

Towards a Phantom for Multimodality Performance Evaluation of Breast Imaging: A 3D Structured Phantom with Simulated Lesions Tested for 2D Digital Mammography

Kristina Tri Wigati^{1,2}, Lesley Cockmartin³, Nicholas Marshall^{1,3},
Djarwani S. Soejoko², and Hilde Bosmans^{1,3}(✉)

¹ Department of Imaging and Pathology, Medical Imaging Research Center,
KU Leuven, Leuven, Belgium

kristinatri.wigati@student.kuleuven.be,

² Department of Physics, Faculty of Mathematics and Natural Sciences,
University of Indonesia, Depok, Indonesia

djarwani@fisika.ui.ac.id

³ Department of Radiology, University Hospitals Leuven,
Leuven, Belgium

{lesley.cockmartin,nicholas.marshall,
hilde.bosmans}@uzleuven.be

Abstract. The aim of this work is to test whether a 3D structured phantom with simulated lesions can be used for performance evaluation of 2D digital mammography, as a step towards a multimodality phantom. A phantom, developed for breast tomosynthesis was therefore applied on 23 digital mammography systems. Ten images were acquired at the clinically used dose and for 11 systems also at half and double dose. The images were read in a four-alternative forced choice (4-AFC) paradigm by 5 readers. CDMAM phantom acquisitions were also performed. It was possible to calculate diameter thresholds of the simulated masses and microcalcifications that guarantee 62.5 % correct response. The results showed the expected sensitivity with mean glandular dose: detectability of microcalcifications improved with dose, whereas the detectability of masses was not affected. Systems of the same manufacturer and operated at similar doses had very similar detectability scores. Percentage correctly detected microcalcifications with average diameter 119 μm correlated with CDMAM based gold thickness thresholds. Present phantom, developed and tested for tomosynthesis, is also a good candidate for 2D mammography, suggesting its use for (future) benchmarking of at least two types of imaging systems.

Keywords: 3D structured phantom · Spiculated masses · Non-spiculated masses · Microcalcifications · Image quality

1 Introduction

With new imaging modalities being introduced for radiological imaging of the breast, a generally accepted phantom for performance testing would be of great use. Current new generation mammography devices include breast tomosynthesis, contrast

enhanced spectral mammography and breast CT. Their (future) role has to be explored and justified. All new modalities aim to enhance particular features of lesions or breast tissue, yet at this point, 2D digital mammography remains the standard in breast cancer screening. Justification of new modalities may therefore be helped from a comparison with 2D mammography.

Test objects that have been used extensively in full field digital mammography (FFDM) are not well suited to judge the advantages of new modalities if the elements of potential benefits are not included in the test. As an example: tomosynthesis is being developed for its ability to reduce overlapping tissue. This aspect cannot be tested with a test object that lacks any 3D structure around a detectability target.

The challenge of testing digital breast tomosynthesis (DBT) had made us develop a new phantom with a 3D structure and simulated lesions for task based performance evaluation in terms of lesion detectability. The choices of number and types of lesions allowed a four-alternative forced choice (4-AFC) evaluation mode. The aim of present study is to test whether the same 3D structured phantom with simulated lesions can be used for performance evaluation of 2D digital mammography, making it a second modality that can be tested with the same phantom and providing basically the same metric as in DBT, namely detectability of lesions. This approach may help in justification cases or for benchmarking of systems.

Present study includes three main phases. First, establish the performance scores that 2D digital mammography obtains with the new phantom when being used under clinical exposure conditions, by calculating diameter thresholds and number of detected targets. This will then automatically lead to minimal scores that a 2D digital mammography unit should reach. Second, study the sensitivity of the 3D structured phantom in response to dose variation. Third, compare the applicability of the new phantom for performance testing with the experience of a routinely used phantom with the same purpose, namely the CDMAM phantom.

2 Materials and Methods

2.1 The 3D Structured Phantom

Figure 1 and Table 1 show the 3D structured phantom and its inserted lesions [1]. The 3D structured background was created with acrylic spheres of 6 different diameters (15.88, 12.70, 9.52, 6.35, 3.18, and 1.58 mm) (United States Plastic Corp., Ohio, USA) that were submersed in a cylindrical container that was otherwise filled with water [2–4]. The container has a semi-circular shape with a thickness of 48 mm and diameter of 200 mm, equivalent with a compressed breast thickness of 60 mm. Five spiculated and 5 non-spiculated 3D printed masses, and 5 microcalcification groups with various sizes, aimed for covering a clinically relevant range, were inserted in the phantom. Spiculated and non-spiculated masses were created from a database of 3D voxel models by Shaheen *et al.* [5]. Two lesions, that, after insertion in DBT, had received a high realism and high malignancy score by the radiologists, were selected for 3D printing. The masses were made of the material NEXT (Materialize, Leuven) in order to approximate the linear attenuation coefficient of cancerous tissue in real breast [1].

The microcalcifications are formed from calcium carbonate (CaCO₃) (Leeds Test Objects, Leeds, UK) and available in 5 size groups. The spheres are not fixed in their position, while the lesions have been glued on a little support in PMMA. This approach generates always new 3D structured backgrounds after a simple shaking of the phantom. This variability was introduced in the phantom on purpose: it is a straightforward way for the application of model observer based evaluations in digital breast tomosynthesis (DBT) and 2D digital mammography. Present study will be followed by more studies developing automated reading, most probably using the model observer approach.

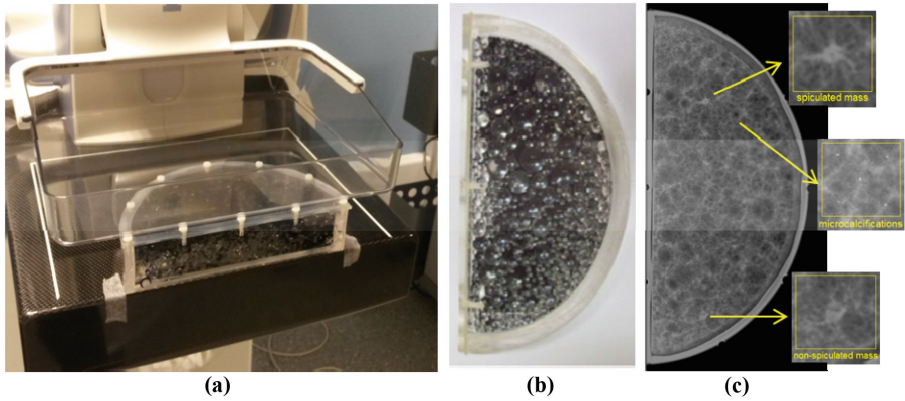


Fig. 1. (a) A semi-circular phantom with structured background filled with spheres of 6 different diameters and inserted target objects representing spiculated masses, non-spiculated masses and microcalcifications [1], (b) Photograph of the phantom from the top, and (c) “For presentation” 2D mammogram of the phantom

Table 1. Diameter of the inserted target object.

Spiculated masses				Non-spiculated masses				Microcalcifications	
X (mm)	Y (mm)	Z (mm)	Average (mm)	X(mm)	Y(mm)	Z(mm)	Average (mm)	Diameter range (µm)	Average (µm)
4.4	4.4	1.6	3.8	1.6	1.4	1.9	1.8	90-100	95
6.1	6.1	2.3	4.8	2.2	2.0	2.7	2.3	112-125	119
8.8	8.8	3.2	6.9	3.2	2.8	3.8	3.3	140-160	150
12.2	12.2	4.5	9.7	4.5	4.0	5.3	4.6	180-200	190
16.6	16.6	6.1	13.1	6.1	5.4	7.2	6.2	224-250	237

2.2 Image Acquisitions

Twenty three digital mammography systems that are used on routine basis in the Belgian breast cancer screening were included in the study. All the systems and their exposure factors are shown in Table 2. Eleven systems were assessed with the 3D structured phantom and the CDMAM phantom at three dose levels: the clinically used dose (also called AEC dose, as this is the setting controlled by the automatic exposure control (AEC) of the system for the phantom), half this dose (also called half AEC dose), and double AEC dose. Twelve systems were only assessed with the 3D structured phantom at AEC dose level. For each dose level, 10 “for presentation” images of the 3D structured phantom and at least eight “for processing” images of the CDMAM phantom were acquired on each system.

Table 2. Digital mammography systems in the present study and their exposure factors.

No	System	A/F	kV	mAs manual		
				Half AEC	AEC	Double AEC
1	Agfa DX-M + CR HM 5.0 (1)	Mo/Rh (25 μ m)	30	32.0	63.0	125.0
2	Agfa DX-M + CR HM 5.0 (2)	Mo/Rh (25 μ m)	28		80.0	
3	Fuji Profect CS + CR HR-BD	Mo/Rh (25 μ m)	28	50.0	100.0	200.0
4	Fuji Amulet S	W/Rh (50 μ m)	30	45.0	90.0	180.0
5	Fuji Amulet Innovality	W/Rh (50 μ m)	30	36.0	71.0	138.0
6	GE Senographe Essential (1)	Rh/Rh (25 μ m)	29	28.0	56.0	110.0
7	GE Senographe Essential (2)	Rh/Rh (25 μ m)	29	36.0	71.0	140.0
8	GE Senographe Essential (3)	Rh/Rh (25 μ m)	29		71.0	
9	GE Senographe Essential (4)	Rh/Rh (25 μ m)	29		63.0	
10	GE Senographe Essential (5)	Rh/Rh (25 μ m)	29		71.0	
11	GE Senographe Essential (6)	Rh/Rh (25 μ m)	29		56.0	
12	Hologic Selenia Dimensions	W/Rh (50 μ m)	31	70.0	140.0	280.0
13	IMS Giotto Image 3DL	W/Rh (50 μ m)	31	99.0	198.0	300.0
14	Philips MicroDose L50	W/Al (50 μ m)	35	6.5	12.5	23.5
15	Philips MicroDose L30 (1)	W/Al (50 μ m)	38		12.0	
16	Philips MicroDose L30 (2)	W/Al (50 μ m)	32		16.0	
17	Siemens Mammomat Inspiration (1)	W/Rh (50 μ m)	30	50.0	100.0	200.0
18	Siemens Mammomat Inspiration (2)	W/Rh (50 μ m)	30		90.0	
19	Siemens Mammomat Inspiration (3)	W/Rh (50 μ m)	30		100.0	
20	Siemens Mammomat Inspiration (4)	W/Rh (50 μ m)	30		90.0	
21	Siemens Mammomat Inspiration (5)	W/Rh (50 μ m)	30		90.0	
22	Siemens Mammomat Inspiration (6)	W/Rh (50 μ m)	31		80.0	
23	Siemens Mammomat Inspiration PRIME	W/Rh (50 μ m)	30	45.0	90.0	180.0

2.3 Image Analysis

The 3D structured phantom images were read in a 4 AFC paradigm: a segment with lesion (signal) was shown next to 3 signal free images. These image compositions were shown in random order to 5 human readers using our in house developed visualisation

program [6]. Since there were 15 inserted target objects (5 spiculated masses, 5 non-spiculated masses, and 5 groups of microcalcifications) in ten “for presentation” images, 150 images of signal and 450 selected images of background had to be prepared for every system. In this study, each image segment of signal and background had a dimension of 20 mm × 20 mm. Two hundred sets of image segments of background were used out of which 450 segments have been randomly chosen, some of them obviously several times. The observers were allowed to adjust contrast, brightness and zoom level to reach the optimum condition for observation. Ultimately, the percentage correctly detected signals (PC) was calculated.

The PC value was then fitted as a function of diameter. As the psychometric curve had shown high correlation coefficients, this function was applied and subsequently a diameter threshold was defined, namely the diameter corresponding to a PC of 62.5 %:

$$PC = 0.25 + \frac{0.75}{1 + \left(\frac{d}{dtr}\right)^{-f}} \quad (1)$$

where d is diameter, dtr is diameter threshold, and f is a parameter of the psychometric curve that is obtained after fitting [1, 7].

The CDMAM phantom acquisitions were analyzed with CDCOM software following the procedure in the European Guidelines, a common practice in our quality control activities [8, 9].

2.4 Mean Glandular Dose Calculation

Mean glandular dose (MGD) was calculated according to the work of Dance *et al.* and copied in the European guidelines [8, 9].

$$MGD = K g c s \quad (2)$$

where K is the incident air kerma (without backscatter, filtered by the compression paddle) at the top surface of the breast. The g -factor transforms the incident air kerma into absorbed dose to the glandular tissue, assuming a glandularity of 50 %. The c -factor is a correction factor that accounts for the glandularity if that is different from 50 %. The s -factor is related with the choice of beam quality used.

3 Results

3.1 The Diameter Threshold

In present study, 14 types of objects were analyzed (4 spiculated masses, 5 non-spiculated masses, and 5 microcalcifications). The phantom contained 5 spiculated masses but one mass was not used for the analysis due to its anomalous characteristics [1]. Figure 2 shows diameter threshold values of spiculated masses (a), non-spiculated

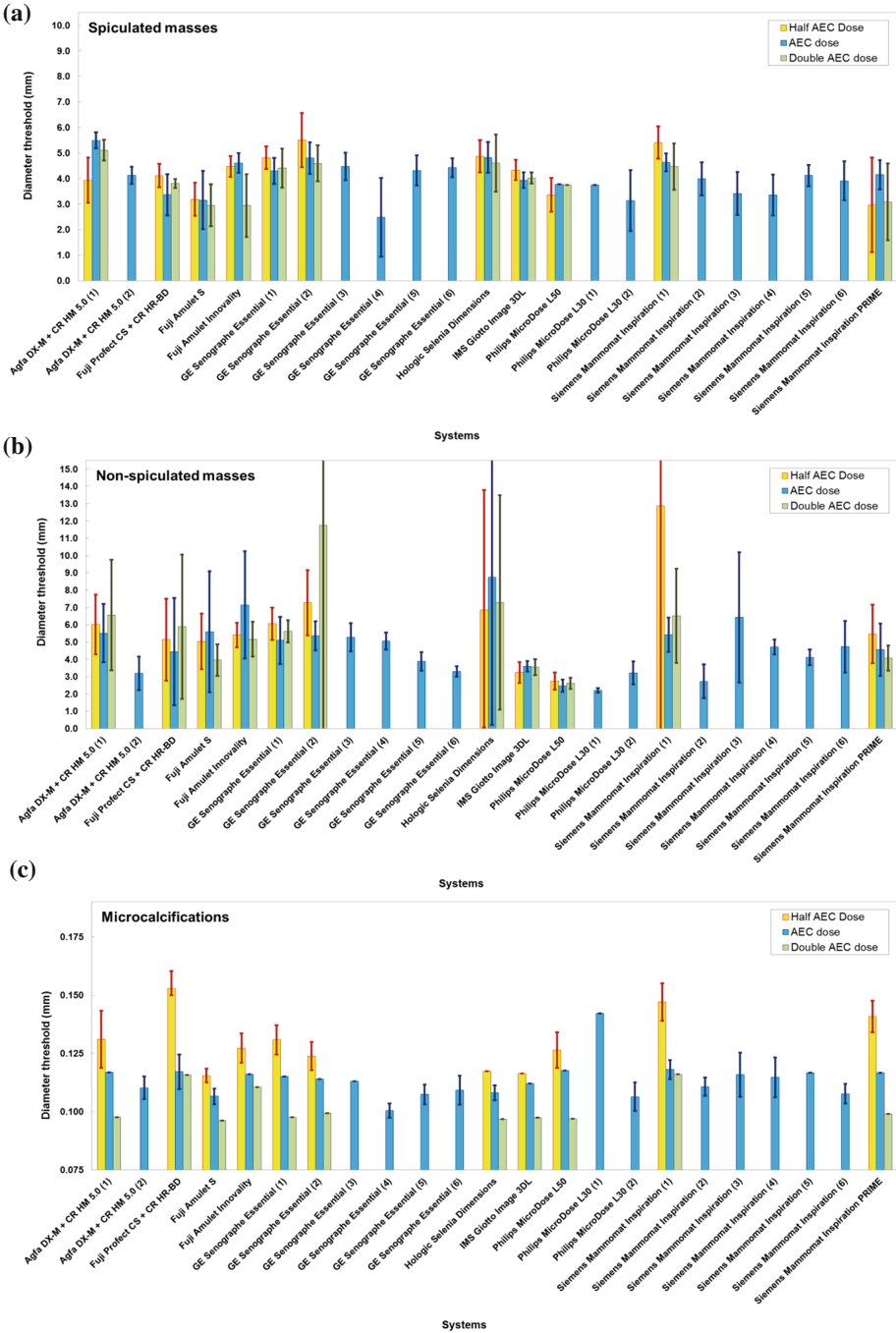


Fig. 2. Diameter thresholds of spiculated masses (a), non-spiculated masses (b) and microcalcifications (c). (Color figure online)

masses (b), and microcalcifications (c) that were obtained from the psychometric curve fits. We could derive such a curve for all systems operating at their clinical working condition. In a next phase, we calculated also the values for the half and double dose conditions.

At AEC dose level, the following system averaged diameter threshold values were obtained: 4.02 mm [range 3.73–4.32 mm] for spiculated masses, 4.64 mm [range 3.98–5.31 mm] for non-spiculated masses and 0.114 mm [range 0.110–0.117 mm] for microcalcifications. According to these values, 3[range 3–4] spiculated masses, 1[range 1–2] non-spiculated mass and 4 groups of microcalcifications were correctly detected for this range of systems.

3.2 Mean Glandular Dose

The MGD at AEC dose level of the 23 systems ranged from 0.75 mGy to 3.32 mGy (Table 3).

Table 3. Mean glandular dose of 23 systems acquired at half AEC dose, AEC dose, and double AEC dose, calculated with the method described by Dance.

No	System	MGD (mGy)		
		Half AEC dose	AEC dose	Double AEC dose
1	Agfa DX-M + CR HM 5.0 (1)	0.96	1.90	3.76
2	Agfa DX-M + CR HM 5.0 (2)		1.68	
3	Fuji Profect CS + CR HR-BD	1.02	2.05	4.10
4	Fuji Amulet S	0.69	1.38	2.76
5	Fuji Amulet Innovality	0.54	1.07	2.08
6	GE Senographe Essential (1)	0.58	1.17	2.30
7	GE Senographe Essential (2)	0.81	1.60	3.16
8	GE Senographe Essential (3)		1.43	
9	GE Senographe Essential (4)		1.39	
10	GE Senographe Essential (5)		1.47	
11	GE Senographe Essential (6)		1.16	
12	Hologic Selenia Dimensions	0.79	1.58	3.16
13	IMS Giotto Image 3DL	1.66	3.32	5.03
14	Philips MicroDose L50	0.41	0.80	1.50
15	Philips MicroDose L30 (1)		0.92	
16	Philips MicroDose L30 (2)		0.75	
17	Siemens Mammomat Inspiration (1)	0.60	1.19	2.38
18	Siemens Mammomat Inspiration (2)		1.08	
19	Siemens Mammomat Inspiration (3)		1.24	
20	Siemens Mammomat Inspiration (4)		1.14	
21	Siemens Mammomat Inspiration (5)		1.14	
22	Siemens Mammomat Inspiration (6)		1.09	
23	Siemens Mammomat Inspiration PRIME	0.54	1.07	2.15

3.3 Dose Sensitivity

Diameter thresholds as a function of MGD are shown for all lesions (spiculated masses, non-spiculated masses, and microcalcifications from 11 systems) in Fig. 3.

The diameter thresholds of spiculated masses and non-spiculated masses are not affected by the dose, although a decreasing trend with dose was observed. On the other hand, diameter thresholds of microcalcifications decreased significantly as dose was increased. Therefore, the change of dose significantly affects the microcalcification detection task performance, but not mass detection. This finding confirms the observations in many other and earlier studies [10–13]. A possible explanation for the difference in the effect of dose on masses versus the effect of dose on microcalcifications must be related with the different effects of anatomical structure noise at the one hand and system (quantum) noise at the other hand. In the study by Burgess *et al.*, it was observed that mammographic backgrounds are dominated by a high magnitude of low spatial frequency components. Lowering dose has a relatively small effect on these low frequency components compared to the higher frequencies in which system noise is known to dominate. Therefore, a variation in dose leads to a smaller effect on the detection of (lower frequency) masses when compared to the detection of microcalcifications that are tiny objects with high frequencies that have to be discriminated from the high (dose affected) frequencies of the small region around the microcalcifications [13].

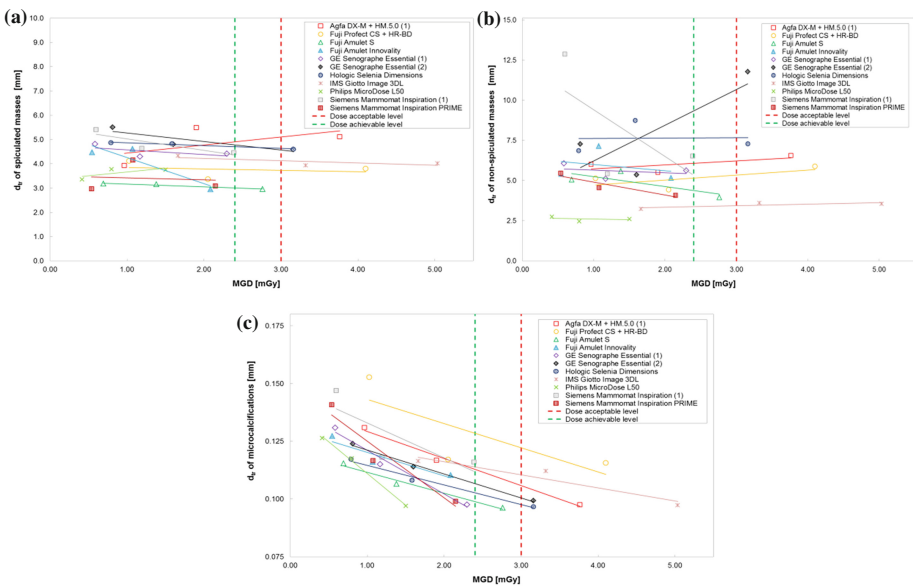


Fig. 3. Correlation between MGD and diameter threshold for spiculated masses (a), non-spiculated masses (b) and microcalcifications (c) from 11 systems tested at three dose levels. (Color figure online)

3.4 Comparison Between the 3D Structured Phantom and the CDMAM Phantom

Correlating the results of the 3D structured phantom and the CDMAM phantom is relevant and interesting for microcalcifications in the 3D structured phantom at the one hand and the smallest gold disks in the CDMAM phantom at the other hand. Figure 4 shows the correlation of the PC values of the group of microcalcifications with average diameter of 119 μm in the 3D structured phantom and the thickness thresholds of the gold disk with diameter 100 μm and present in the CDMAM phantom for all the systems in our study.

Pooled data analysis of all the systems showed that if a digital mammography system would be working at 62.5 % PC for the particular group of microcalcifications with diameter 119 μm , then the CDMAM analysis would show a thickness threshold of 1.50 μm for the 0.1 mm gold disk. This value is very close to the acceptable gold thickness threshold of 1.68 μm in the EUREF guidelines. PC values above 62.5 % for that group of microcalcifications may guarantee compliance with the EUREF criteria.

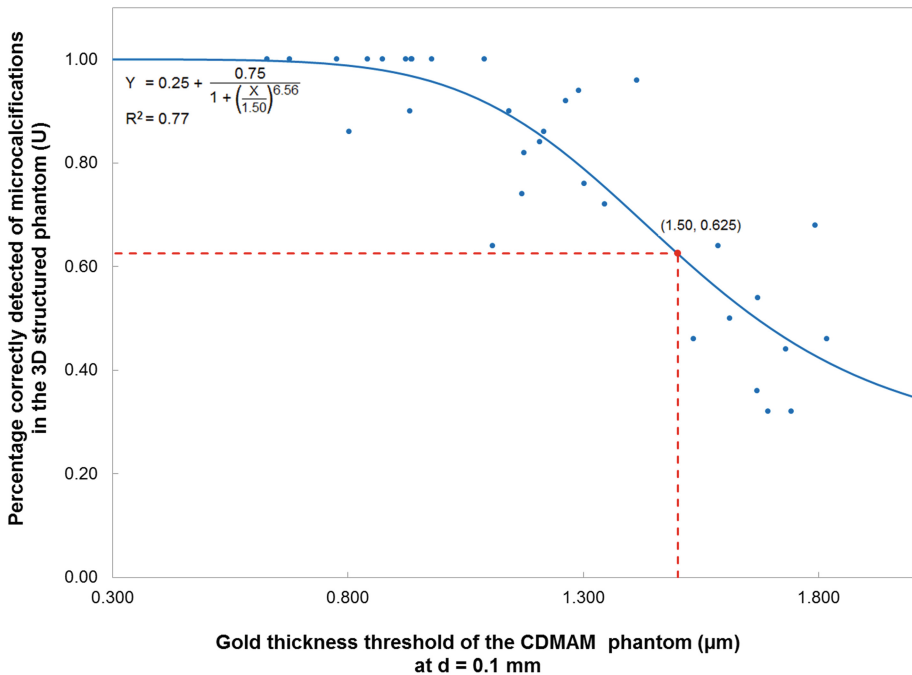


Fig. 4. Correlation between the percentage correctly detected microcalcifications of the group with average size 119 μm of the 3D structured phantom and gold thickness thresholds as obtained from CDMAM analysis of the same systems.

4 Discussions and Conclusions

The performance of each system to display the spiculated masses, non-spiculated masses, and microcalcifications is known to be multifactorial, and depending on the type of image detector, beam quality, dose setting and image processing, as these factors play a role in limiting spatial resolution, contrast, latitude or dynamic range, noise and artifacts. The modulation transfer function has a great impact on the spatial resolution. The X-ray energy spectrum, which is determined by the target material, kVp, and filtration (either inherent in the tube or added), is known to strongly affect the contrast resolution. System performance tests have a role as they allow to summarize all these aspects with a single measure [14]. From the results of present study, it is not clear which of all factors has the largest effect on the performance as tested here but a single performance test phantom will never explain all possible reasons of higher or lower scores. There remains a role for more detailed physics testing.

Detection of microcalcifications is mainly limited by spatial resolution and the effect of quantum and/or detector system noise. The use of fewer quanta (dose reduction) could increase the random noise or quantum mottle (for fixed signal) and therefore decrease SNR and reduce the ability to discern subtle differences in contrast. The visibility of masses is governed more by contrast resolution. The effect of tissue superimposition in 2D projection imaging is thought to merely limit the detection of mass lesions [10, 11, 14]. These characteristics have been confirmed in the dose response curve that was performed with the new phantom.

Present phantom allowed to list threshold diameters for spiculated lesions, non spiculated lesions and microcalcifications, with dose and size having a predictable impact on lesion detectability. These factors, next to the correlation with CDMAM readings, make it a candidate phantom for 2D digital mammography. Obviously, our data base is still limited and we do not have validated explanations for some of the lower or higher scores in our analysis. This could be further explored in the future.

The 3D structured phantom had been successfully applied on DBT. The same phantom being useful for 2D mammography has obvious practical advantages. More importantly, expressing performance in the same way between 2D mammography and DBT triggers comparative studies or bench marking. A next modality to test is breast CT. This must be feasible with similar background material and the same targets, brought together in a different container. Testing contrast enhanced spectral mammography may be possible with iodine including targets. Including breast MRI in this analysis is an ultimate challenge.

Acknowledgements. We would like to acknowledge the 4-AFC readers: Andreas Stratis, Frédéric Bemelmans, Michiel Dehairs, and Xochitl Lopez Rendon.

References

1. Cockmartin, L., Marshall, N.W., Zhang, G., Lemmens, K., Shaheen, E., Van Ongeval, C., Fredenberg, E., Dance, D.R., Salvagnini, E., Michielsen, K., Bosmans, H.: A structured phantom for detection performance comparison between breast tomosynthesis and digital mammography, Ph.D. thesis (2015)

2. Cockmartin, L., Marshall, N.W., Bosmans, H.: Design and evaluation of a phantom with structured background for digital mammography and breast tomosynthesis. In: Maidment, A.D., Bakic, P.R., Gavenonis, S. (eds.) IWDM 2012. LNCS, vol. 7361, pp. 642–649. Springer, Heidelberg (2012)
3. Cockmartin, L., Bosmans, H., Marshall, N.W.: Comparative power law analysis of structured breast phantom and patient images in digital mammography and breast tomosynthesis. *Med. Phys.* **40**(081920), 1–17 (2013)
4. Gang, G.J., Tward, D.J., Lee, J., Siewerdsen, J.H.: Anatomical background and generalized detectability in tomosynthesis and cone-beam CT. *Med. Phys.* **37**, 1948–1965 (2010)
5. Shaheen, E., De Keyzer, F., Bosmans, H., Dance, D.R., Young, K.C., van Ongeval, C.: The simulation of 3D mass models in 2D digital mammography and breast tomosynthesis. *Med. Phys.* **41**(081913), 1–17 (2014)
6. Zhang, G., Cockmartin, L., Bosmans, H.: A four-alternative forced choice (4AFC) software for observer performance evaluation in radiology. Poster presentation at the SPIE Conference, San Diego (2016)
7. Figl, M., Hoffmann, R., Kaar, M., Semturs, F., Brasik, N., Birkfellner, W., Homolka, P., Hummel, J.: Factors for conversion between human and automatic read-outs of CDMAM images. *Med. Phys.* **38**, 5090–5093 (2011)
8. European Commission: European guidelines for quality assurance in breast cancer screening and diagnosis, 4th ed. Europe against Cancer, Office for Official Publications of the European Communities, Luxembourg (2006)
9. van Engen, R., Bosmans, H., Heid, P., Lazzari, B., Schopphoven, S., Thijssen, M., Young, K., Dance, D., Marshall, N.W.: Supplement to the European Guidelines, 4th edn. Health and Consumer Protection - European Communities (2011)
10. Warren, L.M., Mackenzie, A., Cooke, J., Given-Wilson, R.M., Wallis, M.G., Chakraborty, D.P., Dance, D.R., Bosmans, H., Young, K.C.: Effect of image quality on calcification detection in digital mammography. *Med. Phys.* **39**, 3202–3213 (2012)
11. Timberg, P., B ath, M., Andersson, I., Mattsson, S., Tingberg, A.: Visibility of microcalcification clusters and masses in breast tomosynthesis image volumes and digital mammography: a 4AFC human observer study. *Med. Phys.* **39**, 2431–2437 (2012)
12. Ruschin, M., Timberg, P., B ath, M., Hemdal, B., Svahn, T., Saunders, R., Samei, E., Andersson, I., Mattsson, S., Chakraborty, D.P., Tingberg, A.: Dose dependence of mass and microcalcification detection in digital mammography: free response human observer studies. *Med. Phys.* **34**, 400–407 (2007)
13. Burgess, A.E., Jacobson, F.L., Judy, P.F.: Human observer detection experiments with mammograms and power-law noise. *Med. Phys.* **28**, 419–437 (2001)
14. Kanal, K.M., Krupinski, E., Berns, E.A., Geiser, W.R., Karellas, A., Mainiero, M.B., Martin, M.C., Patel, S.B., Rubin, D.L., Shepard, J.D., Siegel, E.L., Wolfman, J.A., Mian, T.A., Mahoney, M.C., Wyatt, M.: ACR-AAPM-SIIM practice guideline for determinants of image quality in digital mammography. *J. Digit. Imaging* **26**, 10–25 (2013)