Modified ideal observer model (MIOM) for high-contrast and high-spatial resolution CT imaging tasks

Juan P. Cruz-Bastida, Daniel Gomez-Cardona, and John Garrett
Department of Medical Physics, University of Wisconsin School of Medicine and Public Health, 1111 Highland Avenue, Madison, WI 53705, USA

Timothy Szczykutowicz
Department of Medical Physics, University of Wisconsin School of Medicine and Public Health, 1111 Highland Avenue, Madison, WI 53705, USA
Department of Radiology, University of Wisconsin School of Medicine and Public Health, 600 Highland Avenue, Madison, WI 53792, USA
Department of Biomedical Engineering, University of Wisconsin-Madison, 1550 Engineering Drive, Madison, WI 53706, USA

Guang-Hong Chen[a] and Ke Li[a]
Department of Medical Physics, University of Wisconsin School of Medicine and Public Health, 1111 Highland Avenue, Madison, WI 53705, USA
Department of Radiology, University of Wisconsin School of Medicine and Public Health, 600 Highland Avenue, Madison, WI 53792, USA

Purpose: Although a variety of mathematical observer models have been developed to predict human observer performance for low contrast lesion detection tasks, their predictive power for high contrast and high spatial resolution discrimination imaging tasks, including those in CT bone imaging, could be limited. The purpose of this work was to develop a modified observer model that has improved correlation with human observer performance for these tasks.

Methods: The proposed observer model, referred to as the modified ideal observer model (MIOM), uses a weight function to penalize components in the task function that have less contribution to the actual human observer performance for high contrast and high spatial resolution discrimination tasks. To validate MIOM, both human observer and observer model studies were performed, each using exactly the same CT imaging task [discrimination of a connected component in a high contrast (1000 HU) high spatial resolution bone fracture model (0.3 mm)] and experimental CT image data. For the human observer studies, three physicist observers rated the connectivity of the fracture model using a five-point Likert scale; for the observer model studies, a total of five observer models, including both conventional models and the proposed MIOM, were used to calculate the discrimination capability of the CT images in resolving the connected component. Images used in the studies encompassed nine different reconstruction kernels. Correlation between human and observer model performance for these kernels were quantified using the Spearman rank correlation coefficient (ρ). After the validation study, an example application of MIOM was presented, in which the observer model was used to select the optimal reconstruction kernel for a High-Resolution (Hi-Res, GE Healthcare) CT scan technique.

Results: The performance of the proposed MIOM correlated well with that of the human observers with a Spearman rank correlation coefficient ρ of 0.88 ($P = 0.003$). In comparison, the value of ρ was 0.05 ($P = 0.904$) for the ideal observer, 0.05 ($P = 0.904$) for the non-prewhitening observer, -0.18 ($P = 0.634$) for the non-prewhitening observer with eye filter and internal noise, and 0.30 ($P = 0.427$) for the prewhitening observer with eye filter and internal noise. Using the validated MIOM, the optimal reconstruction kernel for the Hi-Res mode to perform high spatial resolution and high contrast discrimination imaging tasks was determined to be the HD Ultra kernel at the center of the scan field of view (SFOV), or the Lung kernel at the peripheral region of the SFOV. This result was consistent with visual observations of nasal CT images of an *in vivo* canine subject.

Conclusion: Compared with other observer models, the proposed modified ideal observer model provides significantly improved correlation with human observers for high contrast and high spatial resolution CT imaging tasks. © *2017 American Association of Physicists in Medicine* [https://doi.org/10.1002/mp.12404]

1. INTRODUCTION

Hardware and software advances in modern multidetector detector CT (MDCT) provide new means to improve spatial resolution for MDCT systems. The improvement in spatial resolution may benefit many clinical applications that demand high-spatial resolution, such as the diagnosis of bone fractures or middle and inner ear disease using CT. The High-Resolution (Hi-Res) scan mode (GE Healthcare, Waukesha, WI, USA) is a technique to improve spatial resolution in MDCT. It uses in-plane focal spot deflection, an increased angular sampling rate, and dedicated Hi-Definition (HD) reconstruction kernels to improve the in-plane spatial resolution of CT.[1–4] However, improved spatial resolution does not necessarily mean improved diagnostic performance, since the ultimate diagnostic performance is jointly determined by spatial resolution, noise power, imaging tasks, and the performance of the image reader. Therefore, to optimally use these new technologies in clinical practice to accomplish a clinical diagnostic task, one must comprehensively study the potential tradeoffs between spatial resolution and noise, as well as their joint impact on the overall diagnostic performance for a specific imaging task. Ultimately, it is the human observer who reads, and infers, diagnostic information from a given image for a given task. Therefore, it would be ideal to evaluate the performance of a new technology for a given task by human observers. However, this is usually a time- and resource-consuming process and thus is not easy to implement for every clinical tasks. To mathematically model the human perception process, mathematical observer models were developed to incorporate the quantitative imaging system characteristics into a single figure of merit (FOM) to predict the performance of human observers. When these observer models are properly validated, they can be used to optimize the needed imaging technologies and the associated scanning parameters for a given clinical tasks. The main purpose of this paper was to develop and validate such an observer model for high-contrast and high-spatial resolution discrimination task.

To be more specific, an observer model is a mathematical algorithm that is capable of making decisions about specific visual tasks given an image and *a priori* statistical information.[5–9] Performance of the observer model in decision making can be quantitated by a FOM known as the detectability index d', which can be analytically calculated from the spatial resolution and noise properties of the imaging system as well as the imaging task itself. Compared with other image quality FOMs such as the modulation transfer function (MTF), the observer model-derived d' has the advantage of being task-dependent in addition to taking both spatial resolution and noise into account. Therefore, observer models have been widely used in the development and assessment of many imaging modalities, including mammography,[10–12] dual-energy digital radiography,[13,14] digital tomosynthesis,[15–20] CT,[21–34] cone beam CT[17,18,31,35–38] and differential-phase contrast CT.[39,40]

A variety of observer models have previously been developed, and they differ from each other in the number and nature of hypotheses about the human visual perception that are integrated into the model to obtain a quantitative response variable. Although these observer models have been well validated for low-contrast imaging tasks, a direct application of these observer models to high-contrast and high-spatial resolution discrimination imaging tasks may lead to poor correlation with human observer performance. Therefore, a modified observer model must be developed and validated for high-contrast and high-spatial resolution discrimination tasks.

To facilitate reading, the paper is laid out as follows: First, to demonstrate the failure of conventional observer models for the high-contrast, high-spatial resolution discrimination tasks, human observer studies were performed in this work, and the results were compared with d' generated by several popular observer models. Next, a modified observer model with significantly improved correlation with human observers for the high-contrast and high-spatial resolution task was presented and validated. Finally, the observer model was used to choose the optimal reconstruction kernel for the Hi-Res scan mode as a demonstration of its potential application.

2. METHODS AND MATERIALS

2.A. High-contrast and high-spatial resolution imaging task

The high-contrast and high-spatial resolution CT imaging task used in this work is shown in Fig. 1. The task is to discriminate two hypotheses: the null hypothesis h_0 is a straight 0.3-mm-wide gap in a range-shaped object (inner and outer diameters are 10 and 17 mm, respectively); for the alternative
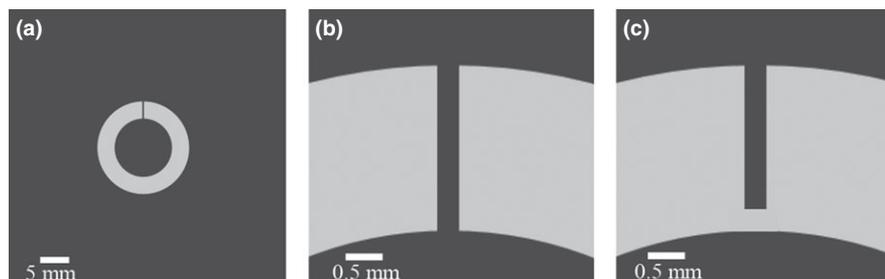

FIG. 1. High-contrast and high-spatial resolution imaging task used in this work. (a) Null hypothesis (h_0). (b) Close-up of the completely disconnected gap in the null hypothesis. (c) Close-up at the connected region across the gap in the alternative hypothesis (h_1).

hypothesis h_1, a 0.3×0.3 mm^2 component connects a portion of the gap. For both hypotheses, the contrast between the ring and the background is 1000 HU. In the spatial domain, this signal-known-exactly, background-known exactly, and binary discrimination task can be written as

$$t(\mathbf{x}) = h_1(\mathbf{x}) - h_0(\mathbf{x}), \tag{1}$$

and the corresponding form in the spatial frequency domain is given by

$$T(\mathbf{k}) = \mathrm{FT}[h_1(\mathbf{x}) - h_0(\mathbf{x})]. \tag{2}$$

This high-spatial resolution imaging task was used in both the observer model and human observer studies, so that results of these studies could be directly compared.

2.B. Human observer studies

2.B.1. Image preparation

To match the experimental conditions between the human observer and observer model studies while having precise control and fine-tuning of the imaging task, the following image preparation approach was used: Both types of observer studies used the same set of experimental noise-only CT images. For the human observer studies, the experimental noise-only images were fused to one of the two hypotheses shown in Fig. 1. For the observer model studies, the experimental noise-only images were used to measure the noise power spectrum (NPS).

To acquire the noise-only CT images, a Catphan 600 phantom (Module CTP591, The Phantom Laboratory, Greenwich, NY, USA) was repeatedly scanned using a 64-slice diagnostic MDCT system (Discovery CT750 HD, GE Healthcare, Waukesha, WI, USA) with the following scan parameters: 120 kV, 400 mAs, axial mode, small focal spot (1.0×0.7 mm), large body bowtie filter, and 20 mm detector collimation width. Both conventional and Hi-Res scan modes were used; for each scan mode, 80 repeated scans were performed. Images acquired using the conventional scan mode were reconstructed using four different kernels, while images acquired using the Hi-Res mode were reconstructed using five high-definition (HD) kernels (Table I). The reconstructed images have a slice thickness of 5 mm, pixel size of 0.1 mm, and a matrix size of 512×512. For each scan mode and reconstruction kernel, the ensemble average across the 80 repeated scans was calculated, which was then used to detrend the background and obtain noise-only images.

TABLE I. Scan modes and reconstruction kernels included in both the human observer and observer model studies.

Conventional mode	Hi-Res mode
Standard	HD Standard
Lung	HD Lung
Bone plus	HD Bone Plus
Edge	HD Ultra
	HD Edge

To prepare images for the reader studies, the imaging task and experimental noise-only background were fused in the following approach: First, each of the two hypotheses in Fig. 1 was filtered by convolving with a point spread function (PSF), which was experimentally measured using the same scan technique and reconstruction kernel as those used to generate the noise-only background.[41] This process can be described by

$$h_j'(\mathbf{x}) = h_j(\mathbf{x}) \otimes \mathrm{PSF}(\mathbf{x}), \tag{3}$$

where $\otimes$ denotes the 2D convolution operator, j took values of 0 and 1 for the null and alternative hypothesis, respectively. Next, the filtered hypothesis was added to each (i^{th}) of the 80 independent realizations of the noise background, namely

$$g^i(\mathbf{x}) = h_j' + n^i(\mathbf{x}), \tag{4}$$

where $i \in [1,80]$. This image fusion process was repeated for each of the nine scan model-kernel combinations listed in Table I, producing a total of $9 \times 80 = 720$ images for the human reader studies. Figure 2 shows some examples of these images.

In order to confirm the hybrid images accurately represented true experimental CT images, we compared hybrid and experimental images of a bar pattern in the CTP528 Module of the Catphan 600 phantom, generated with different reconstruction kernels under the conventional scan mode. As illustrated by Fig. 3, for a given reconstruction kernel, the hybrid image closely resembled the true experimental CT image.

2.B.2. Image reading and scoring

Three medical physicists with 7–14 yr of experience in clinical CT research served as human observers in the study. The reading was conducted in a darkened room using a monitor calibrated for clinical diagnosis. The image display window and level were 2000 and 300 HU, respectively. The observers were instructed to keep a viewing distance of approximately 60 cm from the monitor. Each observer first underwent a training session, in which images and the true state of the hypothesis (either 0 or 1) were presented to the observer, so that the observer had a basic idea of how the image looks like for each hypothesis, each scan mode, and each reconstruction kernel. A total of 18 images (two for each scan mode-reconstruction kernel combination) were used for the training.

After the training session, each observer read a total of 360 images (40 for each scan mode-reconstruction kernel combination) fetched from the pool of 720 images. To reduce observer bias, the order of the images was randomized, and a black screen was displayed between two consecutive images. The observer scored each image using the 5-point Likert scale listed in Table II. The recorded scores were analyzed using a multireader and multicase receiver operating characteristic software (OR-DBM MRMC 2.5, The University of
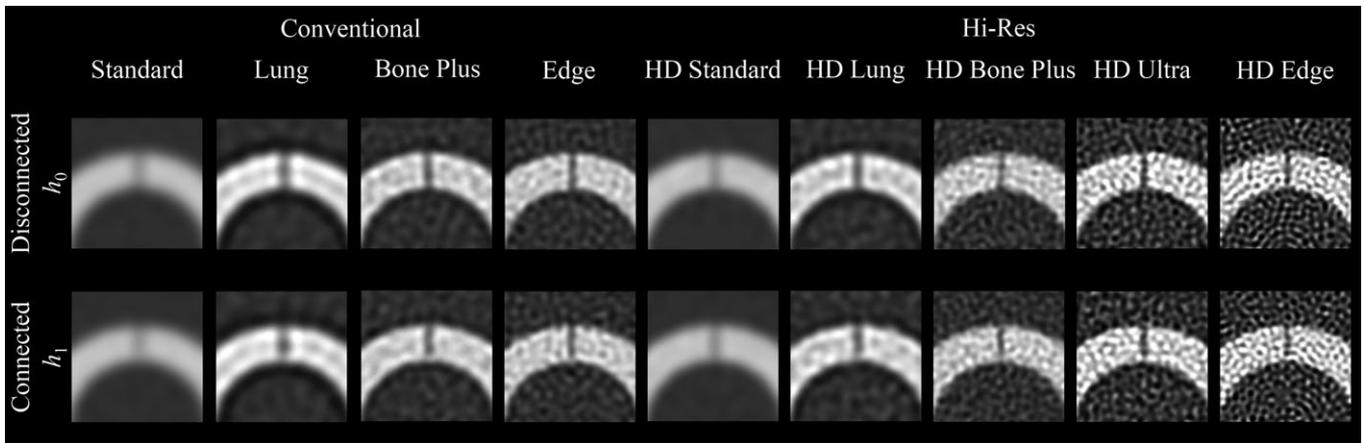
FIG. 2.   Close-up images used in the human observer study. One representative image was selected for each of the nine scan mode-reconstruction kernel combinations and the two hypotheses. The display window and level are 2000 and 300 HU, respectively.
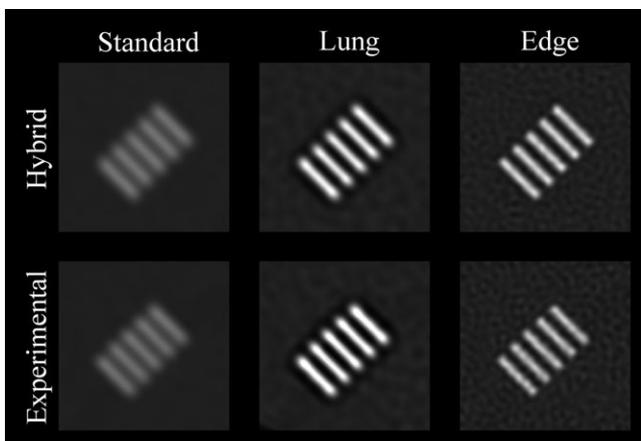

FIG. 3.   Validation of the hybrid image generation method. The top row shows images of a bar pattern (6 lp/cm) generated by the hybrid method (simulated signal + experimental noise background). The bottom row shows experimental images of the bar pattern. The image display window and level are 2000 and 300 HU, respectively.

TABLE II.   5-point image rating scale used in the human observer study.

Score	Criteria
1	The gap is definitely disconnected
2	The gap is probably disconnected
3	Unclear if the gap is disconnected or connected
4	The gap is probably connected
5	The gap is definitely connected

Iowa).[42–46] The interobserver consistency was estimated with the Cohen's kappa coefficient.

2.C. Observer model studies

2.C.1. Conventional observer models

To demonstrate that conventional observer models have poor correlation with human observers for the high-contrast and high-spatial resolution imaging task, a total of five conventional observer models were used to estimate the observer performance for the task.

We first evaluated the ideal observer model, which uses the Bayes theorem to combine all available *a priori* information with the image data to obtain probabilistic estimates of the correctness of a given hypothesis (signal present or absent).[6,7,47] Its performance can be quantitated, without being dependent on the observer's decision criterion, using the ideal observer detectability index d'_{ideal}. The square of d'_{ideal} can be calculated in the Fourier domain as follows:[47]

$$
\begin{aligned}
\left(d'_{ideal}\right)^2 &= \int\int \frac{|T(\mathbf{k})|^2 \mathrm{MTF}^2(\mathbf{k})}{N(\mathbf{k})}\, \mathrm{d}^2\mathbf{k} \\
&= \int\int \frac{|\mathrm{FT}[t(\mathbf{x}) \otimes \mathrm{PSF}(\mathbf{x})]|^2}{N(\mathbf{k})}\, \mathrm{d}^2\mathbf{k}.
\end{aligned}
\tag{5}
$$

In Eq. (5), $\mathbf{k}$ and $\mathbf{x}$ are frequency and spatial vectors, respectively, N denote the NPS and was calculated from the ensemble of experimental noise-only images. A larger d'_{ideal} value indicates better observer performance. In other words, one should expect a strong and positive correlation between d'_{ideal} and human observer performance metrics (such as AUC) if the observer model can accurately predict human observer behavior.

In addition to the ideal observer, several other popular observer models were also used in this work, including pre-whitening observer with eye filter and internal noise (PWEi), nonprewhitening (NPW) observer, and nonprewhitening observer with eye filter and internal noise (NPWEi). The eye filter and internal noise used in PWEi and NPWEi take into account of the nonuniform response of the visual system to different spatial frequencies and the internal fluctuation of the visual system. The parameter c in the eye filter formula, $E(f) = f \exp(-cf)$, was selected such that the maximum response of E occurred at 4 cycles/deg[18,48] for a viewing distance of 70 cm, $c = 3.1$. The static internal noise was implemented as white noise given by $N_{int} = 0.02N_0L^2$, where $L = 0.70$ is the viewing distance in meters, and N_0 is the amplitude of the white noise-equivalent NPS of the

background, given by $N_0 = \int \int N(\mathbf{k})d^2\mathbf{k}/\int \int d^2\mathbf{k}$. The induced internal noise was included by multiplying the NPS by a scaling factor of $(1+\varphi)$, where $\varphi = 0.3$.[7,10,49]

2.C.2. Proposed modified ideal observer model (MIOM)

The ideal observer gives the upper bound of human observer detection performance, as it assumes the human observer can perfectly decorrelate (i.e., prewhiten) the image noise. One of the fundamental assumptions behind the ideal observer is that all the spatial components of the input signal contribute with equal weight to the observer's discrimination performance [Eq. (5)]. While this assumption is valid for low-contrast signal detection tasks, it may be violated for high-contrast and high-spatial resolution discrimination tasks, in which the importance of signal recognition may bias certain spatial components of the signal, for example, high-contrast edges. Therefore, the following modified ideal observer model (MIOM) was proposed in this work:

$$\left(d'_{\mathrm{MIOM}}\right)^2 = \int \int \frac{\left|\mathrm{FT}[t'(\mathbf{x}) \cdot w(\mathbf{x})]\right|^2}{N(\mathbf{k})} d^2\mathbf{k}. \tag{6}$$

The weight function $w(\mathbf{x})$ aims to penalize regions of the signal that are likely to have less contribution to the overall observer performance for high-contrast and high-spatial resolution discrimination tasks. The proposal is based on two assumptions about the human vision that were previously introduced:[50,51] (a) the spatial integration of image information is often limited to a spatial window; (b) the observer generates a tailored discriminator for the signals in question using available information. Throughout the paper this model is referred to as modified ideal observer model (MIOM).

The weighting function used to calculate d'_{MIOM} is shown in Fig. 4(a); it is defined as the alternative hypothesis itself (h_1) (defined in next subsection) bounded by a rectangular window with softened edges and centered on the task function. Both the width of the window and the degree of edge

blurring are optimized, such that the resulting detectability index is as close as possible to the actual human observer performance. The weight function design followed two objectives: (a) to put reduced or zero weight on pixels far away from the task function, since when an actual human observer performs a recognition task, he/she is likely to primarily focus on a region whose size is scaled according to the actual dimensions of the imaging task; (b) to penalize the signal blurring due to the limited spatial resolution of the imaging system, as illustrated in Figs. 4(b)–4(f). The necessity of using this weighting function was confirmed by the results of the human observer experiments in this work.

2.D. Quantitation of correlation between human observer and observer model

The correlation between the human observer performance, described by the AUC, and the observer model performance, described by d', was quantitated using the Spearman rank correlation test, which is a measure of statistical dependence between the ranking of two variables, and it assesses how well the relationship between two variables can be described using a monotonic function, whether their relationship is linear or not.

2.E. Example application of MIOM

After confirming satisfactory correlation between human observer and the proposed MIOM, it was used to select the kernel-scan mode combination that led to the highest performance for a high-contrast, high-spatial resolution CT imaging task. As illustrated by Table III, the introduction of the Hi-Res mode to the 750HD MDCT system provides an additional 15 scan mode-kernel combinations, in addition to the eight kernels provided by previous GE scanner models. Previous studies also demonstrated that the spatial resolution performances of CT systems may also depend on the object position within the scan field of view (SFOV). Optimizing the kernel and scan mode selection for each object position
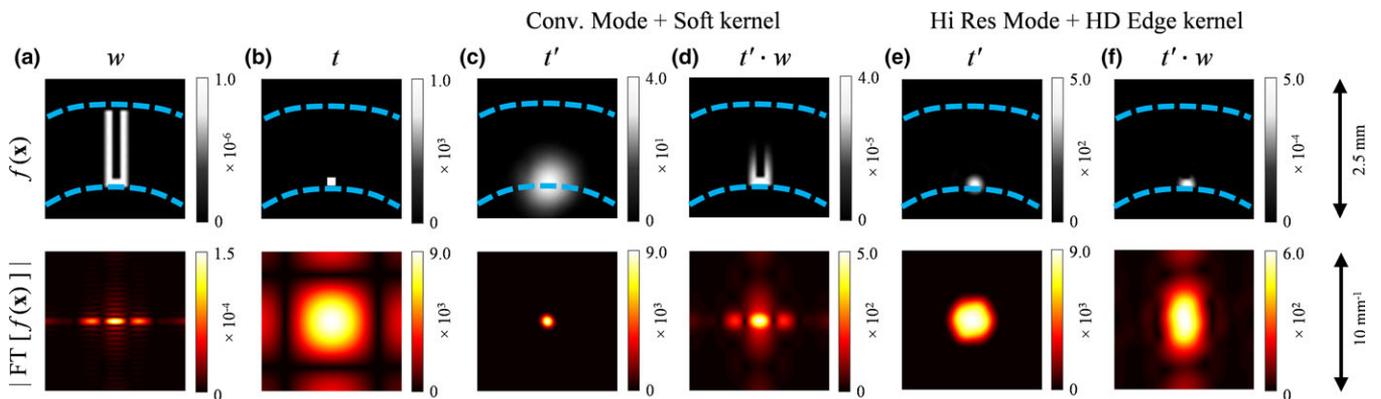

FIG. 4. Functions of interest in spatial and frequency domain representations. (a) Weight model used in the proposed MIOM observer. (b) Task function. (c) and (e) show the task function filtered by the PSF, that is, t', of two different scan modes and reconstruction kernels at the centered position. (d) and (f) show t' weighted by the function w for two different scan modes and reconstruction kernels at the centered position. The dashed lines in the spatial domain images indicate the position of each function relative to h_0 and h_1 (shown in Fig. 1). The units for the spatial domain and frequency domain color bars are HU and HU mm^2, respectively. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE III. Scan modes and the associated reconstruction kernels provided by the GE Discovery CT750 HD CT system.

Conventional mode		Hi-Res mode	
Non-HD kernels	HD kernels	Non-HD kernels	HD kernels
Soft		Soft	HD Standard
Standard		Standard	HD Detail
Detail		Detail	HD Lung
Lung		Lung	HD Bone
Chest		Chest	HD Bone Plus
Bone		Bone	HD Ultra
Bone Plus		Bone Plus	HD Edge
Edge		Edge	

would be time consuming to a human observer. In contrast, the MIOM provides an efficient way to rank the order of kernel-scan mode combinations for nasal cavity CT imaging. In this work, the ranking was performed for each of the eight object positions shown in Fig. 5. To corroborate ranking provided by the MIOM, an in vivo canine subject (2-yr-old 10-kg male adult beagle) was scanned under an Institutional Animal Care and Use Committee (IACUC)-approved protocol. Both conventional and Hi-Res scan modes were used, and other acquisition parameters are listed as follows: tube potential = 80 kV, beam collimation = 20 mm, medium body bowtie filter, axial mode, rotation time = 0.5 s, tube current = 430 mA, and large focal spot. Images were reconstructed with a display field of view of 5 cm, a slice thickness of 0.625 mm, and different reconstruction kernels.
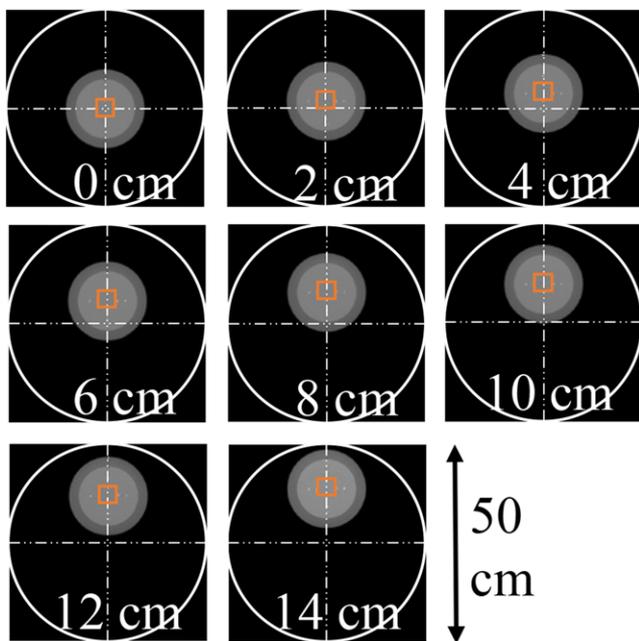

FIG. 5. The proposed MIOM was used to optimize reconstruction kernel and scan mode selection for eight different off-centered positions shown in this figure. The ROI used to calculate d'_{MIOM} was indicated by the solid square. [Color figure can be viewed at wileyonlinelibrary.com]

3. RESULTS

3.A. Human observer results

Each observer took approximately 5–10 s per image including both reading and scoring. The measured areas under the ROC curve (AUCs) for individual observer, scan mode, and reconstruction kernels are listed in Table IV. The Cohen's kappa coefficient was found to be $\kappa = 0.22$, which indicates "fair" interobserver agreement. The average AUC across the three human observers is plotted in Fig. 6. The error bars indicate standard error. Among the nine scan mode-reconstruction kernel combinations, the conventional scan mode with the Standard kernel led to the lowest AUC, since it generated the most blurred image as shown in Fig. 2. The AUC was improved by using the combination of the Hi-Res scan mode and the HD Standard kernel, which improves spatial resolution.[41] Meanwhile, the most aggressive edge-preserving kernel HD Edge did not generate the highest human observer performance. Instead, its AUC is similar to that of the HD Standard kernel. This is due to the dramatic noise amplification associated with this kernel, despite the fact that it led to the highest MTF.[41] This example shows the importance of taking noise and task-based overall image quality into account when evaluating the HD kernels, even if they are designed with high-spatial resolution imaging tasks in mind. Figure 6 also shows that the highest human observer performance was achieved with intermediate kernels such as Bone Plus and HD Lung. These kernels generated higher spatial resolution compared with the Standard (or HD Standard) kernel, and they only led to moderate noise amplification that was still tolerable for the high-contrast and high-spatial resolution imaging task (Fig. 2).

3.B. Correlation between human observers and observer model results

Average AUC values generated by the human observers and the observer models were compared using scatter plots shown in Fig. 7. The ideal observer led to a Spearman's rank correlation coefficient (ρ) of only 0.05 ($P = 0.904$), and the

TABLE IV. Human observer ROC performance for the high-spatial resolution imaging task. AUC: area under the ROC curve; SE: standard error.

| Scan mode	Kernel	Observer 1		Observer 2		Observer 3	
		AUC	SE	AUC	SE	AUC	SE
Conventional	Standard	0.55	0.03	0.53	0.02	0.69	0.06
Conventional	Lung	0.86	0.05	0.81	0.06	0.90	0.04
Conventional	Bone Plus	0.89	0.05	0.89	0.05	0.92	0.05
Conventional	Edge	0.92	0.04	0.75	0.07	0.72	0.08
Hi-Res	HD Std	0.81	0.05	0.58	0.06	0.73	0.08
Hi-Res	HD Lung	0.89	0.06	0.85	0.06	0.92	0.04
Hi-Res	HD Bone Plus	0.78	0.07	0.69	0.08	0.82	0.07
Hi-Res	HD Ultra	0.94	0.04	0.75	0.07	0.80	0.07
Hi-Res	HD Edge	0.69	0.08	0.67	0.07	0.74	0.08
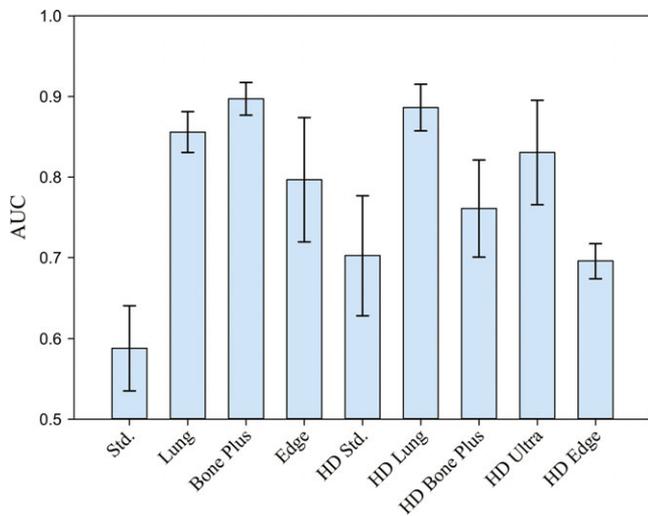
FIG. 6.   Average human observer AUCs for different reconstruction kernels. [Color figure can be viewed at wileyonlinelibrary.com]
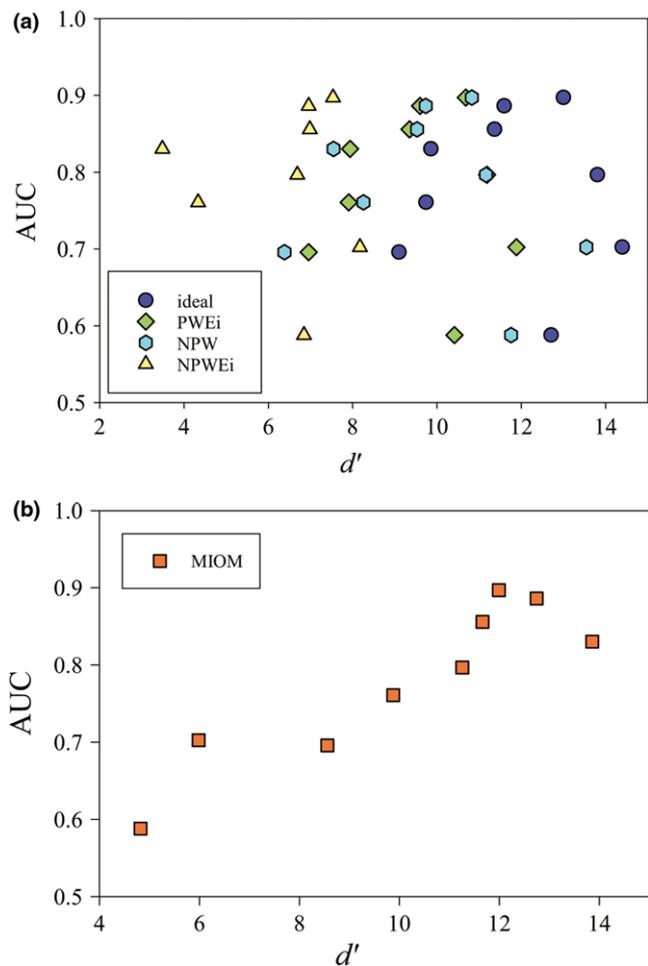

FIG. 7.   Correlation between human observer and observer models. (a) AUC values paired with d' for the four conventional observer models. (b) AUC values paired with d' for the proposed MIOM. [Color figure can be viewed at wileyonlinelibrary.com]

NPW observer actually generated a weak negative correlation ($\rho = -0.18$, $P = 0.0634$). Here, the null hypothesis for the P value is that, the human observer AUC and observer model d'

TABLE V. Spearman rank correlation values for different observer models.

Observer model	ρ	P
Ideal	0.05	0.904
PWEi	0.05	0.904
NPW	−0.18	0.634
NPWEi	0.30	0.427
MIOM	0.88	0.003

are not correlated. Correlation coefficients for other conventional observer models are listed in Table V and are no greater than 0.30. In comparison, MIOM achieved a much stronger correlation with the human observer ($\rho = 0.88$, $P = 0.003$).

3.C. Example application: optimal kernel selection

Having validated the proposed MIOM observer model, we used it to select optimal scan mode and reconstruction kernel as a function of object positions. Figure 8 provides a map of the measured detectability index (d'_{MIOM}) for the imaging task shown in Fig. 1. In general, d'_{MIOM} was very sensitive to the off-centered position; at severe off-centered positions, d'_{MIOM} dropped considerably for a given acquisition mode and reconstruction kernel. For the specific imaging task used in
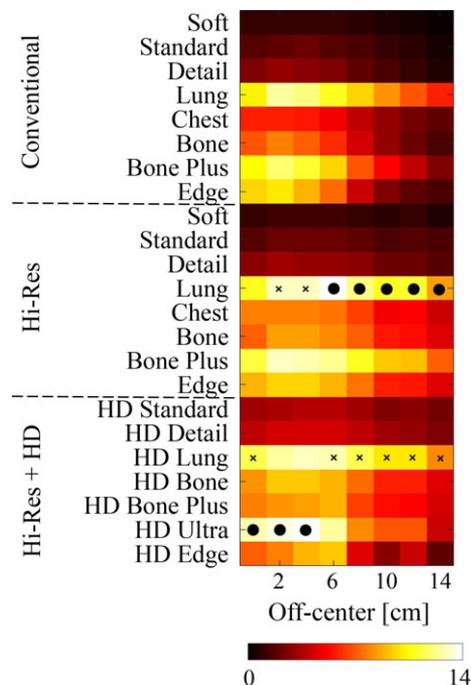

FIG. 8.   Summary of the detectability index (d'_{MIOM}) of the high-spatial resolution imaging task shown in Fig. 1 for all the scan modes, reconstruction kernels, and object positions. For a given object position, the particular scan mode-reconstruction kernel combination that achieved the highest detectability was marked with a black dot. The second highest scan mode-reconstruction kernel combination is also indicated, with a cross. The given color scale corresponds to values of d'_{MIOM}. [Color figure can be viewed at wileyonlinelibrary.com]
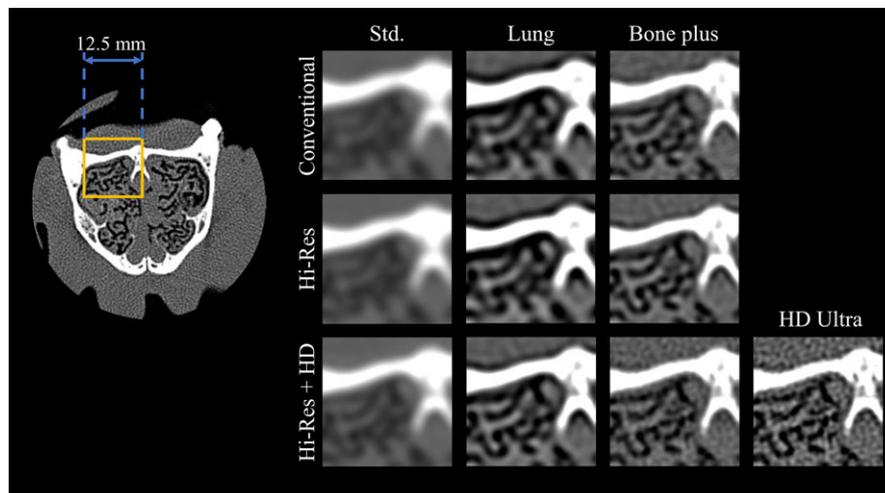
F<small>IG</small>. 9.   Axial images of the nasal cavity of the canine subject. ROI images are displayed for both conventional and Hi-Res modes combined with different kernels. The display window and level are 2000 and 300 HU, respectively. [Color figure can be viewed at wileyonlinelibrary.com]

this work, the Hi-Res scan mode combined with the Lung kernel generates the highest d'_{MIOM} values for most off-centered positions. At the iso-center, Hi-Res mode with the HD Ultra kernel generates the highest d'_{MIOM} value, which is consistent with the human observer results. The intermediate edge-preserving kernels such as Bone Plus and HD Bone Plus showed good balance between spatial resolution and noise. Although the Hi-Res scan mode with the HD Edge kernel led to the highest spatial resolution in all object positions,[41] the large increase of image noise generated by this scan mode-reconstruction kernel combination strongly degraded its overall performance for the high-contrast and high-spatial resolution imaging task.

Images of the nasal cavity of the canine subject generated using the two scan modes (conventional and Hi-Res) and different reconstruction kernels are shown in Fig. 9. Small nasal ducts (or meatuses) were resolved with different levels of detail in these images. The Hi-Res scan mode with the HD Ultra kernel allowed one to clearly identify the convoluted cartilage structure of the nasal conchae. Hi-Res mode with HD Bone Plus kernel also allowed a clear identification of such structures, although with less sharpness. Compared to the Hi-Res mode with HD kernels, Hi-Res mode alone (without HD kernels) led to images of slightly lower sharpness. However, noise levels in such images were considerably lower, which resulted in a better balance between spatial resolution and image noise. This was consistent with the optimization results provided by MIOM.

4. DISCUSSION

This work demonstrated that traditional observer models have limited capability in predicting human observer performance for high-contrast and high-spatial resolution CT imaging task. Therefore, it is necessary to modify conventional observer models for this type of task. By introducing a weight function to the ideal observer framework, a modified ideal

observer model (MIOM) was developed for high-contrast and high-spatial resolution task. Experimental results showed the improved correlation between the proposed observer model and human reader performance.

Results of the observer studies showed that even for high-spatial resolution imaging task, its performance is not entirely determined by spatial resolution, as the kernel with the best MTF did not necessarily lead to the highest observer performance. Noise amplification associated with those kernels may degrade the overall imaging performance. This explains why intermediate kernels such as the Lung and HD Lung kernel generated better performance for the task, since they maintained good balance between spatial resolution enhancement and noise amplification, particularly at off-centered positions.

An *in vivo* animal study was included in this work to corroborate predictions made based on MIOM. Results of the animal study were qualitatively consistent with the predictions. However, it is important to recognize that the animal study was only qualitative rather than quantitative, and observations should not be arbitrarily generalized until evaluated with more subjects, imaging tasks, radiation dose levels, and body sizes.

While the use of the MIOM has been justified by the results of the human observer experiments, there are some important limitations that need to be investigated in future studies: First, the proposed modification of the ideal observer was based on empirical assumptions rather than a formal derivation. In particular, the weight function was determined empirically and heuristically, and a more rigorous theoretical modeling of human observer performance for high-contrast and high-spatial resolution imaging tasks needs to be performed. Second, the observer model and human observer performances were compared in terms of rank correlation coefficient, which did not strictly describe the quantitative agreement for individual data pair. Third, only a single imaging task and a limited number of readers with non-clinical

training were used in this study. Additional validation studies should include a large number of imaging tasks and physician readers. Fourth, high-density metal objects such as metal implants may introduce streak artifacts to CT images. Like other observer models, the proposed MIOM does not take these artifacts into account. Finally, the discussion about MIOM assumed linear and shift invariant CT systems with linear reconstruction algorithms and kernels. Nonlinear CT systems, particularly those with iterative reconstruction (IR) algorithms, are beyond the scope of this work. Although IR is gaining popularity for low-contrast detection tasks, its clinical use for high-spatial resolution applications is yet to be fully explored. If a high-spatial resolution clinical application does involve the use of IR, the applicability of MIOM must be reexamined.

5. CONCLUSIONS

A modified ideal observer model (MIOM) was developed and validated to predict human observer performance for high-contrast and high-spatial resolution CT imaging tasks. Spatial resolution improving technologies such as the Hi-Res scan mode and the associated HD reconstruction kernels can potentially be better utilized in routine clinical practice under the guidance of MIOM.

ACKNOWLEDGMENTS

a)Authors to whom correspondence should be addressed. Electronic mails: gchen7@wisc.edu; kli23@wisc.edu.

REFERENCES

1. Sohval AR, Freundlich D. Plural source computed tomograhy device with improved resolution. U.S. Patent 4637040 A; 1986.
2. Lonn AH. Computed tomography system with translatable focal spot. U.S. Patent 5173852 A; 1990.
3. Hsieh J, Gard MA, Gravelle S. Reconstruction technique for focal spot wobbling. *Proc SPIE.* 1992;1652:175–182.
4. Hsieh J, Saragnese EL, Stahre JE, Dorri B, Kaufman J, Senzig RF. Methods and apparatus for x-ray imaging with focal spot deflection. U.S. Patent 7869571 B2; 2008.
5. Swets J. Measuring the accuracy of diagnostic systems. *Science.* 1988;240:1285–1293.
6. Barrett HH, Yao J, Rolland JP, Myers KJ. Model observers for assessment of image quality. *P Natl Acad Sci.* 1993;90:9758–9765.
7. Burgess AE. Visual signal detection with two-component noise: low-pass spectrum effects. *J Opt Soc Am A.* 1999;16:694–704.
8. Eckstein MP, Abbey CK. Model observers for signal-known-statistically tasks (SKS). *Proc SPIE.* 2001;4324:91–102.
9. Barrett HH, Myers KJ, Hoeschen C, Kupinski MA, Little MP. Task-based measures of image quality and their relation to radiation dose and patient risk. *Phys Med Biol.* 2015;60:R1–R75.
10. Burgess AE, Jacobson FL, Judy PF. Human observer detection experiments with mammograms and power-law noise. *Med Phys.* 2001;28:419–437.
11. Suryanarayanan S, Karellas A, Vedantham S, Ved H, D'Orsi CJ. Detection in compressed digital mammograms using numerical observers. *Proc SPIE.* 2003;5034:513–521.
12. Segui JA, Zhao W. Amorphous selenium flat panel detectors for digital mammography: Validation of a NPWE model observer with CDMAM observer performance experiments. *Med Phys.* 2006;33:3711–3722.
13. Richard S, Siewerdsen JH. Optimization of dual-energy imaging systems using generalized NEQ and imaging task. *Med Phys.* 2007;34:127–139.
14. Richard S, Siewerdsen JH, Tward DJ. NEQ and task in dual-energy imaging: from cascaded systems analysis to human observer performance. *Proc SPIE.* 2008;6913:691311 (12pp).
15. Yoon S, Gang JG, Tward DJ, Siewerdsen JH, Fahrig R. Analysis of lung nodule detectability and anatomical clutter in tomosynthesis imaging of the chest. *Proc SPIE.* 2009;7258:72581M (11pp).
16. Reiser I, Nishikawa RM. Task-based assessment of breast tomosynthesis: Effect of acquisition parameters and quantum noise. *Med Phys.* 2010;37:1591–1600.
17. Gang GJ, Tward DJ, Lee J, Siewerdsen JH. Anatomical background and generalized detectability in tomosynthesis and cone-beam CT. *Med Phys.* 2010;37:1948–1965.
18. Gang GJ, Lee J, Stayman JW, et al. Analysis of Fourier-domain task-based detectability index in tomosynthesis and cone-beam CT in relation to human observer performance. *Med Phys.* 2011;38:1754–1768.
19. Hu Y-H, Zhao W. A 3D linear system model for the optimization of dual-energy contrast-enhanced digital breast tomosynthesis. *Proc SPIE.* 2011;7961:79611C (9pp).
20. Hu Y-H, Zhao W. Experimental quantification of lesion detectability in contrast enhanced dual energy digital breast tomosynthesis. *Proc SPIE.* 2012;8313:83130A (10pp).
21. Wagner RF, Brown DG, Pastel MS. Application of information theory to the assessment of computed tomography. *Med Phys.* 1979;6:83–94.
22. LaRoque SJ, Sidky EY, Edwards DC, Pan X. Evaluation of the channelized Hotelling observer for signal detection in 2D tomographic imaging. *Proc SPIE.* 2007;6515:651514 (9pp).
23. Wunderlich A, Noo F. Image covariance and lesion detectability in direct fan-beam x-ray computed tomography. *Phys Med Biol.* 2008;53:2471–2493.
24. Richard S, Li X, Yadava G, Samei E. Predictive models for observer performance in CT: applications in protocol optimization. *Proc SPIE.* 2011;7961:79610H (6pp).
25. Leng S, Yu L, Chen L, Ramirez Giraldo JC, McCollough CH. Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain. *Proc SPIE.* 2012;8313:83131M (7pp).
26. Yu L, Leng S, Chen L, Kofler JM, Carter RE, McCollough CH. 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.* 2013;40:041908 (9pp).
27. Leng S, Yu L, Zhang Y, Carter R, Toledano AY, McCollough CH. Correlation between model observer and human observer performance in CT imaging when lesion location is uncertain. *Med Phys.* 2013;40:041908 (9pp).
28. Tseng H-W, Fan J, Kupinski MA, Sainath P, Hsieh J. Assessing image quality and dose reduction of a new x-ray computed tomography iterative reconstruction algorithm using model observers. *Med Phys.* 2014;41:071910.
29. Ott J, Becce F, Monnin P, Schmidt S, Bochud F, Verdun F. Update on the non-prewhitening model observer in computed tomography for the assessment of the adaptive statistical and model-based iterative reconstruction algorithms. *Phys Med Biol.* 2014;59:4047–4064.
30. Zhang Y, Leng S, Yu L, Carter RE, McCollough CH. Correlation between human and model observer performance for discrimination task in CT. *Phys Med Biol.* 2014;59:3389–3404.
31. Gang GJ, Stayman JW, Ehtiati T, Siewerdsen JH. Task-driven image acquisition and reconstruction in cone-beam CT. *Phys Med Biol.* 2015;60:3129–3150.
32. Solomon J, Wilson J, Samei E. Characteristic image quality of a third generation dual-source MDCT scanner: noise, resolution, and detectability. *Med Phys.* 2015;42:4941–4953.

33. Samei E, Richard S. Assessment of the dose reduction potential of a model-based iterative reconstruction algorithm using a task-based performance metrology. *Med Phys.* 2015;42:314–323.

34. Christianson O, Chen JJS, Yang Z, et al. An improved index of image quality for task-based performance of CT iterative reconstruction across three commercial implementations. *Radiology.* 2015;275:725–734.

35. Abbey CK, Boone JM. An ideal observer for a model of x-ray imaging in breast parenchymal tissue. In: Krupinski EA, ed. *Digital Mammography: 9th International Workshop, IWDM 2008 Tucson, AZ, USA, July 20–23, 2008 Proceedings.* Berlin, Heidelberg: Springer Berlin Heidelberg; 2008:393–400.

36. Boone JM, Packard NJ, Abbey CK. Ideal observer comparison between tomographic and projection x-ray images of the breast. In: Martí J, Oliver A, Freixenet J, Martí R.eds. *Digital Mammography: 10th International Workshop, IWDM 2010, Girona, Catalonia, Spain, June 16–18, 2010 Proceedings.* Berlin, Heidelberg: Springer Berlin Heidelberg; 2010:591–597.

37. Prakash P, Zbijewski W, Gang GJ, et al. Task-based modeling and optimization of a cone-beam CT scanner for musculoskeletal imaging. *Med Phys.* 2011;38:5612–5629.

38. Gang GJ, Zbijewski W, Webster Stayman J, Siewerdsen JH. Cascaded systems analysis of noise and detectability in dual-energy cone-beam CT. *Med Phys.* 2012;39:5145–5156.

39. Li K, Bevins N, Zambelli J, Chen G-H. Model observer and human observer performance studies in differential phase contrast CT. *Proc SPIE.* 2013;8668:866817 (7pp).

40. Li K, Garrett J, Chen G-H. Correlation between human observer performance and model observer performance in differential phase contrast CT. *Med Phys.* 2013;40:111905 (14pp).

41. Cruz-Bastida JP, Gomez-Cardona D, Li K, et al. Hi-Res scan mode in clinical MDCT systems: Experimental assessment of spatial resolution performance. *Med Phys.* 2016;43:2399–2409.

42. Dorfman DD, Berbaum KS, Metz C. Receiver operating characteristic rating analysis: Generalization to the population of readers and patients with the jackknife method. *Invest Radiol.* 1992;27:723–731.

43. Obuchowski NA, Rockette HE. Hypothesis testing of diagnostic accuracy for multiple readers and multiple tests an anova approach with dependent observations. *Commun Stat Simul Comput.* 1995;24:285–308.

44. Hillis SL, Obuchowski NA, Schartz KM, Berbaum KS. A comparison of the Dorfman-Berbaum-Metz and Obuchowski-Rockette methods for receiver operating characteristic (ROC) data. *Stat Med.* 2005;24:1579–1607.

45. Hillis SL. A comparison of denominator degrees of freedom methods for multiple observer ROC analysis. *Stat Med.* 2007;26:596–619.

46. Hillis SL, Berbaum KS, Metz CE. Recent developments in the Dorfman-Berbaum-Metz - procedure for multireader ROC study analysis. *Acad Radiol.* 2008;15:647–661.

47. Wagner RF, Brown DG. Unified SNR analysis of medical imaging systems. *Phys Med Biol.* 1985;30:489–518.

48. Burgess AE, Li X, Abbey CK. Visual signal detectability with two noise components: anomalous masking effects. *J Opt Soc Am A.* 1997;14:2420–2442.

49. Burgess AE, Colborne B. Visual signal detection. iv. observer inconsistency. *J Opt Soc Am A.* 1988;5:617–627.

50. Näsänen RE, Kukkonen HT, Rovamo JM. A window model for spatial integration in human pattern discrimination. *Invest Ophth Vis Sci.* 1995;36:1855–1862.

51. Burgess AE. High level visual decision efficiencies. In: Blakemore C, ed. *Vision.* New York, NY: Cambridge University Press; 1991:431–440, Cambridge Books Online.