Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain

Shuai Leng,^{a)} Lifeng Yu, and Yi Zhang

Department of Radiology, Mayo Clinic, 200 First Street Southwest, Rochester, Minnesota 55905

Rickey Carter

Department of Biostatistics, Mayo Clinic, 200 First Street Southwest, Rochester, Minnesota 55905

Alicia Y. Toledano

Biostatistics Consulting, LLC, 10606 Wheatley Street, Kensington, Maryland 20895

Cynthia H. McCollough

Department of Radiology, Mayo Clinic, 200 First Street Southwest, Rochester, Minnesota 55905

(Received 8 February 2013; revised 7 June 2013; accepted for publication 9 June 2013; published 10 July 2013)

Purpose: The purpose of this study was to investigate the correlation between model observer and human observer performance in CT imaging for the task of lesion detection and localization when the lesion location is uncertain.

Methods: Two cylindrical rods (3-mm and 5-mm diameters) were placed in a 35×26 cm torso-shaped water phantom to simulate lesions with -15 HU contrast at 120 kV. The phantom was scanned 100 times on a 128-slice CT scanner at each of four dose levels ($\text{CTDI}_{\text{vol}} = 5.7, 11.4, 17.1, \text{ and } 22.8$ mGy). Regions of interest (ROIs) around each lesion were extracted to generate images with signal-present, with each ROI containing 128×128 pixels. Corresponding ROIs of signal-absent images were generated from images without lesion mimicking rods. The location of the lesion (rod) in each ROI was randomly distributed by moving the ROIs around each lesion. Human observer studies were performed by having three trained observers identify the presence or absence of lesions, indicating the lesion location in each image and scoring confidence for the detection task on a 6-point scale. The same image data were analyzed using a channelized Hotelling model observer (CHO) with Gabor channels. Internal noise was added to the decision variables for the model observer study. Area under the curve (AUC) of ROC and localization ROC (LROC) curves were calculated using a nonparametric approach. The Spearman's rank order correlation between the average performance of the human observers and the model observer performance was calculated for the AUC of both ROC and LROC curves for both the 3- and 5-mm diameter lesions.

Results: In both ROC and LROC analyses, AUC values for the model observer agreed well with the average values across the three human observers. The Spearman's rank order correlation values for both ROC and LROC analyses for both the 3- and 5-mm diameter lesions were all 1.0, indicating perfect rank ordering agreement of the figures of merit (AUC) between the average performance of the human observers and the model observer performance.

Conclusions: In CT imaging of different sizes of low-contrast lesions (-15 HU), the performance of CHO with Gabor channels was highly correlated with human observer performance for the detection and localization tasks with uncertain lesion location in CT imaging at four clinically relevant dose levels. This suggests the ability of Gabor CHO model observers to meaningfully assess CT image quality for the purpose of optimizing scan protocols and radiation dose levels in detection and localization tasks for low-contrast lesions. © 2013 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4812430>]

Key words: human observer, model observer, lesion detection and localization, channelized Hotelling observer, LROC

1. INTRODUCTION

The dramatically increased use of CT continues to generate concerns regarding the cancer risks associated with the radiation exposure from CT.^{1–3} Although the magnitude of potential risk associated with low-dose exposures is controversial, and the benefit of a clinically justified CT scan is high, it remains the consensus that radiation dose levels should be as

low as reasonably achievable without sacrificing diagnostic performance. Optimizing CT protocols to achieve adequate diagnostic performance with the lowest reasonable dose has therefore become an important task.^{4,5} Quantitative methods to effectively accomplish this goal, however, are lacking. Currently, evaluation of patient images by interpreting radiologists is the most accepted approach to determining the lowest radiation dose levels for CT protocols.^{6,7} Because this

approach is extremely labor intensive and becomes unmanageable when multiple imaging parameters are to be optimized, there is a need to have more efficient methods that use objective quality metrics appropriate to clinical tasks to represent the performance of interpreting radiologists.

Recently, a large body of research has been performed dedicated to the objective assessment of image quality using model observers, which are mathematical models designed to make decisions on defined tasks based on statistical decision theory.^{8–10} Practical linear observers including the nonprewhitening matched filter (NPW) and Hotelling observer (prewhitening) have been widely studied.¹⁰ It has been demonstrated that each observer correlates well with human observers in certain noise and background scenarios, but not others.⁹ To address these issues, a channelized Hotelling observer (CHO) and a NPW with eye filter were proposed.^{9,11,12} In CHO, images are preprocessed with channel filters that are selected to represent the spatial frequency and orientation response of the human vision system. Different channel filters have been proposed such as square, difference of Gaussian, and Gabor channels.^{10,11,13,14} NPW with eye filter is an extension of the original NPW model, taking into account eye sensitivity at different frequencies.¹⁵ These model observers have been used to evaluate image quality, guide system design, and optimize image reconstruction and post processing in multiple imaging modalities including general radiography,^{14,16} mammography,^{17–20} PET and SPECT,^{21–25} MRI,²⁶ and tomosynthesis and cone beam CT.^{27–29} Despite their applications in many other modalities, very few studies have been performed on clinical CT systems,^{30–32} especially to investigate the performance correlation between human and model observers using real CT images. Most model observer studies have been performed with lesion (signal) properties fully known, i.e., a “signal known exactly” (SKE) task. In previous studies, we have investigated the correlation between human and model observer performance in diagnostic CT using a SKE task.³³ For clinical applications, however, the location of a lesion is unknown to the human observer and a visual searching procedure is required. This location uncertainty degrades the efficiency of lesion detection. There have been several studies using signals with unknown locations, however, mainly in x-ray imaging and emission tomography.^{34–37}

Therefore, the purpose of this study was to investigate the correlation between human observer and model observer performance in lesion detection and localization where the lesion location is unknown. Physical phantoms were scanned to generate signal-present and signal-absent images at different dose levels and for different low-contrast lesion sizes. Human and model observer performance studies were performed and results compared at each dose level and lesion size.

2. METHODS AND MATERIALS

In this study, we scanned a water phantom containing rods representing low-contrast lesions of different sizes. Repeated scans were acquired to generate images with signal present and signal absent that were presented to human and model observers. Human observer studies were performed by ask-

FIG. 1. Photograph of experimental setup. Two cylindrical rods having diameters of 3 and 5 mm (arrows) were attached perpendicularly to an acrylic block (arrow head). The rods/block were placed in the center of a 35 × 26 cm torso-shaped water phantom such that the rods were aligned parallel to the z axis of the scanner.

ing readers to determine the presence or absence of the “lesion” (rod) and identify the corresponding location. The same images were also evaluated with a scanning linear observer based on channelized Hotelling observer (CHO) with Gabor channel filters. ROC and localized ROC (LROC) analyses were performed for both human and model observers and the correlation between these two evaluated.³⁸ Internal noise was inserted into the model observer to represent the internal variability of decision making by human observers.

2.A. Phantom scans

A 35 × 26 cm torso-shaped water phantom was used to simulate the attenuation of a standard size adult abdomen. Two cylindrical rods with 3-mm and 5-mm diameters were inserted into the water phantom to mimic lesions. The rods were made of epoxy resin and had CT number of −9 HU at 120 kV. Iodine solutions were added to the water to generate approximately −15 HU contrast between the rods and the background (6 HU for background). The rods were attached perpendicularly to an acrylic block and the block placed in the center of the water phantom, positioned parallel to the scan plane (Fig. 1). This allowed the rods to be “suspended” in the center of the water phantom with their long axes aligned parallel to the z axis of the scanner (Fig. 1) in a manner similar to that used in our previous study.³³ This phantom was scanned on a 128-slice CT scanner (Definition Flash, Siemens Healthcare, Forchheim, Germany) in the single-source mode of operation, with the phantom centered at scanner isocenter. The scan range was selected to cover the length of the rods and a portion of the water phantom beyond the end of the rods (scanning only through water). The scanning parameters included: 120 kV, 0.5 s rotation time, 0.8 helical pitch, and 128 × 0.6 mm collimation (64 × 0.6 mm physical

FIG. 2. Sample signal-absent and signal-present CT images of the 3- and 5-mm diameter lesion mimicking rods at four different dose levels (CTDI_{vol} values of 5.7, 11.4, 17.1, and 22.8 mGy). The 3- and 5-mm diameter rods were randomly distributed at different locations within the images. All images were displayed with a window width and window level setting of [400, 40] HU.

collimation with z flying focal spot technique). Automatic exposure control (CareDose4D, Siemens Healthcare) was turned on and four dose levels acquired by selecting quality reference mAs values of 120, 240, 360, and 480. The corresponding CTDI_{vol} values were 5.7, 11.4, 17.1, and 22.8 mGy. Images were reconstructed using a medium sharp body kernel (B40) at a 5-mm image thickness. The reconstruction kernel and image thickness were the same as those used in a routine abdomen CT scan protocol in our practice.

2.B. Generating signal-absent and location-unknown, signal-present images

Signal-absent images and location-unknown, signal-present images were obtained from the same water phantom scans. Scans were repeated 100 times to measure the statistical variations in the resultant images at each dose level. Regions of interest (ROIs) with 128×128 pixels were generated around each cylindrical rod to produce signal images for each lesion size at each dose level. The locations of these ROIs were selected so that the relative location of the lesion (rod) was randomly distributed within the realizations of the ROI images, with a uniform probability function at each location. The random number was generated using a MATLAB function (Mathworks, Natick, MA). The same procedure was followed to generate signal-absent ROI images from images without lesion mimicking rods. A total of 50 realizations of the signal-absent images were generated from the 100 scans with randomly distributed locations.

These signal-present and signal-absent ROI images were used in the human and model observer studies described below. Samples of signal-present and signal-absent images of the 3- and 5-mm lesions at four dose levels are shown in Fig. 2. The location of the lesion was randomly distributed inside of each image.

2.C. Human observer studies

Three board certified medical physicists specializing in CT were recruited to independently perform this lesion detection and localization task. To assist the human observer studies, a MATLAB program with a graphic user interface was developed in our lab. After launching the program, a single image (either signal-present or signal-absent image) was displayed to the readers. The signal-present and signal-absent images were displayed in a random order with truth unknown to the readers. For images with signal, the location of signal for each image was also randomized and unknown to the readers. All images were displayed with a standard abdomen window (window width = 400 HU; window center = 40 HU).

For each image, observers were asked to determine the presence or absence of lesion signal, and to identify the location of signal by clicking the mouse when the cursor was located at the center of the perceived lesion. Readers were also asked to rate their confidence in their decision regarding lesion presence or absence using a 6-point ordinal scale (0–5, with 5 as the highest confidence of signal presence). Scores and identified lesion locations for each reader were automatically recorded by the MATLAB program and saved for analysis. Before the reading sessions, signal characteristics, i.e., size and contrast, were shown to all readers using training images. Original images (before ROI extraction) with both size lesions present were shown to the readers. For each dose level, four images were shown to the readers as training data, for a total of 16 images across the four dose levels. These images were not used in the subsequent human observer studies. A total of eight studies were performed, i.e., four dose levels (CTDI_{vol} = 5.7, 11.4, 17.1, and 22.8 mGy.) and two lesion sizes (3 and 5 mm). For each study, the same 150 images (100 with lesion and 50 without) were reviewed by each observer. Therefore, a total of 1200 images were reviewed by each of the three observers.

All studies were performed in a darkened room with consistent ambient lighting (<10 lux) and images were displayed on a monitor that was appropriately calibrated for clinical diagnosis following the ACR Technical Standard for Electronic Practice.³⁹ All three readers performed the study on the same monitor in the same room, with all images displayed in a fixed window width and window level of [400, 40] HU. The observers were asked to sit directly in front of the workstation at a distance of approximately 50–60 cm from the monitor. There was no time limit to review each image. All images were reviewed in multiple sessions, each being limited to a maximum of two hours to avoid fatigue.

ROC and LROC analysis was performed on the confidence scores from the human observer studies. For LROC analysis, a localization was deemed to be correct if the lesion was marked within a certain distance (localization radius) from its true location by the human observer. In this study, we plotted true localization fraction relative to localization radius and the smallest localization radius after which the true localization fraction plateaued was used for the LROC analysis.⁴⁰ Area under the ROC (A) and LROC (A_L) curves were estimated using nonparametric methods and used as figure of merit of the

human observer performance.^{40,41} Variance of A and A_L was also calculated for each reader. We then calculated the averaged area under the curve (AUC) of ROC and LROC and the variance across the three readers at each dose level and lesion size. For the calculation of variance across readers, we incorporated the correlation among the three readers caused by the fact that the same images were read.⁴⁰ Denoting A_1, A_2, A_3 , as the AUC values for the ROC curves from the three readers at a given dose level and lesion size, the averaged AUC ($\bar{A}$) was calculated as $\bar{A} = (A_1 + A_2 + A_3)/3$, and the variance of $\bar{A}$ as

$$\text{Var}(\bar{A}) = \frac{1}{9}(\sigma_{A_1}^2 + \sigma_{A_2}^2 + \sigma_{A_3}^2 + 2\text{cov}(A_1 A_2) + 2\text{cov}(A_2 A_3) + 2\text{cov}(A_1 A_3)), \quad (1)$$

where $\sigma_{A_i}^2$ is the variance of AUC for each reader and $\text{cov}(A_i, A_j)$ is the covariance of AUC between readers i and j , which was calculated using the nonparametric method proposed in Ref. 40. The standard deviation of AUC values across readers was calculated as the square root of the variance determined using Equation (1). The same method was applied to AUC of LROC estimation, with the replacement of A with A_L in the above equation.

2.D. Model observer studies

The same datasets were also analyzed using a model observer derived from CHO.¹¹ As a searching process is involved, the model observer is usually called a scanning-linear observer,⁴² or when considering the localization task as a classification problem, a multiclass observer.³⁵ Following the notations in Ref. 43, we use vector g to represent the ROI images with $N \times N$ pixels ($128 \times 128 = 16384$ in this study), g_s as image with signal present, and g_b as image with signal absent. Vectors of $\bar{g}_s$ and $\bar{g}_b$ represent the mean of signal-present and signal-absent images. The covariance matrices of signal-present and signal-absent images are K_s and K_b , with size of $N^2 \times N^2$ (16384×16384 in this study). For a linear model observer, the decision variable λ is calculated as an inner product of the observer template ω and the image vector g .¹⁰ For a detection task, the final decision of the model observer was achieved by comparing the decision variable λ with a preset threshold t . If $\lambda > t$, the target image is considered to have the signal present; if $\lambda < t$, the target image is considered to have the signal absent.

For a CHO model observer, each image vector is first processed with a set of channel profiles, with channel output of

$$g_c = V^t g, \quad (2)$$

where V is the matrix with columns representing the profiles of each channel. The decision variable is generated by a weighted linear combination of responses from all channels:⁹

$$\lambda = \omega_{\text{CHO}}^t g_c = \sum_{m=1}^M \omega_m g_{cm}, \quad (3)$$

where M is the total number of channels, and ω_{CHO} is the template, which can be obtained by

$$\omega_{\text{CHO}} = S_c^{-1} [\bar{g}_{sc} - \bar{g}_{bc}], \quad (4)$$

where $S_c = \frac{1}{2} [K_{sc} + K_{bc}]$ is the intraclass channel scatter matrix and $\bar{g}_{sc} = V^T \bar{g}_s$ and $\bar{g}_{bc} = V^T \bar{g}_b$ are the means of the channel outputs for images with and without signal, respectively. K_{sc} and K_{bc} are related to the image covariance matrices of K_s and K_b by $K_{sc} = V^T K_s V$, $K_{bc} = V^T K_b V$.

The new study, as signal location changed, the channel center (x_0, y_0) was changed accordingly to coincide with the lesion center. The intraclass channel scatter matrix and template were therefore generated for each signal location using the corresponding signal-present and signal-absent images and Eq. (4). After that, the template was applied to the test images to generate the decision variables. The same set of images was used for training the CHO and estimating the performance. This is consistent with one of the training-testing strategies described in Ref. 44. As signal location was unknown, the model observer was applied to each possible location inside the image ROI and corresponding decision variables were generated. In this study, the possible center location of the lesion was limited to be at least 15 pixels away from the ROI boundaries, the same range as that used in the ROI generation described in the previous section. Decision variables were calculated at each possible location using the corresponding channel profiles (presented below) and templates. The highest value of decision variable in all possible signal locations was selected to represent the final decision variable for a given test image, and the location generating the maximal response was considered as the signal location detected by the model observer.

$$\lambda = \max_{i \in \text{ROI}} (\lambda_i). \quad (5)$$

This scanning linear observer is similar to those derived in Refs. 35 and 42, except that a background dependent term was subtracted in those models. In this study, a uniform background was used, therefore the subtraction was not necessary. However, for studies with more complicated background, a subtraction of the background dependent term is necessary.

2.E. Channel profile selection

Channel profiles are selected to represent the spatial frequency response of the human visual system and multiple choices of channel filters have been proposed.^{10,11,13,14} In this study, Gabor channels were used as the channel profiles. Gabor channels have been demonstrated to be an effective way to represent the response of neurons in the primary visual cortex.^{14,45} The general form of the Gabor function can be expressed as

$$G_a(x, y) = \exp[-4(\ln 2)((x - x_0)^2 + (y - y_0)^2)/\omega_s^2] \cdot \cos[2\pi f_c((x - x_0) \cos \theta + (y - y_0) \sin \theta) + \beta], \quad (6)$$

where ω_s is channel width, f_c is central frequency, θ is the orientation, and β is a phase factor. In this study, we used four passbands, five orientations, and two phases (a total of 40 channels), the same parameters as those used in the study

by Wunderlich and Noo,³² with $\omega_s = 56.48, 28.24, 14.12$, and 7.06 pixels, $f_c = 3/128, 3/64, 3/32$, and $3/16$ cycles/pixel, $\theta = 0, 2\pi/5, 4\pi/5, 6\pi/5$, and $8\pi/5$ radians and $\beta = 0, \pi/2$.

2.F. Internal noise

Internal noise was added to the model observer analysis to simulate the internal variation of the human observer's decision procedure. It explains why different responses are generated when the same images are reviewed by the same human observers, and also the suboptimal performance of human observers compared with model observers. Different methods have been proposed to add internal noise to a model observer to match human observer performance, such as constant additive internal noise, internal noise proportional to the external noise, or internal noise added to the decision variable or channel output.^{46,47} In this study, internal noise was added to the decision variable λ , with the amplitude proportional to the standard deviation of decision variables from the signal-absent images (σ_{λ_bg}). This can be expressed using the following equation:

$$\lambda_{\text{Final}} = \lambda + \alpha \times \sigma_{\lambda_bg} \times \xi, \quad (7)$$

where α is a weighting factor and ξ representing a random number generated in the range of $(-1, 1)$. The weighting factor α was determined through a calibration procedure using the images of the 5-mm lesion that were scanned with $\text{CTDI}_{\text{vol}} = 11.4$ mGy. In this procedure, α values varied from 1 to 9 and AUC of ROC curves were calculated for model observer at each α and compared with that of human observer. The α value that generated the most similar AUC for the model and human observer was selected and subsequently used in all dose levels and lesion sizes.

2.G. ROC and LROC analysis for model observer

For each condition (lesion size and dose level), 150 decision variables were calculated from the 150 realizations (100 with lesion and 50 without). Comparing the decision variables from each image to a given threshold, true positive fraction (TPF) and false positive fraction (FPF) were obtained. In ROC analysis, true positive cases were signal-present images if the final decision variables were higher than the threshold, regardless of whether the highest decision variable was obtained at the correct lesion location or not. ROC curves were then generated by plotting the pairs of TPF (sensitivity) and FPF (1-specificity) while varying the threshold. AUC values were calculated using a nonparametric approach (integrating TPF with respect to FPF). A similar approach was used to generate LROC curves and calculate AUC of LROC. The only difference was that in the LROC analysis, signal-present images were only considered as true positive cases when the decision variable was higher than the threshold and the corresponding location was correctly coincident with the true signal location. A localization was deemed to be a true localization if maximal decision variable occurred within the distance of one localization radius from the true lesion location. We empirically chose the localization radius to be the same as the object ra-

FIG. 3. The weighting factor for internal noise of the model observer was determined by comparing model observer AUC values to those from human observers at one calibration condition. The weighting factor value was selected as the one that yielded the AUC value closest to that of the human observer.

dus, 5 pixels for the 5-mm diameter lesion and 3 pixels for the 3-mm diameter lesion (pixel size ~ 0.5 mm) in this study. For each scenario (dose and lesion size), a total of 200 realizations of the internal noise were generated and added to the decision variables as described in Eq. (7). For each realization, the variance of AUC for ROC and LROC curves was calculated using the same nonparametric procedure as that used in the human observer studies.^{40,41} The averaged value from the 200 realizations was used as the variance of AUC in one scenario (dose and lesion size).

2.H. Evaluation of the correlation between human and model observer performance

The AUC values from the human and model observers for both the ROC and LROC curves were plotted for the four dose levels and two lesion sizes to visually assess their agreement. The Spearman's rank order correlation between the average performance of the human observers and the model observer performance was also calculated for the AUC of both ROC and LROC curves for both the 3- and 5-mm diameter lesions.

3. RESULTS

As shown in Fig. 2, it became more challenging to differentiate signal-present images from signal-absent images as dose decreased due to the increase in image noise. Smaller lesions were generally more difficult to detect and correctly localize compared to larger lesions at the same dose level.

Figure 3 shows the comparison of the average performance of human observers and the model observer performance for a CTDI_{vol} of 11.4 mGy and a 5-mm lesion at different internal noise levels, where the value of α used in Eq. (7) varied from 1 to 9. It was found that $\alpha = 6$ provided the best match between the average AUC of the human observers and the AUC of the model. This value of α was then used for the model observer calculations at all other dose levels and lesion sizes.

All ROC study results are summarized in Tables I and II, in which AUC and its standard deviation for each individual reader at each dose level are reported for the 3- and 5-mm lesions, respectively. The averaged AUC across the three

TABLE I. AUC values (and standard deviation) of ROC curves for the human and model observer studies of the 3-mm diameter lesion.

CTDIvol (mGy)	22.8	17.1	11.4	5.7
Reader 1	0.994 (0.005)	0.916 (0.023)	0.584 (0.048)	0.579 (0.047)
Reader 2	0.995 (0.005)	0.886 (0.024)	0.606 (0.043)	0.521 (0.044)
Reader 3	0.996 (0.003)	0.916 (0.023)	0.573 (0.048)	0.569 (0.048)
All Readers	0.995 (0.003)	0.906 (0.017)	0.587 (0.030)	0.556 (0.032)
(Average)				
Model Observer	0.958 (0.014)	0.948 (0.016)	0.676 (0.045)	0.515 (0.050)

readers and its standard deviation are also reported and compared with the AUC and standard deviation of the model observer. Figure 4 presents the AUC averaged across the human observers at four dose levels and two lesion sizes. At the highest dose levels, i.e., 17.1, and 22.8 mGy, the 5-mm lesion was detected almost perfectly. AUC values decreased as CTDIvol decreased, falling to below 0.6 for the 3-mm lesion at CTDIvol values of 5.7 and 11.4 mGy, indicating that the detection performance approached that of random guessing at these low dose levels.

Figure 5 shows the true localization fraction relative to different localization radii. It can be observed that the true localization fraction plateaus for localization radii larger than 7 pixels for all conditions (lesion sizes and dose levels). Therefore, a value of 7 pixels was used as the localization radius in the LROC analysis of human observer studies.

Sample ROC and LROC curves from the model observer (e.g., at 11.4 mGy for the 3-mm lesion) are shown in Fig. 6. The LROC curve is below the corresponding ROC curve because correct localization was required in the LROC analysis but not in ROC analysis. Another difference between the curves is that ROC curves always reach the point of (1, 1), while this is not a necessary requirement for LROC curves.

All LROC study results are summarized in Tables III and IV, in which the AUC and its standard deviation for each individual reader at each dose level are reported for the 3- and 5-mm lesions, respectively. The averaged AUC across the three readers and its standard deviation are also reported and compared to those of the model observer.

The average AUC values across the three readers for the ROC and LROC curves are shown in Fig. 7. For each dose level and lesion size combination, AUC of LROC is always lower than that of ROC due to the localization requirement.

TABLE II. AUC values (and standard deviation) of ROC curves for the human and model observer studies of the 5-mm diameter lesion.

CTDIvol (mGy)	22.8	17.1	11.4	5.7
Reader 1	1.000 (0.000)	0.993 (0.004)	0.895 (0.025)	0.674 (0.041)
Reader 2	1.000 (0.000)	0.979 (0.010)	0.899 (0.022)	0.698 (0.037)
Reader 3	1.000 (0.000)	0.964 (0.013)	0.927 (0.020)	0.699 (0.042)
All Readers	1.000 (0.000)	0.979 (0.007)	0.907 (0.018)	0.690 (0.025)
(Average)				
Model observer	0.996 (0.003)	0.979 (0.009)	0.888 (0.026)	0.694 (0.044)

FIG. 4. AUC of ROC curves for human observer studies at four dose levels (CTDIvol = 5.7, 11.4, 17.1 and 22.8 mGy) and two lesion sizes (3 and 5 mm). The markers represent the average AUC values from three readers and error bars represent the 95% confidence interval.

FIG. 5. True localization fraction relative to different localization radii in human observer studies of the eight conditions (four dose levels and two lesion sizes), from which it can be observed that the true localization fraction plateaus for localization radii larger than 7 pixels. Data shown are averaged over three readers.

FIG. 6. ROC and LROC curves for the model observer analysis of the 3-mm diameter lesion-mimicking rod scanned at a CTDIvol value of 11.4 mGy.

TABLE III. AUC values (and standard deviation) of LROC curves for the human and model observer studies of the 3-mm diameter lesion.

CTDIvol (mGy)	22.8	17.1	11.4	5.7
Reader 1	0.984 (0.011)	0.851 (0.032)	0.348 (0.043)	0.105 (0.027)
Reader 2	0.990 (0.010)	0.778 (0.040)	0.302 (0.043)	0.076 (0.024)
Reader 3	0.996 (0.003)	0.840 (0.035)	0.310 (0.042)	0.125 (0.030)
All Readers	0.990 (0.005)	0.823 (0.026)	0.320 (0.035)	0.102 (0.020)
(Average)				
Model observer	0.930 (0.017)	0.911 (0.021)	0.473 (0.046)	0.053 (0.019)

Figures 8 and 9 present the comparison between human and model observer performance at the four dose levels and two lesion sizes. Figure 8 shows the comparison of ROC analyses and Fig. 9 shows the comparison of LROC analyses. In both analyses, AUC values for the model observer agreed well with the average values across the three human observers. Spearman's rank order correlation were all 1.0 for the same comparisons indicating perfect rank ordering agreement of the FOMs between the model and mean human observer performance.

4. DISCUSSION

Multiple model observers have been proposed to objectively evaluate image quality and optimize system design. For any model observer, studies are required to demonstrate its correlation with human observer performance before it can be used clinically. Given the substantial difference between imaging modalities (e.g., signal and noise properties), dedicated studies need to be performed for each particular modality. In this study, we investigated the correlation between human observers and a scanning linear observer based on CHO with Gabor filters for CT imaging. The imaging tasks were low-contrast detection and localization, with lesion location randomly distributed inside the image.

Although the majority of model observer studies focus on lesion detection with fixed location, lesion location is usually unknown in clinical practice. The search process adds another layer of complexity to the detection task, which affects the diagnostic performance for lesion detection. One limitation of the conventional ROC curve is that it does not take into account position information, which is considered in LROC analysis. An LROC curve always falls below the ROC curve for the same image data as correct localization is required in

TABLE IV. AUC values (and standard deviation) of LROC curves for the human and model observer studies of the 5-mm diameter lesion.

CTDIvol (mGy)	22.8	17.1	11.4	5.7
Reader 1	1.000 (0.000)	0.984 (0.011)	0.843 (0.033)	0.467 (0.046)
Reader 2	1.000 (0.000)	0.960 (0.020)	0.805 (0.038)	0.445 (0.047)
Reader 3	1.000 (0.000)	0.951 (0.017)	0.867 (0.031)	0.457 (0.046)
All Readers	1.000 (0.000)	0.965 (0.010)	0.839 (0.028)	0.456 (0.036)
(Average)				
Model observer	0.976 (0.003)	0.949 (0.013)	0.835 (0.031)	0.573 (0.046)

FIG. 7. Average AUC values across the three human observers for the detection (ROC curves) and localization (LROC curves) of the 3- and 5-mm diameter lesion-mimicking rods and at four different dose levels.

LROC analysis but not in ROC analysis. Therefore, AUC values of LROC are lower than those of ROC. In our study, the performance of a scanning linear observer based on CHO with Gabor channels was highly correlated with human observer performance in CT images acquired at clinically relevant dose levels, lesion contrast level, and lesion sizes when signal location was uncertain. This was demonstrated in both ROC and LROC analyses. Although the AUC values differed between ROC and LROC (the AUC for LROC was lower), the changes in performance were similar between the model observer and human observers.

Because signal location changed in the ROI images, there might have been some truncation for the Gabor channels when the lesion was close to the boundary of the ROI. The amount of truncation depends on the location and central frequency. More truncation was observed for signal locations closer to the boundary and for channels with lower central frequencies (data not shown). In this study, the size of the channel profile (128×128) was selected to be the same as for the size of the image and the center of the channel profile was selected to be consistent with the center of lesion. These channel profiles were applied to the same images as were shown to the human observers. When

FIG. 8. Human observer (averaged across three readers) and model observer performance at the four dose levels and two lesion sizes, with AUC values for ROC curves calculated at each dose level and lesion size. Error bars represent 95% confidence interval.

FIG. 9. Human observer (averaged across three readers) and model observer performance at the four dose levels and two lesion sizes, with AUC values for LROC curves calculated at each dose level and lesion size. Error bars represent 95% confidence interval.

the lesion was at locations close to the boundary, part of the channel profiles were truncated and hence did not contribute to the response. However, the image itself was “truncated” outside of the ROI boundary (128×128), the same as the Gabor channel profile, and this was reflected in the human observer study in that the human observer saw different amounts of background information when the lesion was close to the boundary compared to the scenario when the lesion was at the center of the ROI. We hypothesize therefore, that the model observer with truncated channel profile represents the scenario as for human observer studies. The effect of this truncation at the ROI boundary, however, merits more detailed study in the future.

This study has several limitations. First, the background used in this study is simplistic. We would like to start with the simplistic background in which the experiments are well controlled. After demonstrating the correlation between model and human observer in this simplistic background, we will investigate this correlation in more complex background in future studies. Second, the human and model observer performances are correlated as the same images were assessed. One method to take this into account is to construct confidence interval between the difference of human and model observer.⁴⁰ However, the human observers used an ordinal rating scale (0–5) and model observer used continuous rating scale (the spread of decision variables). Therefore we could not use this method due to the different rating scales. We did not perform a formal statistical test of the difference between human and model observer performance at a given condition (e.g., dose). Instead, we performed a correlation test between the two performances and showed high degree correlation at different conditions, such as dose levels and lesion sizes as in this study.

In summary, we have demonstrated a strong correlation between performance of human observer and a scanning linear observer based on CHO with Gabor channels in a lesion detection and localization task using real CT images from scans of a phantom simulating an adult abdomen. This suggests that the model observer might be used to quantitatively evaluate CT image quality, efficiently optimize CT protocols, and de-

termine the lowest radiation dose without sacrificing diagnostic performance. This enables a more efficient way to accomplish these tasks without performing labor intensive human observer studies. Our current study focused on detection of different sizes (3 and 5 mm) of low-contrast lesions in a uniform background at an unknown location. More complicated diagnostic tasks, such as lesion detection against a complex anatomical background,^{15,28} are under evaluation, as is the applicability of this method to images created using nonlinear iterative reconstruction methods or nonlinear noise reduction algorithms.

ACKNOWLEDGMENTS

This work was supported in part by NIH grant R01 EB071095 from the National Institute of Biomedical Imaging and Bioengineering. The authors would like to thank Dr. Mathew Kupinski from University of Arizona for his help on model observers, Mr. Mike Bruesewitz and Tom Vrieve for their assistance with data acquisition, Dr. Lingyun Chen and Dr. Juan C. Ramirez Giraldo for inspiring discussion, and Ms. Kristina Nunez and Amy Nordstrom for their assistance with paper preparation. Investigators interested in using the data described in this study should contact the authors.

^aAuthor to whom correspondence should be addressed. Electronic mail: leng.shuai@mayo.edu; Telephone: (507) 293-4233; Fax: (507) 266-1657.

¹D. J. Brenner and E. J. Hall, “Computed tomography—an increasing source of radiation exposure,” *N. Engl. J. Med.* **357**, 2277–2284 (2007).

²R. Smith-Bindman, J. Lipson, R. Marcus, K. P. Kim, M. Mahesh, R. Gould, A. Berrington de Gonzalez, and D. L. Miglioretti, “Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer,” *Arch. Intern. Med.* **169**, 2078–2086 (2009).

³A. Berrington de Gonzalez, M. Mahesh, K. P. Kim, M. Bhargavan, R. Lewis, F. Mettler, and C. Land, “Projected cancer risks from computed tomographic scans performed in the United States in 2007,” *Arch. Intern. Med.* **169**, 2071–2077 (2009).

⁴AAAPM CT Summit, “Scan parameter optimization,” (2010) (available URL: <http://www.aapm.org/meetings/2010CTS/default.asp>). Date accessed 1/10/2011.

⁵W. R. Hendee, G. J. Becker, J. P. Borgstede, J. Bosma, W. J. Casarella, B. A. Erickson, C. D. Maynard, J. H. Thrall, and P. E. Wallner, “Addressing overutilization in medical imaging,” *Radiology* **257**, 240–245 (2010).

⁶A. Apel, J. G. Fletcher, J. L. Fidler, D. M. Hough, L. Yu, L. S. Guimaraes, M. E. Bellemann, C. H. McCollough, D. R. Holmes, 3rd, and C. D. Eusemann, “Pilot multi-reader study demonstrating potential for dose reduction in dual energy hepatic CT using non-linear blending of mixed kV image datasets,” *Eur. Radiol.* **21**, 644–652 (2010).

⁷L. S. Guimaraes, J. G. Fletcher, W. S. Harmsen, L. Yu, H. Siddiki, Z. Melton, J. E. Huprich, D. Hough, R. Hartman, and C. H. McCollough, “Appropriate patient selection at abdominal dual-energy CT using 80 kV: Relationship between patient size, image noise, and image quality,” *Radiology* **257**, 732–742 (2010).

⁸International Commission on Radiation Units and Measurements, “Medical imaging – The assessment of image quality,” Report 54, (International Commission on Radiation Units and Measurements, Bethesda, MD, 1996), pp. 1–88.

⁹H. H. Barrett, J. Yao, J. P. Rolland, and K. J. Myers, “Model observers for assessment of image quality,” *Proc. Natl. Acad. Sci. U.S.A.* **90**, 9758–9765 (1993).

¹⁰J. Beutel, H. Kundel, and R. Van Metter, *Handbook of medical imaging: Physics and psychophysics* (SPIE, Bellingham, WA, 2000).

¹¹K. J. Myers and H. H. Barrett, “Addition of a channel mechanism to the ideal-observer model,” *J. Opt. Soc. Am. A* **4**, 2447–2457 (1987).

- ¹²A. E. Burgess, "Statistically defined backgrounds: performance of a modified nonprewhitening observer model," *J. Opt. Soc. Am. A Opt. Image Sci. Vis* **11**, 1237–1242 (1994).
- ¹³C. K. Abbey and H. H. Barrett, "Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability," *J. Opt. Soc. Am. A Opt. Image Sci. Vis* **18**, 473–488 (2001).
- ¹⁴M. Eckstein, J. Bartroff, C. Abbey, J. Whiting, and F. Bochud, "Automated computer evaluation and optimization of image compression of x-ray coronary angiograms for signal known exactly detection tasks," *Opt. Express* **11**, 460–475 (2003).
- ¹⁵M. P. Eckstein, C. K. Abbey, and J. S. Whiting, "Human vs model observers in anatomic backgrounds," *Proc. SPIE* **3340**, 16–26 (1998).
- ¹⁶S. Richard and J. H. Siewerdsen, "Comparison of model and human observer performance for detection and discrimination tasks using dual-energy x-ray images," *Med. Phys.* **35**, 5043–5053 (2008).
- ¹⁷A. E. Burgess, F. L. Jacobson, and P. F. Judy, "Human observer detection experiments with mammograms and power-law noise," *Med. Phys.* **28**, 419–437 (2001).
- ¹⁸A. S. Chawla, E. Samei, R. S. Saunders, J. Y. Lo, and J. A. Baker, "A mathematical model platform for optimizing a multiprojection breast imaging system," *Med. Phys.* **35**, 1337–1345 (2008).
- ¹⁹A. S. Chawla, E. Samei, R. Saunders, C. Abbey, and D. DeLong, "Effect of dose reduction on the detection of mammographic lesions: A mathematical observer model analysis," *Med. Phys.* **34**, 3385–3398 (2007).
- ²⁰M. L. Hill, J. G. Mainprize, and M. J. Yaffe, "An observer model for lesion detectability in contrast-enhanced digital mammography," *Digit. Mam-mogr.* **6136**, 720–727 (2010).
- ²¹P. Bonetto, J. Qi, and R. M. Leahy, "Covariance approximation for fast and accurate computation of channelized Hotelling observer statistics," *IEEE Trans. Nucl. Sci.* **47**, 1567–1572 (2000).
- ²²H. C. Gifford, M. A. King, D. J. de Vries, and E. J. Soares, "Channelized hotelling and human observer correlation for lesion detection in hepatic SPECT imaging," *J. Nucl. Med.* **41**, 514–521 (2000).
- ²³H. C. Gifford, R. G. Wells, and M. A. King, "A comparison of human observer LROC and numerical observer ROC for tumor detection in SPECT images," *IEEE Trans. Nucl. Sci.* **46**, 1032–1037 (1999).
- ²⁴S. Kulkarni, P. Khurd, I. Hsiao, L. Zhou, and G. Gindi, "A channelized Hotelling observer study of lesion detection in SPECT MAP reconstruction using anatomical priors," *Phys. Med. Biol.* **52**, 3601–3617 (2007).
- ²⁵C. Lartizien, P. E. Kinahan, and C. Comtat, "Volumetric model and human observer comparisons of tumor detection for whole-body positron emission tomography," *Acad. Radiol.* **11**, 637–648 (2004).
- ²⁶M. D. Tisdall and M. S. Atkins, "Using human and model performance to compare MRI reconstructions," *IEEE Trans. Med. Imaging* **25**, 1510–1517 (2006).
- ²⁷D. J. Tward, J. H. Siewerdsen, M. J. Daly, S. Richard, D. J. Moseley, D. A. Jaffray, and N. S. Paul, "Soft-tissue detectability in cone-beam CT: Evaluation by 2AFC tests in relation to physical performance metrics," *Med. Phys.* **34**, 4459–4471 (2007).
- ²⁸G. J. Gang, D. J. Tward, J. Lee, and J. H. Siewerdsen, "Anatomical background and generalized detectability in tomosynthesis and cone-beam CT," *Med. Phys.* **37**, 1948–1965 (2010).
- ²⁹S. Park, R. Jennings, H. Liu, A. Badano, and K. J. Myers, "A statistical, task-based evaluation method for three-dimensional x-ray breast imaging systems using variable-background phantoms," *Med. Phys.* **37**, 6253–6270 (2010).
- ³⁰P. F. Judy, R. G. Swensson, and M. Szulc, "Lesion detection and signal-to-noise ratio in CT images," *Med. Phys.* **8**, 13–23 (1981).
- ³¹K. L. Boedeker and M. F. McNitt-Gray, "Application of the noise power spectrum in modern diagnostic MDCT: Part II. Noise power spectra and signal to noise," *Phys. Med. Biol.* **52**, 4047–4061 (2007).
- ³²A. Wunderlich and F. Noo, "Image covariance and lesion detectability in direct fan-beam x-ray computed tomography," *Phys. Med. Biol.* **53**, 2471–2493 (2008).
- ³³L. Yu, S. Leng, L. Chen, J. M. Kofler, R. E. Carter, and C. H. McCollough, "Prediction of human observer performance in a 2-alternative forced choice low-contrast detection task using channelized Hotelling observer: Impact of radiation dose and reconstruction algorithms," *Med. Phys.* **40**, 041908 (9pp.) (2013).
- ³⁴M. P. Eckstein, B. Pham, and C. K. Abbey, "Effect of image compression for model and human observers in signal-known-statistically tasks," *Proc. SPIE* **4686**, 13–24 (2002).
- ³⁵H. C. Gifford, M. A. King, P. H. Pretorius, and R. G. Wells, "A comparison of human and model observers in multislice LROC studies," *IEEE Trans. Med. Imaging* **24**, 160–169 (2005).
- ³⁶L. M. Popescu and R. M. Lewitt, "Small nodule detectability evaluation using a generalized scan-statistic model," *Phys. Med. Biol.* **51**, 6225–6244 (2006).
- ³⁷Y. Zhang, B. Pham, and M. P. Eckstein, "Evaluation of JPEG 2000 encoder options: Human and model observer detection of variable signals in X-ray coronary angiograms," *IEEE Trans. Med. Imaging* **23**, 613–632 (2004).
- ³⁸R. G. Swensson, "Unified measurement of observer performance in detecting and localizing target objects on images," *Med. Phys.* **23**, 1709–1725 (1996).
- ³⁹ACR Electronic Practice Guideline, "ACR Technical standard for electronic practice of medical imaging," (2007) [See http://gm.acr.org/SecondaryMainMenuCategories/quality_safety/guidelines/med_phys/electronic_practice.aspx].
- ⁴⁰A. Wunderlich and F. Noo, "A nonparametric procedure for comparing the areas under correlated LROC curves," *IEEE Trans. Med. Imaging* **31**, 2050–2061 (2012).
- ⁴¹L. M. Popescu, "Nonparametric ROC and LROC analysis," *Med. Phys.* **34**, 1556–1564 (2007).
- ⁴²M. K. Whitaker, E. Clarkson, and H. H. Barrett, "Estimating random signal parameters from noisy images with nuisance parameters: Linear and scanning-linear methods," *Opt. Express* **16**, 8150–8173 (2008).
- ⁴³C. K. Abbey, F. Bochud, "Modeling visual detection tasks in correlated image noise with linear model observers," in *Handbook of Medical Imaging: Physics and Psychophysics*, edited by J. Beutel, H. Kundel, and R. Van Metter (SPIE, Bellingham, WA, 2000), Vol. 1.
- ⁴⁴H. H. Barrett and K. J. Myers, *Foundations of Image Science* (Wiley, Hoboken, NJ, 2004).
- ⁴⁵Y. Zhang, C. K. Abbey, and M. P. Eckstein, "Adaptive detection mechanisms in globally statistically nonstationary-oriented noise," *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **23**, 1549–1558 (2006).
- ⁴⁶A. E. Burgess and B. Colborne, "Visual signal detection. IV. Observer inconsistency," *J. Opt. Soc. Am. A* **5**, 617–627 (1988).
- ⁴⁷Y. Zhang, B. T. Pham, and M. P. Eckstein, "Evaluation of internal noise methods for Hotelling observer models," *Med. Phys.* **34**, 3312–3322 (2007).