

# Evaluation of Various Mammography Phantoms for Image Quality Assessment in Digital Breast Tomosynthesis

Claudia C. Brunner<sup>1</sup>, Raymond J. Acciavatti<sup>2</sup>, Predrag R. Bakic<sup>2</sup>,  
Andrew D.A. Maidment<sup>2</sup>, Mark B. Williams<sup>3</sup>, Richard Kaczmarek<sup>1</sup>,  
and Kish Chakrabarti<sup>1,\*</sup>

<sup>1</sup> Center for Devices and Radiological Health, U.S. Food and Drug Administration,  
Silver Spring, MD 20910

<sup>2</sup> Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104

<sup>3</sup> Department of Radiology and Medical Imaging, University of Virginia,  
Charlottesville, VA 22908

**Abstract.** We investigated the appropriateness of four different mammography phantoms for image quality evaluation in Digital Breast Tomosynthesis (DBT). We tested the CIRS BR3D phantom, the ACR Prototype FFDM Accreditation Phantom, the Penn anthropomorphic breast phantom and the Quart mam/digi EPQC phantom. This work discusses the advantages and shortcomings of each phantom and concludes that none of them, in their current form, can be considered to be adequate as an image quality evaluation phantom for DBT.

**Keywords:** Digital Breast Tomosynthesis, Quality evaluation, Phantoms.

## 1 Purpose

Digital Breast Tomosynthesis (DBT), recently approved by the FDA for screening and diagnosis of breast abnormalities, improves upon mammography by providing 3-dimensional resolution that allows depth discrimination and overcomes the problem of signal degradation by overlying anatomy. It is regulated under the Mammography Quality Standards Act (MQSA), which requires quantitative image quality evaluation with a human observer as well as objective image quality evaluation. For the quantitative evaluation, human observers have to score the visibility of specific objects in the phantom. The objective evaluation is software-based and provides information about image quality metrics such as resolution, noise and contrast-to-noise ratio (CNR). However, since image quality evaluation in DBT is currently based on quality

---

\* Contact: Center for Devices and Radiological Health, U.S. Food and Drug Administration,  
10903 New Hampshire Avenue, Silver Spring, MD 20993.

Email: [kish.chakrabarti@fda.hhs.gov](mailto:kish.chakrabarti@fda.hhs.gov) or [claudia.brunner@fda.hhs.gov](mailto:claudia.brunner@fda.hhs.gov)

“The mention of commercial products herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services.”

evaluation in projection mammography, specific properties unique to DBT, such as the slice-sensitivity profile, may not be sufficiently captured. To our knowledge no study exists that would present a phantom specifically designed for DBT which can be used to perform complete quality control and acceptance tests. So far studies on quality control phantoms for DBT have only focused on single parameters of image quality such as in-plane resolution and slice thickness [1, 2].

In light of the necessity to characterize clinical image quality in DBT, the purpose of this study was to investigate the appropriateness as well as limitations of four currently available phantoms for image quality evaluation in DBT.

## 2 Material and Methods

For all image acquisitions, we used a Selenia Dimensions Digital Breast Tomosynthesis device (Hologic, Inc., Bedford, MA, USA) located at the Hospital of the University of Pennsylvania (Philadelphia, PA). All phantoms were scanned applying the “AutoFilter” exposure control mode of the system that automatically chooses filter, tube voltage and tube current. For each scan, 15 images were acquired at  $1.07^\circ$  intervals over an angle range of  $\pm 7.5^\circ$ . The device reconstructs images in planes parallel to the breast support in 1 mm increments through the thickness of the phantom. We assessed the reconstructed images in the same way as DBT images are viewed under clinical conditions (i.e. slice-wise evaluation). If the phantom allowed a subjective image evaluation, it was always performed by 4 human observers. The four phantoms tested in this study were:

The **CIRS Model 020 BR3D Mammography phantom** (CIRS, Norfolk, VA, USA) consists of a set of 6 slabs made of two tissue equivalent materials mimicking 100% adipose and glandular tissues “swirled” together in an approximate 50/50 ratio by weight. One of the slabs contains an assortment of speck groups, fibers and masses; its diameters are given in Table 1. The CIRS phantom has been used in former studies for example to investigate the performance of DBT [3] or the potential of an implemented scatter correction in the image reconstruction algorithm [4].

**Table 1.** Diameter of the objects embedded in the target slab of the CIRS phantom

	<b>Fibers</b>	<b>Specks</b>	<b>Masses</b>
<b>1</b>	0.60 mm	0.400 mm	6.3 mm
<b>2</b>	0.41 mm	0.290 mm	4.7 mm
<b>3</b>	0.38 mm	0.230 mm	3.9 mm
<b>4</b>	0.28 mm	0.196 mm	3.1 mm
<b>5</b>	0.23 mm	0.165 mm	2.3 mm
<b>6</b>	0.18 mm	0.130 mm	1.8 mm
<b>7</b>	0.15 mm		

The **ACR Prototype FFDM Accreditation Phantom** (CIRS, Norfolk, VA, USA) is based on the well-known Mammographic Accreditation Phantom (CIRS Model 015) but with a phantom size in the range of the detector size and a finer gradation of the test objects. It contains 6 fibers, 6 speck groups and 6 masses of diameters given in Table 2. The objects are embedded in a homogeneous wax insert positioned within

a PMMA block. Further, the phantom contains a cavity to calculate the CNR and large homogeneous regions to analyze the noise properties of the image.

**Table 2.** Diameter of the objects embedded in the ACR phantom

	<b>Fibers</b>	<b>Specks</b>	<b>Masses</b>
<b>1</b>	0.89 mm	0.33 mm	1.00 mm
<b>2</b>	0.75 mm	0.28 mm	0.75 mm
<b>3</b>	0.61 mm	0.23 mm	0.50 mm
<b>4</b>	0.54 mm	0.20 mm	0.38 mm
<b>5</b>	0.40 mm	0.17 mm	0.25 mm
<b>6</b>	0.30 mm	0.14 mm	0.20 mm

The **Penn anthropomorphic breast phantom** was developed at the University of Pennsylvania specifically for 3-dimensional breast x-ray imaging [5, 6]. It was designed based upon the Penn software anthropomorphic breast phantom [7]. The phantom consists of several slabs of tissue-equivalent adipose and glandular material simulating a dense fibroglandular pattern resulting in images that are qualitatively similar to clinical images with the grayscale range of adipose and fibroglandular elements approximating the pattern seen in a heterogeneously dense breast. An additional, interchangeable slab contains iodinated lesions with 5 different diameters and two different iodine concentrations. The additional slab was designed for the use with contrast enhanced digital breast tomosynthesis [6].

The **Quart mam/digi EPQC phantom** (Quart GmbH, Zorneding, Germany) is a relatively new phantom developed for mammography as well as for breast tomosynthesis [8, 9]. It consists of a PMMA body containing a wedge of 12 steps on the chest wall side to simulate different densities of breast tissue material. A titanium strip, which is equivalent to 100  $\mu\text{m}$  bone material, divides each step and therefore allows the calculation of the CNR for various thicknesses. The MTF can be calculated using the edges of a brass and a lead square and a slot for a suited dosimeter detector enables performing dose measurements according to the EPQC guidelines [10]. Additionally, the phantom features so-called Landolt (broken) rings that are contained within two different layers separated by 20 mm. These rings have a gap at either of 4 different directions (top, bottom, left and right) and are sorted in groups of 6 rings of decreasing size on each step. By focusing on the planes that contain the Landolt rings, it is possible to determine the distance between planes as well as to subjectively quantify object visibility at two different heights. In order to do that, a human observer has to estimate the position of the gaps for the ring groups corresponding to the 7 thickest steps using appropriate zoom and contrast window. The number of correctly perceived gaps gives one of the imaging performance parameters. For mammography, 20 gaps have to be called correctly for the image to pass the test, but for DBT an official threshold has not yet been established.

Further, each of the 12 steps contains a low-contrast number indicating the PMMA thickness corresponding to the step. The thicker the step on which the numbers can be

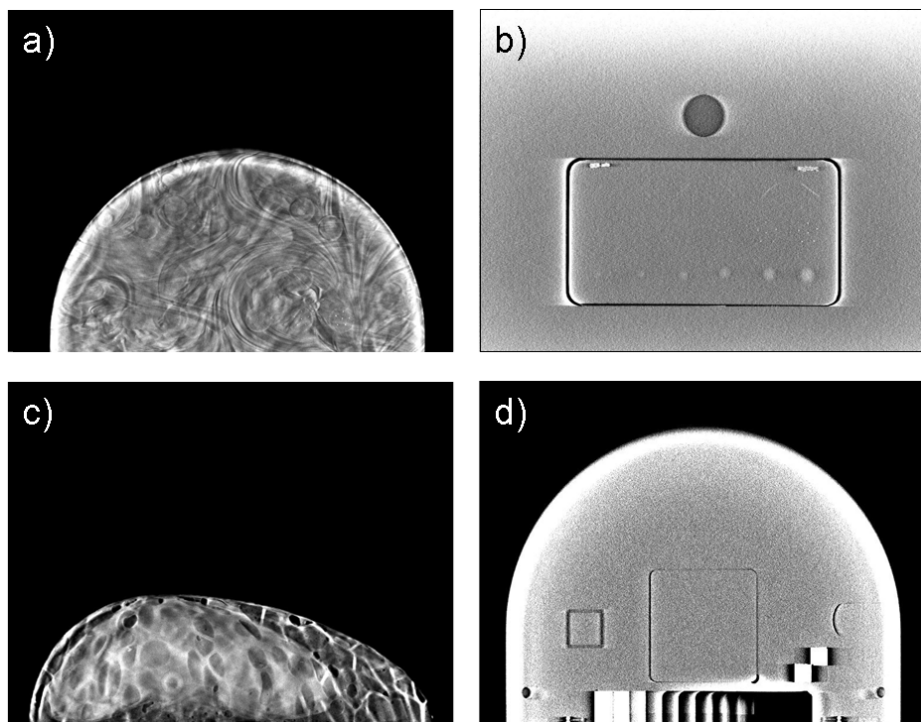
read, the better is the image, so that the visibility of these numbers is an additional imaging performance parameter.

The objective image analysis was done with a user-written Mathematica code, because the manufacturer's software to automatically analyze the image parameters has not yet been released. To calculate the MTF, the brass square was used, since it is further away from the step wedge and therefore suffers less from reconstruction artifacts. Both the MTF perpendicular and parallel to the breast wall were calculated. In order to estimate the noise variance, the noise power spectrum (NPS) was calculated averaging over the homogenous region in the center of the phantom of all available slices.

### 3 Results

Reconstructed images of each phantom under investigation are presented in Fig. 1. All images are shown with the same magnification in order to give the reader an impression about the different sizes of the phantoms and their structures.

Figure 1a shows the slice of the CIRS BR3D phantom that contains the objects. Even though the objects are hardly visible in Fig. 1, the four human observers



**Fig. 1.** Reconstructed images of the mammography phantoms under investigation: a) CIRS BR3D phantom, b) ACR FFDM phantom, c) anthropomorphic phantom and d) Quart phantom

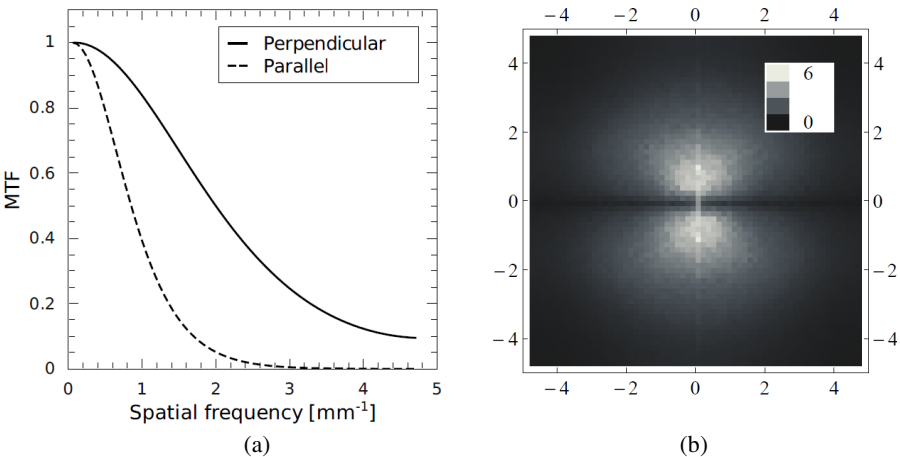
detected on average 4 ½ speck groups, 5 fibers and 3 masses on a suitable screen. However, it turned out that the actual scoring depends not only on the reader, but also slightly on the order of the slabs, which is influenced by the heterogeneous background.

In the ACR phantom, shown in Fig. 1b, the objects are embedded in a homogenous background. The four human observers detected on average all 6 masses, 3 ½ speck groups and 4 ½ fibers.

Figure 1c shows that the anthropomorphic phantom is composed of dense fibroglandular tissue in the inner region in which compartments of adipose tissue are embedded. The outer region is composed of adipose tissue supported by a matrix of Cooper’s ligaments. Since no objects are embedded in the phantom, it can only be stated that the phantom produces a realistic image of a breast, but neither a quantitative nor objective analysis can be performed.

The Quart phantom, shown in Figure 1d, provides means to assess image quality metrics objectively. The brass (top) and lead (bottom) squares on the right can be used to calculate the MTF and the homogeneous area in the center of the phantom allows calculating the NPS. We look at these two parameters to develop means to assess image quality as constancy testing. The MTF was calculated using the edges of the brass square perpendicular and parallel to the chest wall. The resulting MTF curves, shown in Fig. 2a, demonstrate the difference in resolution between the two directions. The 2-dimensional NPS is shown in Fig. 2b as a function of the spatial frequency. As expected, the NPS has the double-cone shape characteristic of a tomosynthesis NPS [11, 12].

The manufacturer recommends calculating the CNR using the step wedge. However, it can be seen in Fig. 1d that the step wedge affords only small ROIs which are very inhomogeneous in the reconstructed images due to edge enhancement effects. This makes it difficult to use it for CNR calculations in DBT. The CNR may be more reliably estimated using the brass and lead contrast squares. Even though they do not allow calculating a thickness specific CNR as the step wedge, they can be used for a simple CNR calculation at one specific height.



**Fig. 2.** MTF calculated from the brass square perpendicular and parallel to the chest wall (a) and the 2-dimensional NPS in mm<sup>2</sup> as a function of the spatial frequency in mm<sup>-1</sup> (b)

Scoring the Landolt rings turned out to be a challenging task, because finding the appropriate zoom and windowing level needs some time and experience. Further, the evaluation is quite time-consuming and wearisome for the eyes. On average the human observers were able to correctly detect 19 gaps which is slightly below the preliminary threshold for the image to pass the test. The manufacturer recommends evaluating more images in case the number of detected gaps is between 18 and 22 to have reliable results. However, since the tomosynthesis device is not on-site and our study focused on testing the phantom for its advantages and shortcomings, we did not acquire any additional images.

## 4 Discussion

Each phantom has its advantages and is useful for the purpose for which it was originally designed. However, we concluded that none of them by itself can be used as either a quantitative or objective phantom that provides information useful for image quality evaluation in DBT images. Further, none of these phantoms under investigation provide means to measure in-plane distance accuracy, slice-sensitivity profiles or the amount of breast tissue missed at the chest wall.

The advantages and shortcomings of each phantom can be summarized as follows:

The CIRS phantom features different objects in a heterogeneous background. It allows the qualitative evaluation of one reconstructed image slice that corresponds to the height in which the objects are positioned. Careful choice of the windowing level as well as high magnification and a trained reader are required to score the image. Also, visibility of the objects is dependent upon the ordering of the slabs, as the structures above and below the plane of reconstruction contribute to the image complexity. Therefore, it is important, in case this phantom is used for constancy tests, to always maintain the same order of slabs. Its shortcomings are that it is neither possible to check the reconstruction depth nor objective image parameters such as resolution, noise variance or CNR.

The ACR phantom allows the scoring of different objects in a homogenous background and enables the calculation of the noise variance and the CNR. However, since all masses are easily detectable, it still has to be proven that the phantom is sufficiently discriminative of differences in image quality and overall system performance for DBT. Moreover, the objects are only arranged at one specific depth so that neither the reconstruction depth nor the object visibility at different depths can be analyzed.

The Penn anthropomorphic phantom allows the evaluation of whether the scanner provides natural looking reconstruction images with breast like structures. It includes a controlled amount of dense tissue, which could be used to validate breast density estimation. Currently it only contains a limited variety of targets for a quantitative scoring of the images and no features that would enable an objective image quality evaluation. However, additional slabs containing defined objects for quantitative analysis or other features such as inserts for dosimeters (e.g., optically stimulated luminescence dosimeters) are feasible. With its realistic structure and 3-dimensional

extension, this phantom has the potential to become useful for image quality analysis as well as direct dose and dose distribution measurements in DBT.

The Quart phantom is the only phantom we tested, that had features to objectively measure the image quality parameters such as MTF, NPS or CNR. Depending on the performance and user-friendliness of the upcoming software to automatically evaluate the image quality, the phantom may potentially be useful for DBT. However, since it was primarily designed for mammography, it also suffers from several shortcomings. One issue is that the step wedge affords only small ROIs which are very inhomogeneous in the reconstructed images due to edge enhancement effects. Even though the manufacturer has informed us that the upcoming software will address this issue, the results of the CNR calculation may not be reliable. Moreover, due to the straight alignment of the inserted objects in this phantom, it is not possible to calculate an oversampled MTF, as required for mammography in the IEC standard [13], from an image that has been acquired according to the manufacturer's instructions. Finally, the Landolt rings allow checking the accuracy of reconstruction depth and object visibility in tomosynthesis images. The subjective evaluation with the Landolt rings however is wearisome and time-consuming, so that this test can hardly be part of a regular quality control in the clinic. Additionally, since the human observer mostly guesses the position of the gap, it has to be investigated whether the observer's memory begins to retain the position of the gaps and therefore achieves better results after repeated scoring.

## 5 Conclusions

Although each phantom under study has its advantages, none of them allows a thorough quality evaluation of reconstructed tomosynthesis images. The phantoms, in their current form, may be still better suited for projection mammography. In some cases (e.g. the Penn anthropomorphic phantom) the inclusion of additional layers permitting 3-dimensional analysis is feasible; while in others (e.g. the ACR FFDM phantom) major phantom redesign would be necessary for use in DBT. For all 4 phantoms tested, neither subjective nor objective evaluations involving all the reconstructed planes are possible. Since there is no other phantom on the market, for our knowledge that includes the features to measure all image properties relevant in DBT, it is necessary to design a new phantom. In order to allow a clinical implementation the new phantom has to allow a quick subjective image evaluation or provide user-friendly software to automatically analyze the objective image parameters that optimally could be measured in only one scan.

**Acknowledgement.** This work was partially supported by the FDA's Critical Path Project.

## References

- [1] Li, B., Saunders, R., Uppaluri, R.: Measurement of slice thickness and in-plane resolution on radiographic tomosynthesis system using modulation transfer function (MTF). In: Proc. SPIE 6142, 61425D (2006)

- [2] Bouwman, R., Visser, R., Young, K., Dance, D., Lazzari, B., van der Burght, R., Heide, P., van Engen, R.: Daily quality control for breast tomosynthesis. In: Proc. SPIE 7622, 762241 (2010)
- [3] Vecchio, S., Albanese, A., Vignoli, P., Taibi, A.: A novel approach to digital breast tomosynthesis for simultaneous acquisition of 2D and 3D images. *European Radiology* 21, 1207–1213 (2011)
- [4] Feng, S.S.J., Sechopoulos, I.: A software-based x-ray scatter correction method for breast tomosynthesis. *Medical Physics* 38, 6643–6653 (2011)
- [5] Carton, A.K., Bakic, P., Ullberg, C., Derand, H., Maidment, A.D.A.: Development of a physical 3D anthropomorphic breast phantom. *Medical Physics* 38, 891–896 (2011)
- [6] Carton, A.K., Bakic, P., Ullberg, C., Maidment, A.D.A.: Development of a 3D high-resolution physical anthropomorphic breast phantom. In: Proc. SPIE 7622, 762206–1 (2010)
- [7] Bakic, P.R., Zhang, C., Maidment, A.D.A.: Development and characterization of an anthropomorphic breast software phantom based upon region-growing algorithm. *Medical Physics* 38, 3165–3176 (2011)
- [8] de las Heras, H., Peng, R., Zeng, R., Freed, M., O'Bryan, E., Jennings, R.J.: A versatile laboratory platform for studying x-ray 3D breast imaging. In: Proceedings of IEEE Medical Imaging Conference MIC12.M-17 (2011)
- [9] de las Heras, H., Schöfer, F., Tiller, B., del Río, M.C., Zwettler, G., Semturs, F.: A new method for dosimetry and image quality assurance in mammography and breast tomosynthesis. In: Proc. IRPA (2012)
- [10] Perry, N., Broeders, M., de Wolf, C., Törnberg, S., Holland, R., von Karsa, L.: European guidelines for quality assurance in breast cancer screening and diagnosis - 4th (edn.) Technical report, Office for Official Publications of the European Communities (2006)
- [11] Richard, S., Samei, E.: Quantitative breast tomosynthesis: From detectability to estimability. *Medical Physics* 37, 6157–6165 (2010)
- [12] Richard, S., Samei, E.: Quantitative imaging in breast tomosynthesis and CT: Comparison of detection and estimation task performance. *Medical Physics* 37, 2627–2637 (2010)
- [13] IEC: 62220-1-2:2007 Medical electrical equipment - Characteristics of digital X-ray imaging devices - Part 1-2: Determination of the detective quantum efficiency - Detectors used in mammography. Technical report, International Electrotechnical Commission (2007)