Published in final edited form as:
*Proc SPIE*. 2013 March 21; 8668: . doi:10.1117/12.2008503.

# Towards Visual-Search Model Observers for Mass Detection in Breast Tomosynthesis

**Beverly A. Lau**, **Mini Das**, and **Howard C. Gifford**
The University of Houston, Houston, TX, USA

## Abstract

We are investigating human-observer models that perform clinically realistic detection and localization tasks as a means of making reliable assessments of digital breast tomosynthesis images. The channelized non-prewhitening (CNPW) observer uses the background known exactly task for localization and detection. Visual-search observer models attempt to replicate the search patterns of trained radiologists. The visual-search observer described in this paper utilizes a two-phase approach, with an initial holistic search followed by directed analysis and decision making. Gradient template matching is used for the holistic search, and the CNPW observer is used for analysis and decision making. Spherical masses were embedded into anthropomorphic breast phantoms, and simulated projections were made using ray-tracing and a serial cascade model. A localization ROC study was performed on these images using the visual-search model observer and the CNPW observer. Observer performance from the two computer observers was compared to human observer performance. The visual-search observer was able to produce area under the LROC curve values similar to those from human observers; however, more research is needed to increase the robustness of the algorithm.

### Keywords

Breast tomosynthesis; acquisition geometries; image quality; mass detection; model observers; task-based assessment; visual search

## 1. INTRODUCTION

There has been considerable research into methods for assessing the image quality of digital breast tomosynthesis (DBT) systems. We are interested in human-observer models that perform clinically realistic tasks, with the expectation that realistic tasks make for easier interpretation of the model-observer results. In previous work, we tested visual-search (VS) observers against human observers for microcalcification (MC) detection and localization in DBT reconstructed images.[1, 2] These VS models attempt to mimic trained radiologists by implementing a two-phase process of initial holistic search followed by directed analysis and decision making. This design was inspired by original research done by Kundel *et al.* on trained radiologists, where he showed that they fixated on the abnormality in the first 1.13 seconds of the search.[3] Within our VS framework, the search phase identifies candidate locations in the image for subsequent analysis with a statistical model observer.

The statistical model observer used for the VS model in this paper is the channelized non-prewhitening (CNPW) scanning observer. In earlier work, we investigated parameter



Further author information: (Send correspondence to H.C.G.) H.C.G.: hgifford@central.uh.edu.

optimization for masses in DBT using the CNPW observer model alone.[4] The CNPW observer performed mass detection and localization, but under a "background-known-exactly" (BKE) task paradigm. Clinical mass detection is heavily impacted by anatomical noise, which is one reason the CNPW scanning observer may not agree with human observers.

The VS framework offers one way to make the scanning model more realistic, by accounting for the effects of anatomical noise on observer performance. The VS observer in our tests uses the same underlying CNPW model, but has a front-end that includes a holistic search to narrow down the search area based on the image without the background subtraction.

The current study investigates a VS observer for mass detection and localization in DBT. Since masses have different properties than MC clusters in number, size, and attenuation, our previous algorithm for the holistic visual search was inappropriate for the task. Instead, a 2D gradient template match was used to identify areas of interest, and the existing scanning observer was used on those areas of interest for analysis and decision making. To the best of our knowledge, this is the first time a 2D template match is being used for lesion detection. A localization ROC (LROC) study was performed to compare the performance of the CNPW and VS observers with human observer performance.

## 2. METHODS

Observer studies were conducted with four (4) human observers, the CNPW scanning observer, and the new VS observer for the detection of spherical lesions in simulated reconstructed digital breast tomosynthesis slices.

The tomosynthesis slice images used for this study were generated for an earlier study on the effect of acquisition parameters with the CNPW observer.[4] Briefly, 8-mm diameter spheres were inserted into the anthropomorphic breast volumes, which were generated by Bakic *et al.* at the University of Pennsylvania.[5] There were six (6) total unique breast phantoms: three (3) with approximately 25% volumetric glandular fraction (VGF) and three (3) with approximately 50% VGF (exact VGF values can be found in Ref.[4]). Eight lesions were inserted into each breast volume, using attenuation data from Johns and Yaffe.[6] The centers of the lesions were separated by a minimum of 4.05 mm of tissue in the slice direction.

Simulated projections were acquired using Siddon's ray tracing method[7] for a circular tomosynthesis unit. A serial cascade model was used to simulate system gain and noise.[8] Focal spot blur and CsI scintillator blur were modeled. Quantum noise was modeled as Poisson noise, and electronic noise was also accounted for. Scatter was not included in the model. Projection view data was reconstructed using the Feldkamp filtered backprojection algorithm, and a postreconstruction 3D Butterworth filter with 0.15 pixel$^{-1}$ cutoff frequency was applied to the images.[2] The reconstructed slices had 0.27 mm × 0.27 mm in-plane resolution.

The six phantoms were each projected five times, using five different acquisition parameter settings. The number of projection views used were $P \in \{3, 7, 11, 15, 19\}$. The rest of the parameters were held constant for the simulated projections, including an angular range of 60 degrees, a filtered molybdenum (Mo) spectrum at 30 kVp, and detector pixel size of 100 microns. The number of photons used corresponded to 4.0-mGy average glandular dose, which was divided equally among the number of projection views. This process was repeated for the six phantoms without inserted masses.

Images for the observer study were produced by extracting slices that contained the center coordinates of the inserted masses. The analogous slices for the corresponding lesion-free phantoms were also extracted to use as paired "normal" cases. This yielded 96 total cases: 48 with lesions and 48 without lesions.

Data from four human observers were collected. The 96 cases were divided into 72 study cases and 24 training cases (keeping paired cases in the same set). Each observer read the images for each $P$ in two sessions of 12 training cases and 36 study cases each. Low ambient light was used during the study. The observer was asked to identify the most likely place where a lesion might be, then to assign a rating based on confidence on the lesion's absence or presence. A scale of 1 to 4 was used, with definitions for the ratings shown in Table 1. For the training cases, the truth was displayed to the observer (either there was no lesion present or the location of the lesion). For the study cases, the truth values were not displayed, and the observer just clicked "Next" to display the next image in the sequence. The observer could change his/her location and ratings an unlimited number of times for a given case, until the "Next" button was pressed. Only the final location and rating were recorded for each case.

An observer's raw data was analyzed based on a radius of correct localization ($R_{cl}$) of 14.8 voxels (4 mm). Observer performance was quantified in terms of area under the LROC curve ($A_L$). The $A_L$ values were calculated using a Wilcoxon-based non-parametric ranking method.[9]

Observer performance data from the CNPW and VS observers were collected and compared with the human observer data.

### 2.1 CNPW Observer

The CNPW observer localization **r** and confidence rating λ for an image **f** were drawn by applying the rules

$$\lambda = \max_{j \in \Omega} Z_j \quad (1)$$

$$\mathbf{r} = \arg \max_{j \in \Omega} Z_j \quad (2)$$

to a set of perception measurements $\{Z_j\}$ made at each voxel in a search region Ω. At voxel $j \in \Omega$, the scanning CNPW observer computed the measurement

$$Z_j = \mathbf{w}_j^t(\mathbf{f} - \mathbf{b}), \quad (3)$$

where **b** is the noise-free background corresponding to **f** and $\mathbf{w}_j$ is the $j^{th}$ channelized observer template. The CNPW template is shift invariant, composed of a slice through the center of the true mass signal (circle with radius 4.0 mm), modulated by a set of frequency selective channels. Three difference-of-Gaussian channels were used, as our group has done previously for CNPW calculations.[4]

The subtraction of **b** in Eq. 3 sets the observer's relative operating point for assessing malignancy at the various locations. The search region Ω for the CNPW observer included all areas of the tomosynthesis slice image that contained breast tissue, except for the edge of the breast, where artifacts from reconstruction produces abnormally high values. The edge of the breast was removed through edge detection and erosion with a radius of 20 pixels,

which was determined empirically. Confidence ratings were discretized to a scale of 1 to 4, to match the human observer responses in Table 1, using a rebinning method.[10]

### 2.2 Visual Search Observer

The VS observer in this work divides the localization task into two steps: a holistic search that returns locations of interest, followed by the CNPW observer applied only to those locations.

The localization and confidence rating for an image **f** are respectively drawn from the arg max and max of a set of perception measurements made at each voxel in a search region $\Omega'$. The search region $\Omega'$ consists of the areas of interest defined by the holistic search.

A 2D gradient template match was used to generate the search region $\Omega'$. For the gradient template match, a [3 × 3] Sobel filter was used on the byte-scaled tomosynthesis slice images in two orthogonal directions. The two gradient images ($\mathbf{G}_x$ and $\mathbf{G}_y$) were normalized by the modulus at each voxel $j$ in the 2D image:

$$G_\alpha(j)=\frac{G_\alpha(j)}{\sqrt{G_x(j)^2+G_y(j)^2}}, \quad (4)$$

for $\alpha = x$ or $y$. Normalized gradients were similarly acquired for the truth image of the circular signal ($\mathbf{w}_x$ and $\mathbf{w}_y$). The pixel size for the truth signal was 0.27 mm × 0.27 mm, chosen to match the reconstructed in-plane voxel size.

The cross-correlation image between each normalized gradient image and the normalized gradient template image was computed at each voxel location $j$:

$$Z_{\alpha,j}=\mathbf{w}_{\alpha,j}^t(\mathbf{G}_\alpha), \quad (5)$$

where $\mathbf{w}_{\alpha,j}^t$ is the $j$th gradient template and $\mathbf{G}_\alpha$ is the gradient image for $\alpha = x$ or $y$. The correlation image was computed for $\mathbf{G}_x$ and for $\mathbf{G}_y$ using the gradient templates $\mathbf{w}_x$ and $\mathbf{w}_y$, respectively. The two correlation images were summed ($\mathbf{Z}_{sum} = \mathbf{Z}_x + \mathbf{Z}_y$), and the whole-breast mask $\Omega$ was used to remove the edge of the breast, where artifacts from the gradient calculation produced abnormally high values. A global threshold was applied to the summed image, using $0.9*\max_{j\in\Omega} Z_{sum,j}$ as the cutoff. Once the global threshold was applied, the search area $\Omega'$ was obtained through morphological dilation using a slice through the center of the true mass signal as the structuring element. To complete the VSO, the CNPW was used (as described in §2.1, with the new restricted search area $\Omega'$) to obtain localization and confidence ratings.

Since threshold cutoff and structuring element radius changed the localization and detection results for the VS observer, a range of threshold values and structuring element radii were investigated. Threshold cutoff values ranged from 50% to 98% of the maximum $Z_{sum}$ value and structuring elements had radii ranging from 0.81 mm to 4.0 mm.

## 3. RESULTS

Area under the LROC curve ($A_L$) data from the two computer model observers were compared to the analogous data from the average human observer, and the results are shown in Figure 3. There are five points per VGF, representing the five different DBT acquisition settings we used (number of views = 3, 7, 11, 15, 19).

The quantity $A_L$ from four human observers were averaged to obtain data for the average human observer (see Fig. 4). For the higher density phantom, a smaller number of views used for data acquisition led to higher $A_L$ values, generally; however, further study is necessary to show significance.

Both the VS and the CNPW observers were compared to the average human observer, across different number of projection views (see Fig. 5). Lastly, $A_L$ for a range of VS observer global threshold values and structuring element radii were calculated (see Fig. 6).

## 4. DISCUSSION

For predicting human performance, the BKE task assumption is an issue with the scanning CNPW observer. With nonlocalization (or location-known-exactly) tasks, Hotelling-type models are routinely used for background-known-statistically (BKS) tasks. The scanning observer can also be applied in a BKS setting, but this calls for the use of an aggregate mean background $\bar{\mathbf{b}}$ in place of the background **b** in Eq. 3. This aggregate is formed as the mean of **b** taken over the various cases used in the study. We note that human observers would likely not use such a reference image in performing the task.

The VS observer offers an alternative approach to handling the effects of anatomical noise. While the results from this study look promising, there are a number of limitations to the study that must be noted. Work to resolve these issues is on-going.

First, the VS observer in our mass study is not yet robust. For other imaging modalities, like SPECT, gradient ascent/descent can be used to obtain the areas of interest mask $\Omega'$, without using a pre-determined global threshold. For this study, however, we used a threshold and dilation procedure to obtain $\Omega'$, both of which can be varied. We have reported that changing the threshold leads to larger variation in $A_L$ than changing the structuring element radius (Fig. 6). One issue with this analysis is that the number of pixels remaining in the mask ($\Omega'$) was not kept constant between the threshold variation and structuring element radius variation. In the future, we will make the same comparison, keeping the total area in $\Omega'$ the same for the range of comparison. We are working on an adaptive thresholding method to narrow down the search area without using global threshold and dilation. One possibility is to include a training phase, just as we train the human observers. The VS observer would use the 24 training cases (with and without lesions) to determine a global threshold to use on the 72 study cases. This type of training is holistically similar to what the humans may be doing.

Second, some of the tomosynthesis slice images in the study were compromised. To expedite the simulated projection and reconstruction, eight (8) lesions were added to each breast phantom. At least 15 voxels (4.05 mm) separated the centers of the lesions in a single breast phantom. Since each of the lesions themselves are 4 mm in radius, this leads to “ghost lesions” in some of the reconstructed slice images, centered at the location of another lesion. It is unclear how these “ghost” lesions affect the VS observer $A_L$ calculation. Since we are trying to model the human observer, one may argue that the computer observer performs realistically like a human, if it is confused (or not confused) the same way as a human, given these artifacts in the image. One problem with this argument is that the CNPW and VSO observers use noise-free background subtraction, which does not include the “ghost” lesion as part of the background.

The last problem to address involves what it means for a computer observer to behave like a human observer. Since humans do not agree, it seems wrong to expect that the computer should agree exactly with one human or another. Instead, we plan to improve the VS

observer such that it creates the mask $\Omega'$ to match false positives marked by many or all human observers. In order to do this, we will need to run our experiment again, but asking the observers to mark the *most likely location of a cancer, even if the observer thinks that one is not in that particular image*. Adding this criterion to the observer study will ensure that the false positive marks are areas of interest for a human observer. To make sure further that we know what the humans are looking at, we plan to do eye-tracking studies to identify areas of interest in the images.

## 5. CONCLUSION

We have begun development of a visual search observer for mass localization and detection in digital breast tomosythesis. We have introduced a 2D gradient template match to locate the areas of interest in a DBT reconstructed slice image. We have shown the VS observer to potentially improve the CNPW observer, because it includes a front-end to the CNPW that takes into account anatomical noise. Future work will involve making the VS observer more robust using eye-tracking experiments.

## Acknowledgments

This research was sponsored in part by NIH K25CA140858 and NIH/NIBIB R01-EB12070. The authors would like to thank Predrag Bakic for use of his anthropomorphic breast phantoms.

## REFERENCES

1. Das M, Gifford HC. Comparison of model-observer and human-observer performance for breast tomosynthesis: effect of reconstruction and acquisition parameters. Proceedings of SPIE. 2011:796118.
2. Das M, Connolly C, Glick SJ, Gifford HC. Effect of postreconstruction filter strength on microcalcification detection at different imaging doses in digital breast tomosynthesis: human and model observer studies. Proceedings of SPIE. 2012:831321.
3. Kundel HL, Nodine CF, Conant E, Weinstein SP. Holistic Component of Image Perception in Mammogram Interpretation: Gaze-tracking Study. Radiology. 2007; 242(2):396–402. [PubMed: 17255410]
4. Gifford H, Didier C, Das M, Glick S. Optimizing breast-tomosynthesis acquisition parameters with scanning model observers. Proceedings of SPIE. 2008; 6917:69170S.
5. Bakic PR, Zhang C, Maidment ADA. Development and characterization of an anthropomorphic breast software phantom based upon region-growing algorithm. Medical Physics. 2011; 38(6):3165. [PubMed: 21815391]
6. Johns PC, Yaffe MJ. X-ray characterization of normal and neoplastic breast tissues. Physics in medicine and biology. 1987; 32(6):675–695. [PubMed: 3039542]
7. Siddon R. Fast calculation of the exact radiological path for a three-dimensional CT array. Medical Physics. 1985; 12:252. [PubMed: 4000088]
8. Vedula AA, Glick SJ, Gong X. Computer simulation of CT mammography using a flat-panel imager. Proceedings of SPIE. 2003; 5030:349–360.
9. Hanley JA, McNeil BJ. The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. Radiology. 1982; 143:29–36. [PubMed: 7063747]
10. Gifford H, Kinahan PE, Lartizien C, King MA. Evaluation of Multiclass Model Observers in PET LROC Studies. IEEE Transactions on Nuclear Science. 2007; 54(1):116–123. [PubMed: 18392119]

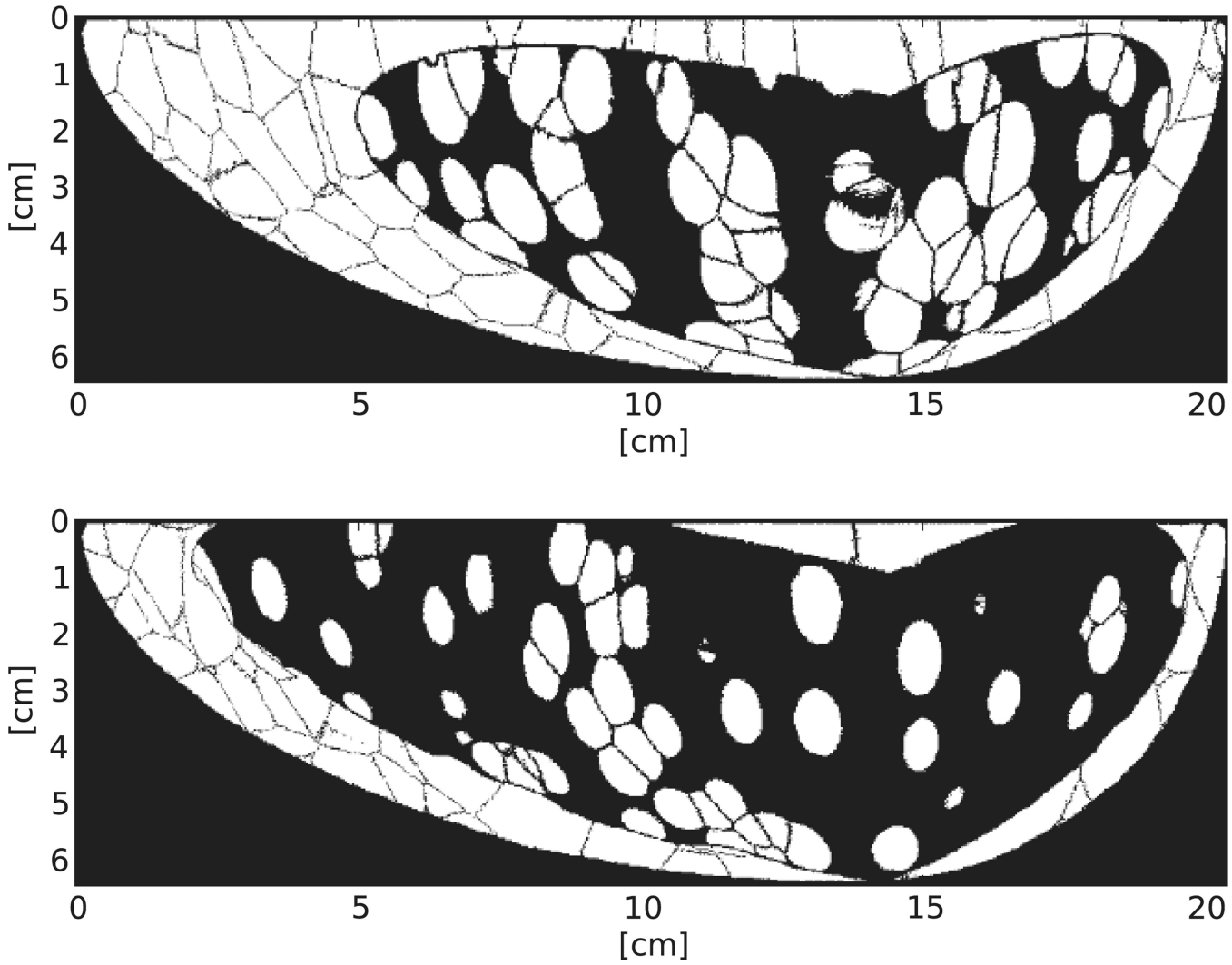


**Figure 1.**
Slice through anthropomorphic breast phantoms from Bakic *et al.*, with isotropic sampling distance equal to 0.2 mm. Top: low-density phantom, with 25% volumetric glandular fraction, bottom: high-density phantom, with 50% volumetric glandular fraction. White is adipose tissue, black is glandular tissue.

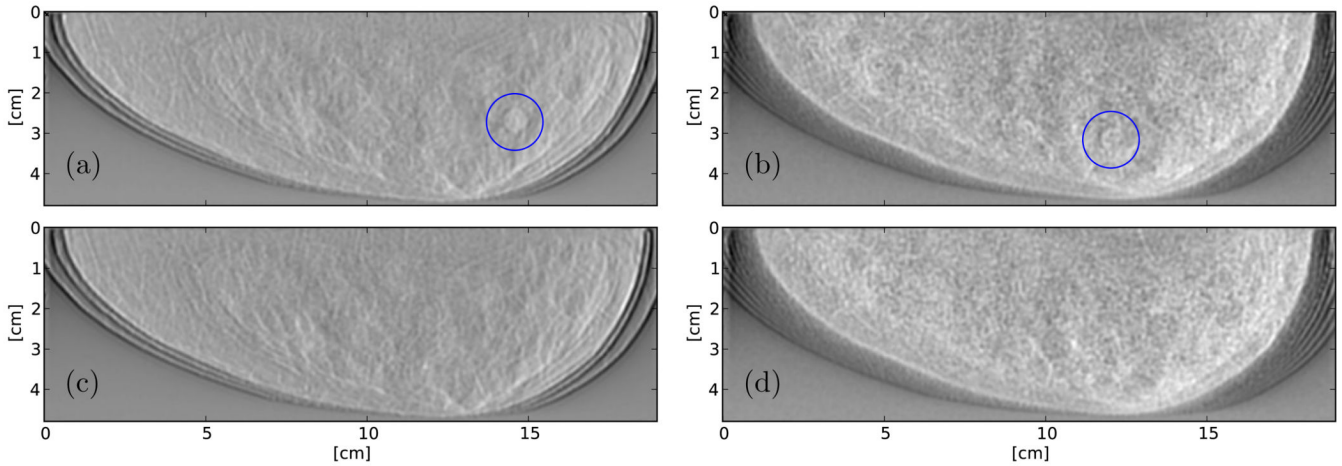


**Figure 2.**
Reconstructed slices through the Bakic phantom. Figures (a) and (b) are reconstructed slices centered at lesion locations, with 25% and 50% VGF, respectively. Figures (c) and (d) show the corresponding phantom slices without lesions.

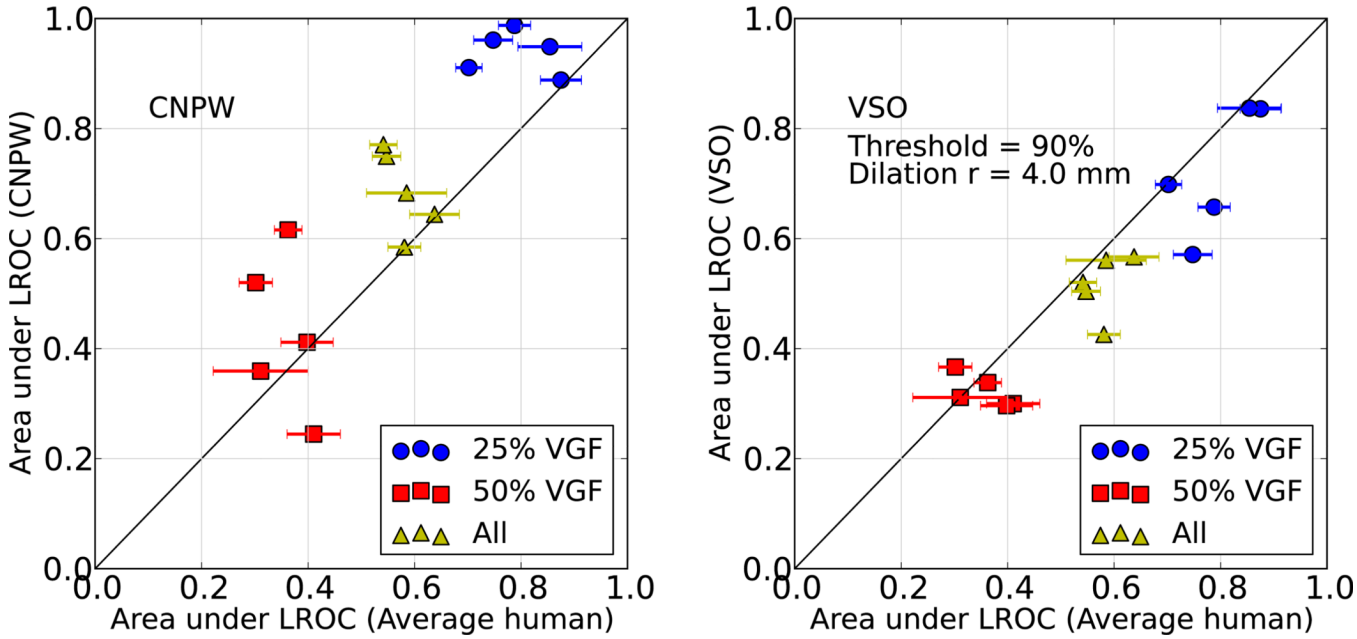


**Figure 3.**
Scatter plot comparing CNPW (left) and VS (right) observer performances with those of human observers.

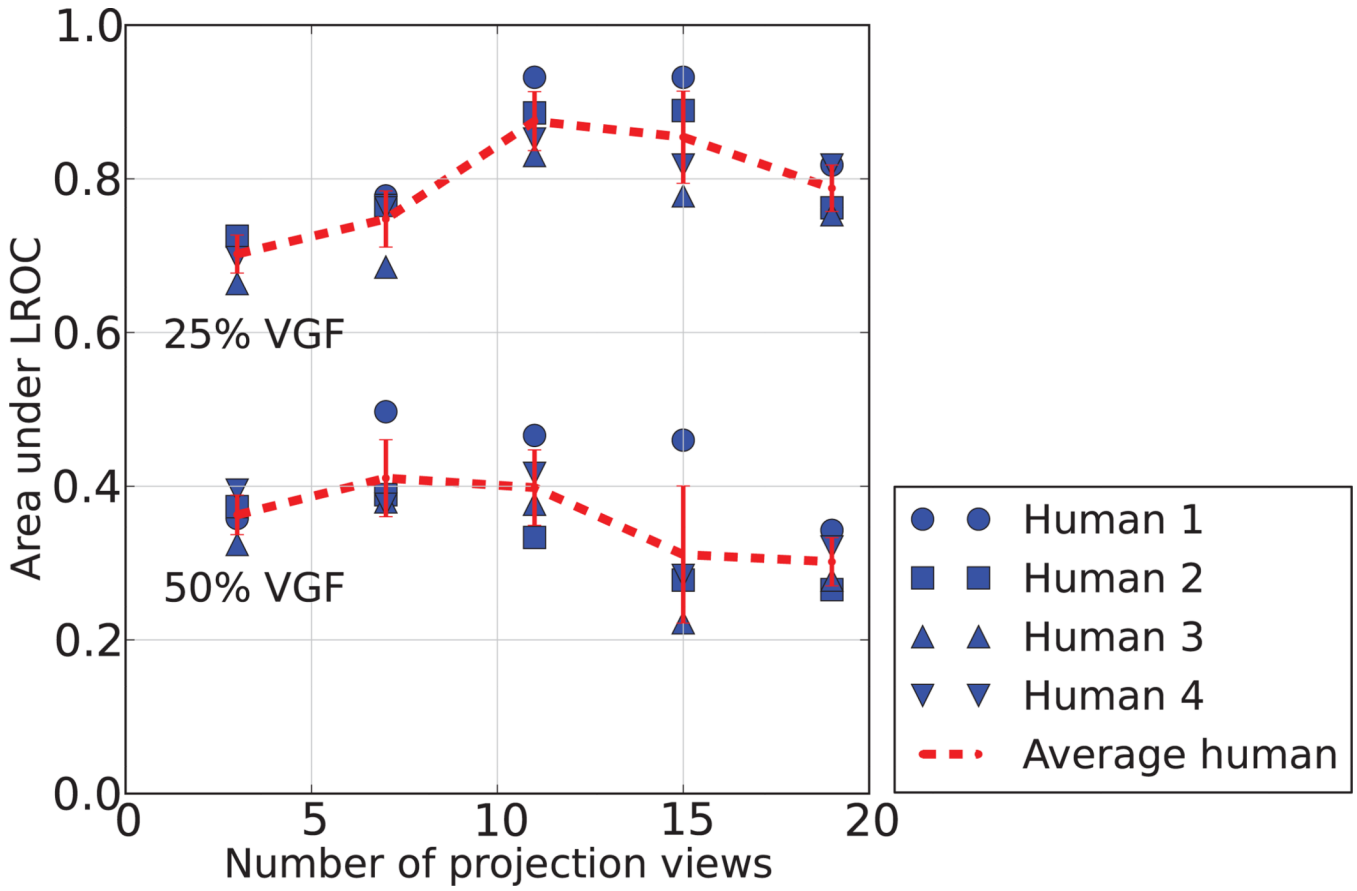


**Figure 4.**
Area under the LROC curve data for four human observers. Data is shown for phantoms with 25% and 50% volumetric glandular fraction, for five different acquisition parameter settings. For each acquisition parameter setting, the angular range was held constant at 60 degrees. The red dashed line shows the mean value, with one standard deviation.

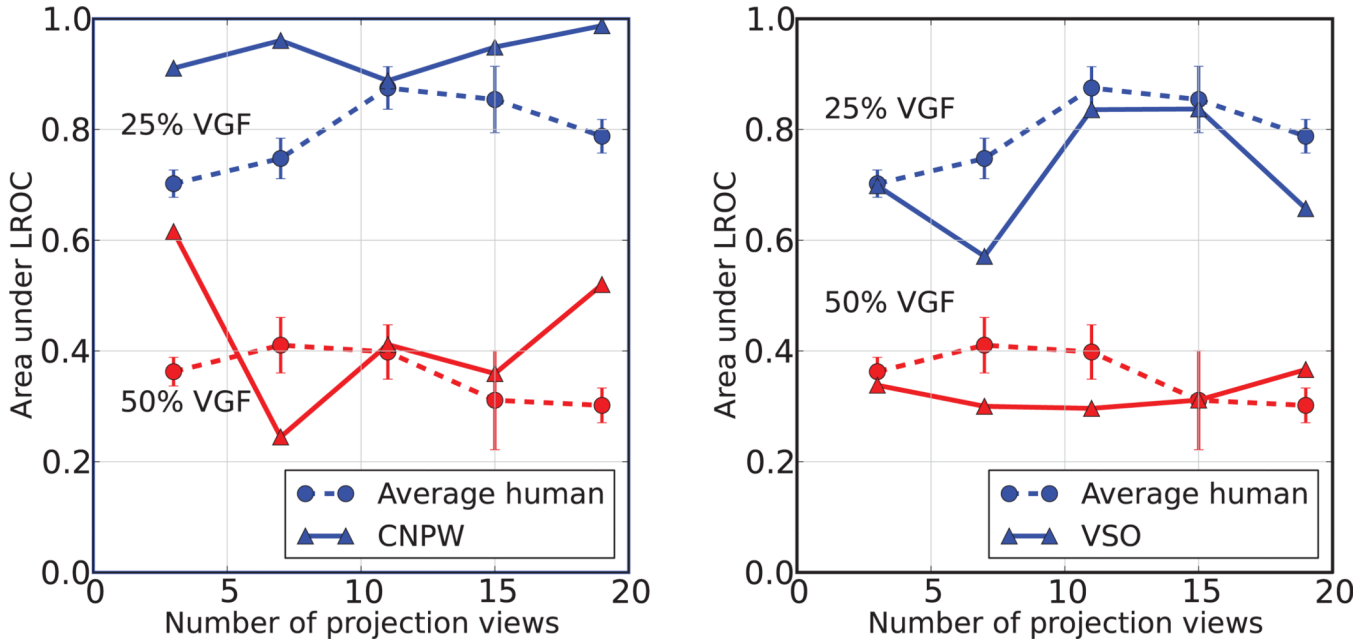


**Figure 5.**
Comparison of $A_L$ between the human observer and the CNPW (left) and the VS observer with thresholding equal to 90% of the maximum value and structuring element radius equal to 4.0 mm (right).

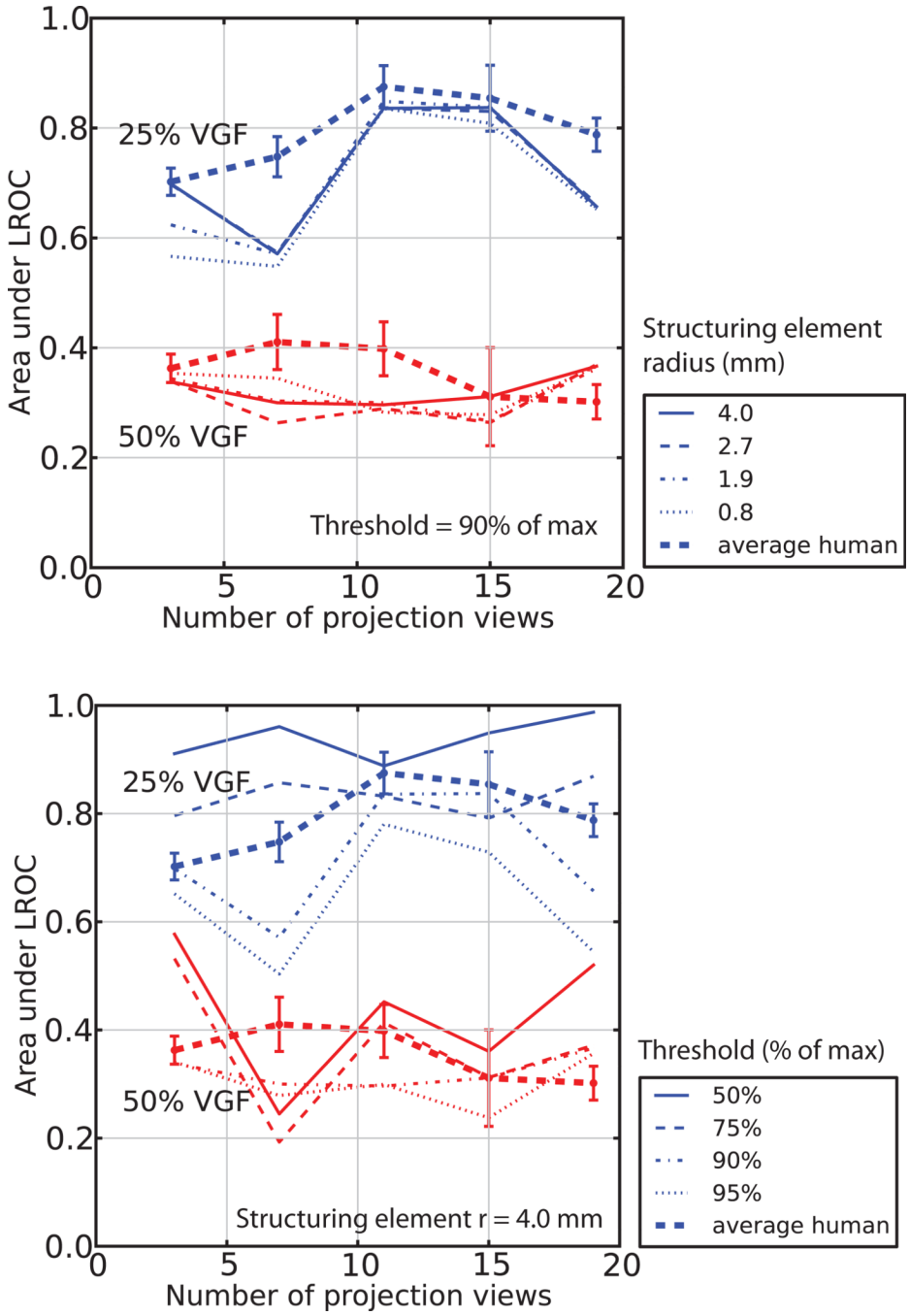


**Figure 6.**

Comparison of $A_L$ for the VS observer, using a range of threshold values and structuring element radii. Structuring element variation (shown on the left) is shown for radii between 0.8 and 4.0 mm, and threshold variation (on the right) is shown for values between 50% and 95% of the maximum pixel value.

**Table 1**

Definitions of ratings for human observer study.

| Scale | Definition |
|---|---|
| 4 | High confidence present |
| 3 | Low confidence present |
| 2 | Low confidence absent |
| 1 | High confidence absent |