Computational and human observer image quality evaluation of low dose, knowledge-based CT iterative reconstruction

Brendan L. Eck and Rachid Fahmi

Department of Biomedical Engineering, Case Western Reserve University, Cleveland, Ohio 44106

Kevin M. Brown, Stanislav Zabic, and Nilgoun Raihani

Philips Healthcare, Cleveland, Ohio 44143

Jun Miao

Department of Biomedical Engineering, Case Western Reserve University, Cleveland, Ohio 44106

David L. Wilson^{a)}

Department of Biomedical Engineering, Case Western Reserve University, Cleveland, Ohio 44106

and Department of Radiology, Case Western Reserve University, Cleveland, Ohio 44106

(Received 5 September 2014; revised 22 June 2015; accepted for publication 6 August 2015; published 30 September 2015)

Purpose: Aims in this study are to (1) develop a computational model observer which reliably tracks the detectability of human observers in low dose computed tomography (CT) images reconstructed with knowledge-based iterative reconstruction (IMRTM, Philips Healthcare) and filtered back projection (FBP) across a range of independent variables, (2) use the model to evaluate detectability trends across reconstructions and make predictions of human observer detectability, and (3) perform human observer studies based on model predictions to demonstrate applications of the model in CT imaging.

Methods: Detectability (d') was evaluated in phantom studies across a range of conditions. Images were generated using a numerical CT simulator. Trained observers performed 4-alternative forced choice (4-AFC) experiments across dose (1.3, 2.7, 4.0 mGy), pin size (4, 6, 8 mm), contrast (0.3%, 0.5%, 1.0%), and reconstruction (FBP, IMR), at fixed display window. A five-channel Laguerre–Gauss channelized Hotelling observer (CHO) was developed with internal noise added to the decision variable and/or to channel outputs, creating six different internal noise models. Semianalytic internal noise computation was tested against Monte Carlo and used to accelerate internal noise parameter optimization. Model parameters were estimated from all experiments at once using maximum likelihood on the probability correct, P_C . Akaike information criterion (AIC) was used to compare models of different orders. The best model was selected according to AIC and used to predict detectability in blended FBP-IMR images, analyze trends in IMR detectability improvements, and predict dose savings with IMR. Predicted dose savings were compared against 4-AFC study results using physical CT phantom images.

Results: Detection in IMR was greater than FBP in all tested conditions. The CHO with internal noise proportional to channel output standard deviations, Model-k4, showed the best trade-off between fit and model complexity according to AIC_c. With parameters fixed, the model reasonably predicted detectability of human observers in blended FBP-IMR images. Semianalytic internal noise computation gave results equivalent to Monte Carlo, greatly speeding parameter estimation. Using Model-k4, the authors found an average detectability improvement of 2.7 ± 0.4 times that of FBP. IMR showed greater improvements in detectability with larger signals and relatively consistent improvements across signal contrast and x-ray dose. In the phantom tested, Model-k4 predicted an 82% dose reduction compared to FBP, verified with physical CT scans at 80% reduced dose.

Conclusions: IMR improves detectability over FBP and may enable significant dose reductions. A channelized Hotelling observer with internal noise proportional to channel output standard deviation agreed well with human observers across a wide range of variables, even across reconstructions with drastically different image characteristics. Utility of the model observer was demonstrated by predicting the effect of image processing (blending), analyzing detectability improvements with IMR across dose, size, and contrast, and in guiding real CT scan dose reduction experiments. Such a model observer can be applied in optimizing parameters in advanced iterative reconstruction algorithms as well as guiding dose reduction protocols in physical CT experiments. © 2015 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4929973>]

Key words: image quality, model observers, iterative reconstruction

1. INTRODUCTION

Computed tomography (CT) is a key medical imaging modality across a broad range of clinical applications. Through 2011, the number of CT exams increased sharply, due to an increase in the number of indications and particularly usage in the emergency room.^{1,2} In recent years in the US, there have been well over 60×10^6 exams per year.³ With increased use, ionizing radiation dose from CT has become a concern.⁴⁻⁶ In fact, CT now contributes the greatest *per capita* medical radiation exposure of all medical imaging procedures.⁷ Fortunately, CT iterative reconstruction algorithms enable lower dose scanning protocols for clinical use.⁸⁻¹¹ Advanced iterative reconstruction algorithms continue to develop, such as model-based¹² and knowledge-based iterative reconstruction,^{13,14} which may further reduce dose. In addition to dose reduction in static, single-volume procedures, advanced iterative reconstruction will encourage the use of CT for dynamic procedures such as cardiac CT perfusion.¹⁵⁻¹⁷

There are several studies aimed at determining the ability to use low dose protocols with iterative reconstruction. This includes a number of clinical studies.¹⁸⁻²⁰ Most assess image noise, contrast-to-noise ratio (CNR), signal-to-noise ratio (SNR), and radiologist ratings with paired analysis or Likert scale. However, it is well known that CNR is insufficient for quantifying medical image quality. Detectability is a firmly established task-based image quality metric which is considered superior to physical measurements such as CNR or SNR.^{21,22} Experimental assessment of detectability requires multiple human observers to review hundreds of images in M-alternatives forced choice (M-AFC) or receiver operating characteristic (ROC) presentations.²² Image data can be clinical,²³ synthetic, or phantom.²⁴ M-AFC is often used in phantom experiments where variables can be tightly controlled. With regards to CT iterative reconstruction, such quantitative image quality studies by us^{25,26} and others have suggested that considerable dose savings are indeed possible, with potential dose reductions up to 86%.²⁷

Detectability studies by trained observers enable quantitative evaluations of image quality. However, the substantial labor precludes usage in large-scale optimization studies.

An alternative is to use computational observers.²⁸⁻³¹ A widely used model in medical imaging is the channelized Hotelling observer (CHO). The CHO models the human observer using spatial frequency-selective filters, or channels, and optimal linear statistics to build a template for signal detection.^{28,32} Channels are selected corresponding to approximations of human perception and appropriateness for signal shape.³³ CHO generally captures overall trends in human detectability, which is useful in the analysis and optimization of CT system parameters, image processing, or reconstruction.^{34,35} However, the CHO largely overestimates human detection,^{36,37} indicating that there is an inefficiency in human observers not accounted for by the CHO.

To fit medical imaging, human detection experiments, it is common to add internal noise to model observers. The addition of equivalent internal noise to a model observer degrades its detection estimate, accounting for inefficiencies in human perception.^{38,39} Early reports in the vision science literature suggest two approaches for including internal noise: additive constant^{40,41} and signal-dependent multiplicative^{42,43} (or proportional⁴⁴) noise. These internal noise models have been added to the CHO in medical imaging to find detection thresholds in x-ray mammography,⁴⁵ optimize JPEG compression settings in coronary angiograms,^{46,47} and optimize cardiac SPECT reconstruction parameters.^{48,49} There are a number of considerations regarding the CHO and internal noise modeling. In x-ray angiograms, the choice of internal noise approach depends upon the CHO filters used.⁴⁷ In cardiac SPECT images, it has been shown that a proportional channel noise best matches human observers.⁵⁰ In CT, internal noise models have only recently been added to the CHO for comparisons to human observers.^{51,52} Because CT has very different characteristics than other medical images, the choice of internal noise model is not obvious. In addition, advanced iterative reconstruction algorithms have much different characteristics than filtered back projection (FBP) (Fig. 1). This motivates us to develop an internal noise model with the CHO which can track human detection in both FBP and iterative reconstruction CT images. Yu *et al.* found that a CHO with a proportional decision variable internal noise model reasonably fits human observers for hybrid iterative reconstruction images.⁵² We will develop a

FIG. 1. Sample images of low dose FBP and IMR image reconstructions from the same projection data. (Left) Standard FBP reconstruction of a low dose abdominal CT scan. (Right) IMR reconstruction. IMR maintains image features while greatly reducing noise. Image characteristics are significantly different in the two reconstructions.

CHO with internal noise for a different class of reconstruction, knowledge-based iterative reconstruction.

In this work, we develop a CHO with an appropriate internal noise model for comparing detectability in low dose CT images reconstructed with FBP and an advanced, knowledge-based iterative reconstruction algorithm, IMR™ (Philips Healthcare). We address details in internal noise implementation which improve model agreement with human observers and test it in a number of applications. CHO model observers with multiple internal noise models were compared against human detection experiments. In the process of optimizing the internal noise model, we compared CHO template formation with and without internal noise. To reduce the computation time for parameter estimation in the tested models, we use a semianalytic approach. Model parameter estimates were obtained using maximum likelihood (ML) which is advantageous compared to the conventional least squares. In particular, maximum likelihood allows for the exact use of the experimental probability density function and enables the use of the Akaike information criterion⁵³ (AIC) as a means to evaluate models of different orders. To stress the model, we considered a range of independent variables, including x-ray dose, size, contrast, and reconstruction. We tested the model with the lowest AIC on blended FBP and IMR images and used it to analyze trends in detectability improvements with IMR. We also use this observer to predict dose savings with IMR and verify with an observer study in physical CT images. Such a model observer, matched to human performance in both conventional FBP and advanced iterative reconstruction, could serve in large-scale reconstruction optimization studies with rapid, automated evaluation of detectability.

2. THEORY

2.A. Channelized Hotelling observer

We used the CHO with six internal noise models (Fig. 2).^{28,32,51,54} Inputs are subimage patches containing a

disk signal plus background noise or background noise only. Five Laguerre–Gauss channels ($L = 5$) were used in the CHO. Such channels are appropriate for rotationally symmetric signals,³³ such as the ones being used in this study. Each image is passed through the channel filters to obtain channel outputs (C_o), a matrix of size $L \times N$ where L is the number of channels and N is the number of images. We refer to the channel outputs for signal and noise images as $C_{O,S}$ and $C_{O,N}$, respectively. For a given image, I , the channel output, $C_{O,I}$, is an $L \times 1$ vector whose entries are multiplied by channel weights and summed to yield the decision variable, λ ,

$$\lambda = \sum_{\ell=1}^L w_{\ell} C_{O,I}^{(\ell)}, \quad (1)$$

where l is the channel index. Channel weights are determined by the Hotelling process which builds the optimal template according to linear statistics.²⁸ Briefly, channel outputs are prewhitened by their covariance and weighted by their relative SNR. A channel with greater SNR thus receives a greater weight. A channel with more noise receives less weight. The weights are calculated by

$$w = S_c^{-1}(\bar{C}_{O,S} - \bar{C}_{O,N}), \quad (2)$$

where S_c is the $L \times L$ intraclass scatter matrix and $\bar{C}_{O,S}$ and $\bar{C}_{O,N}$ are $L \times 1$ channel output means for the signal and noise images, respectively. The scatter matrix is calculated by

$$S_c = \frac{1}{2}(\text{cov}(C_{O,S}) + \text{cov}(C_{O,N})), \quad (3)$$

where $\text{cov}(C_{O,S})$ and $\text{cov}(C_{O,N})$ are $L \times L$ channel output covariance matrices. Detectability (d') is then calculated from the distribution of signal and noise decision variables as follows:

FIG. 2. Schematic of the channelized Hotelling observer with internal noise. For each condition, 200 signal images and 600 noise images are inputs for the model. Each image is correlated with Laguerre–Gauss channel filters to yield scalar channel outputs. Channel weights are determined by the Hotelling process. Channel outputs are summed to give a decision variable for each image. Decision variable distributions yield detectability (d') and the corresponding probability correct (P_C) for 4-alternative forced choice (4-AFC) which can be compared to human observer experiments. Internal noise models add noise to the channel outputs and/or to the decision variable.

$$d' = \frac{\bar{\lambda}_S - \bar{\lambda}_N}{\sqrt{\frac{1}{2}\sigma_{\lambda,S}^2 + \frac{1}{2}\sigma_{\lambda,N}^2}}, \quad (4)$$

where $\sigma_{\lambda,S}^2$ is the decision variable variance for signal images, $\sigma_{\lambda,N}^2$ is the decision variable variance for noise images, and $\bar{\lambda}_S$, $\bar{\lambda}_N$ refer to the means of the signal and noise decision variables, respectively. This is the CHO without internal noise.

Six internal noise models were developed with noise added either to decision variables, λ_S and λ_N , or to channel outputs, $C_{O,S}$ and $C_{O,N}$, or to both (Fig. 2). Considered models are as follows:

- Model-k1.** Noise is proportionally added to the signal or noise decision variables as a linear function of their respective standard deviations, $\sigma_{\lambda,S}$ and $\sigma_{\lambda,N}$: $\sigma_{\lambda,S}^{in} = k_1\sigma_{\lambda,S}$ or $\sigma_{\lambda,N}^{in} = k_1\sigma_{\lambda,N}$.
- Model-k2.** Added noise is constant for both signal and noise decision variables: $\sigma_{\lambda,S}^{in} = k_2$ or $\sigma_{\lambda,N}^{in} = k_2$.
- Model-k1k2.** Noise is a linear function of the standard deviation of the signal or the noise decision variable: $\sigma_{\lambda,S}^{in} = k_1\sigma_{\lambda,S} + k_2$ or $\sigma_{\lambda,N}^{in} = k_1\sigma_{\lambda,N} + k_2$.
- Model-k3.** Noise is uniformly added to all channel outputs: $\sigma_{C_o}^{in} = k_3\mathbf{I}_L$, with $\mathbf{I}_L$ being the $L \times 1$ identity vector and $\sigma_{C_o}^{in}$ an $L \times 1$ vector containing the noise standard deviation added to each C_o .
- Model-k4.** Noise is proportionally added to signal and noise channel outputs: $\sigma_{C_o}^{in} = k_4\sqrt{\text{diag}(\mathbf{S}_c)}$, where $\mathbf{S}_c$ is the scatter matrix given by Eq. (3).

Model-k1k2k3k4.

Noise is added to both the decision variable (signal or noise) and to the channel outputs: $\sigma_{\lambda,S}^{in} = k_1\sigma_{\lambda,S} + k_2$ or $\sigma_{\lambda,N}^{in} = k_1\sigma_{\lambda,N} + k_2$ for the decision variable and $\sigma_{C_o}^{in} = k_3\mathbf{I}_L + k_4\sqrt{\text{diag}(\mathbf{S}_c)}$ for the channel outputs.

We can compute the effect of internal noise in two ways: using Monte Carlo (MC) generated noise or semianalytically added noise. For Monte Carlo generated noise, noise is added to the decision variable by $\lambda_{\text{CHOIN}} = \lambda_{\text{CHO}} + \lambda_{\text{IN}}$, where λ_{IN} is sampled from the normal distribution, $N(0; \sigma_{\lambda}^{in})$, and to the channel outputs by $C_{O,\text{CHOIN},l} = C_{O,\text{CHO},l} + \varepsilon_{\text{IN},l}$, where $\varepsilon_{\text{IN},l}$ is sampled from $N(0; \sigma_{C_o,l}^{in})$ for channel l . This is repeated for each channel and each image. In this approach, each image has hundreds or thousands of noise realizations, and we compute detectability from the decision variable distribution after the addition of internal noise. The time required for Monte Carlo sampling is practical for a single calculation of the model, but it becomes impractical when one is doing iterative parameter estimation when very many model calculations are required. To reduce the time needed to add internal noise, we used a semianalytic approach. In the semianalytic approach, known internal noise statistics are used to compute d' . The internal noise variance is added to the decision variable variance in Eq. (4). For decision variable internal noise, we simply use the terms described in the noise models above ($\sigma_{\lambda,S}^{in}$ and $\sigma_{\lambda,N}^{in}$). For channel output noise, the total internal noise variance to be added is dependent on channel weights,

$$(\sigma_{C_o,\text{total}}^{in})^2 = \sum_{l=1}^L w_l^2 (\sigma_{C_o,l}^{in})^2. \quad (5)$$

With both the decision variable and channel output internal noise variances, one can calculate the CHO d' degraded by internal noise, considering both signal and noise distributions independently,

$$d' = \frac{\bar{\lambda}_S - \bar{\lambda}_N}{\sqrt{\frac{1}{2}(\sigma_{\lambda,S}^2 + (\sigma_{\lambda,S}^{in})^2 + (\sigma_{C_o,\text{total}}^{in})^2) + \frac{1}{2}(\sigma_{\lambda,N}^2 + (\sigma_{\lambda,N}^{in})^2 + (\sigma_{C_o,\text{total}}^{in})^2)}}. \quad (6)$$

The detectability, d' , can be converted to a probability correct, P_C , which corresponds to a 4-AFC experiment and thus compared to human observer data.²² In converting d' to P_C , we assume a Gaussian distribution of decision variables which we observed from the CHO using FBP and IMR images.

Since channel weights are determined by the channel output covariance [see Eqs. (1) and (2)], there are two ways to determine them. One way is that system “training” uses weights computed with internal noise present. That is, noise is included in the covariance calculation.⁵⁰ Another is that the system trains as if there is no internal noise, and noise is excluded from the covariance calculation. We compared

both methods to human observers in order to find which more closely models the human observer training process.

We tested that the semianalytic approach yields the same P_C as the highly sampled Monte Carlo approach. We used 1, 10, 100, 1000, and 10000 noise realizations (per image) and recorded the mean and standard deviation of P_C over 50 runs of the model for a given input image set.

We correct for the known phenomenon of observer lapse where some trials are missed due to a lapse in observer attention and not because of difficulties in perception.⁵⁵ If not accounted for, observer lapse can significantly bias parameter estimates due to overpenalization of high P_C conditions. We

determine the observer lapse rate, $\lambda_{\text{lapse}} = 0.0025$, using the average P_C from the 8 mm, 0.5% contrast IMR condition in which all signals were easily detectable. This is in the range of reported values from 0 to 0.05 for trained observers.⁵⁵ The model observer P_C , with observer lapse correction, is given by

$$P_{C,\text{corrected}} = \gamma + (1 - \gamma - \lambda_{\text{lapse}}) \frac{P_{C,\text{uncorrected}} - \gamma}{1 - \gamma}, \quad (7)$$

where $\gamma = 0.25$ for a 4-AFC experiment. We used corrected values in all parameter estimation results.

2.B. Parameter estimation

We estimated model parameters using ML, which has advantages as compared to typically used least squares (see Sec. 5). We compared every noise model to the average human probability correct results, which we denote P_C , from 4-AFC sessions across individual conditions. The likelihood function was formulated using the binomial probability density function which underlies P_C from 4-AFC experiments. The likelihood of obtaining $k_j = P_{C,\text{human},j} * N$ correct selections for condition j given the model probability $P_{C,j}$ for N trials is given by

$$P(k_j; N, P_{C,j}) = \binom{N}{k_j} P_{C,j}^{k_j} (1 - P_{C,j})^{N - k_j}. \quad (8)$$

The likelihood of obtaining average human 4-AFC results across all J conditions is given by

$$P_{\text{total}} = \prod_{j=1}^J P(k_j; N, P_{C,j}). \quad (9)$$

As is common, we maximize the log likelihood as given below,

$$P_{\text{total},\log} = \sum_{j=1}^J \log(P(k_j; N, P_{C,j})). \quad (10)$$

Finally, we can formulate the cost function via the negative log-likelihood,

$$C = -P_{\text{total},\log} \quad (11)$$

which we minimized using the Nelder–Mead simplex algorithm.⁵⁶ We compared ML results to unweighted and weighted least-squares fitting to P_C as well as unweighted least squares fitting to d' .

The likelihood obtained from ML fits allows one to compute the Akaike information criterion and assess trade-offs between model complexity and goodness of fit. The AIC_c ,⁵³ AIC corrected for a finite number of samples as in this work, is given by

$$AIC_c = 2p + 2C + \frac{2p(p+1)}{n-p-1}, \quad (12)$$

with p being the number of parameters and n the number of conditions. By penalizing the addition of parameters, AIC_c provides a good quality of fit metric for comparing models of different orders.

3. EXPERIMENTAL METHODS

3.A. Phantom and images

CT images from numerical simulations and physical CT scans were used in observer studies. The CT simulator was developed at Philips Healthcare and enables the creation of controlled, reproducible images from a virtual phantom consisting of a homogeneous water cylinder with iodine-containing cylindrical insets of different sizes, contrasts, and spatial locations. The simulator accounts for many CT system properties including x-ray source trajectory, detector arrangement, bowtie filter, finite focal spot size, and a Poisson photon noise model. For the phantom size and dose levels tested, the statistics are dominated by Poisson noise and away from the range where electronic noise and photon starvation are noticeable. Considerations regarding electronic noise are addressed in Sec. 5. The simulator does not include nonideal effects such as beam hardening and scatter, but in a physical scan these effects could be corrected for this circular water phantom with very low-contrast inserts. Physical CT scans were obtained using the MITA IQ LC phantom (model CCTI83) and an iCT scanner (Philips Healthcare).

Images were reconstructed using FBP and IMR. For the simulation, detector collimation was $0.625 \text{ mm} \times 64$, reconstructed slice width was 1 mm, field of view was 250 mm, and in-plane resolution $0.49 \times 0.49 \text{ mm}$. FBP reconstructions used a standard adult body kernel. The IMR algorithm uses a system model, noise statistics, and image regularization to maximize the useful signal from the projection data while minimizing noise effects. A prototype IMR version was used in this study with parameters similar to those found in the IMR “Soft Tissue L3” setting which targets optimal low-contrast detectability through aggressive image noise regularization. The same parameters were used for all IMR reconstructions of simulated data. In the physical CT scans, detector collimation was $0.625 \text{ mm} \times 128$, reconstructed slice width was 0.8 mm, field of view was 220 mm, and in-plane resolution $0.43 \times 0.43 \text{ mm}$. FBP reconstructions used a standard adult body kernel and the IMR reconstructions used the Soft Tissue L3 setting, similar to the simulation but appropriate for the iCT system.

The simulated CT phantom is shown in Fig. 3. It is 15 cm in diameter and contains 12 pins of varied contrast and size, labeled #1–12. Reconstructed images were of size 512×512 with mean background HU equal to zero. Signal plus background subimage patches consisting of 69×69 pixels centered on the pin were used in both human and model observer studies. Background subimage patches were obtained at the corresponding locations in slices without pins. For the human observer dose experiment, pin #6 (6 mm, 0.5% contrast) was used at three CTDIvol dose levels: 4.0, 2.7, and 1.3 mGy. For each dose, 200 signal and 600 noise images were generated. For the size experiment, dose and contrast were fixed at 1.3 mGy and 0.5%, respectively, and pins #5 (4 mm), #6 (6 mm), and #7 (8 mm) were used. For the contrast experiment, dose was fixed at 1.3 mGy and size was fixed at 6 mm. Pins #2 (0.3%), #6 (0.5%), and #10 (1.0%) were used. The results from pin #6 (6 mm, 0.5%) at 1.3 mGy were used

FIG. 3. Phantom used for detection studies with sample 4-alternative forced choice trials. (Left) The simulated phantom has 12 cylindrical signals at varied diameters (mm) and contrasts (%). Signals appear circular in the axial slice. Starting from 12 o'clock and moving clockwise, they are pin #1 (4 mm 0.3%), pin #2 (6 mm 0.3%), pin #3 (8 mm 0.3%), pin #4 (15 mm 0.3%), pin #5 (4 mm 0.5%), pin #6 (6 mm 0.5%), pin #7 (8 mm 0.5%), pin #8 (15 mm 0.5%), pin #9 (4 mm 1.0%), pin #10 (6 mm 1.0%), pin #11 (8 mm 1.0%), and pin #12 (15 mm 1.0%). (Center) A 4-AFC trial for FBP is shown with one signal image and three noise images. (Right) A 4-AFC trial for IMR is shown. All 4-AFC experiments used a window width of 100 and level of 0 HU. Contrast has been enhanced in these images for display purposes.

for the size and contrast experiments. This gives a total of 14 unique conditions for the simulated CT phantom studies.

For the physical phantom observer study, similar scans were performed. Scans were obtained at CTDIvol values of 2, 4, and 10 mGy. The dose was higher in these physical scans due to the larger phantom than the one used in simulations, but the image noise level was similar. Signal plus background subimage patches were 115×115 pixel patches centered on the 10 mm, 0.3% contrast pin. The background subimage patches were taken from the blank section of the MITA phantom. For each dose condition, 300 scans were obtained. These images were used for the physical CT scan observer studies.

3.B. Human observer study

We used 4-AFC experiments to assess human observer detectability.^{22,57} Five trained observers with normal or corrected-to-normal vision performed experiments under an IRB-exempt protocol. Experiments were performed in a darkened room using a Barco[®] MDRC1119 perceptually linear display. Observers performed 4-AFC experiments for each dose, size, and contrast condition for both FBP and IMR reconstructions. In each 4-AFC trial, one signal and three noise subimages were shown in randomized configurations (Fig. 3). A window width of 100 HU and level of 0 HU were used across all conditions. Each subject performed a 200-trial 4-AFC experiment for one IMR condition and one FBP condition in an experimental session. The order of FBP and IMR displays, as well as experimental condition (dose, size, or contrast), was randomized. Subjects performed each condition twice, giving a total of 14 experiment sessions. For each condition, P_C was obtained. Observers were blind to experimental conditions and hypotheses to reduce any bias. To minimize learning effects, all observers performed two full training blocks prior to performing experiments. During experiment blocks, 20 training trials preceded the 200-trial data collection. Fatigue was minimized by allowing subjects to take breaks at any point. Recall was assumed to be minimal given the number of conditions and images. To account for inter- and intraobserver variability in these

studies, we represent the population by its average P_C and standard deviation.

The physical phantom observer study differed slightly from the simulated images. Instead of 200 trials in the 4-AFC experiment, there were 100 trials. Observers performed each condition twice with both FBP and IMR at both dose levels (2 and 10 mGy), and also with IMR at 4 mGy. All other experimental conditions were the same.

3.C. Assessment of blended images

To balance the texture of FBP with the low noise of iterative reconstruction images, FBP and iterative reconstructed images have been blended in the previous studies.⁵⁸ We evaluated detectability in such blended images as a representative case of image processing. Blended images were generated by a weighted average of FBP and IMR reconstructions,

$$f_{\text{Blended}} = w \cdot f_{\text{IMR}} + (1 - w) \cdot f_{\text{FBP}}, w \in [0, 1]. \quad (13)$$

For the human observer 4-AFC study, pin #6 (6 mm, 0.5% contrast) at 1.3 mGy dose was used. The value of w was varied from 0 to 1 at 0.2 increments. These blended images were evaluated first by the model observer with the lowest AIC_c using optimized parameters from the dose, size, and contrast studies, then by human observers. Detection (P_C) of the human observers was compared to that of the model.

3.D. Detectability improvement and dose reduction

Using the model observer with the lowest AIC_c , we evaluated trends in detectability improvement with IMR. Model d' estimates were obtained for all 12 pins at four dose levels (1.3, 2.7, 3.3, 4.0 mGy). Trends in IMR detectability improvement over FBP for a given signal were quantified using the detectability ratio,

$$R_{d'} = d'_{\text{IMR}} / d'_{\text{FBP}}. \quad (14)$$

This ratio enables direct comparison of improvements from IMR to the same signal in FBP. A higher $R_{d'}$ value indicates that IMR yields greater detection improvements than the corresponding FBP reconstruction. $R_{d'}$ was evaluated for

each size and contrast in order to determine trends in IMR detectability improvement.

Using the model observer with the lowest AIC_c , we evaluated potential dose reductions with IMR. To predict dose savings, we used an interpolation/extrapolation function for d' across independent variables, particularly dose. While the Rose model⁵⁹ is applicable for detection in white noise, unlike correlated noise in CT, it provides a rough extrapolation function under the assumption of predominantly Poisson statistics in our dose range. The Rose model calculates detectability by $d'_{\text{Rose}} = C\sqrt{A}\sqrt{n}$, where C is contrast, A is area, and $\sqrt{n}$ is the pixel SNR for a Poisson process. Assuming that n , the number of photon counts, is proportional to dose, and C and A are constants, we obtain $d'_{\text{Rose}} = m\sqrt{\text{dose}}$ where m is a constant. We add a y -intercept parameter, b , which yields better fits and is typically negative predicting minimum dose/contrast values for detection. The result is

$$d' = m\sqrt{\text{dose}} + b. \quad (15)$$

We use this linear model and fit free parameters, m and b , to model d' values as a function of dose for each pin size and contrast combination and each reconstruction. In total, 24 independent curves were fit. With each fitted curve, we can then approximate d' at a lower dose value. Goodness of fit was assessed by the correlation coefficient.

With this simple linear model, we calculated dose reductions with IMR. For each signal, we used the d' from 4.0 mGy FBP as a target for IMR. Using the linear fit described above, we calculated the reduced dose IMR that gave the target detectability (equivalent to 4.0 mGy FBP). We performed this comparison with each of the 12 size and contrast conditions.

Following the predicted dose reductions from the model in simulated CT images, we performed a short observer study using physical phantom scans with similar dose reductions. We used the predicted dose reductions from the linear model in simulated images and compared detectability of low dose IMR images to high dose FBP images. Detectability from

human observer experiments and the model observer was compared in these physical scans.

4. RESULTS

We evaluated detectability as a function of dose, size, contrast, and reconstruction algorithm. At the lowest dose, 1.3 mGy, the 6 mm, 0.5% contrast pin was lost to noise in FBP but was clearly evident in IMR (Fig. 4). Individual observers showed a clear trend of improved detection as dose increased from 1.3 to 4.0 mGy for both reconstructions. Observers 4 and 5 showed similar detection compared to observers 1–3 in the IMR conditions, but much reduced detection in FBP, perhaps being more influenced by image noise. Detection was significantly higher with IMR compared to FBP at all doses, with 1.3 mGy IMR yielding higher detection than 4.0 mGy FBP for all observers (Fig. 5). This indicates more than 67% dose savings are possible with IMR. CHO models with internal noise parameters were fitted to all experiment sessions simultaneously: dose, size, contrast, and reconstruction. P_C in all observers was found to increase monotonically with contrast and size (Fig. 6). Similarly, detection increased with dose for all observers (Fig. 7). In all conditions, IMR provided significantly improved detection over FBP. While each CHO showed the same trend in detection, the CHO with internal noise proportional to channel output standard deviations, Model-k4, had the lowest AIC and showed reasonable agreement with human observer results (see Figs. 6 and 7).

Changes in internal noise sampling and model training improved fits to human experiments. Training the model to determine channel weights without channel output noise improved likelihood values (Fig. 8). This change caused Model-k4 and Model-k1k2k3k4 to improve and outperform Model-k1 in terms of AIC_c . With a large number of noise realizations, the MC internal noise model converged to the semianalytic model result (Table I). From experiments, we

FIG. 4. FBP and IMR sample images from dose experiment. Signal images are shown with dose decreasing from left to right, 4.0 to 1.3 mGy, with (top) FBP reconstruction and (bottom) IMR reconstruction. Noise becomes more aggressive as dose is reduced. Note the drastic difference in noise characteristics between FBP and IMR reconstructed images.

FIG. 5. Subject performance across reconstruction and dose conditions. Five subjects performed 4-AFC experiments on a 6 mm, 0.5% contrast target at reduced dose levels for both FBP and IMR reconstructions. Probability correct (P_C) is reported with error bars indicating observers' standard deviations. From left to right: (1.3 mGy FBP), (2.7 mGy FBP), (4.0 mGy FBP), (1.3 mGy IMR), (2.7 mGy IMR), and (4.0 mGy IMR). Detection increased with dose and also with the use of IMR as compared to FBP, with 1.3 mGy IMR yielding greater detection than 4.0 mGy FBP. Error bars indicate standard deviation of observers' P_C .

noticed that for higher P_C values, fewer MC noise realizations were required. For lower P_C cases (FBP, 6 mm, 0.5%, 1.3 mGy), more realizations, in excess of 10 000, were required. The semianalytic calculation ranged from 5 to 500 times faster than the MC approach, reducing the time required for parameter estimation.

We compared cost functions for parameter optimization including negative log-likelihood (maximum likelihood), least squares, and weighted least squares in the P_C domain, and least squares in the detectability domain. Table II shows the lowest cost after internal noise parameter optimization to human observer experiments for each cost function. AIC_c from ML fits are also given. The observer lapse correction improved fits to data as compared to uncorrected model outputs, with an average cost function reduction of 5%. Using the observer lapse correction slightly reduced internal noise

parameter values, but did not change relative comparisons of models. Model-k4 is the optimal model following from the AIC_c criterion. We, therefore, use Model-k4 for subsequent analyses. The P_C results corresponding to the set of optimal parameters obtained using ML are shown in Figs. 6 and 7.

Figure 9 shows the results of the blended FBP-IMR model and human observer study. The model shows good agreement with human observers across this range of weights and image characteristics.

Using Model-k4, we observed trends in the detectability ratio, $R_{d'}$ [see Eq. (14)], as a function of size and contrast as shown in Fig. 10. $R_{d'}$ was found to remain relatively constant across contrast, with average $R_{d'} = 2.7 \pm 0.4$. Similarly, we found $R_{d'}$ to remain relatively constant across dose, with $R_{d'} = 2.6 \pm 0.3$, 2.7 ± 0.4 , 2.9 ± 0.4 , and 2.8 ± 0.7 for 1.3, 2.7, 3.2, and 4.0 mGy, respectively. The value of $R_{d'}$ did increase

FIG. 6. Size and contrast experiment PC for human observer average and model observers. (Top) Results of FBP and (bottom) IMR detection studies in conditions of varied size and contrast at 1.3 mGy dose. The average of all five human observers in the 4-AFC experiment is shown. Models are organized by increasing AIC_c from left to right. Error bars indicate standard deviation of human observers and standard deviation of the model from the binomial distribution. IMR outperformed FBP within all conditions, including the smallest size and the lowest contrast signals.

FIG. 7. Dose experiment P_C for human observer average and model observers. P_C across model observers is compared within dose and reconstruction conditions. Error bars show standard deviation of human P_C and standard deviation from the binomial distribution for model P_C . CHO with six internal noise models, indicated by their parameters, is compared using optimal parameters from ML fitting. Note that detection in IMR experiments is high, near saturation above 1.3 mGy.

with size, from $R_{d'} = 2.4$ at 4 mm to $R_{d'} = 3.3$ at 15 mm, implying that IMR gives more improvement in detection with larger signals.

Potential dose reductions with IMR were assessed using the model observer. The proportional Rose model with an added y -intercept term showed a good linear relationship with the square root of dose, with correlation coefficient $R^2 = 0.95 \pm 0.08$. With these results, we extrapolated IMR detectability at each size and contrast condition to low dose in order to estimate equivalent 4.0 mGy FBP detection (Fig. 11). We analyzed all pins except one low outlier to predict a 0.8 ± 0.2 mGy dose with IMR, corresponding to approximately 82% dose savings. Larger pins (15 mm) enabled lower dose, 0.6 ± 0.1 mGy, as compared to smaller pins (4 mm), 1.1 ± 0.1 mGy.

The prediction of 82% dose savings was tested with physical CT scans. Figure 12 shows CT images of the full FOV image and subimage patches of high dose FBP (10 mGy), high dose IMR (10 mGy), low dose FBP (4 mGy), and low dose IMR (2 mGy). FBP images at 2 mGy were not evaluated as the signal was virtually indistinguishable from noise. Figure 13 shows the results of this study, as well as results at 4 mGy FBP, 4 mGy IMR, and 10 mGy IMR. Both the human and model observer show greater detection at 2 mGy IMR than 10 mGy FBP, verifying at least an 80% dose reduction. The model closely tracks FBP detectability in these physical images but underestimates IMR detectability, although the trend is captured.

5. DISCUSSION

Our observer studies showed that IMR improved detection over FBP for observers under all tested conditions. We found that IMR was able to improve detection in even the smallest pin (4 mm), the lowest contrast (0.3% contrast), and the lowest dose (1.3 mGy), indicating that small detail, low contrast, and high noise are not problematic for the reconstruction. In the dose study, IMR gave greater detection at 1.3 mGy than FBP at 4.0 mGy, suggesting at least 67% dose savings. Overall, the CHO models with internal noise predicted the same dose savings of over 67% across all 12 size and contrast combinations. From our experiments, we believe it is unlikely that FBP will outperform IMR at the lower limits of size, contrast, and dose.

The addition of internal noise models to the CHO improved their match to human data. Without internal noise, CHO was able to capture appropriate trends in detectability across dose,

FIG. 8. Model error for channel weights computed with and without channel output internal noise. For channel output internal noise models, the optimal negative log-likelihood values are compared for channel weights determined with and without channel output internal noise. When channel noise output is not added to the channel output covariance matrix, the model shows less error as compared to human observer experiments. Negative log-likelihood for Model-k4 with internal noise in channel weight calculation is 248.

TABLE I. Semianalytic P_C calculation compared to MC P_C corresponding to 1, 10, 100, 1000, and 10000 realizations per image using Model-k4. Semianalytic yields the same model outputs as highly sampled MC.

	Semianalytic	MC-1	MC-10	MC-100	MC-1000	MC-10000
Probability correct (P_C)						
FBP 6 mm, 0.5%, 4.0 mGy	0.8037 $\pm$ 0	0.8018 $\pm$ 0.0190	0.8043 $\pm$ 0.0072	0.8046 $\pm$ 0.0020	0.8046 $\pm$ 0.0011	0.8044 $\pm$ 0.0010
IMR 6 mm, 0.5%, 4.0 mGy	0.9990 $\pm$ 0	0.9991 $\pm$ 0.0004	0.9990 $\pm$ 0.0001	0.9990 $\pm$ 0.0000	0.9990 $\pm$ 0.0000	0.9990 $\pm$ 0.0000
FBP 6 mm, 0.5%, 1.3 mGy	0.5741 $\pm$ 0	0.5760 $\pm$ 0.0262	0.5750 $\pm$ 0.0074	0.5735 $\pm$ 0.0028	0.5740 $\pm$ 0.0004	0.5741 $\pm$ 0.0000
IMR 6 mm, 0.5%, 1.3 mGy	0.9165 $\pm$ 0	0.9154 $\pm$ 0.0120	0.9155 $\pm$ 0.0034	0.9157 $\pm$ 0.0013	0.9162 $\pm$ 0.0007	0.9165 $\pm$ 0.0000
Average time (ms)						
FBP 6 mm, 0.5%, 4.0 mGy	1.3	1.9	3.7	25.2	266.7	2645.0
IMR 6 mm, 0.5%, 4.0 mGy	1.3	1.9	3.8	25.5	272.4	2551.2
FBP 6 mm, 0.5%, 1.3 mGy	1.3	1.8	3.8	24.9	271.3	2493.2
IMR 6 mm, 0.5%, 1.3 mGy	1.3	1.8	3.7	25.0	266.0	2602.7

size, contrast, and reconstruction, but estimated P_C values up to 0.26 greater than human observers. Adding internal noise resulted in much better fits to human performance. Parameters proportional to the decision variable variance or channel output variance (i.e., k_1 and k_4 , respectively) most improved model fit to human observers. The improvement afforded by these terms could be due to increased flexibility when considering both low noise IMR and high noise FBP reconstructions. This is in agreement with previous findings,⁵¹ where proportionality internal noise terms were similarly important. Comparing internal noise models, Model-k4 is the optimal model following from the AIC_c criterion, but Model-k1k2k3k4 had the best fit to data, considering all cost functions. If one is concerned about overfitting, Model-k4 would be the best choice. However, we note that using Model-k1k2k3k4 would not have significantly changed results or conclusions. Most of the tested internal noise models fit the human observer data reasonably well except Model-k3 which performed poorly across all objective functions. In fact, the optimal value of k_3 approached zero and converged to the CHO with no internal noise.

Details in channel weight calculation and internal noise addition impacted model results. We found that by training the channel weights without internal noise, the models better agreed with human experiments. This suggests that human observers are able to develop a signal template for detection without internal noise despite having noise in the decision process. As for internal noise computations, we found that the

semianalytic approach gave the same results as Monte Carlo sampling. While both can be used in practice, the semianalytic approach saves significant computation time when estimating model parameters with iterative optimization.

There are observations concerning internal noise parameter estimation. Fitting across multiple variables reduces the likelihood of bias in the model. Fitting to limited conditions in our data causes over- or underestimation of detectability in other conditions. For instance, fitting Model-k1k2k3k4 only to the 8 mm, 0.5% contrast pin in FBP would lead to underestimation of detection in nearly all other conditions. The selection of cost function impacts model fitting, producing different optimal internal noise parameters. Optimizing internal noise parameters using least squared error does not account for measurement variability as a function of P_C , particularly at high values near $P_C = 1$. In fact, this may cause least squares to generate biased parameter estimates. Weighted least squares also does not fully address the problem of the underlying experimental probability function. Maximum likelihood exactly uses the underlying binomial distribution of the 4-AFC experiment, giving an improved estimation procedure. When using ML, we found that observer lapse⁵⁵ should be accounted for, especially when considering P_C values near saturation. The use of ML allows one to analyze the AIC and compare models of different orders. Of course, one can also change the channel filters in the CHO (Gabor, difference of Gaussians, band pass, etc.), which may influence the optimal internal noise

TABLE II. Minimum error from parameter optimization with different cost functions. Cost functions used are in the probability correct (P_C) domain using the negative log-likelihood (maximum likelihood, P_C -ML), weighted least-squares (P_C -WLS), least-squares (P_C -LS), and in the detectability domain using least-squares (d' -LS). AIC_c was calculated from P_C -ML. Italicized values indicate the lowest cost for each cost function.

	P_C -ML	P_C -WLS	P_C -LS	d' -LS	AIC_c
CHO	337.21	359.16	0.646	29.94	337.21
Model-k1	21.21	29.71	0.025	3.94	23.54
Model-k2	22.88	22.22	0.043	3.45	25.21
Model-k1k2	19.23	9.64	0.020	3.00	24.32
Model-k3	336.54	358.80	0.645	29.88	338.87
Model-k4	17.70	11.95	0.015	3.16	20.03
Model-k1k2k3k4	17.57	6.36	0.014	2.90	30.01

FIG. 9. Human and model observer performance in blended FBP-IMR images. Human observer 4-AFC experiments performed on blended FBP-IMR images with weight increments of 0.2 showed good agreement with Model-k4 predictions. Error bars indicate standard deviation of human observer P_C values and of the standard deviation obtained from the binomial distribution for the model observer.

model.⁴⁷ Care should be taken to select channels appropriate for the signal detection task.^{33,50} We found good agreement with Laguerre–Gauss channels for rotationally symmetric cylindrical pins, but it may be insightful to evaluate other filters as well.

We applied the model observer, Model-k4, to evaluate a representative case of image processing as well as detectability improvement trends with IMR reconstruction. Using an image blending technique⁵⁸ to combine FBP and IMR images, we found that Model-k4 reasonably predicted the human results. This suggests that the model can be used over a range of noise spanning the extremes of FBP and IMR. Using Model-k4 to measure the detectability ratio, $R_{d'}$, we found that the value of $R_{d'}$ is relatively constant across contrast and dose. This indicates that IMR is able to improve detection even with very low-contrast targets in significant projection noise, conditions which would normally stress reconstruction algorithms. With regards to size, we find that IMR gives somewhat more improvement with larger signals than smaller ones, perhaps due to the spatial regularization in the reconstruction process. When using IMR in dose reduction studies, larger signals could thus allow more aggressive dose reductions.

Further analysis demonstrates how the CHO model can be used to improve CT imaging, particularly in guiding dose reductions in real CT protocols. Dose reductions with IMR

were estimated using a formula based on the Rose⁶⁰ model. We found a reasonable linear relationship between Model-k4 detectability outputs and the square root of dose. This method predicted an 82% dose saving with IMR as compared to FBP. This prediction was verified in physical CT phantom scans, demonstrating that the model can guide dose reduction experiments.

Regarding the model observers, there are considerations related to this study and various human observer experiment conditions. Model-k4 has proven to reasonably fit data over a substantive range of size, contrast, dose, and reconstruction method. The other tested models, except Model-k3, also showed similar fits to human observer studies. Results will probably apply to similar images and similar reconstruction (e.g., Fig. 1). However, rather than assuming universal generalization, it is probably advisable to evaluate new images with both computational and human observer studies. Additionally, other IMR parameters and settings were not tested. It would be interesting to determine if various settings generate advantages for signals of different sizes and contrasts at low dose levels. Window width and level have been held constant in all conditions here, but this might be another variable to probe when comparing FBP and IMR.

There are some limitations and extensions for this study. First, the anatomical variation in clinical images can affect low-contrast detectability and was not assessed in this work.⁶¹ In addition, the image texture of the iterative reconstruction is much different than that of FBP (see Figs. 1 and 12). Due to these effects, it would be informative to perform observer studies of IMR versus FBP in clinical data with anatomical structure, using, for example, the method proposed by Karantzaev *et al.*⁶² to insert synthetic lesions in CT images. A similar method for lesion insertion in the projection domain has been proposed by Zabic *et al.*⁶³ which would be well-suited for reconstruction comparisons. Second, signal and location known exactly detection might not predict the clinical situation. However, 4-AFC experiments as performed here enable highly controlled studies across single variables and serve as a starting point for more complex image perception tasks. The CHO framework used here may also be extended to localization⁴⁹ or discrimination⁵² tasks as done by others. Third, analyses in this study were limited to the range

FIG. 10. Average detectability ratio of IMR to FBP, $R_{d'}$, for all signals as a function of (a) size and (b) contrast. R increases for larger signals but changes insignificantly as a function of contrast. Error bars are the standard deviation of detectability ratios.

FIG. 11. Predicted dose reduction with IMR. IMR data for each pin are fit using linear regression and used to estimate dose reduction. Target d' for low dose IMR is set to FBP at 4.0 mGy. Shown is the 8 mm, 0.5% contrast pin which has the same detectability as 4.0 mGy FBP as 0.6 mGy IMR.

where Poisson noise dominates and the effects of electronic noise were not assessed. While electronic noise did not appear to be a substantial factor for the phantoms and exposures used in this work (20–63 mAs), Ma *et al.*⁶⁴ have shown that electronic noise needs to be considered for anthropomorphic phantoms or patients at exposures less than 40 mAs. More generally, electronic noise becomes appreciable when photon counts on the detector fall below about 10 counts.⁶⁴ Larger phantoms at the same exposures as this study will likely have appreciable electronic noise. However, currently much CT scanning is performed at higher nominal exposures not near this electronic noise threshold. Clinical CT scans have detector counts on the order of hundreds to thousands, well above the level at which electronic noise becomes significant.⁶⁴ Moving forward, human and model observer experiments in the presence of appreciable electronic noise would be an informative topic. Last, the detectability experiments in this work were performed in the axial plane using cylindrical phantoms which are homogeneous along the z -axis. As such, the results should generally be independent of slice thickness for objects with similar geometry. It would be of interest to apply the methods to objects of spherical geometry and assess gains or limits in detectability along the z -axis for low-contrast objects.

FIG. 12. Example images of the physical phantom used in the dose reduction study. (a) Full phantom image averaged over 100 scans with 10 mm, 0.3% contrast pin FOV outlined in red. (b) FBP signal at 10 mGy. (c) FBP signal image at 4 mGy. (d) IMR signal image at 10 mGy. (e) IMR signal image at 2 mGy. Images (b)–(e) are shown at window-level of WW = 45 HU, WL = 40 HU.

FIG. 13. Dose reduction phantom study using physical CT scans reconstructed with FBP and IMR. Shown is mean probability correct (PC) vs dose and reconstruction. IMR at 2 mGy shows improved detection over FBP at 10 mGy for both human and model observers, representing over 80% dose reduction. Error bars show the standard deviation of human experiments and of the model observer from the binomial distribution.

Using both human and computational observers, we have found that IMR reconstruction substantially improves detectability in low dose CT images. We found that IMR enables at least 67% dose savings over FBP in human and model observer detectability studies. Among the many models tested, we chose a CHO model with internal noise proportional to channel output standard deviations, Model-k4, for analysis of additional images in an automated fashion. Model-k4 is able to account for the effect of drastically different noise characteristics in FBP and IMR images on detectability, making it a valuable tool in dose optimization and iterative reconstruction algorithm optimization. Using an extrapolating function in combination with this model, we found that IMR could reduce dose by 82% on average in the simulated phantom. This was verified by human and model observers in a physical phantom study using a similarly sized phantom and similar x-ray exposures on a Philips iCT system. Further, according to Model-k4, IMR may enable more aggressive dose reductions for larger signals. These human and model observer studies encourage the continued development and application of advanced iterative reconstruction algorithms,

such as IMR, to substantially reduce x-ray dose in CT imaging without sacrificing image quality.

ACKNOWLEDGMENTS

This research project was sponsored by NIH No. R01 EB004070, a NIBIB No. T32EB007509 training grant award to B.E., an Ohio Third Frontier research grant, Cardiac Perfusion with Computed Tomography, from the state of Ohio to CWRU, University Hospitals of Cleveland, Philips Healthcare, and a research contract from Philips Healthcare to CWRU.

^{a)}Author to whom correspondence should be addressed. Electronic mail: dlw@case.edu

¹D. B. Larson, L. W. Johnson, B. M. Schnell, S. R. Salisbury, and H. P. Forman, "National trends in CT use in the emergency department: 1995–2007," *Radiology* **258**, 164–173 (2011).

²V. M. Rao, D. C. Levin, L. Parker, A. J. Frangos, and J. H. Sunshine, "Trends in utilization rates of the various imaging modalities in emergency departments: Nationwide medicare data from 2000 to 2008," *J. Am. Coll. Radiol.* **8**, 706–709 (2011).

³S. R. Baker, "CT utilization and thyroid cancer: Is there a connection?," in *Notes of a Radiology Watcher* (Springer, Cham, Switzerland, 2014), pp. 281–283.

⁴C. H. McCollough, G. H. Chen, W. Kalender, S. Leng, E. Samei, K. Taguchi, G. Wang, L. Yu, and R. I. Pettigrew, "Achieving routine submillisievert CT scanning: Report from the summit on management of radiation dose in CT," *Radiology* **264**, 567–580 (2012).

⁵D. J. Brenner, R. Doll, D. T. Goodhead, E. J. Hall, C. E. Land, J. B. Little, J. H. Lubin, D. L. Preston, R. J. Preston, and J. S. Puskin, "Cancer risks attributable to low doses of ionizing radiation: Assessing what we really know," *Proc. Natl. Acad. Sci. U. S. A.* **100**, 13761–13766 (2003).

⁶A. J. Einstein, "Medical imaging: The radiation issue," *Nat. Rev. Cardiol.* **6**, 436–438 (2009).

⁷F. A. Mettler, Jr., B. R. Thomadsen, M. Bhargavan, D. B. Gilley, J. E. Gray, J. A. Lipoti, J. McCrohan, T. T. Yoshizumi, and M. Mahesh, "Medical radiation exposure in the US in 2006: Preliminary results," *Health Phys.* **95**, 502–507 (2008).

⁸A. C. T. Martinsen, H. K. Sæther, P. K. Hol, D. R. Olsen, and P. Skaane, "Iterative reconstruction reduces abdominal CT dose," *Eur. J. Radiol.* **81**, 1483–1487 (2012).

⁹Y. Funama, K. Taguchi, D. Utsunomiya, S. Oda, Y. Yanaga, Y. Yamashita, and K. Awai, "Combination of a low tube voltage technique with the hybrid iterative reconstruction (iDose) algorithm at coronary CT angiography," *J. Comput. Assisted Tomogr.* **35**, 480–485 (2011).

¹⁰C. H. McCollough, A. N. Primak, N. Braun, J. Kofler, L. Yu, and J. Christner, "Strategies for reducing radiation dose in CT," *Radiol. Clin. North Am.* **47**, 27–40 (2009).

¹¹Q. Ren, S. K. Dewan, M. Li, J. Li, D. Mao, Z. Wang, and Y. Hua, "Comparison of adaptive statistical iterative and filtered back projection reconstruction techniques in brain CT," *Eur. J. Radiol.* **81**, 2597–2601 (2012).

¹²M. Katsura, I. Matsuda, M. Akahane, J. Sato, H. Akai, K. Yasaka, A. Kunimatsu, and K. Ohtomo, "Model-based iterative reconstruction technique for radiation dose reduction in chest CT: Comparison with the adaptive statistical iterative reconstruction technique," *Eur. Radiol.* **22**, 1613–1623 (2012).

¹³D. Mehta, R. Thompson, T. Morton, A. Dhanantwari, and E. Shefer, "Iterative model reconstruction: Simultaneously lowered computed tomography radiation dose and improved image quality," *Med. Phys. Int. J.* **2**, 147–155 (2013).

¹⁴S. Oda, D. Utsunomiya, Y. Funama, K. Katahira, K. Honda, S. Tokuyasu, M. Vembar, H. Yuki, K. Noda, and S. Oshima, "A knowledge-based iterative model reconstruction algorithm: Can super-low-dose cardiac CT be applicable in clinical settings?," *Acad. Radiol.* **21**, 104–110 (2014).

¹⁵M. Cohnen, H. Fischer, J. Hamacher, E. Lins, R. Kötter, and U. Mödder, "CT of the head by use of reduced current and kilovoltage: Relationship between image quality and dose reduction," *Am. J. Neuroradiol.* **21**, 1654–1660 (2000).

¹⁶B. R. Whiting, C. Hoeschen, H. Bruder, R. Raupach, T. Allmendinger, S. Kappler, J. Sunnegardh, K. Stierstorfer, and T. Flohr, "Novel iterative reconstruction method for optimal dose usage in redundant CT-acquisitions," *Proc. SPIE* **9033**, 90330P (2014).

¹⁷R. Fahmi, B. L. Eck, M. Vembar, H. G. Bezerra, and D. L. Wilson, "Dose reduction assessment in dynamic CT myocardial perfusion imaging in a porcine balloon-induced-ischemia model," *Presented at the SPIE Medical Imaging*, 2014.

¹⁸F. Pontana, J. Pagniez, T. Flohr, J.-B. Faivre, A. Duhamel, J. Remy, and M. Remy-Jardin, "Chest computed tomography using iterative reconstruction vs filtered back projection (Part 1): Evaluation of image noise reduction in 32 patients," *Eur. Radiol.* **21**, 627–635 (2011).

¹⁹R. A. Takx, U. J. Schoepf, A. Moscariello, M. Das, G. Rowe, S. O. Schoenberg, C. Fink, and T. Henzler, "Coronary CT angiography: Comparison of a novel iterative reconstruction with filtered back projection for reconstruction of low-dose CT—Initial experience," *Eur. J. Radiol.* **82**, 275–280 (2013).

²⁰A. Moscariello, R. A. Takx, U. J. Schoepf, M. Renker, P. L. Zwerner, T. X. O'Brien, T. Allmendinger, S. Vogt, B. Schmidt, and G. Savino, "Coronary CT angiography: Image quality, diagnostic accuracy, and potential for radiation dose reduction using a novel iterative image reconstruction technique—Comparison with traditional filtered back projection," *Eur. Radiol.* **21**, 2130–2138 (2011).

²¹L. Yu, S. Leng, Y. Zhang, Z. Li, J. M. Kofler, and C. H. McCollough, "Objective assessment of low-contrast performance in the ACR CT accreditation phantom using a channelized Hotelling observer and its correlation with human observers," *Scientific Assembly and Annual Meeting of Radiological Society of North America, Chicago, IL*, 2013.

²²A. E. Burgess, "Comparison of receiver operating characteristic and forced choice observer performance measurement methods," *Med. Phys.* **22**, 643–655 (1995).

²³Y. Hou, X. Liu, S. Xv, W. Guo, and Q. Guo, "Comparisons of image quality and radiation dose between iterative reconstruction and filtered back projection reconstruction algorithms in 256-MDCT coronary angiography," *Am. J. Roentgenol.* **199**, 588–594 (2012).

²⁴L. M. Popescu and K. J. Myers, "CT image assessment by low contrast signal detectability evaluation with unknown signal location," *Med. Phys.* **40**, 111908 (10pp.) (2013).

²⁵B. Eck, R. Fahmi, K. M. Brown, N. Raihani, and D. L. Wilson, "Comparison of computational to human observer detection for evaluation of CT low dose iterative reconstruction," *Proc. SPIE* **9037**, 90370P (2014).

²⁶B. Eck, R. Fahmi, K. M. Brown, N. Raihani, and D. L. Wilson, "Quantitative image quality in CT: Detectability = f (Dose, reconstruction, contrast, size)," in *Scientific Assembly and Annual Meeting of Radiological Society of North America*, 2013.

²⁷F. A. Miéville, F. Gudinchet, F. Brunelle, F. O. Bochud, and F. R. Verdun, "Iterative reconstruction methods in two different MDCT scanners: Physical metrics and 4-alternative forced-choice detectability experiments—A phantom approach," *Phys. Med.* **29**, 99–110 (2013).

²⁸H. 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).

²⁹I. Hernandez-Giron, J. Geleijns, A. Calzado, and W. Veldkamp, "Automated assessment of low contrast sensitivity for CT systems using a model observer," *Med. Phys.* **38**, S25–S35 (2011).

³⁰I. Hernandez-Giron, J. Geleijns, A. Calzado, M. Salvado, R. Joemai, and W. Veldkamp, "Objective assessment of low contrast detectability for real CT phantom and in simulated images using a model observer," *Presented at the IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, 2011.

³¹X. He and S. Park, "Model observers in medical imaging research," *Theranostics* **3**, 774–786 (2013).

³²K. 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).

³³B. D. Gallas and H. H. Barrett, "Validating the use of channels to estimate the ideal linear observer," *J. Opt. Soc. Am. A* **20**, 1725–1738 (2003).

³⁴J. Wang, T. Li, H. Lu, and Z. Liang, "Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose x-ray computed tomography," *IEEE Trans. Med. Imaging* **25**, 1272–1283 (2006).

³⁵A. Wunderlich and F. Noo, "Image covariance and lesion detectability in direct fan-beam x-ray computed tomography," *Phys. Med. Biol.* **53**, 2471–2493 (2008).

- ³⁶A. Burgess, R. Wagner, R. Jennings, and H. B. Barlow, "Efficiency of human visual signal discrimination," *Science* **214**, 93–94 (1981).
- ³⁷A. A. Sanchez, E. Y. Sidky, I. Reiser, and X. Pan, "Comparison of human and Hotelling observer performance for a fan-beam CT signal detection task," *Med. Phys.* **40**, 031104 (9pp.) (2013).
- ³⁸Z. L. Lu and B. A. Doshier, "Characterizing human perceptual inefficiencies with equivalent internal noise," *J. Opt. Soc. Am. A* **16**, 764–778 (1999).
- ³⁹Z.-L. Lu and B. A. Doshier, "External noise distinguishes attention mechanisms," *Vision Res.* **38**, 1183–1198 (1998).
- ⁴⁰D. G. Pelli, "Effects of visual noise" (University of Cambridge, Cambridge, UK, 1981).
- ⁴¹D. G. Pelli, "Uncertainty explains many aspects of visual contrast detection and discrimination," *J. Opt. Soc. Am. A* **2**, 1508–1531 (1985).
- ⁴²A. Burgess and B. Colborne, "Visual signal detection. IV. Observer inconsistency," *J. Opt. Soc. Am. A* **5**, 617–627 (1988).
- ⁴³J. Gold, P. Bennett, and A. Sekuler, "Signal but not noise changes with perceptual learning," *Nature* **402**, 176–178 (1999).
- ⁴⁴C. 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* **18**, 473–488 (2001).
- ⁴⁵A. 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).
- ⁴⁶Y. Z. Y. Zhang, B. T. Pham, and M. P. Eckstein, "Automated optimization of JPEG 2000 encoder options based on model observer performance for detecting variable signals in x-ray coronary angiograms," *IEEE Trans. Med. Imaging* **23**, 459–474 (2004).
- ⁴⁷Y. Zhang, B. T. Pham, and M. P. Eckstein, "Evaluation of internal noise methods for Hotelling observer models," *Med. Phys.* **34**, 3312–3322 (2007).
- ⁴⁸M. Narayanan, H. Gifford, M. King, P. Pretorius, T. Farncombe, P. Bruyant, and M. Wernick, "Optimization of iterative reconstructions of 99m Tc cardiac SPECT studies using numerical observers," *IEEE Trans. Nucl. Sci.* **49**, 2355–2360 (2002).
- ⁴⁹S. Leng, L. Yu, Y. Zhang, R. Carter, A. Y. Toledano, and C. H. McCollough, "Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain," *Med. Phys.* **40**, 081908 (9pp.) (2013).
- ⁵⁰J. G. Brankov, "Optimization of the internal noise models for channelized Hotelling observer," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (IEEE, Chicago, IL, 2011), pp. 1788–1791.
- ⁵¹L. 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).
- ⁵²Y. Zhang, S. Leng, L. Yu, R. E. Carter, and C. H. McCollough, "Correlation between human and model observer performance for discrimination task in CT," *Phys. Med. Biol.* **59**, 3389–3404 (2014).
- ⁵³K. P. Burnham, "Multimodel inference: Understanding AIC and BIC in model selection," *Sociol. Methods Res.* **33**, 261–304 (2004).
- ⁵⁴S. Park, A. Badano, B. D. Gallas, and K. J. Myers, "Incorporating human contrast sensitivity in model observers for detection tasks," *IEEE Trans. Med. Imaging* **28**, 339–347 (2009).
- ⁵⁵F. A. Wichmann and N. J. Hill, "The psychometric function: I. Fitting, sampling, and goodness of fit," *Percept. Psychophys.* **63**, 1293–1313 (2001).
- ⁵⁶J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder–Mead simplex method in low dimensions," *SIAM J. Optim.* **9**, 112–147 (1998).
- ⁵⁷A. E. Burgess, "Visual perception studies and observer models in medical imaging," *Semin. Nucl. Med.* **41**, 419–436 (2011).
- ⁵⁸P. Prakash, M. K. Kalra, A. K. Kambadakone, H. Pien, J. Hsieh, M. A. Blake, and D. V. Sahani, "Reducing abdominal CT radiation dose with adaptive statistical iterative reconstruction technique," *Invest. Radiol.* **45**, 202–210 (2010).
- ⁵⁹A. Rose, "The sensitivity performance of the human eye on an absolute scale," *J. Opt. Soc. Am.* **38**, 196–208 (1948).
- ⁶⁰A. E. Burgess, "The Rose model revisited," *J. Opt. Soc. Am. A* **16**, 633–646 (1999).
- ⁶¹Y. Srinivas and D. L. Wilson, "Quantitative image quality evaluation of pixel-binning in a flat-panel detector for x-ray fluoroscopy," *Med. Phys.* **31**, 131–141 (2004).
- ⁶²K. Karantza, H.-o. Shin, S. Jöndes, B. King, K. Ringe, D. Hartung, F. Wacker, and C. von falck, "Development and evaluation of a software tool for the generation of virtual liver lesions in multidetector-row CT datasets," *Acad. Radiol.* **20**, 614–620 (2013).
- ⁶³S. Zabic, K. M. Brown, and B. Eck, "Statistical framework for synthetic object addition to CT projections," in *Proceedings of the Third International Conference on Image Formation in X-ray Computed Tomography, Salt Lake City, UT* (2014), pp. 372–375, see http://www.ct-meeting.org/?p=meeting_2014.
- ⁶⁴J. Ma, Z. Liang, Y. Fan, Y. Liu, J. Huang, W. Chen, and H. Lu, "Variance analysis of x-ray CT sinograms in the presence of electronic noise background," *Med. Phys.* **39**, 4051–4065 (2012).