Observer study to evaluate the simulation of mammographic calcification clusters

Maria A. Z. Sousa[*a], Karem D. Marcomini[a], Predrag R. Bakic[b], Andrew D.A. Maidment[b], Homero Schiabel[a]

[a]Univ. of São Paulo, Dept. Electrical Engineering, São Carlos, SP, Brazil;

[b]Univ. of Pennsylvania, Dept. of Radiology, Philadelphia, PA, USA.

ABSTRACT

Numerous breast phantoms have been developed to be as realistic as possible to ensure the accuracy of image quality analysis, covering a greater range of applications. In this study, we simulated three different densities of the breast parenchyma using paraffin gel, acrylic plates and PVC films. Hydroxyapatite was used to simulate calcification clusters. From the images acquired with a GE Senographe DR 2000D mammography system, we selected 68 regions of interest (ROIs) with and 68 without a simulated calcification cluster. To validate the phantom simulation, we selected 136 ROIs from the University of South Florida's Digital Database for Screening Mammography (DDSM). Seven trained observers performed two observer experiments by using a high-resolution monitor Barco mod. E-3620. In the first experiment, the observers had to distinguish between real or phantom ROIs (with and without calcification). In the second one, the observers had to indicate the ROI with calcifications between a pair of ROIs. Results from our study show that the hydroxyapatite calcifications had poor contrast in the simulated breast parenchyma, thus observers had more difficulty in identifying the presence of calcification clusters in phantom images. Preliminary analysis of the power spectrum was conducted to investigate the radiographic density and the contrast thresholds for calcification detection. The values obtained for the power spectrum exponent (β) were comparable with those found in the literature.

Keywords: mammography, calcifications, phantom images, observer performance.

1. INTRODUCTION

Mammography is the main imaging strategy for the early detection of breast cancer. Analysis of mammograms involves the detection of lesions (e.g. masses, calcification clusters, or architectural distortion), which require training and experience [1, 2]. Computer-aided detection systems (CAD) are used to assist radiologists with the detection of these lesions [3-5]. The quality control of imaging systems and evaluations of CAD schemes commonly use experiments based on the lesion detection by human observers. However, such validation studies depend on the availability of a large number of images with a particular lesion class with the presence of the lesions confirmed independently, which is not always accessible. In order to address this limitation, physical or computational phantoms are used as alternatives.

The International Commission on Radiation Units and Measurements (ICRU) [6] defines a phantom as any structure that contains one or more human tissues substituted by equivalent tissue with respect to their chemical, physical and attenuation properties. An equivalent tissue is any material that can simulate human tissues and its interaction with radiation.

Polymers such as acrylic, resin, polyurethane, and polyethylene are commonly used to simulate fibroglandular tissue. Gelatinous materials such as paraffin and wax are used to simulate adipose tissue. These materials have densities and radiographic characteristics that are very similar to those of breast tissues.

One of the most commonly used physical mammographic phantoms is the American College of Radiology (ACR) accreditation phantom [7], and it allows validation based on the detection of small structures. More quantitative assessments can be performed with a contrast-detail phantom, such as the Contrast-Detail Mammography Phantom

[*]angelicazucareli@usp.br; phone +55 16 3373 8742; fax #55 16 3373 9372

Medical Imaging 2016: Image Perception, Observer Performance, and Technology Assessment,
edited by Craig K. Abbey, Matthew A. Kupinski, Proc. of SPIE Vol. 9787, 97871D
 · CCC code: 1605-7422/16/$18 · doi: 10.1117/12.2217257

(CDMAM) phantom [8]. The need for developing specific phantoms for different imaging modalities led to more sophisticated designs, such as the anthropomorphic phantom “Rachel” (Gammex, Middleton, WI) [9], and the ones recently developed for digital breast tomosynthesis applications, as the anthropomorphic breast phantom developed at the University of Pennsylvania [10] and the phantom developed at Duke University (Doublet 1.1) [11]. In this study, we insert particles of hydroxyapatite on a physical phantom [12] simulating the breast parenchyma to create a realistic background for the observer’s analysis.

Our phantom was designed for assessing a CAD scheme we previously developed [13]. We have conducted objective (power spectra analysis) and subjective (observer studies) tests with the images produced in order to investigate the radiographic density and the contrast thresholds for the calcification detection.

The power spectrum is frequently used as a descriptor of normal mammographic tissue. Burgess [14] describes it as a power law:

$$PS(f) = \frac{C}{f^{\beta}} \tag{1}$$

where f is the frequency, C is the power spectrum magnitude, and β is the power spectrum exponent. β values range from 1.5 to 3.5. We calculated the power spectrum to evaluate the effect of the fluctuations in the frequencies on the background image produced by the phantom and to determine its relation with the visibility of lesions. For subjective analysis, two observer tasks were performed taking into account the visibility of the calcifications in clinical images from DDSM database and in the images obtained from the phantom exposure to a digital mammography unit. In the first, the observers distinguished between real and simulated ROIs, while in the second they indicated the ROI with calcifications between a pair of ROIs.

2. MATERIALS AND METHODS

2.1 Image Database

The phantom used in our study was created by combining two layers of paraffin gel – saturated hydrocarbon gel, two acrylic plates cut to size of 10 mm X 150 mm X 150 mm, and three layers of the PVC film (Tripack Filmes®). The paraffin was melted in an oven at 90° C to be molded into the same shape as the acrylic plates. The PVC film layers ware prepared in a previous work [12] manually kneading small pieces of the PVC film and putting them together to form a large "tissue" with thickness of 3.28 mm, 2.95 mm and 2.72 mm, simulating different densities of the breast parenchyma. Only one PVC film layer was used in each exposition (Fig. 1).

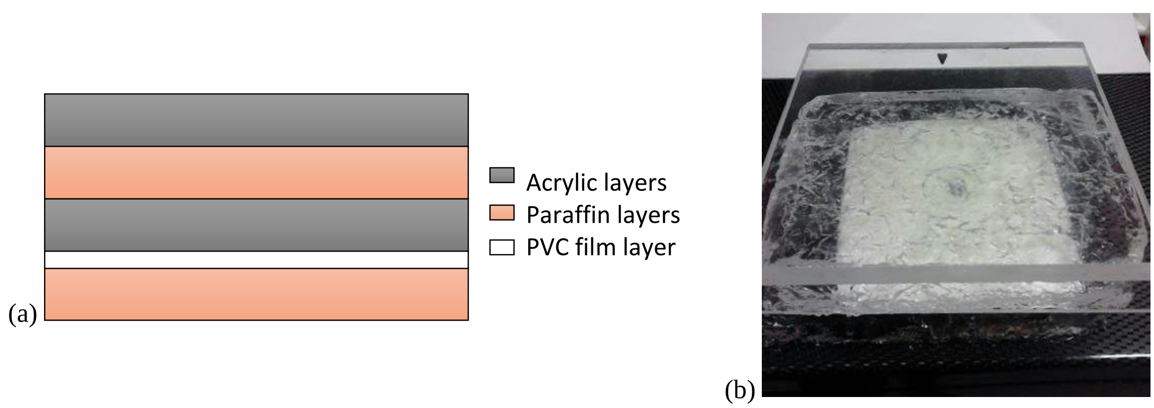


Figure 1. (a) Profile of the materials used in the confection of the phantom. (b) Final configuration of the phantom.

Six calcification clusters were simulated using particles of hydroxyapatite with grain sizes ranging from 0.25 to 0.50 mm. Each cluster has between 7 and 11 calcifications distributed in an area smaller than 1.5 cm^2. As shown in Figure 2, the spatial arrangement of simulated cluster varied from more compact (C1-C2) to irregularly shaped (C3-C5), to elongated, linear cluster shape (C6), covering clinically observed variations [15].

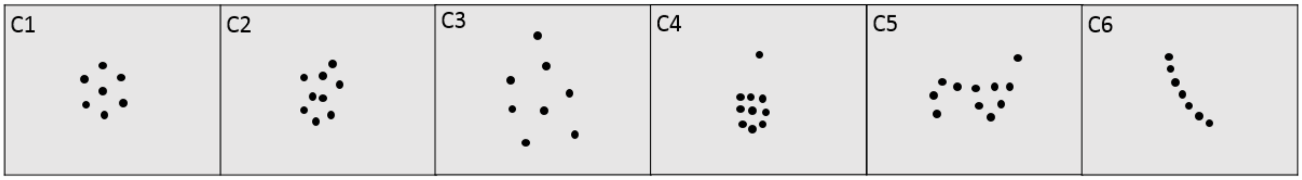


Figure 2. Spatial distributions of individual hydroxyapatite calcifications in the simulated clusters, varying from more compact (C1) to more elongated (C6), covering clinically seen variations.

The clusters were inserted manually above the PVC layer to be in a position as close as possible to the image plate avoiding the calcifications magnification (Fig. 1a). We used quadrants to separate the phantom in four regions (Fig. 3a) and distribute the clusters resulting in one different configuration for each exposure (Fig. 3b). Each PVC layer was exposed seven times to have each cluster exposed once in each quadrant.

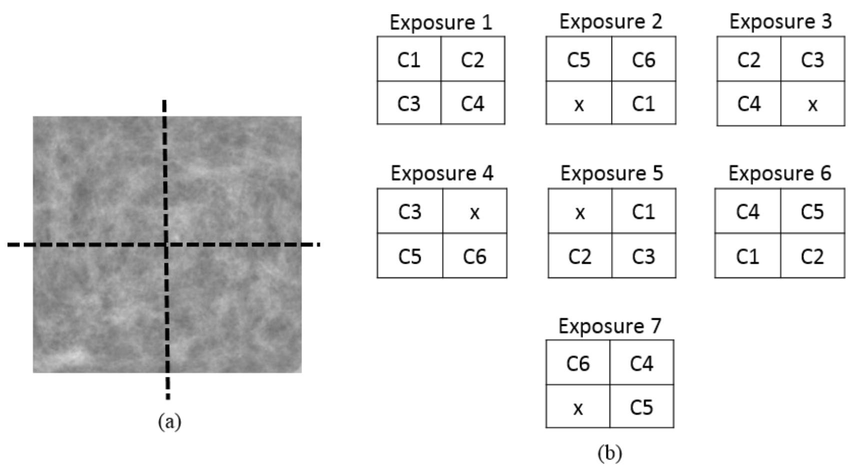


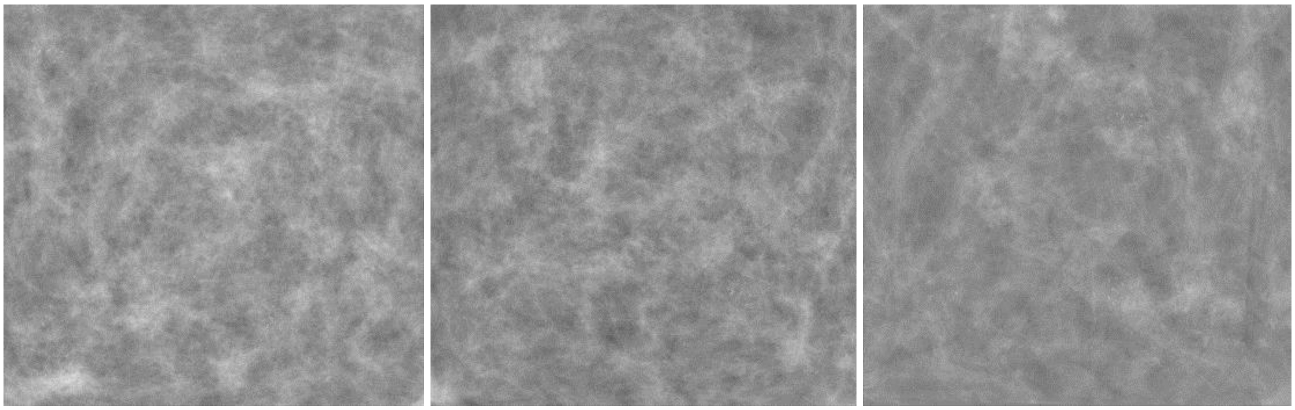
Figure 3. (a) Quadrants to separate the PVC layer in four regions. (b) Distribution of the clusters in the quadrants for each exposure. ‘x’ represents quadrant without a insertion of a cluster.

Images of the phantom were obtained with a digital mammographic equipment GE Senographe 2000D available at the Hospital of São Paulo, in São Paulo, Brazil. For each thickness of PVC layer, we used the automatic exposure control (AEC) in the first exposure and set the same imaging parameters for the following images. Examples of these images are shown on Figure 4.

From the resulting phantom images, a total of 136 regions of interest (ROI) were selected (68 ROIs containing calcification clusters and 68 ROIs without calcification clusters). The size of each ROI is 300 by 300 pixels (30x30 mm). A second group of 136 ROIs (with a half containing clinical calcifications) was selected from real digitized mammograms provided by the University of South Florida’s Digital Database for Screening Mammography (DDSM) [16]. The size of the ROIs varied between 12.5 and 15 mm^2, according to the resolution of the scanner used in the digitalization.

2.2 Power Spectra Analysis

The power spectra for the phantom images were obtained using image regions without calcifications. Mammograms from 9 patient data acquired at University of Pennsylvania with a DR GE Senographe 2000D equipment were used for comparison. 70 ROIs sized by 300 by 300 pixels were obtained from the raw data for each group of images (phantom and mammographic data).



(a) (b) (c)

Figure 4. Examples of images obtained for (a) 3.28 mm, (b) 2.95 mm and (c) 2.72 mm PVC layers, respectively, with different distributions of the calcification clusters.

The spectral analysis method [14] consisted of the calculation of the modulus-squared 2D discrete Fourier transform of each ROI (periodogram). The periodograms were averaged to give an estimate of the 2D spectrum for the collection of ROIs from phantom or clinical images.

One-dimensional line graphs were obtained from the 2D spectrum by plotting it in horizontal and vertical slices (x and y directions, respectively). The power spectrum exponent (β) was calculated as the slope of the linear portion of the log-log plot [17] (Fig. 5).

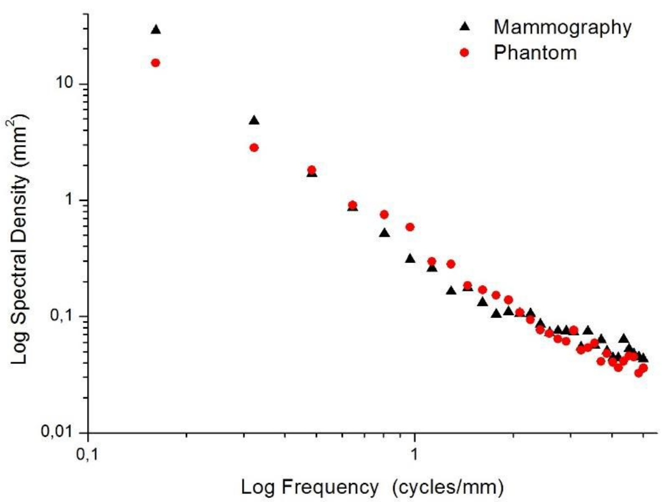


Figure 5. Power spectra for real mammogram and phantom images.

2.3 Observer Studies

Seven trained observers (graduate students and faculty) performed two observer experiments in a darkened room using a high-resolution quality monitor Barco mod. E-3620 (Barco Inc., Duluth, GA 30097). In both experiments, the observers maintained the observing distance between 0.5 and 1.0 m. No image magnification was allowed.

2.3.1 Test #1

In the first observer study, we tested the realism of the phantom images by displaying a random set of 17 ROIs by time to simulate the clinical situation of finding calcifications in the whole image. The observer were asked to identify phantom and real images, and whether they contained a cluster (Fig. 6). All the observers read the same sequence of image to prevent bias.

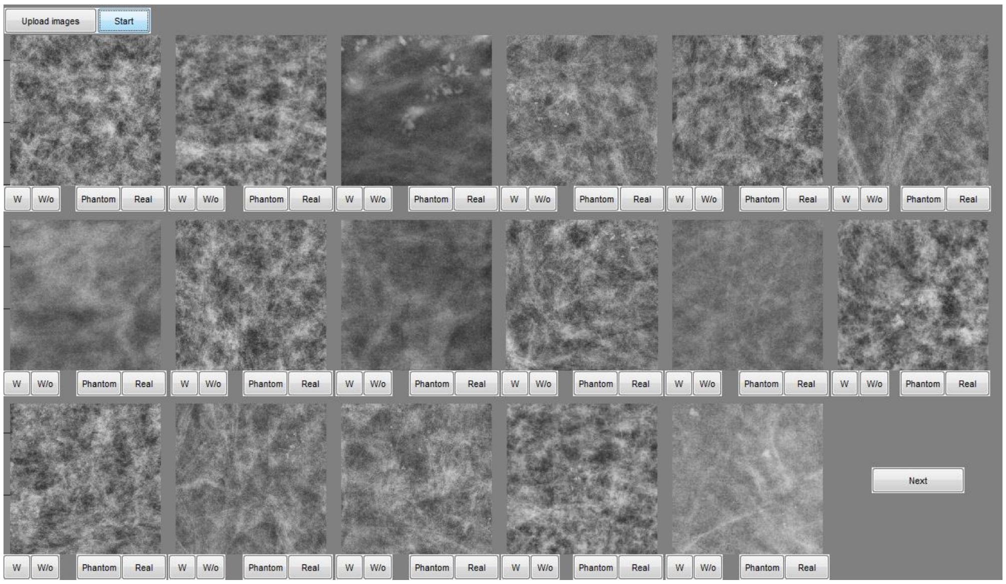


Figure 6. Observer experiment including real and phantom images, with and without calcification cluster.

2.3.2 Test #2

In the second experiment, pairs of ROIs were randomly selected and displayed. For each pair of ROIs, both images were of the same type (real or phantom), and only one of the images contained a calcification cluster. The image containing the cluster was randomized to appear on the left or right side of the display. Observers were tasked with identifying the image containing the cluster (Fig. 7). All the observers read the same sequence of 68 image pairs, to prevent bias.

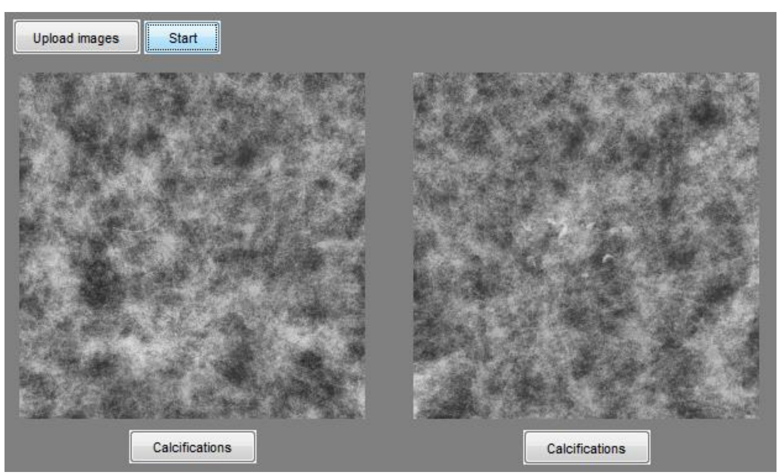


Figure 7. Observer experiment with one of the ROIs having a cluster inserted.

3. RESULTS

3.1 Power Spectra

The analysis of the power spectrum provides insight on how the high frequencies generated by the material of the phantom influence the detection of calcification. The exponent values (*β*) for the phantom images range from 1.79 to 1.84 (Fig. 8), and they are comparable with those obtained for real mammograms images (between 1.68 and 2.5) and consistent with the literature for normal mammograms [14].

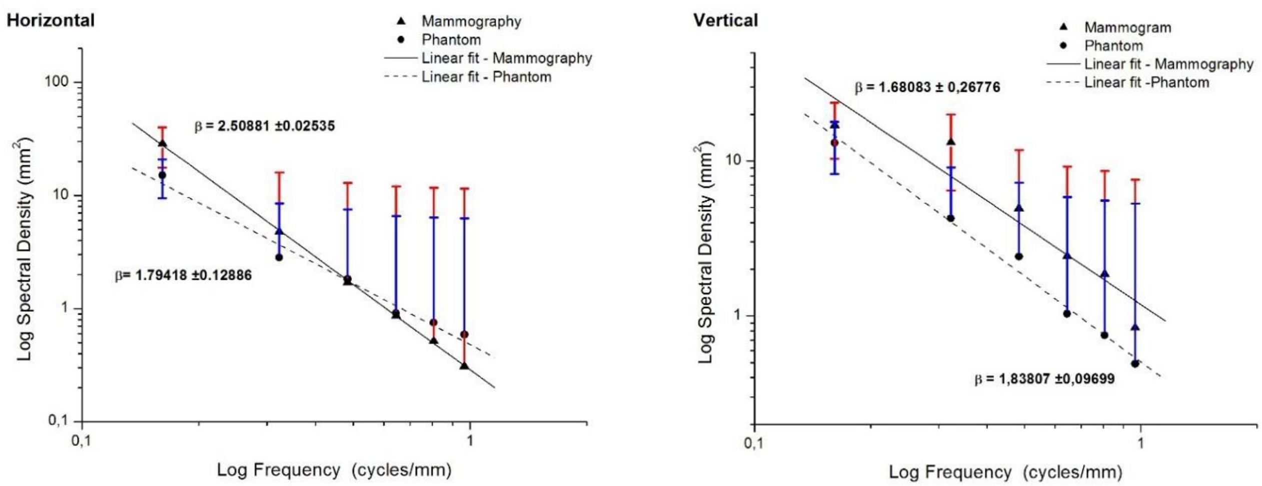


Figure 8. Horizontal and vertical 1D power spectra for real mammogram and phantom images and the respective estimated β values.

3.2 Observer Study

Images obtained from the phantom have a more contrasted background with high frequencies visually more evident than the low frequencies (Fig. 9). The hydroxyapatite calcifications are visually smaller than those that appears in real mammograms and with a poorest contrast, as we can see comparing Fig. 9 and 10.

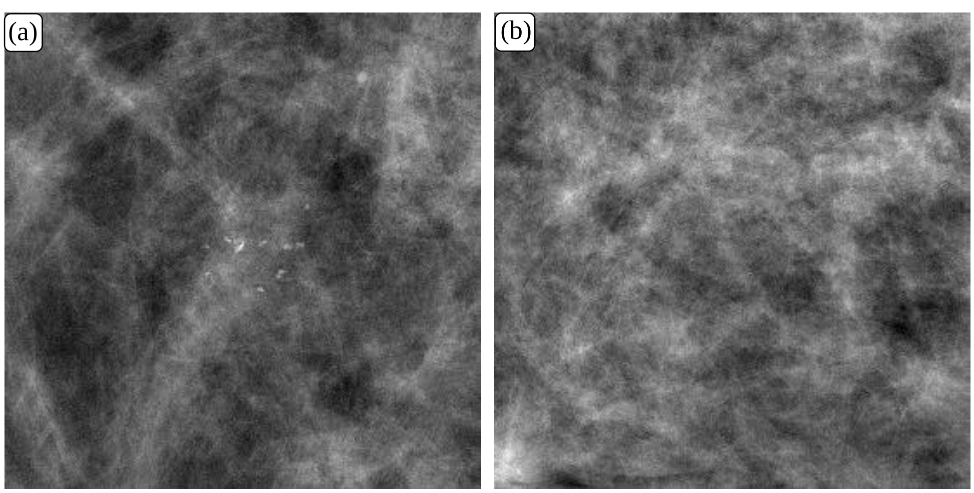


Figure 9. Examples of ROIs of the phantom images with cluster (a) and without cluster (b).

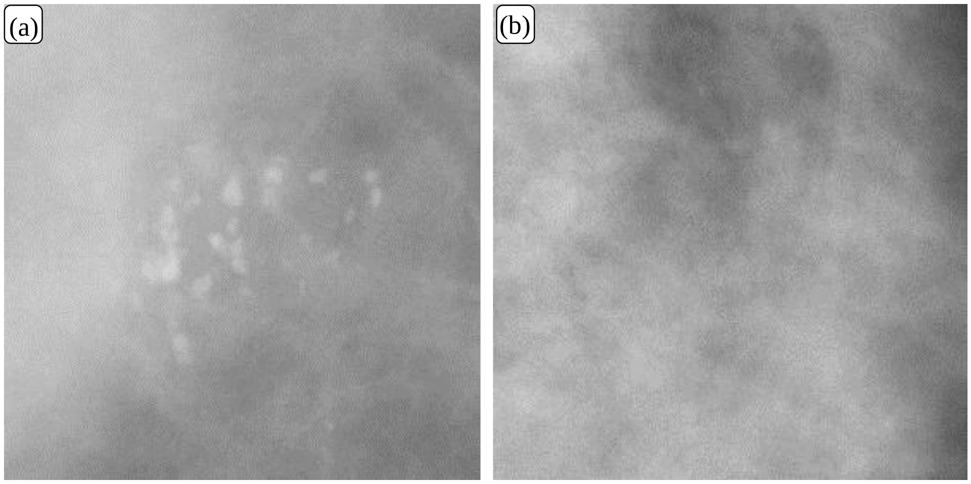


Figure 10. Examples of ROIs obtained from DDSM mammograms with cluster (a) and without cluster (b).

3.2.1 Test #1

In the first experiment, observers could identify if the ROIs were from phantom (P) or real (R) images without much difficulty resulting in an accuracy higher than 86%, as we can see in Table 1.

Table 1. Results for the seven observers in the second experiment.

Task														
Cluster detection	Observer 1		Observer 2		Observer 3		Observer 4		Observer 5		Observer 6		Observer 7	
	Real	Phantom	Real	Phantom	Real	Phantom	Real	Phantom	Real	Phantom	Real	Phantom	Real	Phantom
TN	65	65	64	64	62	62	67	65	65	66	64	59	67	68
TP	68	49	65	58	67	63	68	66	65	55	67	56	57	39
FP	0	19	3	10	1	5	0	2	2	13	1	12	11	29
FN	3	3	4	4	6	6	1	3	3	2	4	9	1	0
Accuracy (%)	97.8	83.8	94.7	89.7	94.7	91.9	99.3	96.33	95.6	89.0	96.3	84.6	91.2	78.7
ROIs identification	Observer 1		Observer 2		Observer 3		Observer 4		Observer 5		Observer 6		Observer 7	
Accuracy (%)	98.9		87.1		100.0		98.5		97.0		92.3		86.8	

The average values for accuracy rates were not statistically significantly different, achieving values of 87.72 % (± 5.83) for phantom images and 95.66 % (± 2.58) for the real mammograms.

Results considering both groups of images were always higher than 86% with a small difference among the observer performances (6% for real images and 3% for the phantom ones).

3.2.2 Test #2

The results for the second experiment comparing the performance of the seven observers are shown in Table 2. We have calculated the true positive (TP), true negative (TN), false positive (FP), false negative (FN) fractions and the accuracy of the observer identified ROIs, separately for phantom and real images.

Table 2. Results for the seven observers in the first experiment.

	Observer 1		Observer 2		Observer 3		Observer 4		Observer 5		Observer 6		Observer 7	
	Real	**Phantom**	**Real**	**Phantom**	**Real**	**Phantom**	**Real**	**Phantom**	**Real**	**Phantom**	**Real**	**Phantom**	**Real**	**Phantom**
TN	34	28	34	34	33	34	34	33	34	33	33	32	34	29
TP	33	30	34	32	34	32	34	33	33	33	34	29	34	28
FP	0	6	0	0	1	0	0	1	0	1	1	2	0	5
FN	1	4	0	2	0	2	0	1	1	1	0	5	0	6
Accuracy (%)	98.53	85.29	100.00	97.06	98.53	97.06	100.00	97.06	98.53	97.06	98.53	89.71	100.00	83.82

The average of accuracy fraction for phantom images was 92.44% (± 6.03) and 99.16% (± 0.79) for the real images. These values are not statistically significantly different.

4. CONCLUSIONS

Analysis of the power spectrum resulted in values for exponent (β) comparable with those found in the literature. However, we understand that it is still necessary to study more about the exponent in terms of the contrast of lesions, as the calcifications in this approach, having a more accurate judgment about the phantom simulations proposed in this work.

Our observer study showed an agreement between the accuracy in determining which images contained a cluster. This is an encouraging result, which together with comparable power spectra between clinical and simulated images, suggesting the potential for our phantom to be used in breast imaging validation. By emphasizing the simplicity and low-cost design of the phantom, we anticipate this phantom can be a widely affordable tool in developing countries. Nevertheless, we are aware of the need for further prepare testing to fully characterize our phantoms design, to optimize the selection of materials and production steps, and to calibrate the phantom for potential clinical use.

Future testing will focus on the quantification of the relative contrast of individual calcifications versus background, the effect of phantom-structured noise (caused by the non-uniform distributions of the paraffin and PVC film or by the presence of air within the phantom), the phantom reproducibility and durability. The effect of cluster spatial arrangements (i.e. compact versus elongated) on the detectability of lesions may also be of interest when comparing phantom and real images.

ACKNOWLEDGMENTS

This work in progress has been sponsored by the National Counsel of Technological and Scientific Development (CNPq 201633/2015-9) and the Coordination for the Improvement of Higher Education Personnel (CAPES). We also acknowledge to Scott Cupp at the University of Pennsylvania for providing the power spectra tool used in this work.

REFERENCES

[1] Johns, P.C. and Yaffe, "M.J. X-ray characterization of normal and neoplastic breast tissues," Phys. Med. Biol. 32(6), 675-95 (1987).

[2] Karellas, A., Giger, M.L, [Advances in Breast Imaging: Physics, Technology, and Clinical Applications: Categorical Course in Diagnostic Radiology Physics], Oak Brook, IL: Radiological Society of North America (RSNA), (2004).

[3] Malich, A., Fischer, D.R., Böttcher, J, "CAD for mammography: the technique, results, current role and further developments", European Radiology. 16, 1449–1460 (2006).

[4] Yoshikawa, R., Teramoto, A., Matsubara, T., Fujita, H, "Automated detection scheme of architectural distortion in mammograms using adaptive Gabor filter" Proc. SPIE, 8670 (2013).

[5] James, J. J. et al, “Mammographic Features of Breast Cancers at Single Reading with Computer-aided Detection and at Double Reading in a Large Multicenter Prospective Trial of Computer-aided Detection: CADET II”, Radiology. 256(2), (2010).
[6] ICRU, Report 44. “Tissue Substitutes in Radiation Dosimetry and Measurements-International Comition on Radiation and Measurements. International Commission on Radiation Units and Measurements”, Bethesda, Md. (1989).
[7] Martin, M.C., Smathers, J. B, “Requirements of the American College of Radiology Mammography Accreditation Program”, Proc. of an American College of Medical Physics Symposium, Medical Physics Publishing, Madison,WI (1992).
[8] Bijker, K. R., Thijssen, M. A. O., Arnoldussen, Th.J.M, “Manual of CDMAM-phantom type 3.4”, University Medical Centre Nijmegen, St. Radboud Department of Diagnostic Radiology, Section of Physics and Computer Science, (2000).
[9] Caldwell, C.B., Yaffe, M.J, “Development of an anthropomorphic breast phantom”, Medical Physics. 17(2) (1990).
[10] Carton, A.K, Bakic, P., Ullberg, C., Maidment, A.D.A, “Development of a 3D high-resolution physical anthropomorphic breast phantom” Proc. SPIE, 7622 (2010).
[11] Kiarashi, N., Sturgeon, G. M., Nolte, L. W., Lo, J. Y., Dobbins III, J. T., Segars, W. P., Samei, E, “Development of matched virtual and physical breast phantoms based on patient data” Proc. SPIE, 8668 (2013).
[12] Sousa, M.A.Z., Siqueira, P.N., Schiabel, H, “Evaluation of a simulation procedure designed to recognize shape and contour of suspicious masses in mammography”. Proc. SPIE, 9416 (2015).
[13] Matheus, B.R.N. and Schiabel, H, “A CADx scheme in mammography: considerations of a novel approach” Proc. 7th International Conference on Advanced Engineering Computing and Applications in Sciences, Porto. p. 15-18 (2013).
[14] Burgess, A. E, “Mammographic structure: Data preparation and spatial statistics analysis” Proc. SPIE, 3661 (1999).
[15] Kopans D. [Breast Imaging] 3rd ed. Philadelphia, PA: LippincottWilliams &Wilkins (2007).
[16] Heath, M., Bowyer, K., Kopans, D., et al, “The Digital Database for Screening Mammography” Proc. 5th International Workshop on Digital Mammography. Madison, Wis: Medical Physics Publishing (2000).
[17] Bakic, P.R. et al, “An Anthropomorphic Software Breast Phantom for Tomosynthesis Simulation: Power Spectrum Analysis of Phantom Projections”, J. Martí et al. (Eds.): IWDM 2010, LNCS 6136, pp. 452–458 (2010).