Impact of Textured Background on Scoring of Simulated CDMAM Phantom

Bénédicte Grosjean* and Serge Muller

Mammography Department, GE Healthcare
283 rue de la Minière – 78530 BUC – France

Abstract. CDMAM phantom scoring is widely used to assess the detectability performance of mammography systems. We propose to study the impact of structured background on this performance assessment, using simulated CDMAM phantom images with flat and textured backgrounds. Three dose levels have been investigated, ranging from -50% to +50% around the reference dose computed by the acquisition system. For textured backgrounds, the simulated projected breast corresponds to a 50mm thick, 60% glandular breast, with a texture generated by a power-law filtered noise model. Images have been scored by four image quality experts. For the smaller insert sizes, Image Quality Factor (IQF) scores obtained in textured backgrounds are lower than and well correlated with those obtained in flat backgrounds. IQF values increased with dose. For the larger insert sizes, detectability performance in textured background is even more degraded and is not as dose dependent as it is in flat backgrounds.

1 Introduction

The CDMAM phantom is widely used to evaluate the detectability performance of mammographic x-ray equipment. This contrast-detail phantom assesses the ability of a system to distinguish objects with very small contrast and small diameter. The task involved with scoring the CDMAM phantom consists of detecting disc-like inserts of various thicknesses and diameters in flat noisy background. However, such a detection task does not reflect the detection task done by radiologists in clinical conditions. One of the main limitations is the use of flat noisy backgrounds that are not representative of backgrounds associated with clinical breast imaging. The structure of clinical backgrounds is due to overlapping projection of the normal breast anatomical structures in 2D mammograms. In terms of detection performance, it has been shown [1, 2] that radiographic abnormalities detection is limited by both imaging system noise and anatomical noise. Observer experiments [3] demonstrated that the breast structure is often the main limiting factor for lesion detection performance. Bochud et al. [4] showed that for a small object, like microcalcifications, the observer performance is limited by the system noise and eventually by anatomical fluctuations depending on the amplitude of these fluctuations. For large objects, like a nodule, the effect of anatomical fluctuations was found more dominant than system noise.

* benedicte.grosjean@ge.com; Phone: +33 1 3070 9737; Fax: +33 1 3070 4140; http:// www. gehealthcare.com

Susan M. Astley et al. (Eds.): IWDM 2006, LNCS 4046, pp. 460 – 467, 2006.

Synthesizing images gives the opportunity to simulate CDMAM phantom images with spatially varying backgrounds in order to simulate real mammographic backgrounds for detection experiments. Moreover, in contrast with real phantoms, it enables to generate images with textured backgrounds for various texture realizations. Such a simulation tool would enable the assessment of mammography systems with different potential design options through textured CDMAM scoring, approaching a more clinically relevant detection task than with standard CDMAM phantom images.

The purpose of this work was to study the impact of structured background on the detectability performance, function of intensity values, assessed by the CDMAM phantom scoring.

2 Method

The CDMAM 3.4 phantom [5] consists of a matrix of 205 cells. Each cell contains two identical gold disks of given thickness and diameter. One is placed in the center and the other in a randomly chosen corner. The observer has to indicate the corner where the eccentric disk is located. The phantom covers a range of object diameters between 60µm and 2mm, and thicknesses between 0.03 and 2µm, including size and contrast ranges for microcalcifications. We simulated images of this contrast-detail phantom with flat and textured backgrounds (Figure 1) using the same acquisition conditions (Mo/Mo, 28kVp and 3 intensity values equal to 50, 100 and 160mAs) for a Senographe 2000D™ system. We already validated [6] the simulation of the digital mammography system when applied to the simulation of standard CDMAM phantom images. In this study, a power-law model [7, 8] of the projected breast structure has been added [9] to the simulation tool in order to generate structured backgrounds. This power-law model of projected breast structure is based on the average power spectrum of real mammograms under an isotropic assumption. No phase information about mammographic images is included in this model since it makes the assumption of random phase. Nevertheless, detection experiments [11] showed that it can be used to investigate perceptual laws in mammography, leading to similar contrast-detail diagrams as in mammographic textured backgrounds.

For flat backgrounds, we simulated the image of the CDMAM phantom inserted between 2 PMMA plates of 20mm thickness. For textured backgrounds, we considered the projection of a breast with the same thickness than the CDMAM phantom assembly and with a power-law exponent equal to 3. This simulated breast was chosen 60% glandular in order to give the same grey level value than the CDMAM phantom assembly when imaged under the standard technique used to image the CDMAM phantom (Mo/Mo, 28kVp, 100mAs). Here the chosen glandularity would correspond to the glandularity derived from the breast thickness and the grey level measured in the most attenuating area of the breast by the automatic exposure of the mammography system [10]. The grey level ratio between the greatest and least attenuating areas of the breast was fixed based on typical grey level distribution in real mammograms.

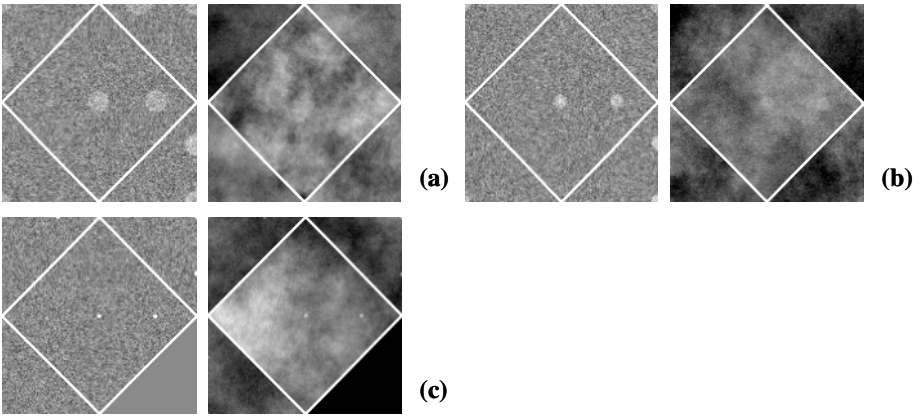

Fig. 1. Simulated images of CDMAM cells with flat (left) and textured (right) backgrounds, at Mo/Mo, 28kVp, 100 mAs. Each cell contains two identical inserts with a diameter of (a) 1.6mm, (b) 1mm, (c) 0.25mm, and a thickness of (a) 0.25µm, (b) 0.36µm, (c) 2µm.

For each acquisition condition and each type of background, four image realizations have been generated and rated by four image quality experts using the mammography-dedicated GE review workstation. The Image Quality Factor (IQF) was calculated on both image sets:

$$\mathrm{IQF} = \frac{16}{\sum_{i=1}^{16} D_i \cdot T_{i,\min}}, \tag{1}$$

where $T_{i,min}$ is the lowest thickness perceived in column i corresponding to an insert diameter D_i. We also defined the $IQF_{smaller\_inserts}$ and the $IQF_{larger\_inserts}$ as the IQF of the eight smaller ($\leq$0.31mm), and eight larger ($\geq$0.4mm) phantom insert sizes:

$$\mathrm{IQF}_{\text{smaller inserts}} = \frac{8}{\sum_{i=1}^{8} D_i \cdot T_{i,min}} \text{ and } \mathrm{IQF}_{\text{larger inserts}} = \frac{8}{\sum_{i=9}^{16} D_i \cdot T_{i,\min}}. \tag{2}$$

These definitions give higher scores for better detection performance. The three IQF values have then been normalized in order to have their variations in the same range for comparison purpose:

$$\mathrm{IQF}_{\mathrm{N}} = \frac{\sum_{i=1}^{16} D_i}{\sum_{i=1}^{16} D_i \cdot T_{i,\min}}, \mathrm{IQF}_{\mathrm{N,\,smaller\ inserts}} = \frac{\sum_{i=1}^{8} D_i}{\sum_{i=1}^{8} D_i \cdot T_{i,min}} \text{ and } \mathrm{IQF}_{\mathrm{N,\,larger\ inserts}} = \frac{\sum_{i=9}^{16} D_i}{\sum_{i=9}^{16} D_i \cdot T_{i,\min}}. \tag{3}$$

3 Results

The values of the IQF_N, $IQF_{N,\ smaller\_inserts}$ and $IQF_{N,\ larger\_inserts}$ are shown for the four human readers and the three dose levels in Figure 2.

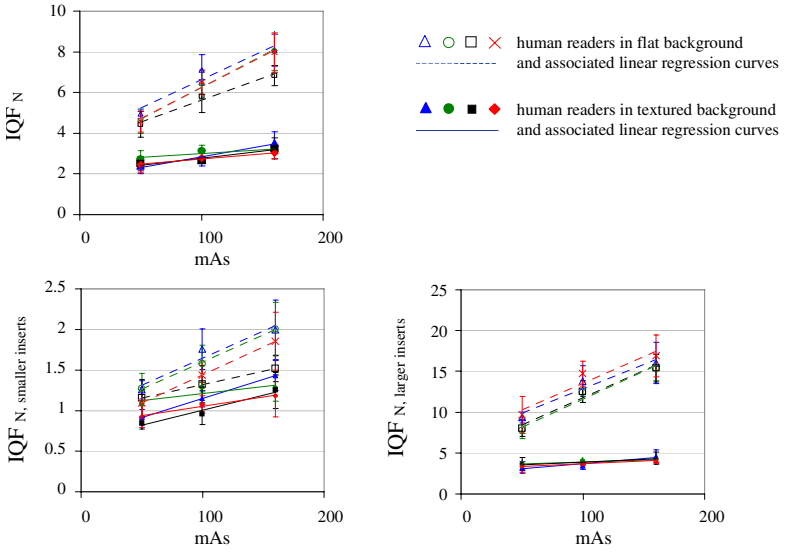

Fig. 2. IQF_N, $IQF_{N,\ smaller\_inserts}$ and $IQF_{N,\ larger\_inserts}$ values derived from the scoring of the 4 readers in flat and textured backgrounds (average over the 4 image realizations), with exposure conditions Mo/Mo, 28kVp and 3 intensity values (50, 100 and 160mAs)

Figure 3 and Figure 4 show contrast-detail curves obtained for the average observer in flat and textured backgrounds for the various intensity mAs values. For each given insert size, the curves indicate the minimal insert thickness needed to reach the detection threshold.

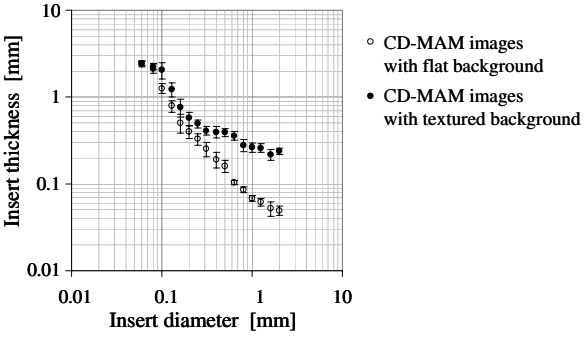

Fig. 3. Contrast-detail curves obtained for the average observer in flat and textured backgrounds, with exposure conditions Mo/Mo, 28kVp, 100mAs

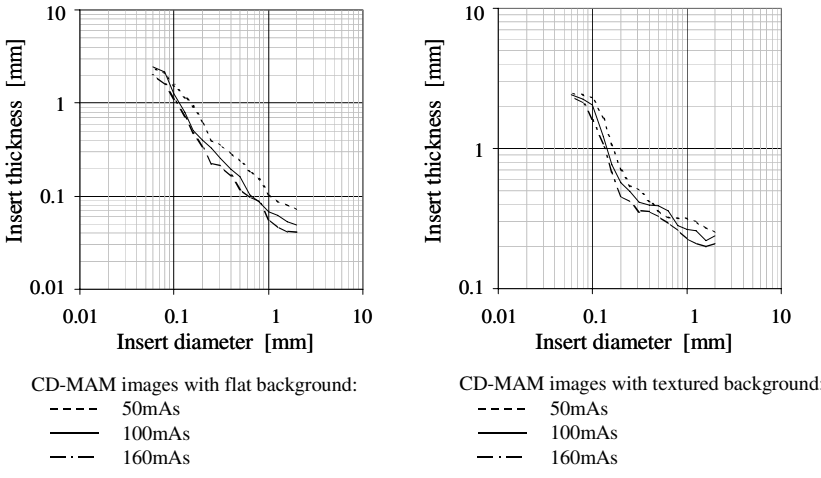

Fig. 4. Contrast-detail curves obtained for the average observer in flat (left graph) and textured (right graph) backgrounds, with exposure conditions Mo/Mo, 28kVp, and intensity values equal to 50, 100 and 160mAs

Figure 2 shows that the resulting IQF_N values are higher in flat compared to textured backgrounds whatever the considered insert size and the mAs intensity value. The contrast-detail curves obtained in Figure 3 highlight that, for a given mAs intensity value, the insert thickness needed to reach the detection threshold becomes much higher in textured backgrounds than in flat backgrounds as the insert size increases. Furthermore, IQF_N values (Figure 2) increase with increasing mAs values in both flat and textured backgrounds. This increasing variation trend is of the same order of magnitude for flat and textured backgrounds when considering the smaller inserts sizes (with a regression slope of $IQF_{N,\ smaller\ inserts}$ versus mAs equal to 0.006 and 0.003 respectively for flat and textured backgrounds). It is much higher for flat than for textured background for the larger inserts sizes (slope of $IQF_{N,\ larger\ inserts}$ versus mAs equal to 0.065 and 0.008 respectively). Regression analysis restricted to the smaller inserts indicates good correlation between $IQF_{N,smaller\ inserts}$ values obtained in flat and textured backgrounds (R-squared=1 for the 3 points corresponding to the 3 intensity values), whereas $IQF_{N,larger\ inserts}$ are weakly correlated for the inserts larger than 0.4mm (R-squared=0.85).

4 Discussion

As expected, we can determine that insert detection performance is degraded in textured backgrounds compared to flat backgrounds. From Figure 3 we can see that the larger the lesion, the higher the degradation. Moreover, detection performance in textured backgrounds, function of the mAs intensity, is poorly correlated to the detection performance in flat backgrounds for the larger insert sizes (Figure 2). For such insert sizes, it is believed that structure noise becomes predominant compared to noise sources induced by the image acquisition processes (quantum noise, scintillator

point spread function, aliasing noise or detector noise) [4, 9, 11]. This can be explained by inspection of power spectra frequency distributions of real mammograms, up to frequencies of about 1cycle/mm [7]. The contribution of the breast structure relative to the other noise sources is more important in the low frequency range. Therefore, as the insert size increases, the mammographic structures amplitude becomes the limiting factor for detection. Taking into account structured backgrounds improves the relevance of the CDMAM phantom scoring, especially for the larger inserts sizes. However, the CDMAM phantom scoring in flat backgrounds remains meaningful for the smaller insert sizes, since we found a good correlation between results obtained in flat and in textured backgrounds. Thus, for comparison of system performance using the CDMAM phantom, we would recommend to restrict the scoring results analysis to the smaller inserts sizes of the phantom.

Furthermore, contrast-detail curves of Figure 4 show negative detection slopes in both flat and textured backgrounds. Burgess showed [11] that in mammographic and also power-law filtered noise backgrounds, for lesion sizes larger than about 1mm, the detection slope is positive, in opposite as intuitively expected. However, this result depends on the shape of the considered signal. Whereas positives slopes were found for shapes corresponding to projected spheres, slopes were negatives for flat-top disc shapes [12]. Thus, the unrealistic disc-like form of the phantom inserts prevents, for the larger insert sizes, from extrapolating the scoring results to clinical performance of large size lesion detection. Furthermore, during the scoring process we noticed that for inserts larger than about 1mm, human readers rely more on the insert edges than on the insert contrast. The simulation of more clinically relevant inserts with smoother edges and with a material composition closer to real microcalcifications would be an additional interesting improvement. The generation of artificial mammographic abnormalities has already been studied in the literature [13, 14]. In future work, it would be interesting to use such inserts as the input to generate simulated CDMAM phantom-like images with textured backgrounds and realistic inserts, in order to provide a contrast-detail test closer to the clinical task.

Van Metter [15] and Young [16] refined the interpretation method for CDMAM phantom scoring results. Indeed, the test suffers from several sources of variability, one of the most significant being the inter- and intra-observer variability. In our study, inter-observer variability and scoring reproducibility must be regarded carefully. The CDMAM phantom scoring procedure [5] applies correction rules, which tend to flatten the variations induced by the presence of texture. Even if the variability seems to be of the same order of magnitude for flat and textured backgrounds, more false positive and false negative detections were found for all the observers in textured backgrounds.

Finally, the opportunity to reduce dose while keeping an acceptable image quality has been described by some authors [4, 17, , 18, 19]. For inserts with contrast in the clinically relevant range 3-30%, the CDMAM provided overlapped curves for dose levels reduced within 40-50% from the reference [18]. These insert contrasts are related to the smaller insert sizes of the CDMAM phantom for the dose levels considered in the study. It has been shown that [19] decreasing dose significantly degrades the detection of microcalcifications, whereas it has minimal effect on the detection of masses. For optimization purposes, if the targeted clinical task consists of detecting the smallest microcalcifications, better performance would be obtained with

higher dose. Moreover, when considering an actual microcalcification, its contrast is often linked to its size. Thus, for a given target size of the smallest microcalcification to be detected, larger microcalcifications will be more easily detectable since they usually lead to a higher contrast in the image. If the optimization task is restricted to mass detection, the dose level could be decreased preserving detection performance since breast structure impairs detection more than any other noise source.

5 Conclusion

CDMAM phantom scoring is very widely used to estimate the detectability performance of mammography x-ray equipment. However, the detection task induced by this contrast-detail phantom is not representative of the clinical task done by radiologists, mainly due to its flat background. We evaluated in this study the impact of structured background on the detectability performance assessed by the CDMAM phantom scoring.

We developed a simulation tool generating quantitative images of CDMAM phantom, depending on the exposure spectrum, and including a model of the projected breast structure based on the average power spectrum of real mammograms. We generated simulated CDMAM phantom images with flat and textured backgrounds for the standard exposure technique of the CDMAM phantom (Mo/Mo, 28kVp) and for three mAs intensity levels (50, 100 and 160mAs). Images have then been scored by four readers. Scoring results show that detection performance in textured backgrounds is degraded compared to flat backgrounds. This degradation increases with increasing insert size. We found that the IQF values obtained in textured and flat backgrounds, function of the mAs intensity level, are well correlated for the smaller insert sizes. This confirms the relevance of the scoring of the CDMAM phantom in flat background for the smaller insert sizes. However, the correlation was weak for the larger insert sizes, since breast structure becomes the limiting detection factor compared to other noise sources for such insert sizes. As a result, for comparison of system performance, the CDMAM phantom scoring analysis should be restricted to the smaller insert sizes.

Furthermore, the simulation tool developed in this study offers the potential to validate new design options with a more clinically relevant detection task than standard CDMAM phantom images.

References

1. G. Revesz, H.L. Kundel, and M.A.Graber, “The influence of structured noise on the detection of radiological abnormalities”, *Invest. Radiol.*, 9(6), pp.479-486, 1974
2. J.P. Rolland, and H.H. Barrett, “Effect of random background inhomogeneity on observer detection performance”, *J. Opt. Soc. Am.*, A(9), pp.649-658, 1992
3. A.E. Burgess, F.L. Jacobson, and P.F. Judy, “Human observer detection experiments with mammograms and power-law noise”, *Med. Phys.*, 28(4), pp.419-437, 2001
4. F.O. Bochud, J.F. Valley, F.R. Verdun, C. Hessler, and P. Schnyder, “Estimation of the noisy component of anatomical backgrounds”, *Med. Phys.*, 26(7), pp.1365-1370, 1999

5. K. Bijkerk, M. Thijssen, and Th. Arnoldussen, *Manual CDMAM-phantom type 3.4*, University Medical Centre, Nijmegen, The Netherlands, 2000
6. B. Grosjean, S. Muller, H. Souchay, R. Rico, X. Bouchevreau, "Automated scoring for CDMAM phantom from simulated images", *Proceedings of IWDM 2004*, 2004
7. A.E. Burgess, "Bach, breasts, and power-law processes", *Proceedings of SPIE*, Vol.4324, pp.103-113, 2001
8. J.J. Heine, and R.P. Velthuizen, "Spectral analysis of full filed digital mammography data", *Med. Phys.*, 29(5), pp.647-661, 2002
9. B. Grosjean, S. Muller, H. Souchay, "Lesion detection using an a-contrario detector in simulated digital mammograms", *Proceedings of SPIE*, to be published in 2006
10. N. Shramchenko, P. Blin, C. Mathey, and R. Klausz, "Optimized exposure control in digital mammography", *Proceedings of SPIE*, Vol.5368, pp.445-456, 2004
11. A.E. Burgess, F.L. Jacobson, and P.F. Judy, "Lesion detection in digital mammograms", *Proceedings of SPIE*, Vol.4320, pp.555-560, 2001
12. A.E. Burgess, "Evaluation of detection model performance in power-law noise", *Proceedings of SPIE*, Vol.4324, pp123-132, 2001
13. A.K. Carton, H. Bosmans, C. Van Ongeval, G. Souverijns, F. Rogge, A. Van Steen, and G. Marchal, "Development and validation of a simulation procedure to study the visibility of microcalcifications in digital mammograms", Med. Phys., Vol.30(8), pp2234-2240, 2003
14. M. Ruschin, A. Tingberg, M. Bath, A. Grahn, M. Hakansson, B. Hamdal, I. Andersson, "Using simple mathematical functions to simulate pathological structures – input for digital mammography clinical trial", *Radiat. Prot. Dosimetry*, Vol.114(1-3), pp.424-431, 2005
15. R. Van Metter, M.D. Health, L.M. Fletcher-Health, "Applying the European protocol for the quality control of the physical and technical aspect of mammography screening to digital systems", *Proceeding of SPIE*, to be published in 2006
16. K.C. Young, J.J. Cook, J.M. Oduko, H.T.Bosmans, "Comparison of software and human observers in reading images of the CDMAM test object to assess digital mammography systems", *Proceeding of SPIE*, to be published in 2006
17. R.P. Highnam, J.M. Brady, and B.J. Shepstone, "Mammographic image analysis", *Eur. J. Radiol*., Vol.24(1), pp.20-32, 1997
18. G. Gennaro, L. Katz, H. Souchay, C. Alberelli, and C. di Maggio, "Are phantoms useful for predicting the potential of dose reduction in full-filed digital mammography?", *Phys. Med. Biol.*, 50, pp.1851-1870, 2005
19. A.S. Chawla, R.S. Saunders and E. Samei, "Effect of dose reduction on the detection of mammographic lesions based on mathematical observer models", *Proceedings of SPIE*, to be published in 2006.