A quantitative metrology for performance characterization of breast tomosynthesis systems based on an anthropomorphic phantom

Lynda Ikejimba a,b, Yicheng Chen c, Nadia Oberhofer d, Nooshin Kiarashi b,e, Joseph Y. Lo a,b,e,g, Ehsan Samei a,b,e,f,g

a Medical Physics Graduate Program, Duke University;  
b Carl E. Ravin Advanced Imaging Laboratories, Department of Radiology, Duke University Medical Center;  
c Department of Engineering Physics, Tsinghua University, China;  
d Health Service South Tyrol, Italy;  
e Department of Electrical and Computer Engineering, Duke University;  
f Department of Physics, Duke University,  
g Department of Biomedical Engineering, Duke University, Durham, NC

ABSTRACT

Purpose: Common methods for assessing image quality of digital breast tomosynthesis (DBT) devices currently utilize simplified or otherwise unrealistic phantoms, which use inserts in a uniform background and gauge performance based on a subjective evaluation of insert visibility. This study proposes a different methodology to assess system performance using a three-dimensional clinically-informed anthropomorphic breast phantom.

Methods: The system performance is assessed by imaging the phantom and computationally characterizing the resultant images in terms of several new metrics. These include a contrast index (reflective of local difference between adipose and glandular material), a contrast to noise ratio index (reflective of contrast against local background noise), and a non-uniformity index (reflective of contributions of noise and artifacts within uniform adipose regions). Indices were measured at ROI sizes of 10mm and 37 mm, respectively. The method was evaluated at fixed dose of 1.5 mGy AGD.

Results: Results indicated notable differences between systems. At 10 mm, vendor A had the highest contrast index, followed by B and C in that. The performance ranking was identical at the largest ROI size. The non-uniformity index similarly exhibited system-dependencies correlated with visual appearance of clutter from out-of-plane artifacts. Vendor A had the greatest NI at all ROI sizes, B had the second greatest, and C the least.

Conclusions: The findings illustrate that the anthropomorphic phantom can be used as a quality control tool with results that are targeted to be more reflective of clinical performance of breast tomosynthesis systems of multiple manufacturers.

Keywords: tomosynthesis, breast phantom, image quality, non-uniformity
I. INTRODUCTION

X-ray based breast imaging is a cornerstone of breast cancer detection and management.\textsuperscript{1-3} At the same time, breast imaging is one of the most challenging tasks in medical imaging, primarily due to the small difference between the attenuation properties of cancerous and fibroglandular tissue being masked by the significant heterogeneity of the normal breast tissue.\textsuperscript{4} This challenge has led to a number of recent advances in breast imaging including contrast mammography and digital breast tomosynthesis (DBT).\textsuperscript{5,6} This challenge has also led to requirements for stringent image quality performance for breast imaging systems. These requirements aim to ensure that the imaging systems can deliver excellent image quality.

An image quality characterization based on realistic phantoms require access to such a tool. Prior work has been done to create volumetric phantoms emulating the heterogeneous texture of breast tissue. Park \textit{et al.}\textsuperscript{7} and Gang \textit{et al.}\textsuperscript{8} developed bead phantoms that consist of acrylic spheres housed in an acrylic volume and filled with water. Others have devised phantoms with internal structures of relevant statistical properties.\textsuperscript{9,11} These efforts have led to models that are much more relevant for mammographic characterization than uniform phantoms. However, two major limitations have remained: First, these models have not been based on properties of three-dimensional breast tissue as present in actual patients. Second, even if such a desire is achieved, there is a need to capture the quality attributes of the images from a single phantom in quantitative terms. Otherwise, relevance in terms of breast heterogeneity cannot be realized in a clinical setting.

The purpose of this study was to devise an objective image quality metrology based on a realistic breast phantom, and further evaluate the metrology in a field trial of three major DBT imaging systems of different makes and models. The phantom was based on a three-dimensional printing of a virtual breast phantom developed at our laboratory.\textsuperscript{12} Formed based on the actual tissue attributes of a real human breast, the phantom was imaged on four DBT systems. The images were characterized in terms of contrast rendition, contrast to local noise ratio, and background non-uniformity in the adipose tissue caused by the out of plane blur. The scope of this paper is not to perform a vendor comparison. Rather, this paper presents and applies new image quality metrics over a cohort of different machine types. The methodology was designed to be generalizable to other systems and anthropomorphic phantoms, and be used for acceptance testing, quality control, and tracking of system performance across sites and over time.

II. METHODS

II.A. Breast Phantom

In previous work, an anthropomorphic physical breast phantom was developed based on dedicated breast CT data of actual human subjects.\textsuperscript{12,14} Starting from the segmentation of a dedicated breast CT acquisition of a human subject, a virtual, voxelized model was developed. The breast was virtually compressed to 42 mm in thickness. Using the voxelized model, a corresponding physical phantom was produced using photopolymers and 3D printing. Two materials, mimicking adipose and glandular tissues, were selected based on their x-ray attenuation properties within the kV range of tomosynthesis radiographs. Hence, this anthropomorphic phantom is referred to as the Doublet. There was graduated mixing during the printing process to create five tissue classes and allow more realistic transitions between adipose and glandular boundaries. After printing, the phantom was imaged on several DBT systems. The reconstructed images may then be compared to the segmented CT image, representing the ground truth. The shape and structuring of the phantom are presented in FIG. 1.
FIG. 1. Axial view of the breast CT scan and Doublet with CC compression. (a) The breast CT scan of a human subject was segmented into five tissue classes to create the voxelized model. The model was 3D printed to create the Doublet phantom. Composite (b) and assembled (c) halves show the intricate anatomical interior detail and overall form.

II.B. Imaging Systems

Three current commercial DBT systems are represented in this study: GE SenoClaire (Chalfont St. Giles, England), Hologic Selenia Dimensions (Bedford, Massachusetts, USA), and Siemens MAMMOMAT Inspiration (Erlangen, Germany). All systems are in clinical use outside the US, but at the current time, only the Hologic and GE systems are approved by the FDA for clinical use in the US. Each machine employed acquisition geometry unique to that vendor, and some have been described in literature\textsuperscript{15}. A summary of the machine specifics is provided in Table 1.

II.C. Image Acquisition

The phantom was imaged with exposure modified to deliver a fixed $1.5 \pm 0.1$ mGy average glandular dose (AGD) as reported by the vendor operator console. Specifics of the tube settings are also provided in Table 1. To reduce uncertainty in the measurements, 6 consecutive scans were acquired and averaged. The results reported are values from the averaged image. Slice through the reconstructed volumes are presented alongside the voxelized model in FIG. 2.

<table>
<thead>
<tr>
<th>Vendor</th>
<th>Siemens (Merano, It)</th>
<th>Hologic (Bolzano, It)</th>
<th>GE (Milan, It)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of projections</td>
<td>25</td>
<td>15</td>
<td>9</td>
</tr>
<tr>
<td>Tube sweep (deg)</td>
<td>45</td>
<td>15</td>
<td>25</td>
</tr>
<tr>
<td>Source to imager distance (mm)</td>
<td>650</td>
<td>700</td>
<td>660</td>
</tr>
<tr>
<td>Image pixel size, x-y plane (mm)</td>
<td>0.085</td>
<td>0.091</td>
<td>0.100</td>
</tr>
<tr>
<td>Image pixel size, z plane (mm)</td>
<td>1</td>
<td>1</td>
<td>0.5*</td>
</tr>
<tr>
<td>Target/filter</td>
<td>W/Rh</td>
<td>W/Al</td>
<td>Rh/Rh</td>
</tr>
<tr>
<td>Tube voltage (kV)</td>
<td>27</td>
<td>28</td>
<td>29</td>
</tr>
<tr>
<td>Tube load (mAs)</td>
<td>125</td>
<td>64</td>
<td>50</td>
</tr>
<tr>
<td>Dose (mGy)</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>
II.D. Image Analysis

The proposed strategy for image quality characterization allows the user to take advantage of having a matched virtual phantom. Introduced are three indices that utilize the ability to compare between the reconstructed DBT images of the phantom and the model, the ground truth: a contrast index, a contrast to noise ratio index, and non-uniformity index. The contrast index represents how truthfully the system reproduces the difference in attenuation coefficients between two regions. Using the virtual phantom, pixels can be analyzed from specific, matched regions of the reconstructed slices to calculate signal differences. The contrast to noise ratio index mitigates the absolute contrast in terms of local background noise. The non-uniformity index indicates the amount of noise within a slice. Out of plane blurring creates a cluttered background against which the image features must be interpreted. The extent of slice cross-talk is captured by comparison with the model, which exhibits perfect slice separation. This package makes it possible to determine how the true glandular-adipose topography is rendered across the reconstructed breast volume.

II.D.1. Preprocessing

Three preprocessing steps were necessary to standardize the images before analysis. First, the images were detrended to eliminate the effect of low-frequency background. A 4th order polynomial fit was applied to the interior of the breast. The fit was then subtracted, producing a “flatter” image. Next, a slice through the image was registered to the corresponding slice in the voxelized model. Registration is a key component of the image analysis in order to be able to reliably associate imaged features to the truth, considering different geometrical and recon attributes of various systems. Registration was performed based on landmark selection using the ImageJ plugin bUnwarpJ. Three different registrations were created, and the one resulting in the highest contrast was selected. Lastly, a linear transformation was applied to the voxel values of the reconstructed images to standardize their dynamic range. In each slice, the voxel in the 0.1 percentile was scaled to 1, while the 99.9th percentile was scaled to 100. This ensured that numerical results of the calculations would not be affected by scalar offsets from different reconstructions. Additionally, this transformed the voxel values to units of percent contrast.
I.D.2. Contrast Characterization

Contrast is a well-established component of image quality. An imaging system that would provide high and consistent level of contrast between tissues regardless of the tissue location would be considered a superior system. The pixel values in the reconstructed DBT slice have undergone log-transformation during image reconstruction. Thus, the signal difference measured in this space is analogous to contrast in projection space or traditional radiographs. However, in DBT, contrast rendition is complicated not only by x-ray absorption properties by also by variability across the imaged volume due to limited angle sampling of the image space. Due to complexity of the phantom and DBT reconstruction artifacts, contrast is highly dependent on anatomical context and location. It is therefore necessary to analyze contrast in terms of a contrast index (CI). While contrast is the local signal difference between adipose vs. fibroglandular voxels within a given ROI, contrast index is the mean contrast value over all ROI locations. An additional metric is used to quantify the location dependence of the contrast. This quantity is called the contrast variability Index (CVI) and is defined as one standard deviation of the mean contrast values.

Local contrast was measured using circular ROIs, which overlapped by 50% in both x and y directions while spanning the central portion of the breast. The ROI diameter or window size was varied between 10 and 37 mm, with the upper bound suggested by a previous psychophysics study\textsuperscript{17}. In each circular window, pixels were classified as glandular or adipose according to the ground truth segmentation mask from the virtual phantom. FIG. 3 illustrates the positioning of the largest ROIs, the overlay of the voxelized model, and the resulting masked ROIs. The slice analyzed was at approximately mid-depth (24 mm above detector). As the diameter of the window increased from 10 to 37 mm, the total number of ROIs within the parenchyma decreased. Windows were excluded if no glandular tissue was present. The contrast index was further defined at two window sizes 10 mm and 37 mm: CI\textsubscript{10} and CI\textsubscript{37}. The index at 10 mm is reflective of the foveal focus, or the contrast in the central portion of the vision. The index at 37 mm is reflective of the adaptation state of the eye; this is the maximum region to which the eye adapts to determine contrast sensitivity.

FIG. 3 Contrast in the phantoms. (a) A circular ROI with 50% overlap is shown here at 37 mm. Dashed outlines indicate the position of the shifted ROIs. (b) Using the voxelized model as ground truth, a mask (overlaid in red) was generated to segment adipose vs. fibroglandular tissue. (c) After masking, ROIs are created consisting of adipose-only and glandular-only signal. For each ROI, local contrast was calculated as the difference in mean values between the two tissues. For comparison, the 37 mm window encompasses a larger area of the phantom, and the 10 mm ROI is shown annotated by the green arrow and overlaid with the green circle.
II.D.3. Contrast to Noise Ratio Characterization

Contrast to noise ratio (CNR) is another established measure of image quality, which takes into account background noise. Within each ROI, the contrast was computed as previously described. Using the masked adipose-only ROIs, noise was measured as the standard deviation of adipose signal. However, for each ROI size, several contrast and noise measurements are made across the entirety of the breast, resulting in a different CNR at each location. These values are then averaged to produce a single, position-independent contrast to noise ratio index (CNI) for each window size. For simplicity of presentation, a scalar noise measurement was also produced for the Results section, computed as the average adipose standard deviation at each window size.

II.D.4. Non-Uniformity Characterization

One of the key image quality attributes of DBT is the “bleeding” of signal into a slice from slices that are above or below the slice. This attribute can be measured by the level of non-uniformity that is present in the adipose compartment of the breast. As such, a non-uniformity index (NI) was used to characterize the extent to which such image artifacts would alter the intensity of an otherwise uniform region. This would include noise, the effects of intra-plane ringing artifacts, and inter-plane ghosting artifacts due to structures above and below a selected plane.

Background non-uniformity is measured from the variability of signal from the adipose material. Because there is no fibrolandular tissue in those regions, in an ideal imaging system the mean adipose value should be uniform across the breast. To compute the non-uniformity, the same masked, sliding ROIs used to compute contrast were used. However, the mean adipose signal was computed in each ROI. We then found the variance of the means across all ROIs of the same size. The NI was defined as the variance of the means with 10 mm ROIs, and the NI with 37 mm.

III. RESULTS

III.A. Contrast Index

The contrast values are provided in FIG. 5 for the three vendors. The data represent the spread of the contrast values calculated from individual ROIs of 10mm. The CI of vendor A represents about 13% of the dynamic range of the whole breast region, B represents about 10%, and C 6%. It can be seen that the greatest contrast values are provided by vendor A with 42, then B with 31, and finally C with 23. The CI for vendor A is about 3% greater than that of vendor B, while the CI of C is about 4% less than B. Although it has the lowest contrast values, vendor C also demonstrates a greater degree of reproducibility since its plot has the narrowest spread. Similarly, vendor A has the widest range of possible contrast values.

In FIG. 5(a) the CI is presented as a function of window size. Power law fits were applied to the data and were used to derive the CI and CVI at 10mm and 37 mm. From the smallest to the largest window size, CI decreased by 26% for vendor B and by 22% for vendor C, but only by 10% for vendor C. Data in part FIG. 5(b) are the contrast variability index, derived from the standard deviation of all CI values from ROIs of a given size. A lower standard deviation of the CI values indicates more consistency in rendering contrast. Measurements from vendors A and B contained greater variability than that of A. Additionally, vendor A showed the strongest dependence of variability on window size as its CVI ranged from 10.6 at 10 mm to 5.3 at 37 mm. The other vendors showed a smaller dependence of CVI on window size. The index values for each vendor are presented in Table 2.
Table 2. Contrast and contrast variability indices for all vendors. Values were obtained from power fits to the data.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI\textsubscript{10}</td>
<td>13.2</td>
<td>11.0</td>
<td>7.9</td>
</tr>
<tr>
<td>CVI\textsubscript{10}</td>
<td>10.6</td>
<td>8.3</td>
<td>6.5</td>
</tr>
<tr>
<td>CI\textsubscript{37}</td>
<td>9.8</td>
<td>8.6</td>
<td>7.1</td>
</tr>
<tr>
<td>CVI\textsubscript{37}</td>
<td>5.3</td>
<td>6.2</td>
<td>3.8</td>
</tr>
</tbody>
</table>

FIG. 4. Contrast for the smallest window size for vendors, at fixed AGD. The mean of these yields CI values across the entire image. This spread of the contrast values suggest changes in signal differentiation based only on location.

FIG. 5 Contrast Index across different window size and different vendors, at fixed AGD (a) The CI at a fixed window diameter of 10 mm yields the range of CI values across the entire image, indicating changes in signal differentiation based only on location. (b) At each window size, the maximum CI was determined from all windows, and results were compared against Vendors A, B, and D. Exponential curves were fit to the data. Solid lines indicate an exponential fit. (c) The mean and (d) standard deviation of all ROIs of the same size were then calculated.
III.B. Contrast to Noise Ratio Index

Results from the noise and CNI computation are provided in FIG. 6. The subfigure FIG. 6(a) presents the standard deviation of adipose voxel values within the ROI as the window was enlarged. A power fit was applied to the data. For all vendors, the standard deviation increased monotonically as more pixels were included. However, beyond a width of about 22 mm, the noise increases at a slower rate. The amount of noise is comparable for each vendor. Figure FIG. 6(b) shows the CNI, averaged contrast to noise ratio, for each window size. Second order polynomial fits were applied to the data. CNI values remained fairly constant regardless of ROI size. Vendor A demonstrated the highest CNI, followed by vendor B, and lastly by vendor C.

![Graphs of Noise and CNI vs. Window Size](image)

FIG. 6 Noise and Contrast to noise ratio index for three vendors. (a) The noise was found from the variance of the adipose signal within each ROI, then averaged to a single data point at each widow size. (b) The CNI was found locally in a similar fashion and averaged across all ROIs of each size.

III.C. Non-uniformity Index

The image non-uniformity was in the spatial domain from the fatty background variability. Regions after adipose-only masking are presented for each vendor in FIG. 7. The inhomogeneity of the background is visible within this region that represents only adipose signals. Results of the NI at each ROI size are plotted in FIG. 8. Power law fits were applied to the data. Vendor A is seen to have the greatest amount of non-uniformity, while vendor C has the least. Furthermore, vendor C shows the least amount of sensitivity with respect to window size. The interpolated NI values from the power fit are organized in Table 3. It can be observed that the background non-uniformity of vendor D was about 30% greater than A, and that of C is about 20% greater than D. Vendor B produced the highest variability overall.
FIG. 7. Non-uniformity of adipose background, shown in a region measuring 25x25 mm. Tissue masking was applied to the slice to exclude glandular and select for signal only from intra- or inter-plane artifacts. The mask appears different for each vendor because a fixed size of 25mm was selected, but pixel sizes vary.

FIG. 8. Variability of adipose signal measurements. The mean adipose signal was found in each ROI as the window width increased. The standard deviation of the signals about the mean decrease as more pixels were included for the larger windows. Vendor A demonstrates the lowest signal variability, while vendors B and C show similar, high values.

| Table 3. Non-uniformity indices measured from adipose ROI mean variability. |
|-----------------|---|---|---|
|                 | A  | B  | C  |
| NI_{10}         | 6.0| 5.7| 3.4|
| NI_{37}         | 4.1| 3.5| 2.8|
IV. DISCUSSION

For a fixed ROI size, the contrast of an image with two materials ideally should not depend location, and its value would remain essentially the same for all ROIs. For a perfect imaging system, the boxplots in FIG. 4 would show little spread with clustering about the mean value. A wider spread further indicates greater separation in contrast values across the ROIs. Specifically, this suggests a greater dependence on location. Variation in the CI values suggests that contrast can be impacted by the proximity of other structures, a phenomenon that may not appreciated from a uniform phantom. It may be seen from FIG. 5 that vendor A had the highest overall CI and vendor C had the lowest. This ranking confirms what is apparent to the eye from the DBT images demonstration images shown. Images from a vendor displayed greater contrast and maintained the highest max and mean CI curves. The image from another vendor looks more “washed out”, and this quality is reflected by the CI metrics. Systems with higher contrast also contained greater contrast variability. This metric is important because it describes the variability of the CI over the image. A better system would have high overall contrast and lower variability, so the lowest curve would correspond to and image with more uniform contrast values. This ordering of results is again likely due to the influence of signal from other slices.

Non-uniformity was measured in the spatial domain from the fatty background variability. Again an ideal imaging system would have a smoother background, resulting in consistent mean values at any window size and location. At all ROI sizes, vendor A demonstrates the greatest amount of non-uniformity, followed in order by vendors B and C. At the smallest window widths, some ROIs will capture the signal of contaminant anatomical structure, while others will not. This will result in a greater range of variability in mean ROI signal. At the larger ROI sizes, each window is likely to contain both true adipose signal and overlying anatomy. In addition a greater number pixels are included in the calculation. Hence, the ROIs are likely demonstrate overall less variability. Overall, the variability decreases as the window width is increased and more pixels were included in each calculation.

Given the significance of global heterogeneity and background variability, CNI can only be computed locally. It would be erroneous to simply divide the curves from FIG. 5(a) by their counterparts in FIG. 6(a). Such an approach would combine all values and disregard the impact of the immediate background noise on the visibility of a region. With this it becomes evident that the CNI changes very little with ROI size, and thus the curves in the contrast and non-uniformity indices were solely due to ROI size. It would also appear that CNI is primarily driven by contrast, since the rank ordering of the curves is similar to that of the CI plot, having received little distinguishing information from the noise plots.

While this work is novel in its scope, it contains some limitations. These preliminary data were based on a single slice of a single phantom. In further studies we will investigate the effects of different depths in multiple phantoms with different densities and tissue patterns. We are currently developing superior materials for further prototyping. Materials with a higher dynamic range of attenuation values may better mimic the appearance of breast tissue in DBT images. Furthermore, this work approached image quality quantification in the spatial domain. In future work imaging noise and resolution will be examined in the frequency domain, which will allow for clearer insight to the interplay of anatomical noise and stochastic noise effects.

V. CONCLUSION

Current methods for QC in breast tomosynthesis are overly simplistic and rely on inconsistent interpretations of the results. This new phantom and methodology would offer a QC metrology that accounts for breast heterogeneity and its 3D depiction. As such, the phantom characterizes system performance in terms that more directly correspond to clinically relevant conditions, as opposed to geometrical phantoms with homogenous background. Visual comparison confirmed that CI, CVI, and NI values correspond with qualitative assessment of contrast and noise from DBT slices. The values of the indices may be used as a baseline for clinical QC measurements.

The proposed technique can characterize diagnostic performance of DBT systems in several important aspects. It provides a platform to evaluate the assortment of existing technologies. It also offers a method of optimizing clinical imaging protocols. Most importantly, it ensures quality control and patient safety beyond vendor-specific quality control procedures.
REFERENCES