Comparison of model and human observer performance for detection and discrimination tasks using dual-energy x-ray images

Samuel Richard

Department of Medical Biophysics, University of Toronto, Toronto, Ontario, M5G 2M9 Canada

Jeffrey H. Siewerdsen^{a)}

Department of Medical Biophysics, University of Toronto, Toronto, Ontario, M5G 2M9 Canada,

Ontario Cancer Institute, Princess Margaret Hospital, Toronto, Ontario, M5G 2M9 Canada,

Department of Radiation Oncology, University of Toronto, Toronto, Ontario, M5G 2M9 Canada,

Department of Otolaryngology-Head and Neck Surgery, University of Toronto, Toronto, Ontario,

M5G 2M9 Canada, Institute of Biomaterials and Biomedical Engineering, University of Toronto,

Toronto, Ontario, M5G 2M9 Canada, Institute of Medical Science, University of Toronto, Toronto,

Ontario, M5G 2M9 Canada

(Received 7 May 2008; revised 2 September 2008; accepted for publication 2 September 2008; published 16 October 2008)

Model observer performance, computed theoretically using cascaded systems analysis (CSA), was compared to the performance of human observers in detection and discrimination tasks. Dual-energy (DE) imaging provided a wide range of acquisition and decomposition parameters for which observer performance could be predicted and measured. This work combined previously derived observer models (e.g., Fisher-Hotelling and non-prewhitening) with CSA modeling of the DE image noise-equivalent quanta (NEQ) and imaging task (e.g., sphere detection, shape discrimination, and texture discrimination) to yield theoretical predictions of detectability index (d') and area under the receiver operating characteristic (A_Z). Theoretical predictions were compared to human observer performance assessed using 9-alternative forced-choice tests to yield measurement of A_Z as a function of DE image acquisition parameters (viz., allocation of dose between the low- and high-energy images) and decomposition technique [viz., three DE image decomposition algorithms: standard log subtraction (SLS), simple-smoothing of the high-energy image (SSH), and anti-correlated noise reduction (ACNR)]. Results showed good agreement between theory and measurements over a broad range of imaging conditions. The incorporation of an eye filter and internal noise in the observer models demonstrated improved correspondence with human observer performance. Optimal acquisition and decomposition parameters were shown to depend on the imaging task; for example, ACNR and SSH yielded the greatest performance in the detection of soft-tissue and bony lesions, respectively. This study provides encouraging evidence that Fourier-based modeling of NEQ computed via CSA and imaging task provides a good approximation to human observer performance for simple imaging tasks, helping to bridge the gap between Fourier metrics of detector performance (e.g., NEQ) and human observer performance. © 2008 American Association of Physicists in Medicine. [DOI: [10.1118/1.2988161](https://doi.org/10.1118/1.2988161)]

I. INTRODUCTION

The development of imaging systems benefits tremendously from the ability to model observer performance from first principles. It enables the development and optimization of medical imaging systems without the requirement of costly prototypes and time consuming human observer studies. Considerable progress has been made in extending resolution and noise analysis to measures of diagnostic accuracy.¹⁻⁵ Still, there is often a gap between basic physical metrics of detector performance [such as modulation transfer function (MTF), noise-power spectrum (NPS), and NEQ] and those that describe the performance of human observers. This work seeks to relate metrics of detector performance (specifically, NEQ, computed theoretically using cascaded systems analysis combined with a Fourier description of imaging task) to human observer performance (measured by alternative forced-choice tests) for a variety of simple imaging tasks over a broad range of imaging conditions.

Whereas detectability in conventional chest radiographs is believed to be limited by anatomical background noise,⁶ DE images significantly reduce this effect. While DE decomposition does not completely remove anatomical background noise, it has been shown to significantly diminish background noise associated with overlying anatomy.⁷ For example, previous work modeling the anatomical background as power-law ($1/f^\beta$) noise quantified the reduction of anatomical background in DE images.⁸ Such work also quantified the amplification of quantum noise in DE imaging, motivating the development of CSA models of quantum noise for DE imaging and the behavior of noise reduction algorithms to better understand the fundamental limits of DE imaging performance.

Dual-energy (DE) imaging with a flat-panel detector (FPD) is a promising technology for improving subtle lesion conspicuity by reducing overlying clutter without increasing dose to the patient and has been the focus of recent work by

several authors.^{9–14} Furthermore, DE imaging presents a wide range of acquisition and decomposition parameters that significantly affect image quality. For purposes of the current work, therefore, DE imaging served as a useful platform for comparing performance as predicted theoretically and as measured with human observers over a wide range on imaging conditions. This work is also motivated by the extension of cascaded systems analysis for modeling observer performance. Such work enables the optimization of DE imaging technique (as investigated previously⁸) as well as the ability to describe observer performance from first principles in terms of CSA.

In the work below, an introduction is first provided for the modeling of DE imaging performance using cascaded systems analysis (CSA) based on previous work,^{7,8,15} followed by a brief overview of the Fisher-Hotelling (FH) and non-prewhitening (NPW) model observers and their extension to include an eye filter and internal noise as described by Burgess.¹⁶ Secondly, the procedure for measuring performance via multiple-alternative forced-choice (MAFC) tests with human observers is presented. Finally, results comparing model and human observer measurements are discussed. This paper expands on preliminary work reported previously¹⁷ by including experimental results involving detection and discrimination in both DE soft-tissue and bone images.

II. THEORETICAL METHODS

II.A. DE imaging performance modeling

II.A.1. Generalized DE image decomposition

In previous work,¹⁵ we introduced a generalized decomposition algorithm for DE imaging, which is briefly summarized below. The generalized DE decomposition algorithm is given by

$$I_{DE}(x, y) = I_L(x, y) * h_L(x, y) + I_H(x, y) * h_H(x, y), \quad (1)$$

where $h_L(x, y)$ and $h_H(x, y)$ are the low- and high-energy convolution filters applied to the low- and high-energy log-projections, denoted $I_L(x, y)$ and $I_H(x, y)$, respectively. Equation (1) provides a practical form for deriving the DE noise-power spectrum (NPS) and modulation transfer function (MTF) as a function of the low- and high-energy NPS, MTF, and decomposition filters. Table I shows the low- and high-energy filters employed for the three decomposition algorithms used in this study: standard log subtraction (SLS), simple-smoothing of the high-energy image (SSH),¹⁸ and anti-correlated noise reduction (ACNR).^{19–21} Details on these DE decomposition algorithms and their generalization to Eq. (1) are found in Refs. 14 and 15. This work investigates the effect of quantum noise on observer performance in DE images under various decomposition techniques. A potential limitation is the absence of anatomical background that might introduce artifacts (e.g., edges) under conditions of excessive filtering. Such artifacts were not presented in the phantom images and were not accounted for in the model.

TABLE I. Summary of low- and high-energy convolution filters for three DE decomposition algorithms (SLS, SSH, and ACNR), which are special cases of the general form in Eq. (1). w denotes the tissue cancellation parameters, ideally given as the ratio of the effective linear attenuation coefficient of the cancelled material (μ_H/μ_L), $\delta(x, y)$ is the Dirac delta function, w_n denotes the noise cancellation parameters, w_c denotes the tissue cancellation parameters for the complementary image, and h_{LPF} and h_{HPF} are low- and high-pass convolution filters, respectively.

Algorithm	$h_L(x, y)$	$h_H(x, y)$
SLS	$-w_s \delta(x, y)$	$\delta(x, y)$
SSH	$-w_s \delta(x, y)$	h_{LPF}
ACNR	$w_n w_c h_{HPF}(x, y) - w_s \delta(x, y)$	$\delta(x, y) - w_n h_{HPF}(x, y)$

II.A.2. Cascaded systems analysis of DE imaging

CSA provides a theoretical framework for modeling Fourier-based performance metrics of imaging systems. Examples of imaging systems modeled using CSA include radiography,²² fluoroscopy,²³ angiography,²⁴ mammography,²⁵ portal imagers,²⁶ and cone-beam CT.²⁷ CSA was extended in previous work,^{15,8} to DE imaging to yield theoretical descriptions of the DE image MTF, NPS, and NEQ. Assumptions inherent to CSA include linearity, shift invariance, and stationarity of the imaging system and are assumed to hold reasonably well over the range of relevant imaging conditions. FPDs have been shown to be highly linear across a large range of incident signal (e.g., $\sim 50\%$ of sensor saturation and appropriate gain correction). Furthermore, Cunningham has shown that discretely sampled digital systems are “cyclically” invariant and that such may be considered sufficient for the application of Fourier-based characterization.²⁸ Albert and Maidment have shown that the degree to which shift-invariance is violated is fairly minor over a broad range of condition and tasks. Further, methods for characterizing non-stationary noise effects is an area of ongoing research. CSA models the imaging chain as a series of stages: seven stages associated with the formation of a projection^{7,8} and a final stage to describe the combination of the low- and high-energy images to yield the DE MTF and NPS. Previous work has shown excellent agreement between the MTF and NPS for DE imaging as predicted by CSA and as measured over a wide range of imaging conditions.¹⁵

II.B. Model observers and the detectability index

A brief description of model observers is presented below, with notation based on that of Burgess *et al.*¹⁶ All of the terms appearing in the four model observers presented below were computed analytically using CSA for DE imaging systems.

II.B.1. The Fisher-Hotelling observer (FH)

The Fisher-Hotelling (FH) observer is modeled as a pre-whitening matched filter incorporating a detection template that decorrelates the noise:²⁹

$$d'_{\text{FH}}^2 = \int \int \frac{\text{MTF}^2(u,v) W_{\text{Task}}^2(u,v)}{\text{NNPS}(u,v)} du dv, \quad (2)$$

where u and v are the spatial frequencies, $\text{MTF}(u,v)$ denotes the modulation transfer function, and $\text{NNPS}(u,v)$ denotes the normalized noise-power spectrum. The two-dimensional (2D) integration is performed over the Nyquist region. The term $W_{\text{Task}}(u,v)$ denotes the task function—a frequency domain representation of the imaging task as described in ICRU Report #54. The task function is given by the Fourier transform of the difference between two hypotheses:²⁹

$$W_{\text{Task}}(u,v) = |F[h_1(x,y) - h_2(x,y)]|, \quad (3)$$

where F denotes the Fourier transform and $h_1(x,y)$ and $h_2(x,y)$ denote the hypothesis functions, which describe the signal in the spatial domain for the two hypotheses, as detailed below in subsection C. In this form, the detectability index can also be understood as a weighted sum of the task function with the noise-equivalent quanta (NEQ). The resulting detectability index has been used as a figure of merit for system optimization in a variety of applications.^{7,30–32} The observer models presented here represent the simple case of a task with signal and background known exactly

II.B.2. The Fisher-Hotelling observer with eye filter (FHE)

The FH observer can be modified to include an eye filter and internal noise to better incorporate factors that potentially degrade performance in human observers

$$d'_{\text{FHE}}^2 = \int \int \frac{\text{MTF}^2(u,v) W_{\text{Task}}^2(u,v)}{\text{NNPS}(u,v) E^2(u,v) + N_i(u,v)} du dv, \quad (4)$$

where $E(u,v)$ is the eye filter and N_i denotes the internal noise. The eye filter employed in this work was the same as that used by Burgess,¹⁶ which was modeled on the contrast sensitivity function curve measured by Barton for the human visual system:³³

$$E(f) = f^n \exp(-cf^2), \quad (5)$$

where f denotes the radial spatial frequency; $f^2 = u^2 + v^2$. This model is also consistent with ones employed by others.^{34,35} A

value of $n=1.3$ has been shown to yield good agreement with measurements and the value of c is determined such that $E(f)$ peaks at 4 cycles/deg for a given viewing distance (set to 50 cm in this work, thus yielding a value of $c \sim 3$). The internal noise was assumed to be uncorrelated (constant NPS) and set to a fraction (0.02 at 100 cm viewing distance) of the zero-frequency NNPS scaled to the observer's viewing distance D :

$$N_i(u,v) = 0.02 \left(\frac{D}{100} \right)^2 \text{NNPS}(0,0). \quad (6)$$

The value of 0.02 was the same as that used by Burgess *et al.*,¹⁶ estimated from published human observer internal noise measurements. The FHE observer has been shown to provide an improved model for human observer performance compared to the FH observer in sphere detection tasks.¹⁶

II.B.3. The Non-prewhitening observer (NPW)

The non-prewhitening observer initially proposed by Wagner³⁶ does not decorrelate the image noise, instead using a template matched to the expected signal as a detection template. Therefore, variations in the signal or background are not compensated, giving²⁹

$$d'_{\text{NPW}}^2 = \frac{[\int \int \text{MTF}^2(u,v) W_{\text{Task}}^2(u,v) du dv]^2}{\int \int \text{NNPS}(u,v) \text{MTF}^2(u,v) W_{\text{Task}}^2(u,v) du dv}. \quad (7)$$

The NPW observer has been shown to better model human observers compared to the FH observer in cases where humans are not able to prewhiten the noise, as in high-pass-filtered (HPF) images (e.g., CT).^{37,38} The NPW observer reduces to the FH observer in the case of uncorrelated (white) noise.

II.B.4. The non-prewhitening observer with eye filter (NPWE)

Similar to the FH observer, the NPW observer can be extended to include an eye filter and internal noise as shown by Burgess.¹⁶

$$d'_{\text{NPWE}}^2 = \frac{[\int \int \text{MTF}^2(u,v) W_{\text{Task}}^2(u,v) E^2(u,v) du dv]^2}{\int \int \text{NNPS}(u,v) \text{MTF}^2(u,v) W_{\text{Task}}^2(u,v) E^2(u,v) + \text{MTF}^2(u,v) W_{\text{Task}}^2(u,v) N_i(u,v) du dv}. \quad (8)$$

This model employs the same eye filter and internal noise functions as described above for the FHE observer.

II.C. Imaging tasks

Three imaging tasks were considered in the current work, as illustrated in Fig. 1.

II.C.1. Sphere detection task

The sphere detection task considered hypotheses of a uniform background and the projection of a sphere. Therefore $h_1(x,y)=0$ and $h_2(x,y)$ corresponded to the 2D projection

FIG. 1. Model hypotheses and task functions investigated in this work. (a) Sphere detection has sphere absent and sphere present, (b) shape discrimination has a sphere and a cylinder, and (c) texture discrimination has a smooth and crenellated cylinder as the two hypotheses, respectively. Task functions adapted from Ref. 17.

profile of a sphere computed analytically from the attenuated signal through a 3 mm diameter sphere. Attenuation data for polyethylene and Teflon[®] were used to model soft-tissue and bony lesions, respectively. Figure 1(a) illustrates the sphere detection hypotheses and the resulting task function, which predominantly weighs very low spatial frequencies. A subtle ring at low-to-mid frequencies associated with the edge of the sphere is also evident.

II.C.2. Shape discrimination task

The shape discrimination task considered hypotheses of a sphere and cylindrical disk. Therefore, $h_1(x,y)$ was the projection of a 6 mm sphere (computed numerically as in the detection task), and $h_2(x,y)$ was the projection profile of a cylinder, as shown in Fig. 1(b). This task describes the discrimination of subtle shape differences that might be present between various types of lesions (e.g., a solid nodule versus a vessel). The signal for each task was normalized such that both hypotheses had the same signal power:

$$\iint h_1^2(x,y) dx dy = \iint h_2^2(x,y) dx dy. \quad (9)$$

This was achieved by setting the height and diameter of the cylinder equal to 87% of the sphere diameter.

II.C.3. Texture discrimination task

The final task was a texture discrimination task, which considered hypotheses of smooth and textured disks. Therefore, $h_1(x,y)$ was the projection profile computed for an 11 mm diameter smooth cylinder with a height of 14 mm, and $h_2(x,y)$ was the projection profile computed for an 11 mm diameter crenellated (i.e., textured) cylinder with minimum and maximum crenella heights of 12 and 16 mm, respectively (i.e., ± 2 mm from the height of the smooth cylinder), as shown in Fig. 1(c). This task roughly corresponds to the discrimination of various textures that might be exhibited by lesions (e.g., a solid nodule versus a spiculated mass or ground-glass opacity). Rings in the task function correspond to the crenellation in the textured cylinder. Again, equal signal power [Eq. (9)] was achieved by setting the diameters equal and setting the height of the smooth cylinder equal to the mean height of the textured cylinder.

II.D. Comparison of d' with human observer performance

The performance of human observers was assessed using MAFC tests as detailed in Sec. III B, for which the outcome measure is the proportion correct (P_C). To compare the detectability index computed for the various model observers with human observers, all results were converted to A_Z , which provides an estimate of the observer signal-to-noise ratio in terms of the area under the ROC. Assuming that the internal decision variable employed in assessing under the receiver operating characteristic (ROC) curve is normally distributed for both hypotheses and has equal variance, the detectability index can be related to A_Z as

$$A_Z = \frac{1}{2} \left(1 + \frac{2}{\sqrt{\pi}} \int_0^{d'/2} \exp(-x^2) dx \right). \quad (10)$$

On the other hand, results from the MAFC tests can be related to the detectability index:

$$P_C(d', M) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-d')^2}{2}\right) [\Phi(x)]^{M-1} dx, \quad (11)$$

where $\Phi(x)$ is the cumulative Gaussian distribution function, $\Phi(x) = \int_{-\infty}^x \exp(-y^2/2) dy$, and M is the number of alternatives presented to the observer. This function was inverted numerically to provide a lookup table between P_C and d' , which was then converted to A_Z using Eq. (10).

Fig. 2. (a) Image of the flat-panel detector and the task phantom used to acquire DE images of the various stimuli corresponding to the different imaging tasks. (b) Radiographic projection of various stimuli. Experimental setup and phantom adapted from Ref. 17.

III. EXPERIMENTAL METHODS

III.A. Imaging setup

III.A.1. DE imaging system

Figure 2 illustrates the prototype DE imaging system developed in our laboratory and the phantom constructed to mimic the idealized tasks of Fig. 1. The system employed a Trixell Pixium 4600 (Moirans, France) flat-panel detector (250 mg/cm², 3000 × 3000 pixels, 0.143 mm pixel pitch, and ~0.8 fill factor). The system also included a 10:1 anti-scatter bucky grid. The low- and high-energy kVp were 60 and 120 kVp, with 2.5 mm Al filtration and 2.5 mm Al + 0.6 mm Ag total filtration, respectively. These acquisition techniques were based on previous theoretical and experimental investigations of DE image quality.^{8,39}

III.A.2. Task phantom

A phantom was built to model the hypotheses described in Sec. II C using a 10 cm slab of PMMA, simulating an average chest thickness,⁴⁰ on which were placed stimuli corresponding to the various hypotheses as shown in Fig. 2. Polyethylene and Teflon® material were used to simulate soft-tissue and bony lesions in the soft-tissue and bone images, respectively. The stimuli were precisely machined to ensure that the objects corresponded closely to the idealized hypotheses. The sphere detection task used 3 mm diameter spheres. The shape discrimination task used a 6 mm diameter sphere and a cylinder with height and diameter equal to 87% of the sphere's diameter to ensure equal signal power between the two hypotheses [Eq. (9)]. For the texture discrimination task, both cylinders had a diameter of 11 mm. The smooth cylinder had a height of 14 mm, and the textured cylinder had minimum and maximum crenellation heights of 12 and 16 mm, respectively, ensuring equal signal power.

III.B. Human observer study

III.B.1. Multiple-alternative forced choice (MAFC) tests

In an MAFC study, the observer is required to identify which of M images contains a specified stimulus, where $M - 1$ images are “normal” (i.e., the first hypothesis), and image

contains an abnormality (i.e., the second hypothesis). The fraction of correct responses, P_C , gives a figure of merit for task performance which can be related to the area under the ROC curve and detectability index as discussed in Sec. II D.

III.B.2. Dependence of performance on acquisition technique

The first study investigated performance as a function of dose allocation, defined as the fraction of dose imparted by the low-energy image:⁸

$$A = \frac{ESD_L}{ESD_L + ESD_H} \quad (12)$$

where A denotes dose allocation (not to be confused with A_Z , the area under the ROC curve) and ESD denotes the entrance surface dose. Therefore, the dose can be allocated between the low- and high-energy projections arbitrarily without changing the total dose. In previous work,⁸ the optimal allocation was shown to be inversely related to the low- and high-energy projection detective quantum efficiency (DQE), tending therefore to optima of $A < 0.5$ (i.e., a majority of dose imparted by the high-energy image). Previous studies also showed the optimal allocation to depend intimately on the choice of decomposition algorithm.⁴¹ Five values of dose allocation were chosen for the study, selected from technique settings that were deliverable by the x-ray generator capabilities such that the total entrance surface dose (i.e., $ESD_L + ESD_H$) was constant: 0.1 mGy, corresponding to the average dose for a digital radiography chest exam.⁴² The resulting allocation values were $A = 0.06, 0.28, 0.45, 0.67$, and 0.85 .

III.B.3. Dependence of performance on decomposition technique

The second study investigated performance as a function of the various decomposition parameters for the SSH and ACNR algorithms. The low- (LPFs) and high-pass filters (HPFs) applied in the Fourier domain were chosen as Gaussian filters:

$$H_{LPF}(u, v) = \exp\left(-\frac{u^2 + v^2}{2d_{LPF}^2}\right) \quad (13)$$

and

$$H_{HPF}(u, v) = 1 - \exp\left(-\frac{u^2 + v^2}{2d_{HPF}^2}\right), \quad (14)$$

which represent the Fourier transforms of h_{LPF} and h_{HPF} , and are characterized by d_{LPF} and d_{HPF} for the SSH and ACNR algorithms, respectively. The scalar variables d_{LPF} and d_{HPF} carry units of mm⁻¹, corresponding approximately to the cut-off frequency of the applied filter. The motivation for this work was not only to identify optimal decomposition parameters, but also to compare the performance of model and human observers as these parameters are varied. Therefore, a nominal set of decomposition parameters (listed in Table II) was selected based on qualitative examination of patient im-

TABLE II. Nominal values for the decomposition parameters associated with each of the decomposition techniques.

Decomposition parameter	Soft-tissue image	Bone image
w_s	0.31	0.59
w_c	0.59	0.31
d_{LPF}	0.90	0.15
w_n	0.90	0.90
d_{HPF}	0.50	0.30

ages acquired in an ongoing clinical study⁴³ such that the decomposed image provided adequate tissue cancellation and reduction of noise without introducing significant artifacts.¹⁵ The nominal parameters were judged appropriate to diagnostic image quality by an expert chest radiologist. The studies below involved variation of decomposition parameters about these nominal values to examine the correspondence of the observer models with human observers.

Images were decomposed by independently varying, the d_{LPF} , d_{HPF} , and w_n parameters while fixing the other parameter values to the nominal values shown in Table II. Dose allocation was held at 0.5 throughout the decomposition study, and the total dose was reduced to 0.05 mGy to reduce the number of cases for which the task was too conspicuous (i.e., $A_Z > 0.95$). The SSH smoothing parameter d_{LPF} and the ACNR high-pass filter parameter d_{HPF} were each varied from 0.1 to 1 in increments of 0.1. The ACNR weighting parameter w_n was varied from 0.2 to 2 in increments of 0.2.

III.B.4. Experimental setup

We chose a 9-AFC test (rather than 2-AFC, 4-AFC, etc.) owing in part to the fairly conspicuous tasks employed in this experiment. Specifically, a sensitivity analysis of P_C evaluated as a function M [i.e., $\partial d' / \partial P_C$ versus P_C evaluated from the inverse of Eq. (11)] showed that the sensitivity of the observer study measurement is maximized at $P_C = 0.5$ for any value of M . Based on analysis of detectability index [Eqs. (2), (4), (7), and (8)], we expected d' to be in range ~ 1 to 2 ($A_Z \sim 0.8$ – 0.9) for the majority of cases in this experi-

ment. This suggested that a value of $M=9$ would give maximum sensitivity at the expected d' for these tests. Therefore, the 9-AFC study was conducted on a dedicated viewing station in a radiology reading room under controlled lighting. Images were displayed on a diagnostic workstation (Dell™ Precision-380, 3 GHz Pentium® 4 with dual-head monochrome Barco displays, 1530×2048 resolution, 8-bit gray-scale). Each image was 300×300 pixels in size (900×900 pixels for the 9-AFC, with a 10-pixel border between each image), and each image pixel corresponded to a single pixel on the display. The background of the display was masked to medium gray, given by the mean pixel value of the images. Observers were not allowed to adjust window, level, or magnification. A comfortable viewing distance of 50 cm was encouraged but not enforced.

Six physicists were used as observers, considered reasonably expert observers for purposes of a simple phantom experiment. Observers were instructed as to the defined “abnormal” (stimulus) and “normal” hypotheses for each task according to the illustrations in Fig. 1. For each case, nine images were presented in a 3×3 matrix, eight representing “normal” and one representing the stimulus, and observers selected the stimulus image by mouse click. The study was separated into three tests, corresponding to the three tasks, conducted in random order, and with images obtained using variable dose allocation and decomposition parameters presented randomly within each test. Each observer was trained using 60 cases for the detection task, 70 cases for the shape discrimination task, and 70 cases for the texture discrimination task. The training data represented distinct cases from the actual study data. As shown in Table III, the study measured observer performance at 5 values of dose allocation (A) and at 10 values of each decomposition parameter (d_{LPF} , d_{HPF} , and w_n). The SLS algorithm was only investigated for the detection task, since the associated decomposition parameter (w_s) does not influence the frequency dependence of the NEQ in a manner that would cause significant task-dependent performance. To ensure statistical independence among the five repeats, one stimulus image and eight normal images were randomly selected from a group of 15 images.

TABLE III. Number of cases in each study for a given observer. Each case was repeated five times (using different projections so that all cases were statistically independent).

Image	Decomposition algorithm	# Cases (dose allocation)	# Cases (decomposition)	# Repeats	# Tasks	Total # cases
Soft-Tissue	SLS	5	—	5	1 (sphere detection)	25
	SSH	5	10 (d_{LPF})	5	3	225
	ACNR	5	$\begin{pmatrix} 10(d_{\text{HPF}}) \\ 10(w_n) \end{pmatrix}$	5	3	375
Bone	SLS	5	—	5	1 (sphere detection)	25
	SSH	5	10 (d_{LPF})	5	3	225
	ACNR	5	$\begin{pmatrix} 10(d_{\text{HPF}}) \\ 10(w_n) \end{pmatrix}$	5	3	375
	Total					1250

FIG. 3. Sphere detection performance as a function of dose allocation in DE imaging (i.e., the fraction of dose allocated to the low-energy projection). Data points represent measurements from human observers, and curves represent model observers. (a–c) Soft-tissue and (d–f) bone image. Results are shown for the three decomposition algorithms (SLS, SSH, and ACNR).

Furthermore, the random image selection was constrained such that a different stimulus image and a different set of normal images were used for each repeat. A total of 1450 cases (including repeats) were presented to each observer (200 training+450 detection+400 shape discrimination+400 texture discrimination). Short breaks were allowed between each test. An average time of ~ 3 h was required to complete all three tests, with an average of ~ 8 s per case.

IV. RESULTS

IV.A. Dependence of performance on acquisition technique

Figure 3 plots theoretical and measured observer performance in terms of A_Z as a function of dose allocation. Results pertain to the sphere detection task, with separate plots shown for the three decomposition algorithms (SLS, SSH, and ACNR) as well as the soft-tissue [Figs. 3(a)–3(c)] and bone [Figs. 3(d)–3(f)] image. The results in Fig. 3 and below expand on preliminary studies in Ref. 17 and include detection/discrimination tasks in both DE soft-tissue and bone images. The data points represent measurements obtained from the 9-AFC study, with error bars corresponding to the 95% confidence interval computed across the six observers. The curves in each plot correspond to four observer models (FH, FHE, NPW, and NPWE). Results agree with the previously observed dependence of image quality on dose allocation,^{8,15,39} with optimal performance exhibited in the

range $A \sim 0.3$ – 0.7 depending on the applied decomposition algorithm and task. As expected from previous work,¹⁵ ACNR yielded the greatest performance for the soft-tissue sphere detection tasks, and SSH is best for the bone image.

The FH and NPW models perform similarly in the sphere detection task, since the task function weighs mostly low spatial frequencies [Fig. 1(a)] for which there is little correlation in the image data and therefore little difference between a prewhitening and non-prewhitening observer. Both of these models appear to overestimate the performance of the human observers. The addition of an eye filter and internal noise in the FHE and NPWE models yields better agreement with human observer performance. For the soft-tissue image, although the experimental error bars are fairly large, the NPWE model is seen to agree reasonably well with measurement in terms of both overall magnitude and trend for all three decomposition algorithms. For the bone image, the FHE model appears to give the best agreement with measurement, although the fairly high conspicuity ($A_Z > 0.95$) for the SSH and ACNR algorithms make it difficult to differentiate results.

Figure 4 shows similar results for the shape and texture discrimination tasks. Results are shown for the soft-tissue images only, since (similar to Fig. 3), the conspicuity in the bone images was fairly high ($A_Z > 0.95$, as predicted by theory and borne out in measurement). Despite fairly large error bars, the results suggest that the NPWE model exhibits the closest correspondence with measurements for both the

FIG. 4. (a–c) Shape and (d–f) texture discrimination performance as a function of dose allocation for the soft-tissue image. Data points represent measurements from human observers, and curves represent model observers. Results are shown for the three decomposition algorithms (SLS, SSH, and ACNR).

shape and texture discrimination tasks. Again, the ACNR algorithm provides the highest performance for the soft-tissue images. For the bone images (not shown), the SSH and ACNR algorithm provided the highest performance, in agreement with theory [similar to Figs. 3(c)–3(f)]. Greater discrepancies were observed between human and model observers for the discrimination task than for the detection task. This may be related at least in part to the fact that discrimination is a more complex psychophysical process than detection, and the idealized binary hypothesis framework oversimplifies the task; furthermore, discrimination requires accurate description of two hypothesis functions compared to just one for detection.

IV.B. Dependence of performance on decomposition technique

Figure 5 plots theoretical and experimental performance for the SSH and ACNR algorithms evaluated as a function of decomposition parameter. Results are shown only for the soft-tissue image, since the bone images were too conspicuous to differentiate trends ($A_Z > 0.95$), although theory and measurement were seen to agree. For the SSH algorithm [Figs. 5(a) and 5(c)], the d_{LPF} parameter was seen to impart a significant effect on observer performance for both imaging tasks, showing a fairly sharp performance drop at $d_{\text{LPF}} \sim 0.5$ (see Section V below). Both the overall magnitude and trend in performance was seen to agree with the NPWE model. For the ACNR algorithm [Figs. 5(b) and 5(e)], the

parameter w_n exhibits a clear optimum in the region $w_n \sim 0.8$ – 1.0 . In contrast with results of the allocation study (Figs. 3 and 4), the application of an eye filter and internal noise not only reduced the overall magnitude of A_Z (compared to the FH and NPW models) but also significantly affected the trend (shape) of the curves in a manner that brings model observer performance more in line with that of human observers. Larger discrepancies were observed in these experiments for the FH and NPW observers, possibly due to the inclusion of very low spatial frequencies in the detectability index. The overestimation compared to human performance is consistent with the notion that human observers do not to use very low spatial frequencies for simple discrimination tasks. The NPWE and FHE models, which suppress low spatial frequencies by way of the eye filter, correlated better with human observers. Some cases [e.g., Figs. 5(a) and 5(d)] exhibited large variations attributed to random error and observer variability, but agreement was still better for the eye filter models.

V. DISCUSSION AND CONCLUSION

Perhaps the most interesting example where Fourier metrics of imaging performance (NEQ, computed analytically using CSA) combined with a description of imaging task provides a good approximation to human observer performance is seen in Fig. 5(d) (i.e., the texture discrimination task under SSH decomposition). In that case, the algorithm effectively zeroes the NEQ at a given spatial frequency, de-

FIG. 5. Experimental and theoretical observer performance for the shape and texture discrimination tasks evaluated as a function of decomposition parameters in the SSH and ACNR algorithms.

pending on the choice of d_{LPF} . As illustrated further in Fig. 6(b), we see that if d_{LPF} is chosen such that the “zeroing” of the NEQ coincides with the spatial frequencies of interest in the imaging task (i.e., the frequency of the crenellation shown in Fig. 1), then the detectability index plummets. Such is predicted theoretically and borne out experimentally, demonstrating that Fourier-based modeling from first principles of signal and noise transfer characteristics provides a reasonable approximation to human observer performance for these types of simple detection and discrimination tasks.

Results also indicated that inclusion of an eye filter and internal noise (as in the FHE and NPWE models) yielded higher correspondence to human observer performance due primarily to the removal of low spatial frequencies. In the current work, internal noise was not tuned to match human observer performance (estimated instead as a constant as described above). The simple model for internal noise could be a source of discrepancy between measured and theoretical results. Future work could improve on the modeling by more refined estimates of internal noise that bring model and human performance to closer agreement.

Despite the large number of cases, uncertainty in some of the measurements was large. Figures 5(a) and 5(d), for example, exhibited a high degree of intra- and inter-observer variability owing to higher noise for the SSH algorithm in soft-tissue decomposition. (SSH is known to have lower performance than ACNR for soft-tissue decomposition.⁴¹) Future work investigating the interesting spatial-frequency-

dependent effects in these noisier cases (e.g., the zeroing of the NEQ at a specific spatial frequency as evident in Figs. 5(a) and 5(d)] will require a greater number of cases to reduce such variability.

The influence of anatomical background noise is minimized in DE images compared to a conventional radiograph, whereas quantum noise is typically amplified. Still, residual anatomical background (e.g., clutter associated with a material of the same type as the lesion) may diminish conspicuity. Future work could therefore include anatomical background (e.g., as power-law $1/f^\beta$ noise) in generalized description of NEQ and detectability index.⁸ Including anatomical noise in this form in the detectability index degrades model observer performance at low spatial frequencies in a manner similar to the eye filter (i.e., reduction of the integrand at low frequencies). Anatomical background noise diminishes low-frequency performance by means of high noise-power, whereas the eye filter models accomplish the same by means of the ramp-like eye transfer function. Therefore, the NPW and FH models modified to include power-law anatomical background noise will likely behave similar to the FHE and NPWE observers and will have a smaller effect on the FHE and NPWE observers (which already reduce low-frequency components of the integrand).

Another potential limitation of this work is the requirement that the imaging system is linear, shift invariant, and stationary. As discussed above, such assumptions are believed to be valid for FPDs under conditions dominated by

FIG. 6. Illustration of basic agreement between Fourier-based descriptions of NEQ, task, and observer model with human observer performance in the discrimination of a textured and smooth disk. (a–c) Plot of NEQ and task function (weighted by the eye filter) for three values of the d_{LPF} parameter employed in the SSH decomposition algorithm. Corresponding images of the (d–f) crenellated and (g–i) smooth cylinder. Since the value of d_{LPF} determines the zero in the NEQ, the value for which the zero coincides with frequencies of interest for this task causes the discriminate performance to plummet. Such is clearly observed in (b), where the value $d_{LPF}=0.4$ places the zero of the NEQ near the peak of the task function, and the ability to discriminate images (e) and (h) is lost.

x-ray quantum noise. Future work more closely examining anatomical background noise and/or artifacts associated with DE decomposition may violate stationarity assumptions to a greater degree and may therefore be less well suited to a simple frequency domain approach. Furthermore, the linear shift-invariance assumption is not believed to have a significant effect on the ability to predict human observer performance in this work.

In this work, the performance of human observers was compared to several model observers computed from a theoretical framework using cascaded systems analysis. Overall, results demonstrated that image acquisition parameters, decomposition parameters, and the imaging task all affect the performance of human observers. Furthermore, this work provides initial validation that the theoretical, Fourier-based modeling of NEQ, imaging task, and simple model observers can reasonably predict the performance of human observers. Results also identified that inclusion of an eye filter and internal noise yielded higher correspondence to human observer performance due to the removal of low spatial frequencies; i.e., the FHE and NPWE models agreed best with measurements. Such findings help to bridge the gap between metrics of detector performance (e.g., NEQ, either computed theoretically or measured experimentally) and the perfor-

mance of human observers, linked by a Fourier description of the imaging task.

ACKNOWLEDGMENTS

The authors thank Drs. R. Van Metter and J. Yorkston (Carestream Health Inc., Rochester, NY) for stimulating discussions and collaboration in creating the DE imaging prototype. Thanks also to D. J. Tward, C. A. Varon, and H. Kashani (Ontario Cancer Institute, Toronto, ON) for assistance with the experiments and analysis. This work was supported by a Cunningham Fellowship Award, a University of Toronto Open Scholarship Award, and National Institute of Health Grant No. R01-CA-112163.

^{a)} Author to whom correspondence should be addressed. Electronic mail: Jeff.Siewerdsen@uhn.on.ca. Telephone: 416-946-4501 (x5516); Fax: 416-946-6529.

¹M. P. Eckstein, Y. Zhang, and B. T. Pham, "Metrics of medical image quality: Task-based model observers vs. image discrimination/perceptual difference models," *Proc. SPIE* **5372**, 42–52 (2004).

²M. P. Eckstein, C. K. Abbey, and J. L. Bartoff, "Model observer based optimization of JPEG image compression," *Proc. SPIE* **3981**, 106–115 (2000).

³M. Ruschin, P. Timberg, M. Bath, B. Hemdal, T. Svahn, R. Saunders, E. Samei, I. andersson, S. Mattsson, D. P. Chakraborty, and A. Tingberg, "Dose dependence of mass and microcalcification detection in digital mammography: Free response human observer studies," *Med. Phys.* **34**, 400–407 (2007).

⁴R. Saunders, J. A. Baker, D. M. Delong, J. P. Johnson, and E. Samei, "Does image quality matter? Impact of resolution and noise on mammographic task performance," *Med. Phys.* **34**, 3971–3981 (2007).

⁵E. A. Krupinski, J. P. Johnson, H. Roehrig, J. Nafziger, J. Fan, and J. Lubin, "Use of a human visual system model to predict the effects of display veiling glare on observer performance," *Proc. SPIE* **5372**, 423–429 (2004).

⁶E. Samei, M. J. Flynn, and W. R. Eyler, "Detection of subtle lung nodules: Relative influence of quantum and anatomic noise on chest radiographs," *Radiology* **213**(3), 727–734 (1999).

⁷S. Richard, J. H. Siewerdsen, D. Jaffray, D. J. Moseley, and B. Bakhtiar, "Generalized DQE analysis of radiographic and dual-energy imaging using flat-panel detectors," *Med. Phys.* **32**, 1397–1413 (2005).

⁸S. Richard and J. H. Siewerdsen, "Optimization of dual-energy imaging systems using generalized NEQ and imaging task," *Med. Phys.* **34**(1), 127–139 (2007).

⁹R. E. Alvarez, J. A. Seibert, and S. K. Thompson, "Comparison of dual energy detector system performance," *Med. Phys.* **31**(3), 556–565 (2004).

¹⁰H. Bornefalk and M. Lundqvist, "Dual-energy imaging using a photon counting detector with electronic spectrum-splitting," *Proc. SPIE* **6142**, 523–533 (2006).

¹¹J. L. Ducote, T. Xu, and S. Y. Molloy, "Optimization of a flat-panel based real time dual-energy system for cardiac imaging," *Med. Phys.* **33**(6), 1562–1568 (2006).

¹²F. Fischbach, T. Freund, R. Rottgen, U. Engert, R. Felix, and J. Ricke, "Dual-energy chest radiography with a flat-panel digital detector: revealing calcified chest abnormalities," *AJR, Am. J. Roentgenol.* **181**(6), 1519–1524 (2003).

¹³S. C. Kappadath and C. C. Shaw, "Dual-energy digital mammography for calcification imaging: Scatter and nonuniformity corrections," *Med. Phys.* **32**(11), 3395–3408 (2005).

¹⁴R. J. Warp and J. T. Dobbins III, "Quantitative evaluation of noise reduction strategies in dual-energy imaging," *Med. Phys.* **30**(2), 190–198 (2003).

¹⁵S. Richard and J. H. Siewerdsen, "Cascaded systems analysis of noise reduction algorithms for dual-energy imaging," *Med. Phys.* **35**(2), 586–601 (2008).

¹⁶A. E. Burgess, X. Li, and C. K. Abbey, "Visual signal detectability with two noise components: anomalous masking effects," *J. Opt. Soc. Am. A* **14**(9), 2420–2442 (1997).

¹⁷S. Richard, J. H. Siewerdsen, and D. J. Tward, "NEQ and task in dual-

- energy imaging: From cascaded systems analysis to human observer performance," *Proc. SPIE* **6913**, 691311-1-691311-12 (2008).
- ¹⁸P. C. Johns and M. J. Yaffe, "Theoretical optimization of dual-energy x-ray imaging with application to mammography," *Med. Phys.* **12**(3), 289-296 (1985).
 - ¹⁹C. H. McCollough, M. S. Van Lysel, W. W. Pepler, and C. A. Mistretta, "A correlated noise reduction algorithm for dual-energy digital subtraction angiography," *Med. Phys.* **16**(6), 873-880 (1989).
 - ²⁰D. L. Ergun, C. A. Mistretta, D. E. Brown, R. T. Bystryanyk, W. K. Sze, F. Kelcz, and D. P. Naidich, "Single-exposure dual-energy computed radiography: Improved detection and processing," *Radiology* **174**(1), 243-249 (1990).
 - ²¹W. Kalender, E. Klotz, and L. Kostaridou, "An algorithm for noise suppression in dual-energy CT material density images," *IEEE Trans. Med. Imaging* **7**, 218-224 (1988).
 - ²²J. H. Siewerdsen, L. E. Antonuk, Y. El Mohri, J. Yorkston, W. Huang, J. M. Boudry, and I. A. Cunningham, "Empirical and theoretical investigation of the noise performance of indirect detection, active matrix flat-panel imagers (AMFPIs) for diagnostic radiology," *Med. Phys.* **24**(1), 71-89 (1997).
 - ²³S. Vedantham, A. Karellas, and S. Suryanarayanan, "Solid-state fluoroscopic imager for high-resolution angiography: Parallel-cascaded linear systems analysis," *Med. Phys.* **31**(5), 1258-1268 (2004).
 - ²⁴A. Ganguly, S. Rudin, D. R. Bednarek, and K. R. Hoffmann, "Micro-angiography for neuro-vascular imaging. II. Cascade model analysis," *Med. Phys.* **30**(11), 3029-3039 (2003).
 - ²⁵W. Zhao and J. A. Rowlands, "Digital radiology using active matrix read-out of amorphous selenium: theoretical analysis of detective quantum efficiency," *Med. Phys.* **24**(12), 1819-1833 (1997).
 - ²⁶J. P. Bissonnette, I. A. Cunningham, D. A. Jaffray, A. Fenster, and P. Munro, "A quantum accounting and detective quantum efficiency analysis for video-based portal imaging," *Med. Phys.* **24**(6), 815-826 (1997).
 - ²⁷J. H. Siewerdsen and D. A. Jaffray, "Three-dimensional NEQ transfer characteristics of volume CT using direct and indirect-detection flat-panel imagers," *Proc. SPIE* **5030**, 92-102 (2003).
 - ²⁸C. E. Metz and K. Doi, "Transfer function analysis of radiographic imaging systems," *Phys. Med. Biol.* **24**(6), 1079-1106 (1979).
 - ²⁹ICRU Report 54. *Medical Imaging—The Assessment of Image Quality*, (International Commission on Radiation Units and Measurements, Bethesda, MD, 1996).
 - ³⁰J. H. Siewerdsen and D. A. Jaffray, "Optimization of x-ray imaging geometry (with specific application to flat-panel cone-beam computed tomography)," *Med. Phys.* **27**(8), 1903-1914 (2000).
 - ³¹M. J. Tapiovaara, "Efficiency of low-contrast detail detectability in fluoroscopic imaging," *Med. Phys.* **24**(5), 655-664 (1997).
 - ³²A. S. Chawla, E. Samei, R. S. Saunders, J. Y. Lo, and J. A. Baker, "A mathematical model platform for optimizing a multiprojection breast imaging system," *Med. Phys.* **35**(4), 1337-1345 (2008).
 - ³³P. G. J. Barten, *Contrast Sensitivity of the Human Eye and Its Effect on Image Quality* (SPIE Bellingham, WA, 1999).
 - ³⁴M. A. Webster and E. Miyahara, "Contrast adaptation and the spatial structure of natural images," *J. Opt. Soc. Am. A* **14**(9), 2355-2366 (1997).
 - ³⁵M. P. Eckstein, C. K. Abbey, and F. O. Bochud, "Visual signal detection in structured backgrounds. IV. Figures of merit for model performance in multiple-alternative forced-choice detection tasks with correlated responses," *J. Opt. Soc. Am. A* **17**(2), 206-217 (2000).
 - ³⁶R. F. Wagner, D. G. Brown, and M. S. Pastel, "Application of information theory to the assessment of computed tomography," *Med. Phys.* **6**(2), 83-94 (1979).
 - ³⁷A. E. Burgess, R. F. Wagner, R. J. Jennings, and H. B. Barlow, "Efficiency of human visual signal discrimination," *Science* **214**(4516), 93-94 (1981).
 - ³⁸K. J. Myers, H. H. Barrett, M. C. Borgstrom, D. D. Patton, and G. W. Seeley, "Effect of noise correlation on detectability of disk signals in medical imaging," *J. Opt. Soc. Am. A* **2**(10), 1752-1759 (1985).
 - ³⁹N. A. Shkumat, J. H. Siewerdsen, A. C. Dhanantwari, D. B. Williams, S. Richard, N. S. Paul, J. Yorkston and R. Van Metter, "Optimization of image acquisition techniques for dual-energy imaging of the chest," *Med. Phys.* **34**(10), 586-601 (2007).
 - ⁴⁰AAPM Report No. 31, *Standards Methods for Measuring Diagnostic X-Ray Exposures*, (American Institute of Physics, Melville, NY, 1991).
 - ⁴¹S. Richard and J. H. Siewerdsen, "Cascaded systems analysis of noise reduction algorithms in dual-energy imaging," *Med. Phys.* **35**(2), 586-601 (2008).
 - ⁴²Nationwide Evaluation of X-Ray Trends (NEXT): "Tabulation and Graphical Summary of 2001 Survey of Adult Chest Radiography" (2005).
 - ⁴³H. Kashani, G. Gang, N. A. Shkumat, C. A. Varon, J. Yorkston, R. Van Metter, N. S. Paul, and J. H. Siewerdsen, "Development of a high-performance dual-energy chest imaging system: Initial investigation of diagnostic performance," *Med. Phys.* (submitted).