221–4238 doi:10.1088/0031-9155/56/14/003

Image quality assessment in digital mammography: part II. NPWE as a validated alternative for contrast detail analysis

P Monnin[1,2], N W Marshall[3], H Bosmans[3], F O Bochud[1] and F R Verdun[1]

[1] Institute of Radiation Physics (IRA), CHUV–UNIL, Rue du Grand Pré 1, 1007 Lausanne, Switzerland
[2] Haute Ecole Cantonale Vaudoise de la Santé (HECVSanté), Filière TRM, Avenue de Beaumont 21, 1011 Lausanne, Switzerland
[3]Department of Radiology, UZ Gasthuisberg, Herestraat 49, 3000 Leuven, Belgium

E-mail: pascal.monnin@chuv.ch

Received 1 March 2011, in final form 4 May 2011
Published 23 June 2011
Online at stacks.iop.org/PMB/56/4221

Abstract

Assessment of image quality for digital x-ray mammography systems used in European screening programs relies mainly on contrast-detail CDMAM phantom scoring and requires the acquisition and analysis of many images in order to reduce variability in threshold detectability. Part II of this study proposes an alternative method based on the detectability index (d') calculated for a non-prewhitened model observer with an eye filter (NPWE). The detectability index was calculated from the normalized noise power spectrum and image contrast, both measured from an image of a 5 cm poly(methyl methacrylate) phantom containing a 0.2 mm thick aluminium square, and the pre-sampling modulation transfer function. This was performed as a function of air kerma at the detector for 11 different digital mammography systems. These calculated d' values were compared against threshold gold thickness (T) results measured with the CDMAM test object and against derived theoretical relationships. A simple relationship was found between T and d', as a function of detector air kerma; a linear relationship was found between d' and contrast-to-noise ratio. The values of threshold thickness used to specify acceptable performance in the European Guidelines for 0.10 and 0.25 mm diameter discs were equivalent to threshold calculated detectability indices of 1.05 and 6.30, respectively. The NPWE method is a validated alternative to CDMAM scoring for use in the image quality specification, quality control and optimization of digital x-ray systems for screening mammography.

1. Introduction

The current fourth edition of the European Guidelines for Quality Assurance in Mammography Screening (European Commission 2006) specifies minimum performance in terms of image quality and radiation dose. An optimal balance between these two requirements must be ensured by choosing a detector with high detective quantum efficiency (DQE) and correctly setting the automatic exposure control (AEC) on the mammography systems. Image quality is determined from the threshold detectability of small defined structures (gold discs of various sizes and thicknesses) contained in the CDMAM contrast-detail phantom. Readers determine the minimal detectability performance in terms of threshold gold thickness for given disc sizes.

This procedure is routinely used in European quality control programmes and suffers from several disadvantages. Subjective image quality evaluation methods based on visual interpretation depend on the variable and subjective nature of human-observer decisions. Previous studies have revealed inter- and intra-observer variability in scoring (Young *et al* 2008, Van Metter *et al* 2006). In addition, partial volume effects may also affect the efficiency of visual scoring of small details compared to the pixel size. This problem may affect the detection of the smallest CDMAM discs for a large number of digital mammography systems. The acquisition of several identical images with slight shift between each shot is therefore necessary to partially overcome this problem. Moreover, observer performance relies on the viewing environment. Finally, the CDMAM test object itself is difficult to manufacture due to the tight tolerances required to generate accurate gold thickness at the micron level. Errors in deposited gold thickness can lead to systematic differences between different CDMAM test objects (Young *et al* 2008). As a consequence CDMAM scoring is time consuming and can be problematic for efficient routine quality control. This variability may affect the result of the threshold visibility test and thus limits the ability of accurately assessing system performance. These limitations can be partially overcome if several observers are used. Nevertheless, the use of multiple observers is impractical for routine quality assurance work. Therefore, software tools have been developed to suppress inter-observer errors and automatically evaluate CDMAM images with an increased scoring efficiency. Processing applied to the automated scoring output can greatly influence the established threshold gold thickness. The complete methodology to obtain threshold gold thicknesses therefore has to be described and thoroughly validated if present limiting values have to be used (Young *et al* 2008).

The lack of reproducibility and a potential bias provide a strong reason to search for an alternative objective method to replace CDMAM scoring. Digital mammography systems provide inherent digital data that can be directly used for quantitative evaluations of image quality which do not rely on human-observer decisions (Bosmans *et al* 2005, Monnin *et al* 2007). In this context, the use of a model observer which links physical image parameters and imaging conditions to object detectability is expected to provide an increase in precision compared to the image scores from human observers and seems attractive (Gagne *et al* 2006). The purpose of this study was to test the validity of a non-prewhitening model observer weighed for the response of the eye (NPWE) to predict CDMAM phantom scoring with the automated method that had been validated against human readings (Young *et al* 2008). The NPWE model calculates the perceived signal-to-noise ratio (SNR) for a specific detection task using basic image parameters and the human visual transfer function (VTF) (Wagner *et al* 1979). In principle, the model observer concept can be used to predict the object detectability of any contrast-detail phantom (Marshall 2006). Its validity has already been investigated for screen-film systems and digital detectors (Desponds *et al* 1991, Aufrichtig 1999, Segui and Zhao 2006).

This study calculated a detectability index for the NPWE model observer for 11 different digital mammography systems, based on systematic standardized measurements of the modulation transfer function (MTF), normalized noise power spectrum (NNPS) and image contrast. Correlation between these NPWE detectability indices and threshold contrasts from CDMAM images following the method in Young *et al* (2008) was established. Then we studied whether a specific threshold NPWE detectability index (d') for given disc diameters would predict a system achieving the acceptable image quality levels of the current European Guidelines.

2. Materials and methods

2.1. Mammography systems and data acquisition

Eleven standard digital mammography systems used in a clinical environment were included in the study. They consisted of six flat-panel-based units and one scanned photon counter, here referred to as a group of seven digital radiography (DR) systems, and four computed radiography (CR) systems. Technical parameters for the systems are given in part I of this study (table 1) (Marshall *et al* 2011). The images required for the image quality metrics are briefly listed here. Threshold thickness was measured using the CDMAM test object together with 4 cm poly(methyl methacrylate) (PMMA), which is approximately equivalent to a total thickness of 5 cm PMMA. The beam quality of an AEC-controlled acquisition of 5 cm PMMA was used. The proposed NPWE method requires the pre-sampling MTF, the NNPS and a measure of image contrast at the AEC setting for 5 cm PMMA. MTF and NNPS acquisitions are described in part I of the study. NNPS and contrast were measured from a 5 cm PMMA block containing an aluminium square. Validation of the detectability index d' of the NPWE method against threshold thickness from CDMAM acquisitions was performed as a function of air kerma at the detector; tube voltage and anode/filter were set using the AEC selection for 5 cm PMMA and the tube current–exposure time product (mAs) was varied to give a range of detector air kerma settings.

With regard to image acquisition for the CR units, the plates were erased before conducting exposures in order to avoid background radiation effects on the measurements. All images were stored as 'for processing' DICOM data: images for the DR units were corrected for detector blemishes and gain variations without post-processing. All image data were converted to air kerma values on a pixel-wise basis for the contrast, MTF and NPS calculations using the detector response curves listed in part I of this study (table 1). This produced linearized pixel value data and removed any offset. The air kerma measurements for the detector response curves were made with two dosemeters referenced in part I of this study.

2.2. Contrast-detail acquisitions

Contrast-detail measurements were made with the CDMAM phantom consisting of a $16 \times 24 \times 0.5$ cm^3 Al plate with a matrix of gold discs arranged in 16 rows and 16 columns. Each cell contained two identical discs of a given thickness and diameter, one at the centre and the other at a random corner. Disc thicknesses and diameters decreased logarithmically, respectively, within the ranges of 0.2 to 0.03 mm and 2.0 to 0.006 mm. Two sets of 2 cm thick PMMA plates were used, one positioned below and one above the phantom to give a total attenuation approximately equivalent to 5 cm of PMMA (or 6 cm thickness of 50% glandular breast tissue). The tube potential and anode/filter combination selected by the AEC for the acquisitions are listed in table 2 of part I. Assessing the effect of dose on the

measured threshold thickness required several sets of CDMAM images acquired as a function of dose; one measurement made at the AEC operating point (considered the centre of the air kerma range studied), and additional measurements at air kerma settings above and below the operating point, which were acquired by manual selection of mAs. The mAs operating point was selected by the AEC for the acquisition of 5 cm PMMA since the CDMAM thickness (45 mm) would have made the height of the compression paddle different from the corresponding attenuation thickness. In order to avoid a summation of identical partial volume effects on scoring, the phantom was slightly repositioned between the different images acquired for every condition (defined x-ray beam and dose). Eight CDMAM images were acquired at each dose setting. Antiscatter grids were used for all systems except the scanned photon counting unit.

Threshold thicknesses for the CDMAM images were automatically obtained with the Erica[2] software (version 2.2.3) based on the CDCOM module available as a download on the EUREF website (Verbrugge 2007). The main advantage of automatic reading is the elimination of potential observer error. Other studies have already demonstrated good correlation between the threshold gold thickness determined by human and the automated methods (Young *et al* 2006b, 2008). The software selects the cell corner in which the disc is most likely to be located. A psychometric curve fitting procedure for the scored CDMAM data is then used to calculate a probability of detection for a given disc. The probability levels are fitted by means of a least mean square procedure. The resulting threshold thickness is deduced from the gold thickness which corresponds to a 62.5% correct detection probability. Threshold gold thicknesses for a given size were calculated by averaging results from an analysis of eight similar images, as suggested for determining threshold detail detectability (Veldkamp *et al* 2003).

2.3. NPWE model observer

Human-observer response can be modelled by considering detection scores for a given signal on an image, which is then represented by a detectability index (d'). The NPWE model observer is a mathematical model that gives similar results compared to human observers for detection tasks in the presence of low-pass noise (Wagner *et al* 1979). It assumes spatial stationarity in order to be expressed in the Fourier domain. The detectability index of the NPWE model for a case in which a signal is known with Gaussian, stationary noise and background can be expressed as a function of the relative object contrast (C), the MTF, the NNPS, the object thickness function (S) and the VTF (Carlson *et al* 1977, Burgess 1995). The disc signal was assumed to be isotropic in the spatial domain, and the following polar equation was used:

$$d' = \frac{\sqrt{2\pi}C\int_0^\infty S^2(u)\,\mathrm{MTF}^2(u)\,\mathrm{VTF}^2(u)u\,\mathrm{d}u}{\sqrt{\int_0^\infty S^2(u)\,\mathrm{MTF}^2(u)\,\mathrm{VTF}^4(u)\mathrm{NNPS}(u)u\,\mathrm{d}u}}. \tag{1}$$

The signal spectrum for a disc with radius R was calculated using the first-order Bessel function of the first kind (J_1):

$$S(u) = \frac{R}{u}J_1(2\pi Ru). \tag{2}$$

The pre-sampled MTF for the different mammography systems was measured as described and can be seen in part I of this study (figures 1(a) and (b)).

The VTF represents the eye contrast sensitivity governed by the relationship $k_1u_c^2 \cdot \exp(-u_c/k_2)$, where u_c is the spatial frequency in cycle/degree (Barten 1992). The model for a typical human observer proposed by Kelly (1979) was adopted as it has been found to

give good correspondence for contrast-detail experiments (Aufrichtig 1999, Aufrichtig and Xue 2000, Borasi *et al* 2006). The VTF bandwidth depends on the viewing distance from the image display and on the magnification factor. A typical viewing distance of 40 cm with a magnification factor of 1.5 was used. The VTF curve converted in mm^{-1} was finally

$$\mathrm{VTF}(u) = 29.5\, u^2 \cdot \exp(-4u)\,. \tag{3}$$

2.4. Theoretical contrast-detail analysis and link between T, d′ and CNR

Following initial work by Rose in 1948 and later work by Wagner and Brown in 1985, the ideal observer contrast-detail function has the general form

$$C^2 d^2 K = \text{constant}, \tag{4}$$

where C is the relative signal contrast, d the object diameter and K the mean photon density, proportional to the air kerma at the detector. When the relative contrast is small compared to unity, it is approximately equal to the product of the object linear attenuation coefficient (μ) and its thickness (T) : $C \approx \mu T$. Equation (4) becomes

$$T^2 d^2 K = \text{constant}. \tag{5}$$

According to this result, the barely visible discs should have the same thickness–diameter product for a given energy and imaging system. Under the assumption that there is exclusively uncorrelated Poisson-distributed (quantum) noise, the normalized standard deviation of pixel values (σ_N) varies with the inverse square root of the detector air kerma. For a given disc size, equation (5) would be equivalent to

$$\log T - \log \sigma_N = \text{constant}. \tag{6}$$

For a well-designed detector, the largest noise component is the quantum noise, but there are also other (constant and proportional) noise components that do not vary as functions of the square root of detector air kerma and therefore equation (6) does not hold under all conditions in practice. The total noise is similar in shape to that representing the quantum noise, and a more general power relationship is therefore suitable:

$$\sigma_N = a \cdot K^{-\beta}. \tag{7}$$

This relationship applies to the case where image noise (linearized standard deviation) is normalized by the signal. Without normalization, the noise will increase with increasing air kerma, as shown in part I of this study: $\sigma = a \cdot K^b$. The link between b and β is therefore simply $\beta = 1 - b$. Equation (6) can now be generalized for a non-strictly quantum noise limited situation:

$$\log T - \frac{1}{2\beta} \log \sigma_N = \text{constant}. \tag{8}$$

The theoretical model predicts the type of correlation function between the detectability d' from the NPWE model and CDMAM-based threshold contrasts. To do so equation (1) can be expressed for a given disc size as a function of the image contrast-to-noise ratio (CNR):

$$d' = \kappa \frac{C}{\sigma_N} = \kappa \cdot \mathrm{CNR},$$

with

$$\kappa = \frac{\sqrt{2\pi} \int_0^\infty S^2(u)\mathrm{MTF}^2(u)\mathrm{VTF}^2(u) u \,\mathrm{d}u}{\sqrt{\int_0^\infty S^2(u)\mathrm{MTF}^2(u)\mathrm{VTF}^4(u) \frac{\mathrm{NNPS}(u)}{\sigma_N^2} u \,\mathrm{d}u}}, \tag{9}$$

Table 1. Log–log linear interpolations of the normalized standard deviation of pixel values as a function of detector air kerma.

Detector name	Log(a) (σ)	β (σ)	Correlation coefficient (R^2)
Agfa MM 3.0R	−1.0892 (0.0277)	0.3939 (0.0136)	−0.9982
Agfa HM 5.0	−0.9976 (0.0168)	0.4606 (0.0081)	–0.9995
Carestream EHR-M3	−0.6326 (0.0198)	0.6003 (0.0103)	−0.9994
Fuji Amulet	−0.7649 (0.0398)	0.5890 (0.0217)	−0.9953
Fuji Profect	−0.8508 (0.054)	0.4764 (0.0292)	−0.9944
GE Senographe 2000 D	−0.7435 (0.0261)	0.5711 (0.0135)	−0.9981
GE Senographe DS	−0.8618 (0.014)	0.5296 (0.0070)	−0.9994
GE Senographe Essential	−1.0606 (0.0181)	0.5374 (0.0093)	−0.9986
Hologic Selenia Dimensions	−0.6700 (0.0353)	0.5540 (0.0177)	−0.9985
Sectra MDM	−0.6906 (0.002 21)	0.5137 (0.0013)	−1.0000
Siemens Inspiration	−0.4817 (0.0245)	0.6198 (0.0128)	−0.9992

where C is the signal contrast and σ_N is the standard deviation of pixel values normalized by the mean pixel value. Under the reasonable assumption that the spectral compositions of the MTF and NPS do not change within a detector exposure range around the threshold gold thickness, κ can be considered as a constant specific to an imaging system for a given x-ray beam and viewing conditions. We therefore expect a simple proportionality between d' and image CNR.

For a given x-ray beam spectrum, details (shape and size) and viewing conditions, the threshold gold thickness for a mammography system depends only on the image noise considering equation (8). Under the same conditions, equation (9) predicts that d' is inversely proportional to image noise as the aluminium contrast is kept constant. As a result the threshold gold thickness and the detectability index are related to each other according to a power law

$$\log T + \frac{1}{2\beta} \log d' = \text{constant}. \tag{10}$$

This relationship indicates that a threshold gold thickness should correspond to a threshold d' value for a defined detail. Moreover the proportionality between d' and CNR clearly establishes the correspondence between the threshold contrast visibility and a specific CNR for a given object (shape and size) and a given detector (spatial resolution and noise properties):

$$T \cdot \text{CNR}^{\frac{1}{2\beta}} = \text{constant}. \tag{11}$$

The theoretical relationships developed in this section were applied to the threshold detail detectability measured using the CDMAM test object.

2.5. Contrast measurements

The contrast (C) was assessed with a 10 × 10 × 0.2 mm^3 aluminium foil (purity of 99.5%) positioned at the reference region of interest (ROI), sandwiched between a slab of 2 cm thick PMMA (image detector side) and a slab of 3 cm thick PMMA (x-ray tube side). The contrast between aluminium and PMMA was determined on linearized images from an approximately 5 × 5 mm^2 ROI in the middle of the aluminium foil and four identical ROIs on all four sides of the aluminium object. Small ROIs were used to minimize variability due to large heterogeneous signals in contrast measurements such as the Heel effect (Alsager *et al* 2008). This was most

Table 2. Threshold d' values obtained from log–log linear interpolations and corresponding to the acceptable European limits (0.10 mm diameter discs).

Imaging system	Image contrast	Fitted slope (σ)	Correlation coefficient (R^2)	Threshold d'	Difference with a target $d' = 1.05$ (%)	Threshold detector air kerma (μGy)
Agfa MM 3.0R	0.125	−0.993 (0.099)	−0.9854	1.055	0.47	100.3
Agfa HM 5.0	0.131	−1.111 (0.116)	−0.9842	1.063	1.23	65.2
Carestream EHR-M3	0.130	−0.906 (0.061)	−0.9866	1.049	−0.09	83.2
Fuji Amulet	0.124	−0.862 (0.061)	−0.9832	0.992	−5.48	40.9
Fuji Profect	0.139	−0.900 (0.100)	−0.9820	1.081	2.99	65.9
GE Senographe 2000 D	0.133	−0.903 (0.037)	−0.9942	0.996	−5.1	57.5
GE Senographe DS	0.133	−0.970 (0.039)	−0.9945	0.983	−6.38	45.6
GE Senographe Essential	0.139	−0.911 (0.056)	−0.9833	0.985	−6.2	32.0
Hologic Selenia Dimensions	0.129	−0.906 (0.061)	−0.9866	1.077	3.12	42.0
Sectra MDM	0.085	−0.743 (0.080)	−0.9511	1.065	1.41	41.3
Siemens Inspiration	0.140	−0.988 (0.053)	−0.9943	1.066	1.51	45.3

important for systems without non-uniformity (full flat field) correction, especially for CR systems. The image contrast was calculated as the relative variation of the image signal produced by the 0.2 mm aluminium foil. It was averaged over three identical images per dose according to equation (12) and is listed for every system in table 2:

$$C = \frac{|\mathrm{PV}_{\mathrm{PMMA}} - \mathrm{PV}_{\mathrm{Al}}|}{\mathrm{PV}_{\mathrm{PMMA}}}. \tag{12}$$

Contrast data were used to calculate the CNR as defined in the Guidelines according to equation (13):

$$\mathrm{CNR} = \sqrt{2} \cdot \frac{|\mathrm{PV}_{\mathrm{PMMA}} - \mathrm{PV}_{\mathrm{Al}}|}{\sqrt{\sigma^2_{\mathrm{PMMA}} + \sigma^2_{\mathrm{Al}}}}. \tag{13}$$

2.6. Noise assessment

The 50 mm thick block of PMMA with aluminium square images was also used for noise measurements (three images per air kerma level). The NPS used for d' calculation was assessed according to the IEC protocol 62220-1-2 (IEC 2005) and the NNPS was found by dividing the NPS by the squared mean pixel value of the linearized image. For the calculation, square areas of 512 × 512 pixels were selected on the linearized images positioned 6 cm from the chest wall, adjacent to the aluminium foil. From each image, nine 2D NPS were calculated from 256 × 256 sub-areas, each of them overlapping by 128 pixels in both horizontal and vertical directions. Since background trends may artificially increase the 2D NPS along the two axes (Dobbins *et al* 1995, Bradford *et al* 1999), low-frequency trends of the signal caused by the Heel effect and x-ray beam inhomogeneity were removed by subtracting a two-dimensional second-order polynomial fitting from the 512 × 512 ROI. This function was subtracted so that the mean pixel value of the ROI did not change. The 2D NPS calculation was then performed using a Fourier transformation algorithm applied to each 256 × 256 ROI. Data from the 2D NPS were then rearranged in a single 1D vector as a function of spatial frequency and these 1D NPS curves were then linearly interpolated (without pre-averaging) to obtain a constant interval sampling.

3. Results and discussion

3.1. Threshold detail detectability analysis

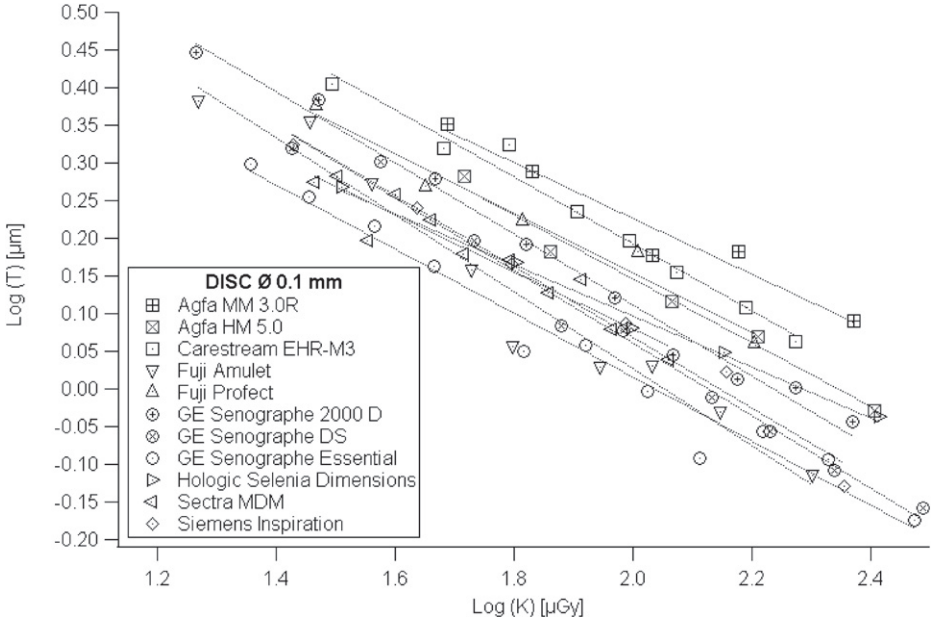
Threshold gold thicknesses obtained from the CDMAM images assessed using the Erica[2] software are shown in figures 1(a) and (b). Threshold thicknesses for the 0.10 and 0.25 mm diameter discs as functions of detector air kerma show a linear dependence in log–log coordinates, confirming the relationship given in equation (4). The results of log–log linear interpolations of the measured normalized standard deviation as a function of detector air kerma are given in table 1. As expected most β coefficients were found to be close to 0.5, and all were found to lie within a range between 0.39 and 0.62. Linear correlation coefficients R^2 were systematically found to be over 0.99. These results are consistent with the noise decomposition conducted in part I for detector noise measured directly from flood images: quantum noise was found to be the dominant noise source for the detector air kerma range under consideration for the contrast-detail measurements. For the normalized (relative) noise form used here (Borasi *et al* 2003) and applied to the contrast-detail data, a β coefficient lower than 0.5 indicates that image noise decreases more slowly than is expected in pure quantum noise as air kerma at the detector is increased. The inverse is true when the β coefficient is higher than 0.5. The former could correspond to a unit for which structured noise becomes increasingly important at high air kerma, while the latter could reveal the relative importance of electronic noise at low air kerma. We note that systems with higher electronic or structured noise according to β from the threshold thickness analysis corresponded to systems with higher electronic or structured noise according to the noise separation presented in part I.

For all mammography units, the power relationship of equation (8) between image noise and threshold gold thickness was confirmed for the 0.10 and 0.25 mm diameter details. The relative error considering equation (8) for the whole detector exposure range investigated in this study remained within $\pm 7\%$. Figures 2(a) and (b) report the relative deviations as a function of dose for the 11 mammography systems.

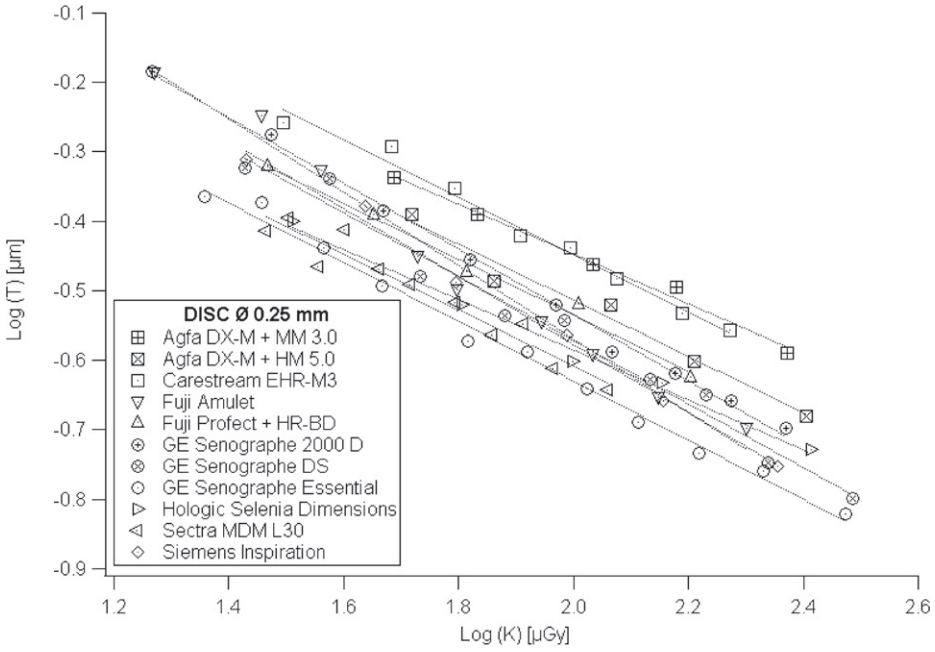
Note that a simple proportionality between T and CNR is assumed in the European Guidelines to calculate the limiting value of CNR from the measured T and CNR to ensure image quality at other thicknesses. According to equation (11), this methodology is correct only if $\beta = 0.5$, i.e. in the case of a strictly Poisson-distributed quantum noise, or within a range of small dose variations around the threshold thickness. Neglecting deviations for systems such as the Agfa MM 3.0 ($\beta = 0.39$) or Siemens Inspiration ($\beta = 0.62$) can result in non-negligible errors in the limiting CNR value assessment, i.e. a case where the CNR no longer relates to the true detectability in the image.

3.2. NPWE model observer performance

The NPWE model observer was used to calculate the detectability index (d') from equation (1) using parameters that correspond to those of the CDMAM acquisitions in terms of the x-ray spectrum and detector air kerma. The pre-sampled MTFs of the mammography systems used in the NPWE model are presented in part I of this study. The MTF averaged over the two pixel matrix directions was used for the d' calculation. It was especially important for systems with non-isotropic resolution properties such as the Fuji Profect or the Sectra MDM. In the same way, the NPS for detectors with non-isotropic noise were also taken as mean curves between the two axial directions whereas radial averaged NPS were used for other systems. This procedure was necessary to ensure the consistency of the d' calculation in the NPWE model since an MTF averaged over all directions (radial MTF) was not measured for non-isotropic systems.



(a)



(b)

Figure 1. (a) Linear log–log interpolations of CDMAM threshold gold thicknesses as functions of detector air kerma (0.10 mm diameter discs). (b) Linear log–log interpolations of CDMAM threshold gold thicknesses as functions of detector air kerma (0.25 mm diameter discs).

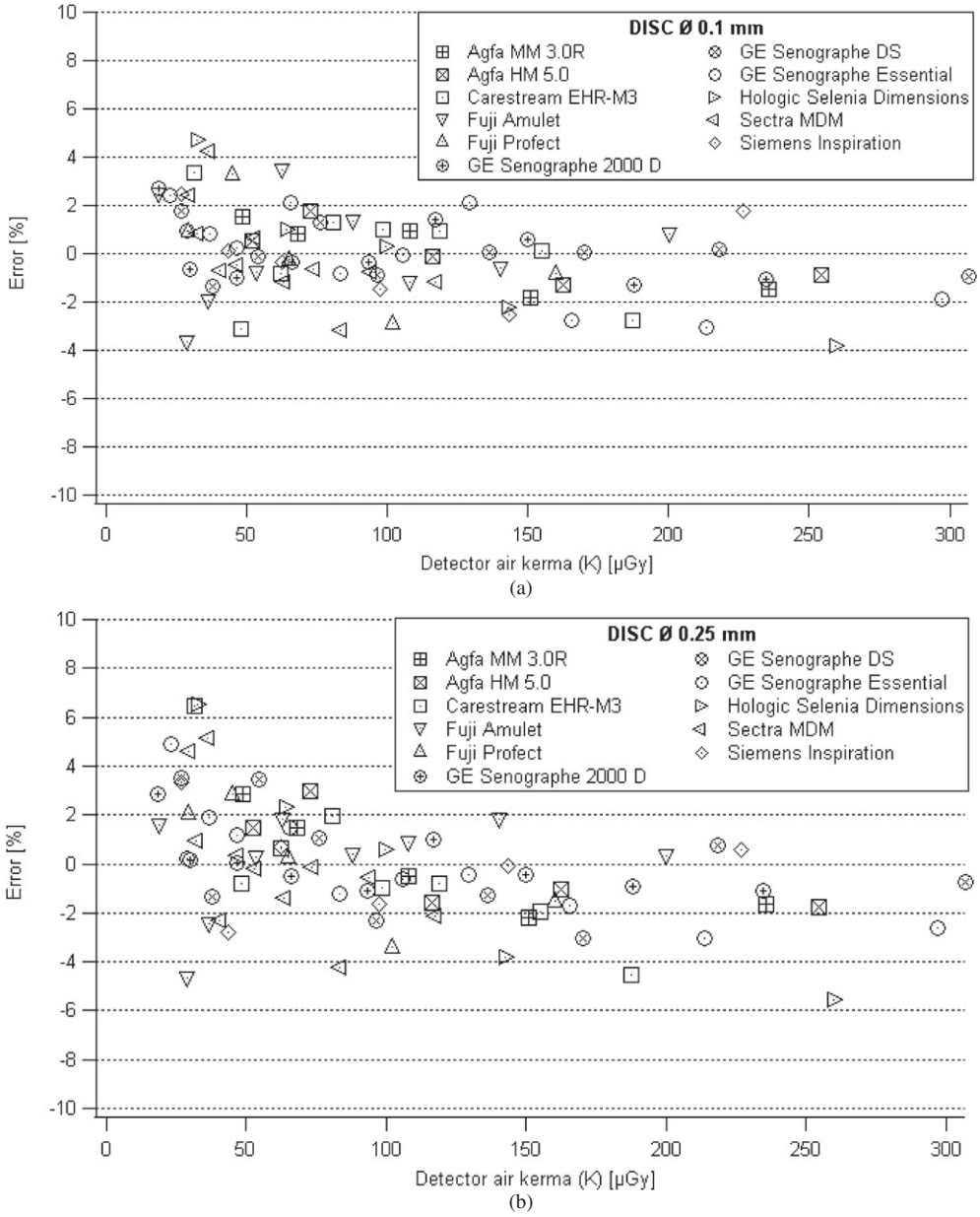


Figure 2. (a) Relative error by considering a linear log–log interpolation between threshold gold thicknesses and pixel standard deviation (0.10 mm diameter discs). (b) Relative error by considering a linear log–log interpolation between threshold gold thicknesses and pixel standard deviation (0.25 mm diameter discs).

Figures 3(a) and (b) plot the CDMAM threshold gold thicknesses for 0.10 and 0.25 mm discs, respectively, as functions of the NPWE detectability index (d') over the entire dose range investigated for the 11 mammography systems involved in this study. The acceptable European threshold gold thickness limit for screening programs is represented with dotted

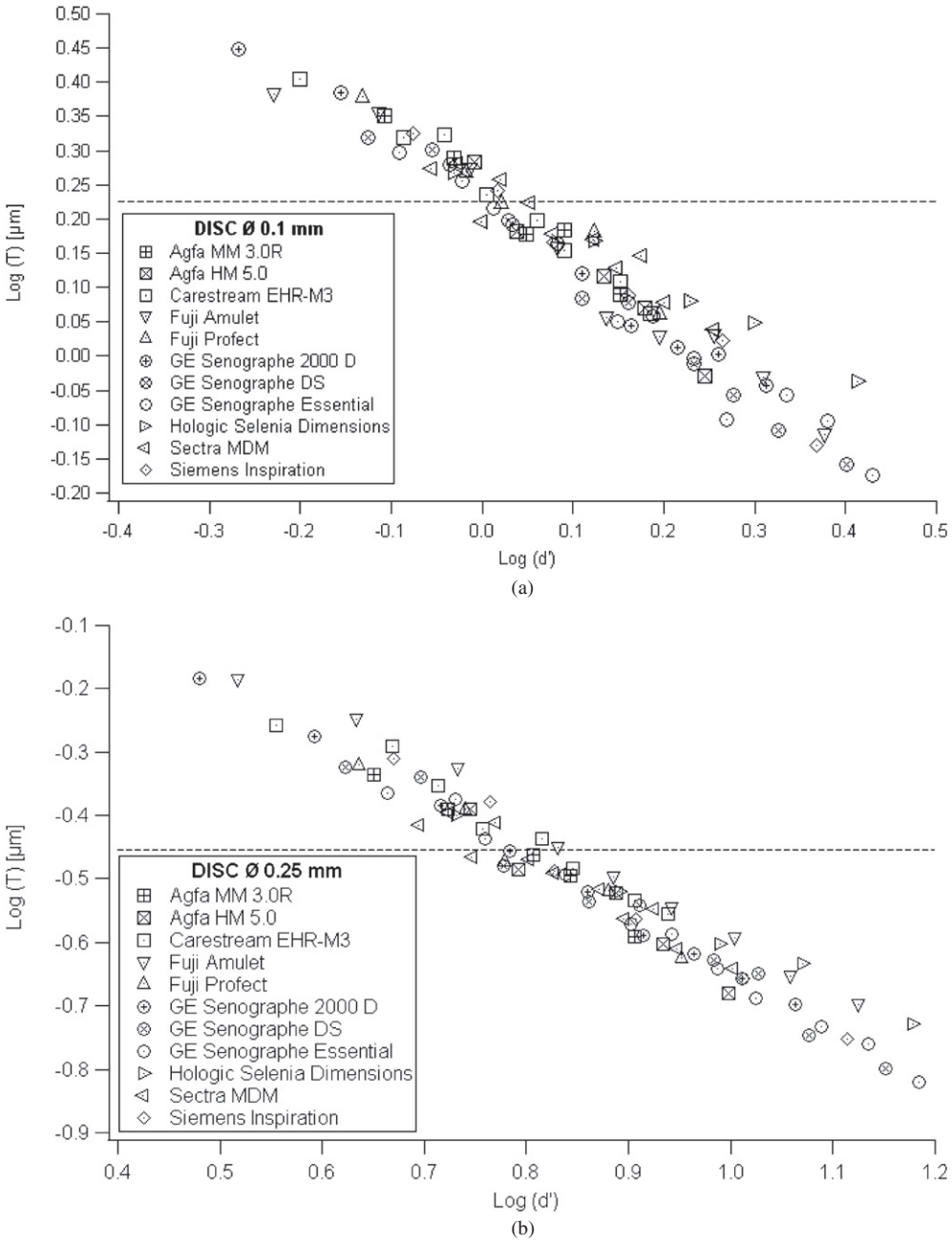


Figure 3. (a) Linear log–log relationship between the threshold gold thicknesses T and the detectability indices d' (0.10 mm diameter discs). (b) Linear log–log relationship between the threshold gold thicknesses T and the detectability indices d' (0.25 mm diameter discs).

Table 3. Threshold d' values obtained from log–log linear interpolations and corresponding to the acceptable European limits (0.25 mm diameter discs).

Imaging system	Fitted slope (σ)	Correlation coefficient (R^2)	Threshold d'	Difference with a target $d' = 6.30$ (%)	Threshold detector air kerma (μGy)
Agfa MM 3.0R	−0.959 (0.087)	−0.9878	6.085	−3.41	103.9
Agfa HM 5.0	−1.053 (0.138)	−0.9751	6.209	−1.44	70.4
Carestream EHR-M3	−0.839 (0.062)	−0.9839	6.592	4.63	95.1
Fuji Amulet	−0.889 (0.030)	−0.9961	6.992	10.98	60.4
Fuji Profect	−0.934 (0.102)	−0.9825	6.073	−3.60	62.7
GE Senographe 2000 D	−0.902 (0.015)	−0.9990	6.074	−3.59	67.3
GE Senographe DS	−0.917 (0.053)	−0.9886	6.085	−3.42	55.4
GE Senographe Essential	−0.911 (0.029)	−0.9954	6.087	−3.39	38.7
Hologic Selenia Dimensions	−0.721 (0.034)	−0.9966	6.333	0.52	43.0
Sectra MDM	−0.792 (0.075)	−0.9619	5.960	−5.40	41.0
Siemens Inspiration	−1.016 (0.051)	−0.9950	6.486	2.95	54.6

lines. Threshold gold thicknesses versus d' indices were fitted with a linear function in a log–log relationship. As suggested by equation (10) slopes close to unity were found as the factor $(2\beta)^{-1}$ was expected to be close to 1 (tables 2 and 3). Correlation coefficients (R^2) over 0.95 were obtained for all systems, indicating only small deviations from the expected logarithmic relationship. The threshold d' values obtained from these log–log linear interpolations and corresponding to the acceptable European limits are summarized in table 3. Acceptable performance in the European Guidelines for 0.10 and 0.25 mm diameter discs is equivalent to target d' of 1.05 and 6.30, respectively, whereas achievable performance in the Guidelines corresponds to target d' of 1.70 and 9.50. Differences between the threshold d' calculated individually for the 11 systems and the target d' values proposed above remain largely within an uncertainty limit of $\pm 10\%$ for both 0.10 and 0.25 mm diameter discs, except for the 0.25 mm detail with the Fuji Amulet system (difference of 11%). This larger difference can be explained in this particular case by the signal pre-processing that gives strongly non-isotropic image noise: the MTF and NPS axial curves used for the d' calculation might not exactly match overall image characteristics, resulting in potential bias with CDMAM results.

Table 3 shows the detector air kerma calculated to be necessary to reach the target d' values of 1.05 and 6.30 corresponding to the threshold gold thicknesses for the 0.10 and 0.25 mm diameter discs, respectively. These dose values were linearly interpolated from the log–log relationships between target d' and detector air kerma, as suggested by equations (5) and (10). The minimum dose to reach the target d' varies substantially between 40 and 120 μGy depending on the system. This large dose range represents considerable differences in detector air kerma for a given detectability task. However, this result was expected regarding the large differences in DQE between the systems, with variations in DQE peaks between 34% and 81%. Clearly, the minimal detector air kerma needed to detect the 0.10 and 0.25 mm details is inversely related to the DQE measured in part I for the mammography systems.

The simple proportionality between d' and CNR suggested by equation (9) was confirmed by CNR measurements for both 0.10 and 0.25 mm diameter discs. Figure 4 presents data just for the 0.10 mm discs, although a similar trend was found for the 0.25 mm discs. Linear interpolations systematically gave correlation coefficients (R^2) over 0.999 for the whole range of exposure settings investigated. A target d' will therefore correspond to a target CNR value. Quality controls based on a target CNR approach have already been proposed as

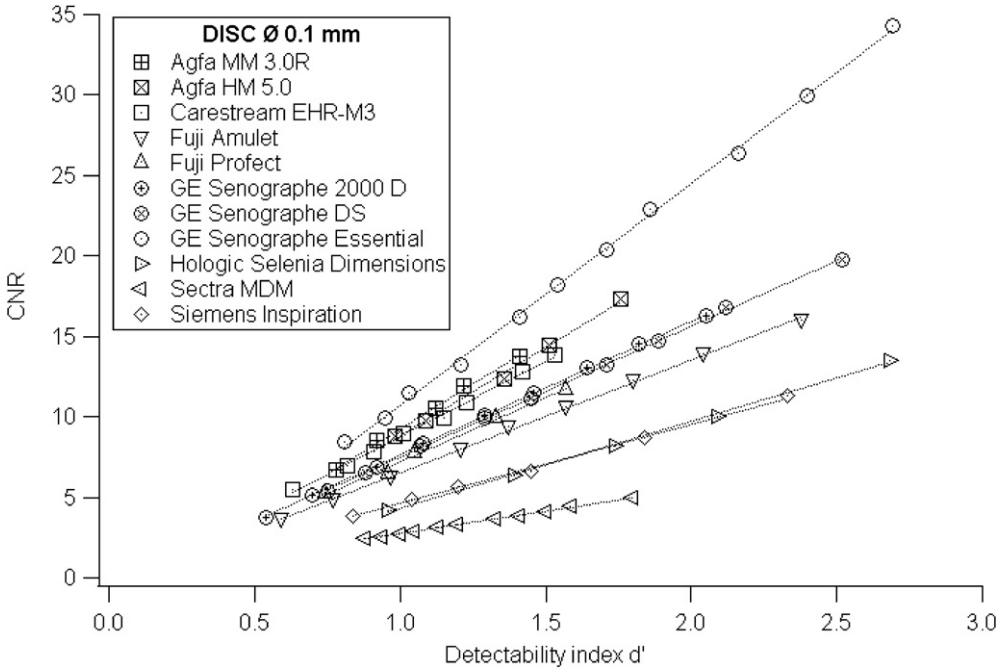


Figure 4. Linear interpolations of image CNR versus detectability indices d' (0.10 mm diameter discs).

an alternative method to contrast-detail measurements. Target CNR values were used by Young *et al* (2006a) as an experimental method to find the minimal dose necessary to achieve the minimum image quality defined in the European Guidelines for a given beam quality and mammography system. The CNR is already well established as means quantifying and optimizing performance (specific to a given system) (Bernhardt *et al* 2006, Williams *et al* 2008); however, a major drawback to the CNR method includes the requirement to set system-specific values and these may need updating following software updates. The linear relationship found between CNR and d' (estimated from homogeneous images) suggests that similar results would be found if d' were to be used instead of the CNR. The d' method depends only on the object shape and size and hence offers a general approach to performance specification and optimization.

Global agreement between threshold gold thicknesses and target d' demonstrates that the NPWE observer model can predict with enough precision the 0.10 and 0.25 mm diameter disc detection from the CDMAM phantom. These results are in accordance with another NPWE model study that provided reasonable prediction of object detectability for an a-Se mammography system with theoretically calculated contrasts and observer performance experiments (Segui and Zhao 2006). Their precision is better than that obtained with the much simpler Rose model under similar conditions (Marshall 2006). The use of d' as an alternative method to CDMAM scoring represents an advantage in terms of rapidity and robustness for efficient routine quality controls. The number of images which have to be acquired for d' calculation is greatly reduced compared to the amount of CDMAM data required. A major advantage of the NPWE model is that the final image quality parameter (d') is separated into the established image quality metrics of contrast, resolution and noise. In the case of poorly performing systems, the potential cause of an image quality problem can be investigated by

examining the contrast, MTF and NNPS results directly. This can be contrasted with the contrast-detail method, in which several system and detector performance characteristics are combined into a single parameter (threshold thickness at some diameter); determining the reasons behind a failing system can be difficult in this case. Moreover, a direct link between d' and the objective image quality metrics such as noise can be used to easily indicate target mAs values and optimal AEC designs of mammography systems (Salvagnini *et al* 2011, Oberhofer *et al* 2008).

3.3. Limitations on use of the NPWE model observer

Agreement between the NPWE model and CDMAM disc visibility was lower for discs larger than 0.25 mm: differences up to 20% between the threshold d' required to achieve the visibility of 0.50 mm diameter discs were found between the different mammography systems. This higher variability is likely due to the fact that large disc spectra contain only low-frequency signals where NNPS and MTF curves are less precise. As a consequence, the NPWE model should not be used to predict detectability for discs larger than 0.25 mm. However, this is a relatively minor limitation within the framework of quality controls because the two detail sizes of 0.10 and 0.25 mm are the most difficult to detect. Other disc sizes specified in the European protocol (0.5, 1.0 and 2.0 mm diameter) give much larger threshold d' than those obtained for discs larger than 0.25 mm (a target d' of 16.8 was found for the 0.50 mm disc). This confirms their minor importance in terms of detection challenge.

One must be careful when calculating d' for systems with non-isotropic noise and resolution properties. The MTF and axial NPS averaged over both axial directions were used in this study for such systems, but the result does not necessarily represent a global image quality over the whole image plane. This can make matching the objective measurements to the CDMAM analysis difficult if the MTF and NPS are not reasonably isotropic within the considered frequency bandwidth. All CR systems plus the Fuji Amulet and Sectra MDM were found to have more or less marked non-isotropic properties in terms of resolution and/or noise. Except for the scanning movement of the Sectra, non-isotropy can result from a signal lag caused by the luminescence decay time that can specifically affect signal sharpness in the laser scan direction for CR systems or from some signal pre-processing applied along a preferential direction (Monnin *et al* 2007, Rowlands 2002, Mackenzie and Honey 2007). Image pre-processing applied to 'for processing' data was found to be the cause of NNPS non-isotropy and a higher maximal difference of +11% between the threshold d' found for the Fuji Amulet and the other systems. A filter designed to remove banding non-uniformities along both the vertical and horizontal directions is integrated in the reading step of this system. Noise contribution along the axes is lowered and slight differences in the resolution and noise behaviours were measured between the two directions (Rivetti *et al* 2009). It is important to note that the filtering effect observed in the NNPS measurements is not necessarily visible in the MTF measurements. This is because the filters are generally applied during the analogue readout process, between the logarithmic or square root data conversion and the A/D converter, in a nonlinear domain. The MTF analysis is based on an edge image containing a large exposure range, from nearly zero up to relatively high values, where nonlinearity effects cannot be neglected. However, the situation for noise analysis is different. The images used for NNPS measurements are uniformly exposed and contain only small variations due to noise which make the pixel values mimic a linear response. As a result, NNPS measurements will clearly show the effects of the filters on the data, whereas a clear representation of the filter effect is not necessarily expected in the MTF measurements (Bloomquist *et al* 2006). The pre-processed filtering can therefore introduce different distortional effects on the MTF

and NNPS curves, and can result in a less precise d' evaluation in regards to CDMAM scoring.

Another difficulty comes from the proportional increase of d' with object contrast which makes precision in contrast evaluation of great importance. A precise contrast measurement was indeed not straightforward: large heterogeneities in the image such as the Heel effect can affect the precision of d' estimation for the CR systems that do not have a non-uniformity (flat field) correction. Therefore, it is important to put the aluminium square at a consistent distance from the bucky side (6 cm) and to average the background pixel value on the four sides of the aluminium foil (Alsager *et al* 2008). Variability in aluminium thicknesses between different machined plates can also strongly affect contrast values and must be avoided. This inter-object variability may also affect CDMAM phantoms with variations in the thickness of the gold discs (Young *et al* 2008). A precise d' evaluation also requires a precise and reproducible method for MTF and NPS curve calculation. Objective performance assessment in terms of MTF and NPS is a well documented area for digital mammography systems. The use of standardized methods for image quality metrics is therefore absolutely necessary.

3.4. Use of the NPWE model observer with nonlinear systems

It is interesting to note that CDMAM was scored using unprocessed but un-linearized images whereas d' was computed with contrast, MTF and NNPS data that had been linearized to air kerma values. This means that CDMAM images acquired using systems with log (or square root) detector response curves might not match the same target detectability index. An aluminium-to-PMMA contrast of about 15% (depending on beam energy) is used to obtain d' whereas threshold CDMAM contrasts vary respectively around 3% and 23% for the 0.10 and 0.25 mm diameter discs (depending on the x-ray beam spectra). However, high contrasts (thicknesses) are potentially more affected by nonlinear phenomena than low (threshold) contrasts.

To test this proposition for systems with a nonlinear response, CDMAM data at low, medium and high doses (eight images at each dose) were linearized to air kerma using the detector response curve and then scaled to 12-bit images (4096 levels). This was done for the Agfa MM 3.0R, the Agfa HM 5.0, the Fuji Amulet and the Fuji Profect systems, giving 12 sets of linearized CDMAM results. These were then compared against the results from un-linearized images. Figure 5 shows data for the Fuji Profect; similar trends were seen for the other systems examined. No systematic trend was seen between linearized and un-linearized data as a function of disc diameter across the four systems, even though contrasts are higher at small detail diameters. The difference in threshold gold thickness between linearized and un-linearized data was therefore averaged over disc diameters for a given dose; the average difference was small and ranged from −2.9% to 5.5%, with a median value of 0.2%. This result can be easily understood by considering that CDMAM analysis is based on barely visible discs containing only a small exposure range around the background value, where nonlinearity effects are expected to be small. It can also be found from equation (9); the threshold thickness for a given disc size corresponds to a target d' which is proportional to the image CNR expressed in relative values. A first-order approximation of contrast and standard deviation for small signal variations (contrasts) based on log and square root response functions gives respectively

$$C \cong \frac{\mathrm{PV}_{\log}}{b \cdot \log e} C_{\log} \qquad \text{and} \qquad \sigma \cong \frac{\mathrm{PV}_{\log}}{b \cdot \log e} \sigma_{\log}, \tag{14a}$$

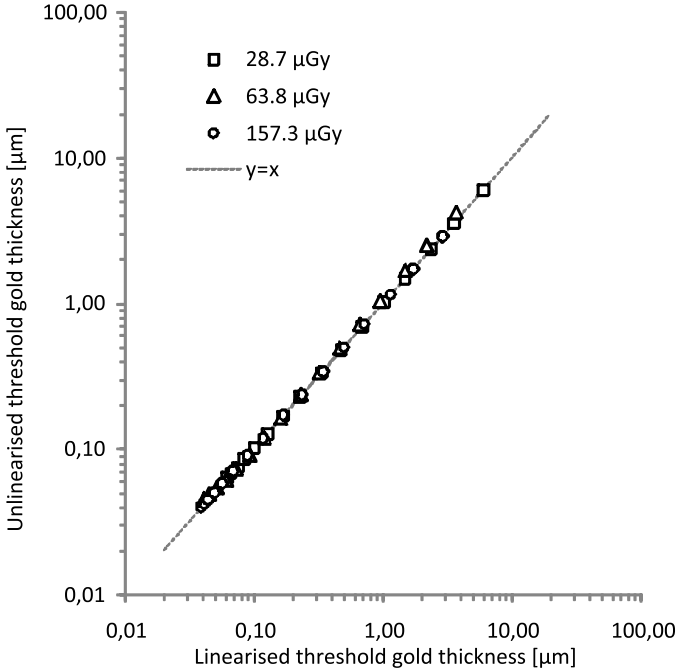


Figure 5. A comparison of the threshold gold thickness scored from linearized versus un-linearized (logarithmic response) CDMAM images at three detector air kerma levels for the Fuji Profect CR system.

$$C \cong \frac{b\sqrt{Q}\cdot \mathrm{PV_{SR}}}{2}\cdot C_{\mathrm{SR}} \qquad \text{and} \qquad \sigma \cong \frac{b\sqrt{Q}\cdot \mathrm{PV_{SR}}}{2}\cdot \sigma_{\mathrm{SR}}, \tag{14b}$$

where $C_{\log}$ and C_{SR} are relative contrasts, $\mathrm{PV}_{\log}$ and $\mathrm{PV_{SR}}$ mean pixel value and $\sigma_{\log}$ and σ_{SR} normalized standard deviation of pixel values computed on logarithmic and square root data, respectively. For weak contrasts, the CNR remains constant to a first-order approximation, and the CNR measured from images with a log or square root signal response are close to the CNR computed from images linearized to air kerma. As a result, the threshold thicknesses do not vary significantly between linearized and pre-processed data.

4. Conclusions

The threshold detail detectability performance measured using the CDMAM test object for 11 different digital mammography systems could be predicted by a NPWE model observer expressed in terms of contrast, MTF and NNPS. This approach utilizes the contrast and NNPS computed from an image of 5 cm PMMA containing a 0.2 mm aluminium contrast, along with the measured pre-sampling MTF. The model predictions were shown to be reliable for the 0.10 and 0.25 mm diameter discs; precision can be affected by MTF and NPS anisotropy. The threshold gold thickness values used to specify acceptable performance in the European Guidelines for 0.10 and 0.25 mm discs can be replaced by a target threshold calculated detectability indices of 1.05 and 6.30, respectively. The target d' methodology proposed here can be used as an alternative to CDMAM scoring in the framework of mammography quality controls.

References

Alsager A, Young K C and Oduko J M 2008 Impact of heel effect and ROI size on the determination of contrast-to-noise ratio for digital mammography systems *Proc. SPIE* **6913** 69134I

Aufrichtig R 1999 Comparison of low contrast detectability between a digital amorphous silicon and screen-film based imaging system for thoracic radiography *Med. Phys.* **26** 1349–58

Aufrichtig R and Xue P 2000 Dose efficiency and low-contrast detectability of an amorphous silicon x-ray detector for digital radiography *Phys. Med. Biol.* **45** 2653–69

Barten G J 1992 Physical model from the contrast sensitivity of the human eye *Proc. SPIE* **1666** 57–72

Bernhardt P, Mertelmeier T and Hoheisel M 2006 X-ray spectrum optimization of full-field digital mammography: simulation and phantom study *Med. Phys.* **33** 4337–43

Bloomquist A K *et al* 2006 Quality control for digital mammography in the ACRIN DMIST trial: part I *Med. Phys.* **33** 719–36

Borasi G, Nitrosi A, Ferrari P and Tassoni D 2003 On site evaluation of three flat panel detectors for digital radiography *Med. Phys.* **30** 1719–30

Borasi G, Samei E, Bertolini M, Nitrosi A and Tassoni D 2006 Contrast-detail analysis of three flat panel detectors for digital radiography *Med. Phys.* **33** 1707–19

Bosmans H, Carton A-K, Rogge F, Zanca F, Jacobs J, Ongeval C V, Nijs K, Steen A V and Marchal G 2005 Image quality measurements and metrics in full field digital mammography: an overview *Radiat. Prot. Dosim.* **117** 120–30

Bradford C D, Peppler W W and Dobbins J T 1999 Performance characteristics of a Kodak computed radiography system *Med. Phys.* **26** 27–37

Burgess A E 1995 Comparison of non-prewhitening and Hotelling observer models *Proc. SPIE* **2436** 2–9

Carlson C R, Cohen R W and Gorog I 1977 Visual processing of simple two-dimensional sine-wave luminance gratings *Vis. Res.* **17** 351–8

Desponds L, Depeursinge C, Grecescu M, Hessler C, Samiri A and Valley J-F 1991 Image quality index (IQI) for screen-film mammography *Phys. Med. Biol.* **36** 19–33

Dobbins J T, Ergun D L, Rutz L, Hinshaw D A, Blume H and Clark D C 1995 DQE(f) of four generations of computed radiography acquisition devices *Med. Phys.* **22** 1581–93

European Commission 2006 The European protocol for the quality control of the physical and technical aspects of mammography screening: part B. Digital mammography *European Guidelines for Breast Cancer Screening* 4th edn (Luxembourg: European Commission)

Gagne R M, Gallas B D and Myers K J 2006 Toward objective and quantitative evaluation of imaging systems using images of phantoms *Med. Phys.* **33** 83–95

IEC 2005 Medical electrical equipment—characteristics of digital x-ray imaging devices: part 1–2. Determination of the detective quantum efficiency—mammography detectors *IEC 62220-1-2* (Geneva: International Electrotechnical Commission)

Kelly D H 1979 Motion and vision II stabilized spatio-temporal threshold surface *J. Opt. Soc. Am.* **69** 1340–9

Mackenzie A and Honey I D 2007 Characterization of noise sources for two generations of computed radiography systems using powder and crystalline photostimulable phosphors *Med. Phys.* **34** 3345–57

Marshall N W 2006 A comparison between objective and subjective image quality measurements for a full field digital mammography system *Phys. Med. Biol.* **51** 2441–63

Marshall N W, Monnin P, Bosmans H, Bochud F O and Verdun F R 2011 Image quality assessment in digital mammography: part I. Technical characterization of the systems *Phys. Med. Biol.* **56** 4201–20

Monnin P, Gutierrez D, Bulling S, Guntern D and Verdun F R 2007 A comparison of the performance of digital mammography systems *Med. Phys.* **34** 906–14

Oberhofer N, Paruccini N and Moroder E 2008 Image quality assessment and equipment optimisation with automated phantom evaluation in full field digital mammography (FFDM) *IWDM* (*Lecture Notes in Computer Science* vol 5116) pp 235–42

Rivetti S, Lanconelli N, Bertolini M, Borasi G, Golinelli P, Acchiapati D and Gallo E 2009 Physical and psychophysical characterization of a novel clinical system for digital mammography *Med. Phys.* **36** 5139–48

Rose A 1948 The sensitivity performance of the human eye on an absolute scale *J. Opt. Soc. Am.* **38** 196–208

Rowlands J A 2002 The physics of computed radiography *Phys. Med. Biol.* **47** R123–66

Salvagnini E, Bosmans H, Monnin P, Struelens L, Verdun F R and Marshall N W 2011 The use of detectability indices as a means of automatic exposure control for a digital mammography system *Proc. SPIE* **7961** 79615J

Segui J A and Zhao W 2006 Amorphous selenium flat panel detectors for digital mammography: validation of a NPWE model observer with CDMAM observer performance experiments *Med. Phys.* **33** 3711–21

Van Metter R, Heath M and Fletcher-Heath L 2006 Applying the European protocol for the quality control of the physical and technical aspects of mammography screening threshold contrast visibility assessment to digital systems *Proc. SPIE* **6142** 614205

Veldkamp W J H, Thijssen M A and Karssemeijer N 2003 The value of scatter removal by a grid in full field digital mammography *Med. Phys.* **30** 1712–8

Verbrugge B 2007 Validation of analysis methods for automated CDMAM reading http://www.kuleuven.be/radiology/lucmfr/erica/downloads/documents/BeatrijsVerbruggeThesisCdmamComparison.pdf

Wagner R F and Brown D G 1985 Unified SNR analysis of medical imaging systems *Phys. Med. Biol.* **30** 489–518

Wagner R F, Brown D G and Pastel M S 1979 Application of information theory to the assessment of computed tomography *Med. Phys.* **6** 83–94

Williams M B *et al* 2008 Optimization of exposure parameters in full field digital mammography *Med. Phys.* **35** 2414–23

Young K C, Alsager A, Oduko J M, Bosmans H, Verbrugge B, Geertse T and Van Engen R 2008 Evaluation of software for reading of the CDMAM test object to assess digital mammography systems *Proc. SPIE* **6913** 69131C-1–11

Young K C, Cook J J H and Oduko J M 2006a Use of the European protocol to optimise a digital mammography system *Proc. 8th Int. Workshop on Digital Mammography (Lecture Notes in Computer Science* vol 4046*)* (Berlin: Springer) pp 362–9

Young K C, Cook J J H, Okudo J M and Bosmans H 2006b Comparison of software and human observers in reading images of the CDMAM test object to assess digital mammography systems *Proc. SPIE* **6142** 614206