Comparison of human and Hotelling observer performance for a fan-beam CT signal detection task

Adrian A. Sanchez,^{a)} Emil Y. Sidky,^{b)} Ingrid Reiser, and Xiaochuan Pan^{c)}
Department of Radiology, The University of Chicago, Chicago, Illinois 60637

(Received 1 August 2012; revised 11 December 2012; accepted for publication 19 December 2012; published 28 February 2013)

Purpose: A human observer study was performed for a signal detection task for the case of fan-beam x-ray computed tomography. Hotelling observer (HO) performance was calculated for the same detection task without the use of efficient channels. By considering the full image covariance produced by the filtered backprojection (FBP) algorithm and avoiding the use of channels in the computation of HO performance, the authors establish an absolute upper bound on signal detectability. Therefore, this study serves as a baseline for relating human and ideal observer performance in the case of fan-beam CT.

Methods: Eight human observers participated in a two-alternative forced choice experiment where the signal of interest was a small simulated ellipsoid in the presence of independent, identically distributed Gaussian detector noise. Theoretical performance of the HO, which is equivalent to the ideal observer in this case (see Sec. 13.2.12 in Barrett and Myers [*Foundations of Image Science* (Wiley, Hoboken, NJ, 2004)]), was also computed and compared to the performance of the human observers. In addition to a reference FBP implementation, two FBP implementations with inherent loss of HO signal detectability (e.g., by apodizing the ramp filter) were also investigated. Each of these latter two implementations takes the form of a discrete-to-discrete linear operator (i.e., a matrix), which has a nontrivial null-space resulting in the loss of detectability.

Results: Estimated observer detectability index ($\hat{d}_A$) values for the human observers and SNR values for the HO were obtained. While Hanning filtering in the FBP implementation with a cutoff frequency of 1/4 of the Nyquist frequency reduces HO SNR (due to the reconstruction matrix's nontrivial null-space), this filtering was shown to consistently improve human observer performance. By contrast, increasing the image pixel size was seen to have a comparable effect on both the HO and the human observers, degrading performance.

Conclusions: These results, which characterize HO and human observer performance for a signal detection task in fan-beam FBP noise, form a basis for applying model observer metrics to fan-beam CT when knowledge of the full image-domain noise statistics is important. Further, by calculating HO performance without relying on channels, these results are particularly relevant when an information theoretic approach is considered, e.g., in optimization of the image reconstruction algorithm with respect to preservation of signal detectability. Finally, the HO (which is here equivalent to the ideal observer) provides an absolute upper bound on detection performance, and our results therefore provide insight into the performance of human observers relative to the optimum for two different cases wherein ideal observer performance is compromised through degradation of the data quality. In one case (regularization), human performance is improved to practically ideal performance, and in the other (larger pixel size), ideal and human observer performance are approximately degraded equivalently. © 2013 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4789590>]

Key words: image reconstruction, model observers, computed tomography, Hotelling observer, signal detection

I. INTRODUCTION

Lesion detection is one prominent clinical task for which x-ray computed tomography (CT) is employed. Therefore, lesion detectability has been frequently investigated as a useful metric for systems optimization and evaluation in CT.²⁻⁷ The ultimate goal of CT in terms of image quality is task-specific, and for the case of lesion detection, human observer detectability is arguably the most meaningful metric of image quality. However, human observer studies needed to evaluate

detectability are too expensive and burdensome to be useful for systems optimization. Therefore, mathematical model observers are often employed in order to either estimate human observer performance [e.g., channelized Hotelling observers¹ (HO)] or provide a theoretical upper bound on human observer performance (the Bayesian ideal observer⁸).

In keeping with current interest in utilizing the ideal observer as a means of establishing an upper bound on observer performance,⁵ this work investigates human and Hotelling observer performance for a lesion detection task in fan-beam

CT. For the present case, the HO is equivalent to the ideal observer. Seeking to provide a link between existing studies that have extensively investigated human and model observer performance for systems that mimic CT (Ref. 7) and those which simplify the computation of observer performance via channels,^{5,6} this work investigates exact Hotelling observer performance for a binary lesion detection task and compares this result to measured human observer performance for the case of detecting a small, high-contrast lesion in noise correlated via the filtered backprojection (FBP) algorithm.

Three separate implementations of fan-beam FBP are investigated. The first serves as a reference case and utilizes the conventional ramp filter in reconstruction. The remaining two reconstruction implementations are intended to degrade HO SNR and illustrate the corresponding effect on human observer performance. One implementation accomplishes this through regularization via a Hanning filter, and the other increases the reconstruction image pixel size. Each of these implementations of the FBP algorithm takes the form of a matrix operator whose null-space determines the specific loss of inherent signal detectability from the sinogram domain to the reconstructed image domain. In the case of regularization, some components of the filtered data in the discrete Fourier transform domain are set to zero, thereby introducing a null-space corresponding to these components. In the case of increased pixel size, the $M \times N$ reconstruction matrix acquires a null-space by virtue of the fact that $M < N$.

The paper is organized as follows: Sec. II provides necessary background regarding fan-beam CT reconstruction, two-alternative forced choice (2AFC) studies, and HO methodology. Section III outlines the methods for generating the images used in the 2AFC studies and performing the psychophysical and model observer studies. The results of the studies are presented in Sec. IV, and a brief discussion of these results and conclusion are given in Sec. V.

II. BACKGROUND

II.A. Fan-beam CT reconstruction

The CT configuration considered here is fan-beam, in which a source and detector pair rotate about a central field-of-view, acquiring discrete projection data at a fixed number of views. In particular, we consider a flat detector with uniform detector bin spacing as in Fig. 1. The fan-beam CT data model relates the continuous object function $f(\vec{r})$ to the discrete projection data vector g_i via a line integral over $f(\vec{r})$ along the i th ray, defined by the source position s_i and the ray direction $\hat{\theta}_i$

$$\mathbf{g}_i = \int_L f(s_i + l\hat{\theta}_i) dl + \mathbf{n}_i, \text{ where } i \in [1, N], \quad (1)$$

where $\mathbf{n}_i$ represents data noise and $\mathbf{g}_i$ and $\mathbf{n}_i$ are random variables representing the i th elements of the data and noise vectors, respectively. In other words, $\mathbf{g} = [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_N]$ and $\mathbf{n} = [\mathbf{n}_1, \mathbf{n}_2, \dots, \mathbf{n}_N]$. The domain of integration L is defined as the intersection of the i th ray with the compact support of the object f . Note that the additive noise term $\mathbf{n}_i$ is added in the projection data domain as opposed to being considered

FIG. 1. Fan-beam geometry with a flat detector and uniform detector element spacing.

in the incident photon fluence, hence, the term $\mathbf{n}_i$ is not contained in the integral of Eq. (1). Also observe that since this data model produces a discretized projection data vector $\mathbf{g}$, all subsequent image reconstruction steps map an intermediate discrete data vector to the final discretized image and are, therefore, discrete-to-discrete transformations. Further, since the FBP algorithm is a linear algorithm, the reconstruction operator constitutes a linear, discrete-to-discrete operator which can be represented by a matrix A .

Following the FBP algorithm outlined in Ref. 9, reconstruction follows a three-step process: First, each $\mathbf{g}_i$ is weighted by the factor $w(n) = D/\sqrt{D^2 + n^2 du^2}$, where D is the source-to-detector distance, n is an integer denoting the detector element where $n = 0$ corresponds to the central ray, and du is the detector element spacing. Numerically, the implementation of this weighting is equivalent to operating on the discrete projection data with a diagonal weighting matrix such that

$$\mathbf{g}' = \text{diag}(w(n)) \mathbf{g}, \quad (2)$$

where $\text{diag}(w(n))$ is the diagonal weighting matrix. Second, the modified projection data vector $\mathbf{g}'$ is transformed to the frequency domain using a discrete FFT algorithm, also representable by a matrix which we shall label F . In this domain, the conventional ramp filtering can be performed via multiplication. Alternatively, one may apply an apodization window such as a Hanning window to perform regularization in the reconstruction. Finally, one applies the matrix F^{-1} and performs a weighted backprojection of each of the filtered projections onto a discrete image array via a backprojection matrix operator B , resulting in the reconstructed image pixels $\mathbf{y}_i$. The full reconstruction operation can then be written as

$$\mathbf{y} = \mathbf{A}\mathbf{g} = \mathbf{B}\mathbf{F}^{-1}\text{diag}(aw)\text{diag}(rk)\mathbf{F}\text{diag}(w(n))\mathbf{g}, \quad (3)$$

where $\text{diag}(aw)$ and $\text{diag}(rk)$ are diagonal matrices with an apodization window and the ramp kernel along their

diagonals, respectively, and B , F , and $\text{diag}(w(n))$ are again matrices representing backprojection, the discrete FFT operation and data weighting. In general, this reconstruction matrix A is singular, and its null-space can play a direct role in degrading ideal observer SNR for a given signal detection task. One goal of this study is, therefore, to evaluate human observer performance in a 2AFC study in light of this degradation.

II.B. Binary detection

In order to elucidate the two hypotheses which a binary signal detection task comprises, we rewrite the data model of Eq. (1) as a continuous to discrete operator $\mathcal{H}$, which maps the continuous object function $f(\vec{r})$ to the discrete $M \times 1$ data vector $\mathbf{g}$

$$\mathbf{g} = \mathcal{H}f + \mathbf{n}, \quad (4)$$

where $\mathbf{n}$ is an additive measurement noise term in the projection data domain, and the sole source of noise considered in this study. As described in Sec. III.A.2, we consider the N different $\mathbf{n}_i$ to be independent and drawn from identical Gaussian distributions.

The two hypotheses relevant to a signal detection task are the signal-absent and signal-present hypotheses, which we shall label H_0 and H_1 , respectively. Since we consider detection in the reconstructed image domain, we must apply the reconstruction operator A discussed in Sec. II.A to the data vectors representing each hypothesis. The resulting hypotheses may then be expressed mathematically as

$$\begin{aligned} H_0 : \mathbf{y} &= A(\mathcal{H}f_b + \mathbf{n}), \\ H_1 : \mathbf{y} &= A(\mathcal{H}(f_b + f_s) + \mathbf{n}), \end{aligned} \quad (5)$$

where f_b and f_s represent the background object and signal object, respectively. It should be noted that the statistics of the additive noise term in each hypothesis is assumed to be the same. The task we consider is a signal-known-exactly, background-known-exactly (SKE/BKE) task, meaning that the Hotelling observer has full knowledge of the reconstructed signal and the image noise statistics. Further, the human observer study is also in the SKE paradigm since the reconstructed signal is presented to the observer for each trial of the 2AFC study.

II.C. 2AFC study

In a 2AFC experiment, an observer is presented with two images $\mathbf{y}$ and $\mathbf{y}'$, where $\mathbf{y}$ is drawn from $\text{pr}(\mathbf{y}|H_0)$ and $\mathbf{y}'$ is drawn from $\text{pr}(\mathbf{y}|H_1)$.¹ In the case of a signal detection task, the hypotheses H_0 and H_1 will represent a signal-absent and signal-present case (see Sec. II.B). The observer computes a test statistic $t(\mathbf{y})$, commonly referred to as the decision variable, for each of the two images. The observer then assigns the signal-present decision to the image that produces the higher value of the decision variable. For any decision variable, the probability of a correct decision in a 2AFC process is then

equal to the area under the ROC curve (AUC) for that observer and detection task. The mean number of correct decisions in a human observer study can then be taken as an estimator for the AUC. AUC can then be related to an effective SNR by the equation

$$\text{SNR} = 2\text{erf}^{-1} [2(\text{AUC}) - 1]. \quad (6)$$

Two observers can then be compared by an efficiency metric equal to the ratio of the squares of their two SNRs

$$\epsilon = \frac{\text{SNR}_{\text{Obs1}}^2}{\text{SNR}_{\text{Obs2}}^2}. \quad (7)$$

II.D. Hotelling observer

The HO is the optimal mathematical model observer that is linear in the data; i.e., the HO is the optimal observer whose test statistic is computed as a linear combination of the data elements. The Hotelling SNR is the figure of merit used in this work, and its computation is performed as follows: First, one must solve for w , the Hotelling template, which is the image with which the HO masks a reconstructed image $\mathbf{y}$ in order to make the decision between signal-absent and signal-present. w is found by solving the linear equation

$$K_y w = \Delta \bar{y}, \quad (8)$$

where K_y is the image covariance matrix and $\Delta \bar{y}$ is the mean difference between the hypotheses H_0 and H_1 of Eq. (5). Since the only source of statistical variability in Eq. (5) is the additive noise term, the quantity $\Delta \bar{y}$ can simply be computed as

$$\Delta \bar{y} = (A\mathcal{H}(f_b + f_s) + A\bar{n}) - (A\mathcal{H}f_b + A\bar{n}) = A\mathcal{H}f_s, \quad (9)$$

where $\bar{n}$ is a vector whose elements represent the mean detector element noise. Note that the quantities f_b and f_s are not statistically variable in this case, as the task chosen is a SKE/BKE task. Equation (9), therefore, contains no statistically variable quantities.

Although the form of Eq. (8) is compact, obtaining a direct solution for w_y is computationally nontrivial, since the matrix K_y can be too large to be stored in computer memory and is nondiagonal, as discussed in Sec. III.A.2. In fact, the computational burden of solving Eq. (8) is the basis for much active research. For instance, approximate inversions of the image covariance in data space either by efficient channels^{1,5,10,11} or assumptions of stationarity in the image covariance are sometimes made to ease the computational burden. In the present case, 1024 view angles and 256 detector bins are used, implying that $\mathbf{g}$ and K_g have dimensions 262, 144 $\times$ 1 and 262, 144 $\times$ 262, 144, respectively. Similarly, $\mathbf{y}$ and K_y have dimensions 147, 456 $\times$ 1 and 147, 456 $\times$ 147, 456, respectively. These covariance matrices are too large to be directly stored in memory on many systems, implying the need for methods that require only a function which outputs matrix-vector

products, without direct access to the matrices. Here, we solve the linear system in Eq. (8) for the Hotelling template iteratively via the method of conjugate gradients, as described in Sec. 5.1 of Ref. 12. We find this method preferable to the use of channels (either efficient or psychophysical) because we wish to accurately determine the performance of the HO unhindered by information loss through channels. Further, although efficient channels might exist which can be found using a method like that in Witten *et al.*,¹¹ most work involving efficient channels imposes circular symmetry on the signal and template, an assumption violated by the templates in our case. The calculated HO performance can then provide an absolute upper bound on measured human observer performance for several cases of information loss (see discussion of null-space in Sec. II.A).

Finally, after obtaining a solution to Eq. (8), the HO SNR for a signal-detection task is computed by applying the Hotelling template w to the mean difference between the image vectors for the hypotheses H_0 and H_1 , $\Delta\bar{y}$

$$\text{SNR}_{\text{Hot}}^2 = w^T \Delta\bar{y}, \quad (10)$$

where $\Delta\bar{y}$ is again given by Eq. (9). This quantity, SNR_{Hot} can then be compared to the effective SNR for human observers computed via Eq. (6). Note that $\Delta\bar{y}$ is derived analytically, i.e., none of the quantities in Eq. (9) are statistically variable. Further, the covariance matrix K_y is also defined analytically, and does not depend on statistical realizations. The original derivation of Eq. (10) in Barrett and Myers¹ relies on the assumption that w and $\Delta\bar{y}$ are derived from a large number of statistical realizations, so that the resulting SNR is close to that which would result from using w and $\Delta\bar{y}$ computed from independent sets of realizations. However, since no statistical realizations are used here, we avoid this complication and need not consider any statistical bias resulting from the use of Eq. (10).

III. METHODS

III.A. Generation of images

III.A.1. The reconstructed signal

The signal of interest was chosen to be an ellipse with major and minor axes of lengths six detector bin widths and three detector bin widths, respectively, and its location was fixed for each of the studies to the center of the field of view for reconstruction. We investigate a small high-contrast signal, which can be seen as a model of microcalcification detection in breast CT. The signal was defined in the continuous object domain and discretized by a continuous-to-discrete forward projection operator, as in Eq. (1).

The reconstruction algorithm used was the FBP algorithm discussed in Sec. II.A. Projection data for the reference case were acquired over a full 2π rotation at 1024 evenly spaced angles with 256 detector elements. The reconstruction was performed onto a 384×384 pixel image grid with pixels roughly twice the size of a single detector element. The images were then cropped to a central 96×96 pixel ROI, which

FIG. 2. (Top left) The reconstructed reference signal with ramp filtering and no added noise. Display window: $[0 \ 1.0]$ Visible artifacts are a result of the coarse level of discretization in the sinogram and image domains necessary to allow tractability of the relevant linear systems. (Top right) The Hotelling template for the reference reconstruction. Display window: $[-0.2 \ 0.2]$ (Middle left) The reconstructed reference signal with Hanning filtering and no added noise. The window used here is narrower than that used in the experiment for printing purposes. Display window: $[0 \ 0.45]$ (Middle right) The Hotelling template for the filtered reconstruction. Display window: $[-1.7 \ 1.7]$ (Bottom left) The reconstructed reference signal with ramp filtering, increased pixel size, and no added noise. Display window: $[0 \ 1.0]$ (Bottom right) The Hotelling template for the large-pixel reconstruction. Display window: $[-0.26 \ 0.26]$.

was then displayed at $4 \times$ magnification. Black and white lines were then superimposed on the image in order to aid the observers in localizing the signal at the center of the ROI. The reference images for the human and model observer study were reconstructed without regularization, i.e., with only the ramp kernel used for filtration. The ratio of the source-to-detector distance to the source's radius of rotation was 8:5. The reference reconstructed signal is shown on the top left of Fig. 2. Visible artifacts in the reconstructed signal are a result of the discretization in the sinogram and image domains and the small size of the signal. This small signal size is desirable in this case because we can then expect greater variability in observer performance with respect to the reconstruction algorithm with the given noise model. The corresponding Hotelling template is pictured on the top right of Fig. 2.

III.A.2. Noise model

After computing the discrete projection data as in Eq. (1), zero-mean, independent, identically distributed Gaussian noise was added to the projection data. The standard deviation of the additive noise was uniform across the detector elements and equal to roughly ten times the maximum value of the signal in the projection data domain. This implies that the data covariance vector K_g is diagonal such that

$$(K_g)_{i,j} = \begin{cases} \alpha & i = j \\ 0 & i \neq j \end{cases}, \quad (11)$$

where α is a constant. We then have that the reconstructed image covariance matrix will be given by

$$K_y = AK_gA^T = \alpha AA^T, \quad (12)$$

where A is the reconstruction matrix described in Sec. II.A, and K_y is again the reconstructed image covariance matrix. Inspecting this expression for the image covariance matrix, it becomes obvious why the inversion of this matrix is non-trivial. K_y has dimensionality $M \times M$, where M is the number of pixels in the reconstructed image. Further, various components of the matrix A such as the matrix representing the weighted backprojection step in the FBP algorithm make K_y nondiagonal. Selected components of the matrices K_y for the three reconstruction implementations considered here are shown in Fig. 3.

The noise model was assumed to be invariant under the two hypotheses. This assumption is particularly reasonable in our case in light of the fact that the signal of interest is small, with size on the order of only a few detector bins. An example of a pair of images used in the human observer study is presented in the top row of Fig. 4.

III.A.3. Regularization and image domain sampling

As mentioned in Sec. II.A, regularization can be performed in fan-beam FBP by applying a multiplicative apodization window to the ramp kernel in the Fourier domain. A case of heavy regularization is considered here, wherein a Hanning window with cutoff frequency equal to 1/4 of the Nyquist frequency is used. For this case, we expect a drop in HO SNR relative to the reference (unmodified ramp kernel) case since this multiplicative window will place high-frequency components of the weighted projection data in the null-space of the reconstruction operator. The mean reconstructed signal and Hotelling template for the regularized study are shown in the middle row of Fig. 2.

Next, we considered the case of reconstruction onto pixels which are a factor of 4 larger than in the reference case, i.e., approximately eight detector bin widths square. Here, the expected loss in HO SNR arises from the fact that the reconstruction matrix A has a more significant null-space due to the lower number of output image pixels relative to the number of input data elements. In other words, the matrix A is now farther from being square, transforming an N -element data vector to an M -element image array where $M < N$. The mean reconstructed signal and Hotelling template for the study involving larger pixels is shown in the bottom row of Fig. 2.

FIG. 3. These images are the result of right multiplying the matrices K_y with column vectors containing only a single nonzero element, i.e., images of a single nonzero pixel. The nonzero pixel (valued 1) was placed either halfway to the edge of the FOV (left) or in the center of the FOV (right). The images shown are cropped ROIs showing the pixels which covary with the given pixel. Therefore, each image represents a portion of a single row of the matrix K_y . The nonzero components away from the pixel of interest demonstrate the nondiagonality of K_y , while the change in shape between the images in the two columns demonstrate the shift-variance of image covariance. The images correspond to the reference FBP implementation (top), Hanning filtration (middle), and large image pixels (bottom). Note that the “Mexican hat” shape and near shift-invariance of the Hanning filtered case lead to ripples in the corresponding Hotelling template in Fig. 2. The physical extent of each ROI is identical. The display window of the top and bottom rows is centered on 0 with a width equal to 0.04 times the maximum covariance element. The display window of the middle row is centered on 0 with a width equal to 0.4 times the maximum covariance element.

Example pairs of images used in the regularization and large pixel studies are shown in the middle and bottom of Fig. 4.

III.B. Psychophysical studies

Seven human observers participated in 100 trials of a 2AFC procedure for each of the three reconstructions investigated: reference FBP, heavy regularization, and $4 \times$ larger pixel size. In order to achieve improved confidence limits, an eighth human observer (hereafter referred to as Observer 8)

FIG. 4. (Top) An example of a 2AFC trial used in the human observer study for the reference case of ramp filter FBP reconstruction. The signal amplitude here and in the other two subfigures is ten times greater than in the experiment for the sake of visualization. In addition to two images such as these, the reconstructed signal (left of Fig. 2) was shown for each trial. Display window: $[-2.5 \ 2.5]$ (Middle) An example of a 2AFC trial used in the human observer study for the filtered case. Display window: $[0.3 \ 0.7]$ (Bottom) An example of a 2AFC trial used in the human observer study for the large pixel case. The window used here is narrower than that used in the experiment for printing purposes. Display window: $[0.3 \ 0.7]$.

performed 1500 trials of a 2AFC procedure for the same three experimental conditions. Observer 8 rested after every 250 images in order to ameliorate observer fatigue. In addition to the two image choices presented to the observers, the reconstructed signal was also shown accompanying each pair of images (i.e., the signals on the left of Fig. 2). Ten trials were performed preceding the 100 (or 1500) recorded trials for each set of conditions as training for each observer. The monitor used for the psychophysical experiments was a standard grayscale-calibrated DICOM 14 mammography monitor, and the ambient light and monitor brightness were fixed across all observers and reconstruction implementations. The images were windowed and leveled so that the level was roughly centered on the mean pixel value and the window width roughly

corresponded to two-thirds of the full range of pixel values. The exact window settings used were $[-10, 10]$, $[-2.5, 2.5]$, and $[-10, 10]$, for the reference, filtered, and large pixel case, respectively. While the window and level were changed for each of the three experimental conditions, they were held fixed for the trials and observers within a given experimental condition. However, it should be noted that, in general, the results of preliminary investigation did not appear to be sensitive to the window and level used. The software used for the image display and data acquisition was written in the MATLAB programming language. None of the eight observers had received medical training. Each was a graduate student in medical physics with a basic understanding of the goals of the studies.

For each observer and experimental condition, the proportion of correct decisions was taken to be the unbiased estimator of the area under the ROC curve, $\widehat{AUC} = \frac{1}{n} \sum_{i=1}^n \lambda_i$, where λ_i denotes the outcome of the i th trial (1 for a correct response and 0 for an incorrect response). In order to compute a confidence interval (CI) for the true AUC, we use Papoulis's¹³ expression for estimating the CI when the variance is unknown and the number of trials is large enough that the central limit theorem applies (i.e., the estimator $\widehat{AUC}$ is approximately normal). Namely, we find that

$$P \left\{ \widehat{AUC} - \frac{s}{\sqrt{n}} z_{1-\delta/2} < AUC < \widehat{AUC} + \frac{s}{\sqrt{n}} z_{1-\delta/2} \right\} > 1 - \delta = \gamma, \quad (13)$$

where δ is the confidence level, γ is the confidence coefficient, s is the sample standard deviation, z_u denotes the u th percentile of the standard normal density, and n is the number of trials (always 100 except for Observer 8). The sample standard deviation s is computed using the unbiased estimate of variance defined by Eq. 9-13 of Papoulis.¹³ This estimate of standard deviation is also predicated upon the assumption that n is large enough for the central limit theorem to apply.

Since by Eq. (6) we have that

$$AUC = \frac{1}{2} \left[\operatorname{erf} \left(\frac{1}{2} \operatorname{SNR} \right) + 1 \right], \quad (14)$$

we can rewrite Eq. (13) as

$$P \left\{ \widehat{AUC} - \frac{s}{\sqrt{n}} z_{1-\delta/2} < \frac{1}{2} \left[\operatorname{erf} \left(\frac{1}{2} \operatorname{SNR} \right) + 1 \right] < \widehat{AUC} + \frac{s}{\sqrt{n}} z_{1-\delta/2} \right\} > \gamma \quad (15)$$

or equivalently

$$P \left\{ 2 \operatorname{erf}^{-1} \left[2 \left(\widehat{AUC} - \frac{s}{\sqrt{n}} z_{1-\delta/2} \right) - 1 \right] < \operatorname{SNR} < 2 \operatorname{erf}^{-1} \left[2 \left(\widehat{AUC} + \frac{s}{\sqrt{n}} z_{1-\delta/2} \right) - 1 \right] \right\} > \gamma. \quad (16)$$

The 95% confidence intervals for AUC and SNR for each observer and experimental condition were then found by setting $\gamma = 0.95$ and evaluating Eqs. (13) and (16).

III.C. Model observer studies

For each of the three experimental conditions considered here, HO SNR was computed as in Eq. (10), where the image covariance matrix K_y is given by Eq. (12). The Hotelling template was computed iteratively via the method of conjugate gradients. One important distinction between the HO SNR or AUC and the corresponding human AUC or d_A is that, while the human observer estimation of human AUC and d_A are statistically variable, and hence have associated uncertainties, the HO SNR is a deterministic quantity derived directly from the statistical properties of the two hypotheses.

IV. RESULTS

Figure 5 shows the AUC results obtained for the eight human observers and the HO for each of the three experimental conditions. The corresponding results in terms of SNR (or $\hat{d}_A$ for the human observers) are shown in Fig. 6. These results are also summarized in Tables I and II, respectively. Clearly, there is a marked improvement in human observer performance for the case of regularization, whereas the ideal observer performance degrades with regularization due to the loss of information through smoothing. By contrast, increasing the pixel size in the reconstructed image domain degrades both the human and Hotelling observers similarly. Note that Observer 4 achieved 100% correct on the regularized 2AFC study, so that a reliable CI for AUC or estimate of d_A is not possible. However, we include the results for Observer 4 under the other two experimental conditions for completeness.

The error bars shown are 95% confidence intervals obtained through Eqs. (13) and (16). Although the upper limit of the confidence interval exceeds the deterministic HO SNR for the filtered case, the HO performance should be considered an absolute upper bound. The statistical error bars are only shown to provide intuition regarding the nonlinearity of the confidence intervals when converting from AUC to SNR or d_A .

FIG. 5. AUC results obtained for the eight human observers and the HO for each of the three experimental conditions: reference fan-beam FBP, FBP with Hanning filtration, and FBP with increased pixel size. The error bars shown are 95% confidence intervals obtained through Eq. (13), however, the HO performance should be considered an absolute upper bound. The confidence interval for Observer 4 in the filtered case is unknown because this observer achieved 100% correct on this 2AFC study.

FIG. 6. SNR results obtained for the eight human observers and the HO for each of the three experimental conditions: reference fan-beam FBP, FBP with Hanning filtration, and FBP with increased pixel size. The error bars shown are 95% confidence intervals obtained through Eq. (16), however, the HO performance should be considered an absolute upper bound. For instance, when the CI on AUC extends to 1, the SNR error bars extend to infinity, and are therefore truncated. Observer 4 has an unknown SNR for the filtered case because this observer achieved 100% correct on this 2AFC study.

It should be noted that the result that regularization improves human observer performance to the extent that (in the low statistics case of 100 trials) it is statistically indistinguishable from the ideal observer performance is a consequence of the particular detection task investigated. In other words, in order to have any statistical power in the large pixel, human observer study, a signal was chosen that would necessarily be at a low-difficulty operating point for the humans in the regularization study. For this reason, the performance of Observer 8, the high statistics observer was investigated, clarifying the conclusions of the regularization study. Namely, in keeping with the findings of Abbey and Barrett,⁷ regularization was found to enable human observers to function at roughly 50% efficiency with respect to SNR^2 of the Hotelling observer (the square of the values presented in Table II). The fact that these results are at the high end of those found in Abbey and Barrett⁷ (efficiency of humans ranging from 40% to 50% of HO detectability) is likely due to the more complex noise model employed in both the Hotelling and human

TABLE I. AUC of the ideal observer and eight human observers under each of the three experimental conditions. The intervals shown are the 95% confidence intervals. The confidence interval for Observer 4 in the filtered case is unknown because this observer achieved 100% correct on this 2AFC study.

Observer	Reference AUC	Filtered AUC	Large pixel AUC
Ideal	0.9987	0.9980	0.8126
Obs. 1	0.80 ± 0.0791	0.97 ± 0.0337	0.66 ± 0.0937
Obs. 2	0.82 ± 0.0759	0.99 ± 0.0197	0.76 ± 0.0844
Obs. 3	0.89 ± 0.0619	0.98 ± 0.0277	0.77 ± 0.0832
Obs. 4	0.85 ± 0.0706	$1.0 \pm ?$	0.68 ± 0.0922
Obs. 5	0.89 ± 0.0137	0.99 ± 0.0197	0.72 ± 0.0888
Obs. 6	0.84 ± 0.0725	0.99 ± 0.0197	0.76 ± 0.0844
Obs. 7	0.83 ± 0.0743	0.97 ± 0.0337	0.73 ± 0.0878
Obs. 8	0.9207 ± 0.0137	0.9853 ± 0.0061	0.7273 ± 0.0226

TABLE II. SNR of the ideal observer and eight human observers under each of the three experimental conditions. The intervals shown are the 95% confidence intervals. Observer 4 has an unknown SNR for the filtered case because this observer achieved 100% correct on this 2AFC study.

Observer	Reference SNR	Mean val.	Filtered SNR	Mean val.	Large pixel SNR	Mean val.
Ideal	4.259	N/A	4.070	N/A	1.255	N/A
Obs. 1	(0.828, 1.655)	1.190	(2.156, ∞)	2.660	(0.236, 0.970)	0.583
Obs. 2	(0.928, 1.780)	1.295	(2.667, ∞)	3.290	(0.644, 1.432)	0.999
Obs. 3	(1.339, 2.353)	1.735	(2.358, ∞)	2.904	(0.688, 1.485)	1.045
Obs. 4	(1.089, 1.993)	1.466	Unknown	Unknown	(0.314, 1.055)	0.661
Obs. 5	(1.636, 1.843)	1.735	(2.667, ∞)	3.290	(0.474, 1.235)	0.824
Obs. 6	(1.033, 1.918)	1.406	(2.667, ∞)	3.290	(0.644, 1.432)	0.999
Obs. 7	(0.979, 1.848)	1.349	(2.156, ∞)	2.660	(0.515, 1.283)	0.867
Obs. 8	(1.870, 2.134)	1.993	(2.882, 3.371)	3.082	(0.761, 0.953)	0.855

observer studies. In particular, whereas Abbey and Barrett⁷ approximate regularization in tomographic reconstruction by applying a discrete Fourier transform to a white noise image, apodizing, and applying the corresponding inverse transform, the noise structure in this study is the result of applying a full tomographic reconstruction algorithm. It is reasonable, therefore, that the precise efficiency results for humans detecting a signal in regularized noise could differ somewhat between the two investigations.

While the results of the regularization study provide insight into a case where the upper bound provided by the HO moves in the direction opposite to the mean human observer performance, the case of large pixels, in some sense, provides a simpler result. Namely, decreasing the number of image pixels destroys information in the reconstruction by increasing the null-space of the reconstruction matrix A , and this loss of information appears to have no underlying benefit to the human observer. Rather, the impairment of the HO in the case of larger image pixels is mirrored by a similar impairment of human observer performance, suggesting that the modeling of human performance in this case may not need to be as complex as in the case of regularization.

V. CONCLUSION

The purpose of this work has been to evaluate human observer performance for a signal-detection task in fan-beam CT and to compare the results to the HO performance. In order to establish a clear upper limit on detectability for each case, HO SNR values were computed, using knowledge of the statistics underlying the signal-present and signal-absent hypotheses. Meanwhile, human observer performance was computed using a 2AFC experiment, wherein approximate AUC values were obtained. These values were compared to the corresponding AUC values for the HO. Further, the detectability indexes d_A for each observer and reconstruction implementation were computed from the approximate AUC values. These detectability estimates were then also related to Hotelling SNR.

Three cases for reconstruction were considered. First, a reference case of unfiltered, ramp-spectrum noise was investigated. Next, heavy regularization was implemented by means

of a Hanning filter in the fan-beam FBP algorithm. Finally, the reconstructed pixel size was increased. The last two implementations introduce signal detectability loss by virtue of the fact that they expand the null-space of the reconstruction matrix which constitutes the discrete implementation of the FBP algorithm.

While the regularization impairs the performance of the HO, it was shown to result in a marked improvement for the human observers, making their performance statistically indistinguishable from that of the ideal observer. The general trend of this result is consistent with the results of Abbey and Barrett.⁷ Increasing the reconstructed image pixel size, however, degraded both HO and human observer performance. Future work could investigate the optimum degree of regularization in the reconstruction algorithm with respect to human observer performance.

ACKNOWLEDGMENTS

This work was supported in part by National Institutes of Health (NIH) R01 Grant Nos. CA120540, EB000225, and CA158446 and NIH T32 Grant No. EB002103-22. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

^{a)}Electronic mail: adrianalberto@uchicago.edu

^{b)}Electronic mail: sidky@uchicago.edu

^{c)}Also at Department of Cellular and Radiation Oncology, The University of Chicago, Chicago, Illinois 60637.

¹H. H. Barrett and K. J. Myers, *Foundations of Image Science* (Wiley, Hoboken, NJ, 2004).

²S. J. LaRoque, E. Y. Sidky, D. C. Edwards, and X. Pan, "Evaluation of the channelized Hotelling observer for signal detection in 2D tomographic imaging," *Proc. SPIE* **6515**, 651514 (2007).

³E. Y. Sidky, S. J. LaRoque, and X. Pan, "Accurate computation of the Hotelling template for SKE/BKE detection tasks," *Proc. SPIE* **69170W**, 69170W (2008).

⁴E. Y. Sidky and X. Pan, "In-depth analysis of cone-beam CT image reconstruction by ideal observer performance on a detection task," *NSS/MIC Conference Record*, IEEE, 2008.

⁵S. Park, J. M. Witten, and K. J. Myers, "Singular vectors of a linear imaging system as efficient channels for the Bayesian ideal observer," *IEEE Trans. Med. Imaging* **28**(5), 657–668 (2009).

- ⁶A. Wunderlich and F. Noo, "Image covariance and lesion detectability in direct fan-beam x-ray computed tomography," *Phys. Med. Biol.* **53**, 2471–2493 (2008).
- ⁷C. 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**(3), 473–488 (2001).
- ⁸H. L. V. Trees, *Detection, Estimation, and Modulation Theory, Part I* (Academic, New York, 1968).
- ⁹A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging* (IEEE, New York, 1988).
- ¹⁰J. Beutel, H. L. Kundel, and R. L. V. Metter, *Handbook of Medical Imaging: Physics and Psychophysics* (SPIE, Bellingham, Washington, 2000), Vol. 1.
- ¹¹J. Witten, S. Park, and K. Myers, "Partial least squares: A method to estimate efficient channels for the ideal observers," *IEEE Trans. Med. Imaging* **29**(4), 1050–1058 (2010).
- ¹²J. Nocedal and S. Wright, *Numerical Optimization* (Springer, New York, 1999).
- ¹³A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes* (McGraw-Hill, New York, 2002).