Analysis of Fourier-domain task-based detectability index in tomosynthesis and cone-beam CT in relation to human observer performance

Grace J. Gang
Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Ontario M5G 2M9, Canada and Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland 21205

Junghoon Lee
Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, Maryland 21205

J. Webster Stayman, Daniel J. Tward, and W. Zbijewski
Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland 21205

Jerry L. Prince
Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, Maryland 21205

Jeffrey H. Siewerdsen[a)]
Department of Biomedical Engineering. Department of Computer Science, and Department of Radiology, Johns Hopkins University, Baltimore, Maryland 21205

(Received 2 September 2010; revised 8 February 2011; accepted for publication 9 February 2011; published 8 March 2011)

Purpose: Design and optimization of medical imaging systems benefit from accurate theoretical modeling that identifies the physical factors governing image quality, particularly in the early stages of system development. This work extends Fourier metrics of imaging performance and detectability index (d') to tomosynthesis and cone-beam CT (CBCT) and investigates the extent to which d' is a valid descriptor of task-based imaging performance as assessed by human observers.

Methods: The detectability index for tasks presented in 2D slices (d'_{slice}) was derived from 3D cascaded systems analysis of tomosynthesis and CBCT. Anatomical background noise measured in a physical phantom presenting power-law spectral density was incorporated in the "generalized" noise-equivalent quanta. Theoretical calculations of d'_{slice} were performed as a function of total angular extent (θ_{tot}) of source-detector orbit ranging 10°–360° under two acquisition schemes: (i) Constant angular separation between projections (constant-$\Delta\theta$), giving variable number of projections (N_{proj}) and dose vs θ_{tot} and (ii) constant number of projections (constant-N_{proj}), giving constant dose (but variable angular sampling) with θ_{tot}. Five simple observer models were investigated: Prewhitening (PW), prewhitening with eye filter and internal noise (PWEi), nonprewhitening (NPW), nonprewhitening with eye filter (NPWE), and nonprewhitening with eye filter and internal noise (NPWEi). Human observer performance was measured in 9AFC tests for five simple imaging tasks presented within uniform and power-law clutter backgrounds. Measurements (from 9AFC tests) and theoretical calculations (from cascaded systems analysis of d'_{slice}) were compared in terms of area under the ROC curve (A_z)

Results: Reasonable correspondence between theoretical calculations and human observer performance was achieved for all imaging tasks over the broad range of experimental conditions and acquisition schemes. The PW and PWEi observer models tended to overestimate detectability, while the various NPW models predicted observer performance fairly well, with NPWEi giving the best overall agreement. Detectability was shown to increase with θ_{tot} due to the reduction of out-of-plane clutter, reaching a plateau after a particular θ_{tot} that depended on the imaging task. Depending on the acquisition scheme, however (i.e., constant-N_{proj} or $\Delta\theta$), detectability was seen in some cases to decline at higher θ_{tot} due to tradeoffs among quantum noise, background clutter, and view sampling.

Conclusions: Generalized detectability index derived from a 3D cascaded systems model shows reasonable correspondence with human observer performance over a fairly broad range of imaging tasks and conditions, although discrepancies were observed in cases relating to orbits intermediate to 180° and 360°. The basic correspondence of theoretical and measured performance supports the

 0094-2405/2011/38(4)/1754/15/$30.00

application of such a theoretical framework for system design and optimization of tomosynthesis and CBCT. © *2011 American Association of Physicists in Medicine*. [DOI: 10.1118/1.3560428]

Key words: cone-beam CT, tomosynthesis, image quality, detectability index, cascaded systems analysis, observer study, imaging task

I. INTRODUCTION

Tomosynthesis and cone-beam CT (CBCT) using flat-panel detectors (FPDs) offer the potential for improved lesion conspicuity and localization in a wide range of diagnostic and image-guided procedures, including breast imaging,[1–4] chest imaging,[5,6] and surgical interventions.[7,8] The development of new tomosynthesis and CBCT systems for such applications stands to benefit significantly from an understanding of the factors that govern image quality and a theoretical framework for the assessment and optimization of imaging performance.[9]

Image quality assessment commonly involves human observer-based measurements [e.g., receiver operating characteristic (ROC) or alternative forced-choice (AFC) tests] or observer-independent modeling or measurement [e.g., modulation transfer function (MTF), noise-power spectrum (NPS), detective quantum efficiency (DQE), and noise-equivalent quanta (NEQ)]. The former, although time-consuming and requiring careful attention to experimental design to minimize bias, has been applied to evaluate 3D imaging systems.[10] However, due to the broad parameter space associated with system design, acquisition techniques, and reconstruction methods, human observer studies may be impractical as a rigorous approach for designing and optimizing medical imaging systems. Alternative approaches with mathematical observer models have been an active area of research.[11–13] The latter have been commonly used in characterization of radiographic imaging systems,[14–19] with recent research extending the approach to include dual-energy radiography,[20] tomosynthesis,[21–24] cone-beam CT,[25–27] and the incorporation of anatomical background noise.[28] Despite the widespread use of both these broad approaches, there is comparatively little rigorous understanding of the connection between the two for real imaging systems, e.g., how improvement in DQE might (or might not) relate to improvement in ROC, underscored primarily by the fact that these prevalent Fourier metrics do not, in themselves, account for how observers interpret the image data or, more specifically, how detector performance relates to a given imaging task.

A connection between prevalent, practical metrics such as NEQ and human or model observer performance metrics such as ROC offers significant value in system development, e.g., in identifying low-dose performance limits and guiding design improvements (e.g., x-ray converter efficiency, electronics noise, etc.). As generally acknowledged, imaging performance is best defined with respect to an intended task and quantitative frameworks based on statistical decision theory have been proposed to incorporate task in the assessment of image quality.[9,29–33] For example, the detectability index proposed by Green and Swets[34] has been adapted in the context of medical imaging[35,36] in terms of the NEQ and a spatial-frequency-dependent template (task function) corresponding to an ideal observer. Such ideal observer models have proven useful to system optimization in some contexts.[31,37] Other observer models that aim to better describe human observer performance under various imaging conditions (e.g., statistical backgrounds) have been an active area of research.[29,38–44]

The work described below extends a theoretical cascaded systems model[25,26,45] for the 3D NEQ of FPD based tomosynthesis and CBCT to include: (i) Spatial-frequency-dependent task descriptors to yield the detectability index for a variety of idealized imaging tasks and observer models; (ii) background power-law noise, shown to be a major factor in affecting detectability over the continuum of angular extent from low-angle tomosynthesis to CBCT; and (iii) comparison of theoretical calculations of detectability with the performance measured for human observers in real image data. The first is a straightforward interpretation of task-based detectability index (d') outlined in ICRU Report 54 (Ref. 9) in the context of 3D imaging. The second yields so-called "generalized" Fourier metrics [viz., generalized NEQ (GNEQ) and generalized d'], where the term "general" refers specifically to the incorporation of background noise in the NEQ, as described by Barrett *et al.*[38] (distinct from what might be termed "system" NEQ factors such as focal spot size, x-ray scatter, etc., which may also be included[46]). The third directly compares theoretical calculations of model observer performance (based on GNEQ and task function) to the performance measured using human observers, thereby investigating the extent to which task-based detectability index provides a meaningful figure of merit for observer performance and, ultimately, system optimization.

Note that the intent of this work is not to advance the extent to which observer models in themselves are descriptive of human observer performance; rather, the work utilizes a variety of well known observer models and investigates the extent to which a theoretical framework for 3D NPS and NEQ gives correspondence with human observers. The significance of the work lies in the potential to predict imaging performance in a manner that corresponds to that of human observers, e.g., as a function of angular extent, number of projections, and total dose in tomosynthesis and CBCT, from first principles of signal and noise propagation in a 3D cascaded systems model of the imaging system.

II. METHODS

II.A. Generalized detectability index

II.A.1. Cascaded systems analysis

Cascaded systems analysis has been widely used in modeling detector performance in 2D radiography[14,15,19] and

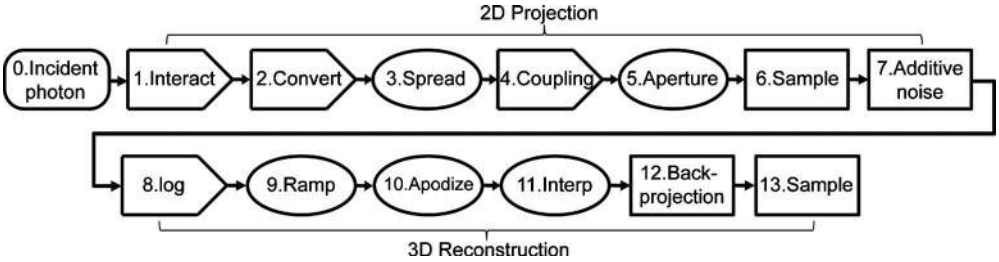

FIG. 1. Cascaded systems analysis consisting of the 2D projection formation (stages 1–7) and 3D reconstruction cascade (stages 8–13).

more recently extended to describe 3D imaging performance in tomosynthesis and CBCT.[24–26] Details of the 3D cascaded systems analysis model have been described in previous works,[26,27] with a summary of the model illustrated in Fig. 1. Briefly, stages 1–7 model the propagation of signal and noise in the formation of the 2D projection image and stages 8–13 model the 3D reconstruction. Specifically, stage 8 describes the log transform of projections, stage 9 the application of the ramp filter, stage 10 the smooth apodization filter, stage 11 the interpolation of projection data for voxel-driven reconstruction, stage 12 the backprojection according to the central slice theorem, and stage 13 the sampling of data in the 3D reconstruction domain. The accuracy and utility of this approach has been demonstrated with measurements of NPS and NEQ in CBCT and tomosynthesis,[23,26,27] providing a general framework for modeling imaging performance for a wide range of system parameters, acquisition techniques, and reconstruction settings. More recently, the model was extended to include background power-law noise ("anatomical noise") in the generalized NEQ.[47]

The detectability index provides a task-based performance metric that combines the NEQ with a task function describing the spatial frequencies of interest[9]

$$d'^{2}_{3D} = \int\int\int \frac{(T \cdot W_{\mathrm{Task}})^2}{S_Q + S_E} df_x df_y df_z, \tag{1}$$

where T is the MTF and S_Q and S_E are the quantum NPS and electronic NPS, respectively. The MTF and NPS are derived directly from cascaded systems analysis as described in previous works.[14,15,26] The term W_{Task} is a task function given by the difference of the Fourier transforms of the spatial representations (spatial domain object functions) of the two hypotheses, e.g., for a detection task, a signal-present hypothesis represented by a volume-of-interest (VOI) containing the signal and a signal-absent hypothesis represented by a background only VOI. Linearity of the Fourier transform suggests this is equivalent to the Fourier transform of the difference of the object functions.[9] Spatial-frequency-dependence (f_x, f_y, f_z) is implicit in T, W_{Task}, S_Q, and S_E corresponding to a 3D detectability index denoted as d'_{3D}.

To include the effect of anatomical background on task performance, the detectability index can be generalized to include anatomical noise as an additional noise source as follows:

$$d'^{2}_{3D} = \int\int\int \frac{(T \cdot W_{\mathrm{Task}})^2}{T^2 S_B + S_Q + S_E} df_x df_y df_z, \tag{2}$$

where S_B is the anatomical background power spectrum, often modeled according to a power-law characteristic[40]

$$S_B(f) = \frac{\kappa}{(af)^{\beta}}, \tag{3}$$

where κ denotes the magnitude of background variations and β the degree of correlation. The term a is a scale factor (taken as 1) with units inverse to frequency (e.g., $a=1$ mm for f in units of mm^{-1}), effectively making the denominator dimensionless despite different values of β. Generalized in this way, detectability index provides a theoretical framework for understanding the tradeoffs among quantum noise, electronic noise, and anatomical background as a function of acquisition parameters (e.g., angular range, number of projections, and dose) and reconstruction methods (e.g., reconstruction filter, sampling, etc.).

II.A.2. Model observers

The detectability index in Eq. (2) corresponds to a prewhitening (PW) matched filter observer that is able to decorrelate image noise. The model can be extended to the nonprewhitening (NPW) matched filter observer[48] that does not estimate the background, but instead applies a detection template in the form of the signal

$$d'^{2}_{3D} = \frac{[\int\int\int (T \cdot W_{\mathrm{Task}})^2 df_x df_y df_z]^2}{\int\int\int (T^2 S_B + S_Q + S_E) \cdot (T \cdot W_{\mathrm{Task}})^2 df_x df_y df_z}. \tag{4}$$

These two models can be extended to include an eye filter $E(f)$ and internal noise N_i to account for response characteristics of the human visual system.[49,50] The PW model extended in this way is denoted PWEi (PW model with eye filter and internal noise) and written as

$$d'^{2}_{3D} = \int\int\int \frac{E^2 (T \cdot W_{\mathrm{Task}})^2}{E^2 (T^2 S_B + S_Q + S_E) + N_i} df_x df_y df_z. \tag{5}$$

Similarly, the NPW model with the eye filter is denoted NPWE and written as

$$d'^{2}_{3D} = \frac{[\iiint E^2(T \cdot W_{\text{Task}})^2 df_y df_x df_z]^2}{\iiint E^4(T^2 S_B + S_Q + S_E) \cdot (T \cdot W_{\text{Task}})^2 df_y df_x df_z} \quad (6)$$

and can be extended further to include internal noise, denoted NPWEi, and written as

$$d'^{2}_{3D} = \frac{[\iiint E^2(T \cdot W_{\text{Task}})^2 df_x df_y df_z]^2}{\iiint \{E^4(T^2 S_B + S_Q + S_E) \cdot (T \cdot W_{\text{Task}})^2 + N_i\} df_x df_y df_z}. \quad (7)$$

The equations above for the 3D detectability index are drawn from a previous work[26] and could, in principle, be related to task performance in 3D images. However, there are currently no established 3D eye filter models, there are no models shown to correspond to a human observer scrolling slices or viewing a slice montage, and there is little established experimental methodology for reliable human observer performance assessment in fully 3D images. The exact form of eye filters and internal noise appropriate to fully 3D data (e.g., slice scrolling, multislice montage presentation, or volumetric viewing) are subjects of ongoing and future works in modeling of image perception and are therefore not invoked in this work. For these reasons, we derived a 2D "slice" detectability index (denoted as d'_{slice}) from the fully volumetric detectabilty index (denoted as d'_{3D}) to correspond to a slice image extracted from the 3D image. As shown in Sec. II A 3, the slice detectability is not derived simply from 2D analysis; rather, one must derive the fully 3D detectability index and then integrate (in the Fourier domain) over the direction corresponding to slice extraction. Just as the 3D image NPS is not correctly analyzed from 2D analysis of a slice from the 3D image (rather, 3D Fourier analysis is required to account for noise correlation in all three dimensions),[51] analysis of slice detectability from a 2D NPS and NEQ is fraught with error and fully 3D analysis (followed by integration along a given direction in the Fourier domain) corresponds appropriately to performing a task within a slice extracted from the 3D image.

II.A.3. Slice detectability vs "3D" detectability

The detectability index in Eqs. (2)–(7) is written in a form in which the observer is assumed to fully perceive the volumetric image information. Although 3D detectability can be derived from the cascaded systems analysis model as in Eq. (1), the human observer tests described below involved reading of coronal slices extracted from the 3D image for purposes of simplicity. To enable comparison with human observer performance, detectability index was derived in a form pertaining to a single 2D slice extracted from the volume. Slice extraction corresponds to integration across the direction orthogonal to the slice (taken as f_y).[51] For the purpose of this work, axial images correspond to the x-y plane, with y corresponding to the depth direction. Coronal slices (i.e., images in the x-z plane) were used in the observer study, corresponding to the usual tomosynthesis view. Therefore, 2D slice metrics can be obtained by integrating 3D metrics in Sec. II A 2 over the f_y direction. It is important to acknowledge that the slice detectability, chosen here simply for comparison to human observer tests, is not a complete metric for optimization of a 3D imaging system. While the fully 3D detectability index may be a suitable optimization metric, the rationale for analyzing slice detectability below was simply for purposes of measuring correspondence to human observer performance and not as a basis for system optimization. The slice detectability corresponding to the 3D observer models above are therefore

PW

$$d'^{2}_{\text{slice}} = \int\int \frac{(\int T \cdot W_{\text{Task}} df_y)^2}{\int (T^2 S_B + S_Q + S_E) df_y} df_x df_z. \quad (8)$$

PWEi

$$d'^{2}_{\text{slice}} = \int\int \frac{E^2(\int T \cdot W_{\text{Task}} df_y)^2}{E^2 \int (T^2 S_B + S_Q + S_E) df_y + N_i} df_x df_z. \quad (9)$$

NPW

$$d'^{2}_{\text{slice}} = \frac{[\iint (\int T \cdot W_{\text{Task}} df_y)^2 df_x df_z]^2}{\iiint (T^2 S_B + S_Q + S_E) df_y \cdot (\int T \cdot W_{\text{Task}} df_y)^2 df_x df_z}. \quad (10)$$

NPWE

$$d'^{2}_{\text{slice}} = \frac{[\iint E^2(\int T \cdot W_{\text{Task}} df_y)^2 df_x df_z]^2}{\iint [E^4 \int (T^2 S_B + S_Q + S_E) df_y] \cdot (\int T \cdot W_{\text{Task}} df_y)^2 df_x df_z}. \quad (11)$$

NPWEi

$$d'^{2}_{\text{slice}} = \frac{[\iint E^2(\int T \cdot W_{\text{Task}} df_y)^2 df_x df_z]^2}{\iint \{[E^4 \int (T^2 S_B + S_Q + S_E) df_y] \cdot (\int T \cdot W_{\text{Task}} df_y)^2 + N_i\} df_x df_z}. \quad (12)$$

The eye filter employed in this study was a simple approximation of Barten's[52] contrast sensitivity curve of the human eye consistent with that used in the study by Burgess[39]

$$E(f) = f \exp(-cf), \quad (13)$$

where f is the spatial radial frequency. The eye filter was implemented such that its maximum response occurred at 4 cycles/deg. For a typical viewing distance of 50 cm, c equals 2.2. The internal noise was implemented as uncorrelated white noise dependent on the magnitude of variation in background power spectra[53]

$$N_i = 0.001 \left(\frac{D}{100}\right)^2 \text{NPS}_{\text{eq}}(0,0), \quad (14)$$

where D is the viewing distance estimated as 50 cm for this study and NPS_{eq} is the white NPS equivalent in total power to the image noise (sum of S_B, S_Q, and S_E). The scale factor 0.001 was fixed following variation as a free parameter as in Burgess[39] to give coarse overall agreement to measurements.

The slice detectability index in Eqs. (8)–(12) formed the basis for theoretical calculations performed in comparison to human observer performance for a variety of imaging tasks described below. Although slice detectability describes imaging performance on a 2D slice, the derivation of such is only achieved via the fully volumetric analysis of the 3D NPS, NEQ, and d'_{3D}. The 2D slice detectability index is derived simply to evaluate correspondence to measurements in the observer study and is not intended as an objective function for system optimization.

II.B. Imaging tasks and physical imaging phantoms

The sections below describe physical experimentation conducted to allow direct comparison of theoretical d'_{slice} (derived above) with the performance measured in human observers. A variety of imaging tasks were implemented in real phantoms in a manner that imparted a range of conspicuity (from imperceptible to obvious) over a broad range of experimental conditions (number of projections, dose, and source-detector orbital extent).

II.B.1. Detection in uniform background

Detection in a uniform background was investigated as the simplest case in which the NEQ was governed by quantum and electronic noise only (no background clutter). A physical phantom was used that consisted of acrylic spheres of various sizes embedded in a uniform polyurethane cylinder. As described below, the task corresponded to the detection of a (110 HU) sphere against a uniform (90 HU) background. For tomosynthesis and CBCT, a sphere of 6.4 mm diameter was selected as the signal to present strongly varying conspicuity across a range of source-detector orbital extent (θ_{tot}).

II.B.2. Detection/discrimination in background clutter

A previous work[47] described a phantom designed from principles of fractal self-similarity that contained different diameter spheres randomly mixed to give power-law spectral density. Power law parameters κ and β can be adjusted according to the contrast of the spheres and the proportion of various diameters, respectively. In this work, an equal volume of acrylic spheres of five diameters (15.9, 12.7, 9.5, 6.4, and 3.2 mm) were placed in an acrylic box of dimensions $(20\times20\times12.5)$ cm^3. As previously reported, κ and β were measured to be $\kappa=3.72\times10^{-7}$ (μ^2mm^3) and $\beta=2.76$ and were taken as empirical parameters in the generalized detectability index. The variation in background power spectrum (S_B) with θ_{tot} is accounted through the product of the fully 3D $S_B(f_x,f_y,f_z)$ with the tomosynthesis "double-wedge" corresponding to the angular range of sampled frequencies in the Fourier domain.

II.B.3. Imaging tasks

Imaging tasks were conceived that could be physically implemented in either the uniform or clutter phantoms and modeled according to a simple binary hypothesis-testing model. Six identical objects were inserted in the phantom, giving six statistically independent trials for each imaging task. "Signal-present" images were taken from the central coronal slice through each object and ROIs were selected such that signals were at the center (see Fig. 2). On the other hand, "noise-only/signal-absent" images were taken from the same or neighboring slices without the signal. Five imaging tasks emphasizing different regions of the frequency domain (i.e., various spatial-frequency contents) were investigated in this study. Task functions are plotted in Fig. 2 and described in the following section.

II.B.3.a. Sphere detection on uniform background. Signal detection in an otherwise uniform phantom such as the one described in Sec. II B 1 corresponds to the case in which H_1 is simply the signal represented by the object function, denoted $O(x,y,z)$ (i.e., a sphere in the 3D image and a disk in a 2D slice), and H_2 is a constant. The task function W_{Task} is given by the product of the difference in attenuation coefficient between the signal and background ($\Delta\mu$) and the Fourier transform of the signal (FT$\{O\}$). For the phantom described in Sec. II B 2, the task function was computed with $\Delta\mu$ taken as the measured signal difference between the acrylic sphere (110 HU) and polyurethane background (90 HU) in a full 360° CBCT reconstruction and FT$\{O\}$ computed numerically as the Fourier transform of a 6.4 mm diameter sphere in the 3D image

$$W_{Task} = \Delta\mu \cdot \mathrm{FT}\{O\} = (\mu_{Acrylic} - \mu_{Polyurethane}) \cdot \mathrm{FT}\{O_{6.4\ mm\_Sphere}\}. \quad (15)$$

Signal values used in calculation of $\Delta\mu$ were measured as an average of multiple ROIs at various positions on the central coronal slice of the reconstruction.

II.B.3.b. Large sphere detection on cluttered background. Imaging tasks in cluttered background were modeled as signals (i.e., physical objects) embedded in the clutter phantom of Sec. II B 2. For tasks presented on cluttered background, the signal-present images contain the signal at the center of the ROI, whereas "signal-absent" images present a background sphere of equivalent size (e.g., 12.7 mm for the large sphere) at the same location. The signal-present hypothesis was formulated as

$$\mathrm{FT}\{H_1\} = \mu_1\mathrm{FT}\{O_1\} + \mathrm{FT}\{B\}, \quad (16)$$

where μ_1 and O_1 are the attenuation coefficient and object function of the signal, respectively, and B corresponds to background clutter (random collection of acrylic spheres), the power spectrum of which obeys the power-law relationship [Eq. (2)]. The signal-absent hypothesis is

$$\mathrm{FT}\{H_2\} = \mu_2\mathrm{FT}\{O_2\} + \mathrm{FT}\{B\} = \mu_2\mathrm{FT}\{O_1\} + \mathrm{FT}\{B\}, \quad (17)$$

where μ_2 is the attenuation coefficient of the background (acrylic) and O_2 (equal to O_1 in this case) is the object function corresponding to an (acrylic) sphere of the same size in place of the signal within the background clutter. Note that the Fourier transform describes magnitude only (disregarding phase), so FT$\{B\}$ in theory may pertain to various inde-

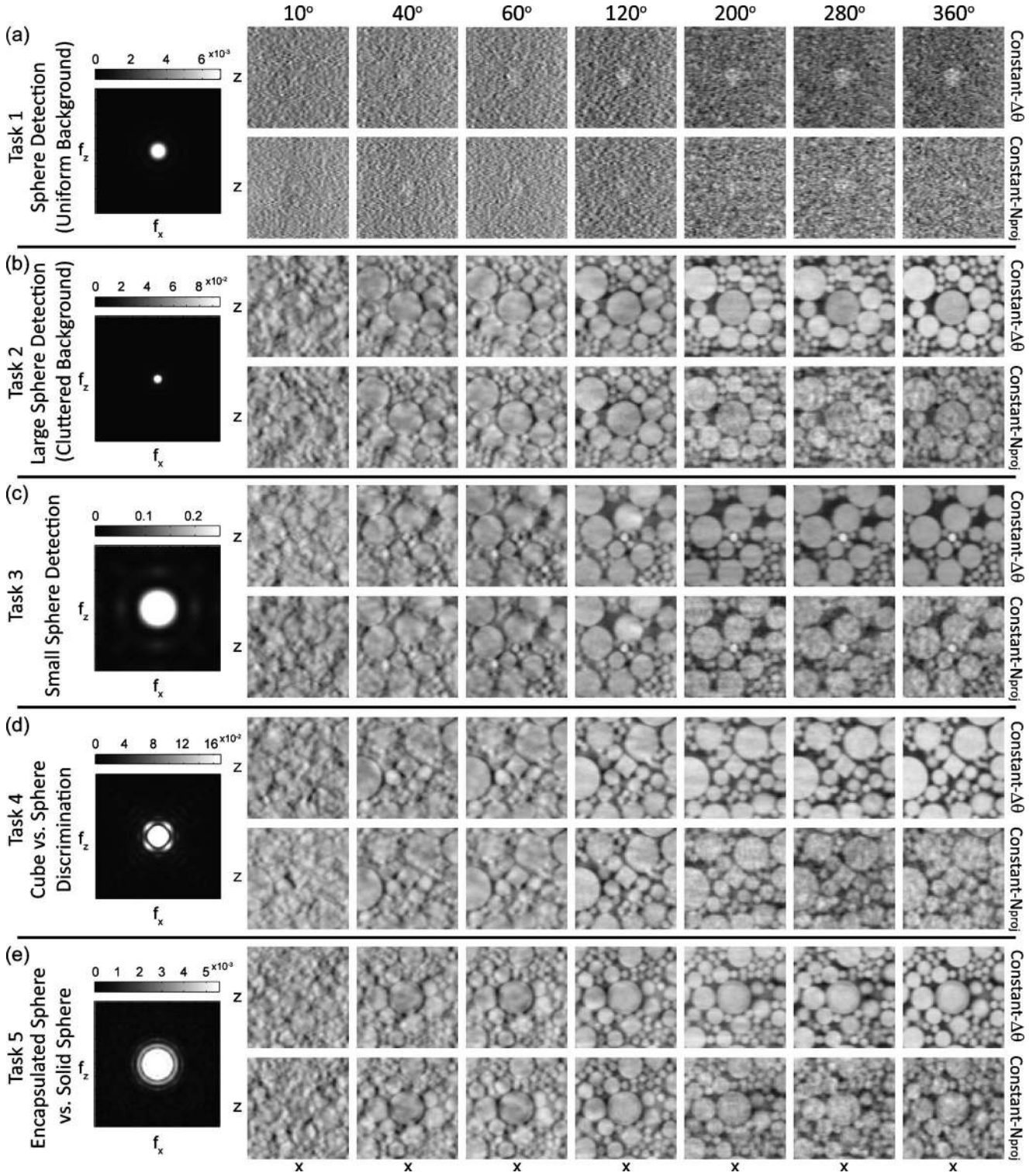

FIG. 2. Fourier-domain task functions (left column) and coronal image ROIs (x-z) (images at the right) for varying angular extent under the constant-$\Delta\theta$ and constant-N_{proj} cases: (a) Sphere detection on uniform background; (b) large sphere detection in clutter; (c) small sphere detection in clutter; (d) cube vs sphere discrimination in clutter; and (e) encapsulated sphere vs solid sphere discrimination in clutter.

pendent realizations of the background, provided they have the same (e.g., power-law) noise characteristic. Different realizations of the background therefore cancel out in theory, such that discrimination of the signal (i.e., O_1 in clutter) from a clutter-only image corresponds to

$$\bar{W}_{\text{Task}} = \text{FT}\{H_1\} - \text{FT}\{H_2\} = (\mu_1 - \mu_2) \cdot \text{FT}\{O_1\} = \Delta\mu \cdot \text{FT}\{O_1\}. \quad (18)$$

For the large sphere on a cluttered background task, the sig-

nal was a 12.7 mm diameter polypropylene sphere (−84 HU), which presented a lower contrast signal in comparison to an equivalent size acrylic sphere (110 HU) in the background clutter. The contrast $\Delta\mu$ was taken as the (absolute value) difference in attenuation between polypropylene and acrylic as measured from a full 360° CBCT acquisition. The task function is thus given by

$$W_{\text{Task}} = (\mu_{\text{Acrylic}} - \mu_{\text{Polypropylene}}) \cdot \text{FT}\{O_{12.7\ \text{mm\_Sphere}}\}. \quad (19)$$

As illustrated in Fig. 2, this task presented primarily low and midfrequency components.

II.B.3.c. Small sphere on cluttered background. Similar to the previous task, detection of a small Teflon sphere (710 HU) in a cluttered background of acrylic spheres (110 HU) follows Eq. (18), where $\Delta\mu$ is the measured contrast of Teflon and acrylic and O_1 is the object function for a small (3.2 mm diameter) sphere

$$W_{\text{Task}} = (\mu_{\text{Teflon}} - \mu_{\text{Acrylic}}) \cdot \text{FT}\{O_{3.2\ \text{mm\_Sphere}}\}. \quad (20)$$

As illustrated in Fig. 2, this task presented higher-frequency components.

II.B.3.d. Cube vs sphere discrimination on cluttered background. Shape discrimination was similarly modeled as the Fourier transform of the difference of two hypotheses

$$\text{FT}\{H_1\} = \mu_1 \text{FT}\{O_1\} + \text{FT}\{B\}, \quad (21)$$

$$\text{FT}\{H_2\} = \mu_1 \text{FT}\{O_2\} + \text{FT}\{B\}, \quad (22)$$

where the attenuation coefficient (μ_1) is the same for the signal and background, but object functions of the signals to be discriminated (O_1 and O_2) vary. The task function is therefore

$$W_{\text{Task}} = \text{FT}\{H_1\} - \text{FT}\{H_2\} = \mu_1 \cdot (\text{FT}\{O_1\} - \text{FT}\{O_2\}). \quad (23)$$

A "cube vs sphere" discrimination task was formulated such that O_1 representing an acrylic cube (of side length 6.4 mm) is discriminated from O_2 representing an acrylic sphere (of diameter 6.4 mm), i.e.,

$$W_{\text{Task}} = \mu_{\text{Acrylic}} \cdot [\text{FT}\{O_{6.4\ \text{mm\_Cube}}\} - \text{FT}\{O_{6.4\ \text{mm\_Sphere}}\}]. \quad (24)$$

As illustrated in Fig. 2, this task consists of middle and high-frequency components.

II.B.3.e. Encapsulated vs solid sphere on cluttered background. A second shape discrimination task involved a 6.4 mm diameter acrylic sphere (110 HU) encapsulated by a 3.2 mm shell of paraffin wax (giving 12.8 mm total diameter, −50 HU) as the signal, which is discriminated from a solid acrylic sphere (110 HU) in the noise-only image. The two hypotheses can be similarly written as Eqs. (21) and (22), with O_2 representing the paraffin-encapsulated sphere and O_1 representing the 12.7 mm diameter acrylic sphere. The task function can be derived as

$$W_{\text{Task}} = (\mu_{\text{Acrylic}} - \mu_{\text{Wax}}) \cdot [\text{FT}\{O_{3.2\ \text{mm\_Shell}}\}], \quad (25)$$

where the difference in object functions yields the 3D encapsulating shell (an annulus on a 2D slice). As shown in Fig. 2, this task emphasizes higher frequencies compared to tasks in Secs. II B 3 a, II B 3 b, and II B 3 d. The high-frequency content is related to the fine detail associated with the encapsulating layer.

II.B.4. Imaging bench and acquisition parameters

Images were acquired on an experimental imaging bench for tomosynthesis and CBCT. As described in previous works,[26,47] the system includes an x-ray tube (Rad 94 in a sapphire housing; Varian Medical Systems, Salt Lake City, UT), an FPD (RID-1640A, 0.4 mm pixel pitch, 1024×1024 pixels; Perkin Elmer Optoelectronics, Santa Clara, CA) with ~250 mg/cm^2 CsI:Tl scintillator, and a motion control system (6K series translation stages, Parker Daedal, Harrison, PA, and Dynaserv rotation motor, Parker Hannifin, Rohnert Park, CA) that sets the system geometry to that approximating CBCT-guided radiotherapy (93.5 cm source-to-axis distance and 144 cm source-to-detector distance) with the phantom rotated at isocenter. Acquisition techniques were held fixed at 120 kVp (1.53 mm Al+0.1 mm Cu added filtration) and 0.63 mA s per projection, imparting a constant in-air exposure per projection of 0.49 mR at the detector.

Projections were acquired for 12 orbital extents (θ_{tot}) spanning a continuum of low-angle tomosynthesis to full CBCT: 10°, 20°, 40°, 60°, 90°, 120°, 160°, 200° (180° +fan), 240°, 280°, 320°, and 360°. For each orbital extent, two general acquisition schemes were adopted. The first was a "constant-$\Delta\theta$" case in which a constant angular separation of 0.45° was fixed between projections, giving a variable number of projections (and total dose) for each setting of θ_{tot}. In this scheme, therefore, as θ_{tot} increases (and background clutter is reduced), quantum and electronics noise vary, while view-sampling effects are constant. The second scheme was a "constant-N_{proj}" case in which the number of projections was fixed at 89 for all θ_{tot}. This scheme represents the practical case where total dose is a fixed constraint and θ_{tot} needs to be chosen in a manner that optimally manages the tradeoffs among background clutter, quantum noise, and view-sampling effects. Together, these schemes allow investigation of the fairly complex tradeoffs among competing noise sources and a broad set of experimental conditions against which to test the theoretical model.

Image reconstruction was performed using the FDK algorithm for 3D filtered backprojection (FBP), with a Hann apodization filter and no additional interslice filter for tomosynthesis. Projections were binned 2×2 pixels and images were reconstructed at isotropic voxel size $(0.52\times0.52\times0.52)$ mm^3, chosen to adequately resolve the smallest sphere in the phantom. Modified Parker weights[54] were applied to redundant views of angle greater than 180°−fan, with the exception of a full 360° acquisition where a uniform weighting of 0.5 was multiplied to all projections.

II.C. Experimental validation: Human observer performance

II.C.1. Human observer study

As a simple measure of observer-based imaging performance, multiple-alternative forced-choice tests were performed in which an array of ROIs was shown to observers, with one ROI containing the signal and $M-1$ ROIs containing background only. A 9AFC test ($M=9$, displayed as a 3×3 array of ROIs) was chosen to give optimal statistical power for the selected tasks according to the table formulated by Elliot,[55] which relates sensitivity of measurements of detectability and proportion correct (P_{corr}) to the number of choices. Each signal ROI was a 70×70 pixel subimage cropped from a coronal slice of a 3D reconstruction with the signal at the center of the ROI. The noise ROIs were cropped from the same or neighboring slices and care was taken to avoid out-of-plane shadow of the signal or other artifacts. Different noise realizations were used for each imaging task, but the same regions of interest were used for various θ_{tot} in both the constant-$\Delta\theta$ and constant-N_{proj} schemes. In addition, both the signal and noise ROIs were randomly flipped up/down and left/right to minimize observer familiarity with the images. The grayscale window was fixed to a range of 90% of the minimum to 110% of the maximum pixel value and the level was set to the mean. Observers were not allowed to adjust the window/level or zoom of the images, and a constant viewing distance of ~50 cm was encouraged but not strictly enforced.

Observer studies were conducted in a darkened reading room using a monochrome diagnostic quality display (Image Systems, Richardson Electronics Ltd., Plymouth, MN). Prior to each test, observers were trained using images acquired at the same experimental conditions as the test data, typically 24 images for a given task, requiring ~10 min to gain familiarity with the task. The test data for each task included images acquired at 12 levels of θ_{tot} (detailed above) for both the constant-$\Delta\theta$ and constant-N_{proj} schemes, with five statistically independent samples acquired for each case, giving (12 angles) × (2 schemes) × (5 images) = 120 readings. The order of the five tasks was randomized for each observer, as was the order of images presented for each task. At ~5 s per choice, a complete study required (5 tasks) × (~10 min)(training) + (5 tasks) × (12 angles) × (2 schemes) × (5 images) × (~5 s)(test) ≅ 100 min for each observer.

For the fairly simple (nonclinical) imaging tasks involved in these phantom studies, medical physicists/engineers were considered suitable observers. Eight observers were involved, yielding (8 observers × 5 independent images =) 40 responses for each data point on the graphs below. Assuming independence among observers, all responses were pooled and the fraction of correct responses was computed to yield the mean proportion correct (P_{corr}).

Measurements of P_{corr} follow a binomial distribution with mean equal to P_{corr} and standard deviation $P_{corr}(1-P_{corr})$. Statistical error in P_{corr} can be estimated from the standard deviation or confidence interval. As mentioned below, P_{corr} measurements were also interpreted in terms of d'_{slice} and area under the ROC curve (A_z). Based on the binomial distribution of P_{corr}, the distribution of A_z was derived using the relations between A_z and P_{corr} (below). The corresponding measurement error in A_z was expressed as the 95% confidence intervals calculated from the distribution of A_z.

II.C.2. Comparison of theoretical and experimental results

For direct comparison of theoretically derived detectability index [d'_{3D} or d'_{slice} in Eqs. (4)–(12)] and experimentally measured P_{corr}, we used the basic relationships among d', P_{corr}, and A_z that follow from the simplifying assumptions of normal, equal variance distributions in the underlying decision variables, consistency in observer response over the course of the tests, etc.[9] The detectability index and A_z are related by

$$A_z=\frac{1}{2}+\frac{1}{\sqrt{\pi}}\int_0^{d'/2}e^{-x^2}dx=\frac{1}{2}\left(1+erf\left(\frac{d'}{2}\right)\right), \tag{26}$$

$$d'^2=4inverf^2\left[2\left(A_z-\frac{1}{2}\right)\right], \tag{27}$$

which in turn are related to P_{corr} as

$$P_{corr}(d',M)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{(x-d')^2}{2}\right)[\phi(x)]^{M-1}dx, \tag{28}$$

where M is the number of alternatives in the AFC test (9 herein) and ϕ is the cumulative Gaussian distribution. Note the usual relationship: $P_{corr}=A_z$ for $M=2$. A lookup table relating P_{corr}, d', and A_z was constructed using these relations. Theoretical and experimental results could therefore be directly compared in terms of any of these three performance metrics. In selecting a metric by which to display the results below, d' (which is unbounded from 0 to ∞) and P_{corr} (which is less general in its interpretation due to a particular choice of M) were felt to be somewhat less meaningful at a glance. While any of the three metrics would suffice for purposes of comparison, results below are reported in terms of A_z. Compared to P_{corr}, A_z provides more general representation of results that can be compared to other studies (e.g., ROC performance); compared to d', the A_z metric is bounded and easily interpreted within the limits of pure guessing (A_z

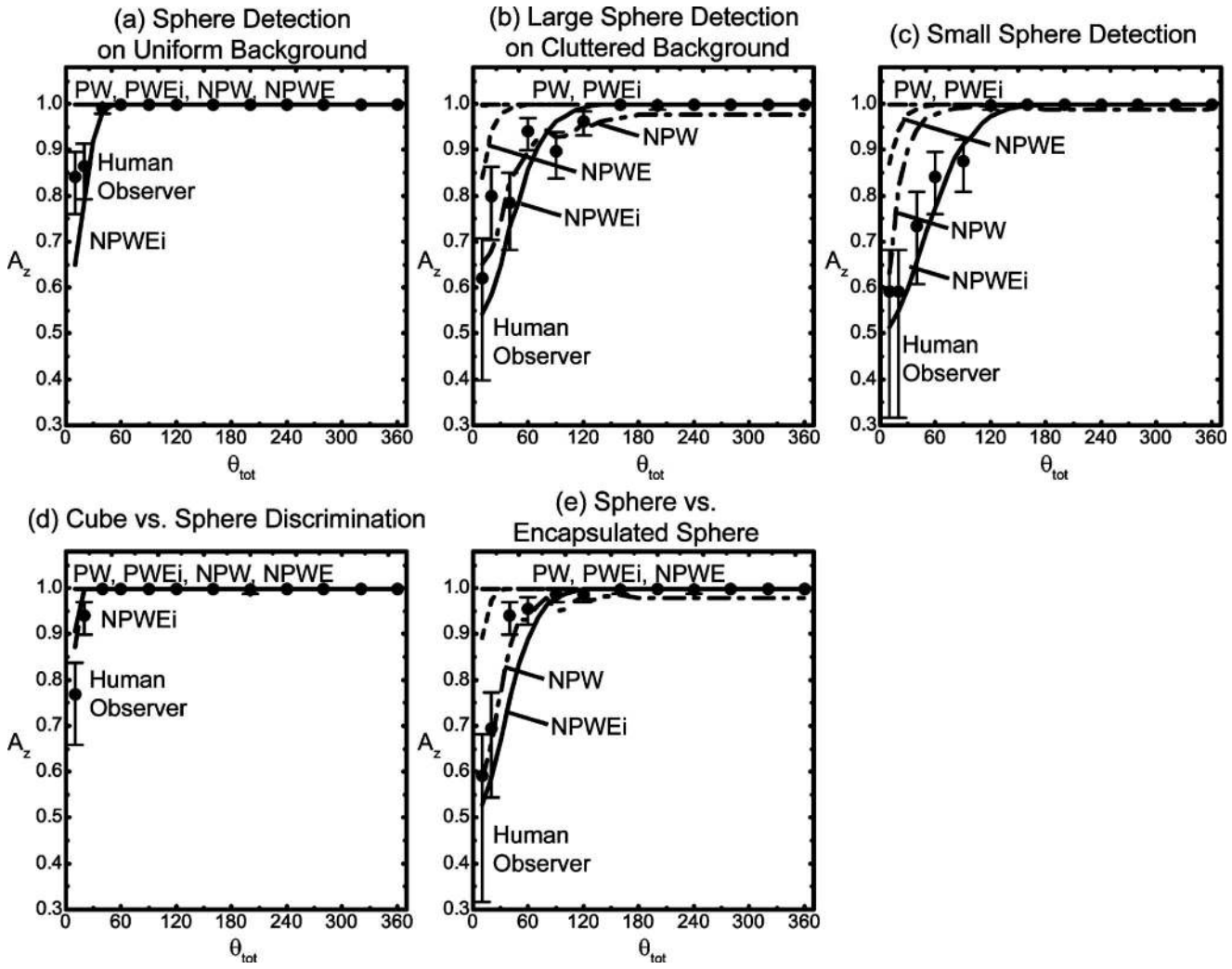

FIG. 3. Comparison of theoretical and measured performance for five imaging tasks in the constant-$\Delta\theta$ acquisition scheme. Curves correspond to theoretical calculations for the five observer models (PW, PWEi, NPW, NPWE, and NPWEi) of Eqs. (8)–(12). Reasonable correspondence between theoretical and experimental results is observed, with NPWEi showing the best agreement overall.

$\rightarrow 0.5$) and "completely obvious" ($A_z \rightarrow 1.0$), which the physical phantoms and experimental conditions were constructed to span. Comparison in terms of d' or P_{corr} (not shown) exhibited the same level of agreement and does not affect the conclusions of the work.

III. RESULTS

III.A. Comparison of theoretical detectability and human observer performance

Figures 3 and 4 plot A_z vs total angular extent (θ_{tot}) for all five imaging tasks under the constant-$\Delta\theta$ and constant-N_{proj} cases, respectively. In each graph, theoretical calculations from the cascaded systems model are plotted as solid or dashed curves and human observer measurements are represented by individual data points with error bars corresponding to 95% confidence intervals.

The constant-$\Delta\theta$ acquisition scheme (Fig. 3, $\Delta\theta$=0.45° for all cases) shows a monotonic increase in A_z with angular extent as contrast is improved for task on uniform background (task 1) and as out-of-plane clutter is gradually removed for tasks on cluttered background (tasks 2–5). Furthermore, in the constant-$\Delta\theta$ case, total dose increases with θ_{tot}, further supporting a monotonic increase in performance. For each task, a maximum A_z of 1 is reached at a certain value of θ_{tot}, beyond which all model and human observers identify the signal as conspicuous. Of the five tasks considered, task 1 (sphere detection on a uniform background) and task 4 (cube vs sphere discrimination) were the easiest, with conspicuity predicted and realized in the region $\theta_{tot} < \sim 30°$. Conversely, task 2 (large sphere detection in clutter), task 3 (small sphere detection in clutter), and task 5 (encapsulated sphere discrimination in clutter) were more challenging, as shown by both theoretical calculations and observer results peaking in the range $\theta_{tot} \sim 100° - 150°$. Task 2 achieved better performance than task 3, seen by a steeper increase in A_z with θ_{tot}. The interpretation is that low-frequency tasks associated with detection of signals of larger spatial extent reside in a similar frequency range as the power-law background and thus experience faster increase in detectability with the

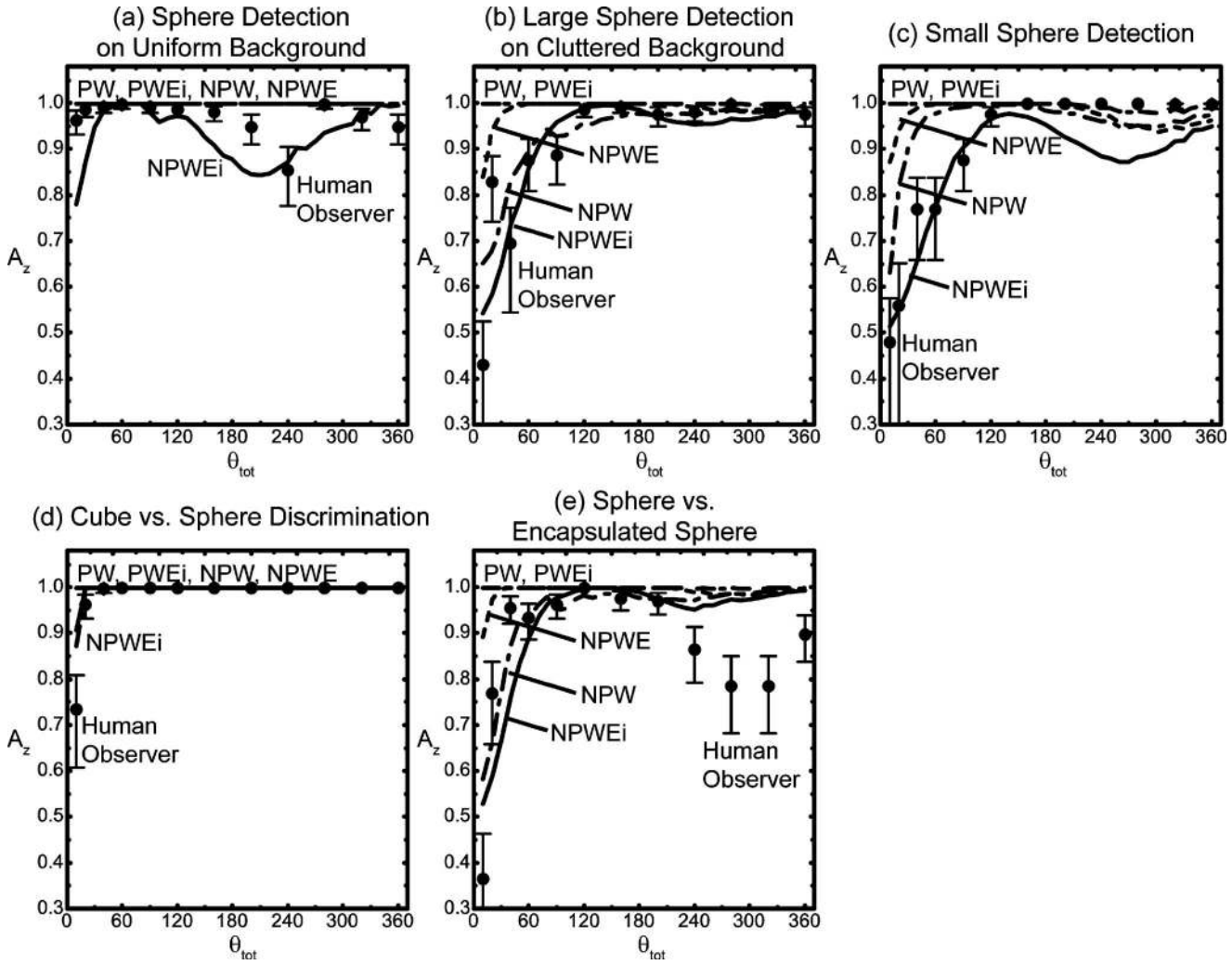

FIG. 4. Comparison of theoretical and measured performance for five imaging tasks in the constant-N_{proj} acquisition scheme. Labeling is the same as in Fig. 3. Fair correspondence is observed between theoretical and experimental results, including a complex nonmonotonic trend in performance at large angles above $\theta_{tot} \sim 180°$ associated with finite sampling effects (view aliased noise).

rejection of out-of-plane clutter (i.e., increase in θ_{tot}). High-frequency tasks, on the other hand, associated with finer details exhibit frequency content in a higher range than the background and thus do not benefit as significantly from the rejection of out-of-plane clutter. The prewhitening models (PW and PWEi) tended to overestimate observer performance in all cases, while the nonprewhitening models (NPW, NPWE, and NPWEi) exhibited reasonable agreement with measurement for all tasks over the entire continuum of θ_{tot}. Not surprisingly, the NPWEi model gave closer correspondence to the measured human performance than NPWE, since it includes additional terms related to inefficiency imparted by internal noise. We examined the observer efficiency implied by the ratio of measured and theoretical d'_{slice} (squared) and found a complicated dependence on task (as might be expected from Abbey *et al.*[56]) and angular extent (θ_{tot}). Efficiency varied in the range $\sim 0.1-0.5$ but did not exhibit any clear trends in retrospective analysis as a function of task or θ_{tot}.

Figure 4 summarizes theoretical and experimental results for the constant-N_{proj} acquisition scheme ($N_{proj}=89$ projections for all cases). In this case, the total dose is fixed for settings of θ_{tot}, giving a fixed level of quantum noise amid varying levels of background clutter and view-aliasing noise. Similar to the constant-$\Delta\theta$ case, the PW and PWEi observer models overestimated observer performance, whereas the various NPW models yielded reasonable agreement with experimental results. An exception was observed for the small sphere detection task, where human observer performance agreed with NPWEi model for $\theta_{tot} < 180°$, but was closer to the PW observers beyond $\theta_{tot} \sim 180°$. This result is evident also in Fig. 2(c), where due to the high contrast of the sphere, the signal was sufficiently different from the background for $\theta_{tot} > 180°$ for conspicuous discrimination and noise (including view-sampling effects) does not appear to "masquerade" as signal. An interesting nonmonotonic trend in A_z vs θ_{tot} was predicted and observed: A_z initially increases with θ_{tot} in a manner similar to the constant-$\Delta\theta$ case (due to reduced out-

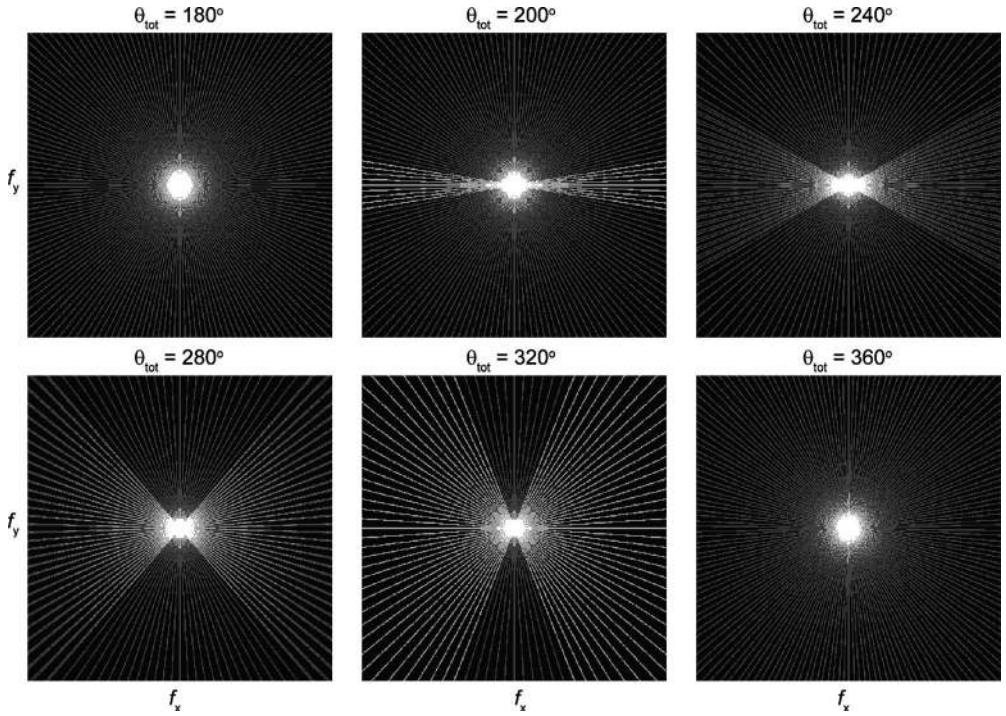

FIG. 5. Illustration of Fourier-domain projection distribution on the axial (f_x-f_y) plane according to the central slice theorem. A total of 89 projections are distributed across a range of θ_{tot}. Under the assumption of parallel-beam geometry, redundant projection views are sampled in the θ_{tot}=200°, 240°, 280°, and 320° cases (i.e., competing effects of increased view aliasing and reduced stochastic noise), while a 360° acquisition results in interleaved projection views and finer sampling (i.e., reduced view aliasing and an increase in detectability). The sampling distribution depends on the number of projections and angular extent. For example, evenly distributing 89 projections over 320° gives redundant, overlapping projections under the parallel-beam approximation ("brighter" spokes); however, at 360°, projections do not overlap and result in finer angular sampling.

of-plane clutter); however, as the total angle increases further, performance is seen to degrade somewhat, particularly for imaging tasks involving higher spatial-frequency components (tasks 3 and 5), followed by an increase toward θ_{tot} =360°. The nature of the effect is detailed more completely below, owing to the distribution of a limited number of projections over wide angular ranges. While the nonmonotonic trend was predicted and observed to occur in the range $180° < \theta_{tot} < 360°$, agreement between theory and measurement is far from perfect: The effect appears to be overestimated for task 3 (small sphere detection) and underestimated for task 5 (encapsulated sphere discrimination). In task 1 (sphere detection on uniform background), theoretical calculations for the NPWEi observer model are consistent with the general trend of the observer measurement, but do not show exact correspondence in A_z values. The other tasks (involving lower spatial-frequency tasks) were in better agreement.

III.B. Number of projections and orbital extent: Effect on quantum noise and view aliasing

The initial decrease in detectability starting at $\theta_{tot} \approx 150°$ under the constant-N_{proj} scheme can be attributed to view-aliasing artifacts arising for large angular separation between projections, visible to the eye as granular mottle in the coronal images of Fig. 2 (evident also as familiar radial streaks in the correspondent axial images, not shown). As θ_{tot} is further increased to 360°, detectability is seen to recover as a result of two possible scenarios of projection distribution illustrated in Fig. 5: (1) A range of projections overlap, forming a region of redundant sampling (to which Parker weights are appropriately applied) or (2) projections do not overlap, but instead result in a region of finer sampling. The first scenario almost always applies for the constant-$\Delta\theta$ case because the finely spaced projections approximate a continuum in which projections at angles above 180° fan constitute redundant samples. In the constant-N_{proj} case, however, the first scenario may only occur under certain combinations of N_{proj} and θ_{tot}, in which case the summing and averaging of redundant rays passing through a particular voxel reduces stochastic noise components (quantum and electronics noise) associated with the projections, but does not affect the magnitude of "deterministic" noise (background clutter) or the signal (task function). The number of redundant projections increases as the angular extent increases. Thus, greater reduction in stochastic noise competes with deteriorating view-aliasing artifacts as angular extent increases, resulting in the nonmonotonic reduction in detectability observed in Fig. 4. An alternative scenario may also arise due to the large $\Delta\theta$ in the constant-N_{proj} scheme, where projections acquired above 180° fan interleave between previously acquired projections, resulting in a region of finer sampling. In such cases, reduction of view-aliasing artifacts also causes detectability to recover.

The distribution of projections in the Fourier domain (axial f_x-f_y plane) according to the central slice theorem is plotted in Fig. 5 for six values of θ_{tot}. Note that the figure

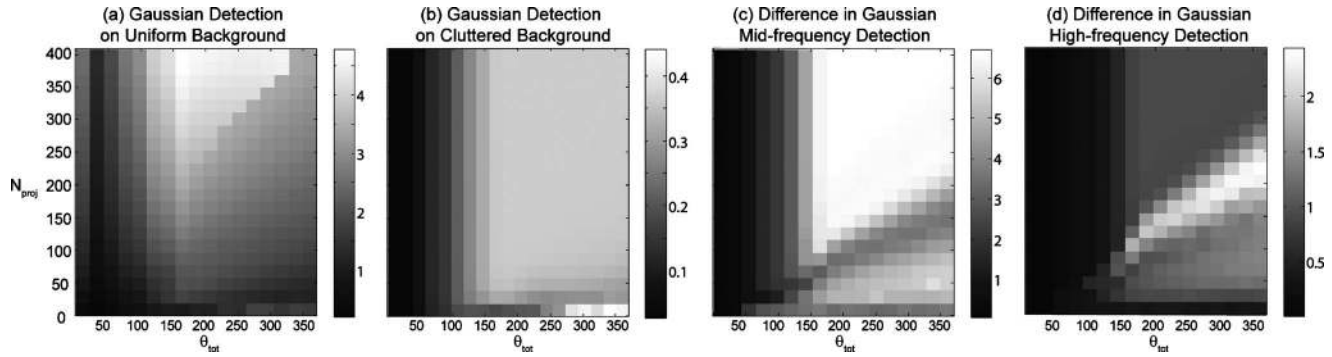

FIG. 6. Surface plots of d'_{slice} [in a coronal slice as computed by Eq. (12)] for a wide range of θ_{tot} and N_{proj} for 4 imaging tasks: [(a and b)] A low-frequency task corresponding to detection of a 3.6 mm Gaussian on (a) uniform and (b) cluttered backgrounds; [(c and d)] a higher-frequency task corresponding to discrimination of two Gaussians of size (c) 3.1 and 3.7 mm and (d) 1.3 and 1.8 mm, each on a cluttered background.

illustrates the angular sampling distribution associated with a given number of projection over a given angular range and does not depict the actual noise-power spectrum. Redundant, overlapping projections appear as brighter spokes. The intermediate angular ranges (θ_{tot}=200°, 240°, 280°, and 320°) exhibit the behavior associated with scenario 1, in which projections overlap across a range of angles above 180° [illustrated in Fig. 5 by increased intensity (brightness) in overlapping rays]. For this specific number of projections (N_{proj} =89), the 360° case actually corresponds to finer sampling and gives a reduction in view-aliasing artifacts due to interleaving of projections. In relation to the valley observed in Fig. 4, detectability begins to decrease above $\theta_{tot} \sim 180°$ due to increased view-aliasing artifacts, followed by an increase beyond $\theta_{tot} \sim 280°$ due to the reduction of stochastic noise with a greater number of redundant projections (corresponding to scenario 1). At 360°, detectability further increases due to finer sampling (corresponding to scenario 2).

III.C. Implications for task-based system design in tomosynthesis and CBCT

Given the reasonable correspondence observed between theoretical and experimental results in Figs. 3 and 4, we computed detectability for a variety of conditions as could apply to the design and understanding of tomosynthesis and CBCT systems. A spectrum of tasks were considered to elucidate the distinction of low-frequency and high-frequency tasks: (a) Low-frequency Gaussian detection task (Gaussian width, σ=3.6 mm) on a uniform background; (b) low-frequency Gaussian detection (σ=3.6 mm) in a cluttered background; (c) midfrequency discrimination of two Gaussian signals (σ_1=3.1 mm vs σ_2=3.7 mm); and (d) high-frequency discrimination of two Gaussians signals (σ_1 =1.3 mm vs σ_2=1.8 mm). In each case, d'_{slice} was computed as a function of N_{proj} and θ_{tot} (with the dose per projection view fixed at a level corresponding to 0.066 mR in-air exposure to the detector) to examine tradeoffs among background clutter and view aliasing. The NPWEi model was chosen, since it demonstrated the best overall agreement with human observer response (Figs. 3 and 4). For simplicity in these calculations, a parallel-beam geometry was assumed, implying that projections 180° apart were considered redundant and were multiplied by a Parker weight of 0.5.

Results are shown in Fig. 6. For any of the tasks, one may consider a "horizontal" slice of the d'_{slice} surface as the case in which the total dose is fixed and the angular range of the tomosynthesis/CBCT system is varied. In practical terms, this might correspond to a rotational C-arm system, in which one seeks to determine what value of θ_{tot} provides a desired level of detectability. Conversely, a "vertical" slice of the d'_{slice} surface corresponds to a fixed angular range, with the number of views and total dose varied. This might correspond to a tomosynthesis system with fixed θ_{tot} (e.g., a clinical breast or chest tomosynthesis system) and one seeks to determine what number of views (total dose) gives a desired level of detectability. The alternative case (not shown) in which d'_{slice} (N_{proj}, θ_{tot}) was computed at fixed total dose shows a complementary set of tradeoffs among quantum noise, background clutter, and view aliasing.

As shown in Fig. 6, for all four tasks, when the number of projections is low (constant-N_{proj} with $N_{proj} < \sim 200$), the trends observed in the experiments of Figs. 3 and 4 are once again observed: d'_{slice} increases with angle, reaches a maximum, then decreases due to competing effects of view aliasing and distribution of projections described above. As the number of projections increases (and view-aliasing artifacts are reduced), a distinct difference is observed among the tasks in uniform and cluttered backgrounds. After reaching a maximum, d'_{slice} decreases with θ_{tot} for the uniform background task while remaining constant for the cluttered background tasks. In the absence of clutter, increasing the angular arc only increases quantum noise, therefore causing detectability to decrease; however, in a cluttered background, this effect is less pronounced due to the preponderance of background noise outweighing quantum noise. Note also the overall reduction in the magnitude of d'_{slice} in Fig. 6(b) compared to Fig. 6(a) despite equivalent task and signal power, attributable to background noise. Another difference between the uniform and cluttered background tasks can be seen from profiles at a fixed angle: For a given number of projections, the uniform background task exhibits continued improvement in detectability with N_{proj} as quantum noise is driven

down by increasing dose, whereas the cluttered background tasks reach a background-noise-limited region, beyond which increasing N_{proj} or dose brings no further improvement in detectability.

Figures 6(b)–6(d) illustrate tasks with the same signal power but different frequency characteristics. The detectability index for the low-frequency task (b) is much lower than that for midfrequency and high-frequency tasks in (c) and (d), respectively, due to the frequency components of the task function coinciding with those of background noise. This corresponds to the observation of Myers *et al.*[57] regarding noise masquerading as signal. For low-angle tomosynthesis (i.e., low θ_{tot}), increasing the acquisition angle and thereby removing out-of-plane clutter demonstrates more pronounced improvement for low-frequency tasks than for higher-frequency tasks.

On the other hand, detectability index for higher-frequency tasks is more sensitive to the choice of θ_{tot} and N_{proj}, since they reside in the same frequency region as view-aliasing noise and are therefore more prone to the complex tradeoffs among noise components described above. View-aliasing artifacts within the low N_{proj} and high θ_{tot} region degrades performance for midfrequency (c) and high-frequency (d) tasks, but barely affects the low-frequency (a) task. For the midfrequency task in (c), a sharp transition in the magnitude of d'_{slice} can be observed along a diagonal with an approximately constant choice of $\Delta\theta$. Similarly for the high-frequency task in (d), there is a narrow selection of θ_{tot} and N_{proj} with an approximately constant choice of $\Delta\theta$, which marks optimal performance. The trend observed is highly dependent on the frequency range of the imaging tasks, choice of observer model, and relative magnitude of signal and noise. Overall, it is clear that low-frequency tasks benefit more from increasing acquisition angle and are mainly limited by background noise, while tasks involving midfrequency and high-frequency components require careful selection of acquisition parameters to minimize the influence of view-aliasing artifacts and quantum and electronics noise. The framework provided by 3D cascaded systems analysis combined with idealized task functions and observer models provides a quantitative foundation from which specific trends in performance can be more rigorously investigated with respect to specific imaging systems and applications.

IV. DISCUSSION AND CONCLUSIONS

This work derived the generalized detectability index as a task-based performance metric for tomosynthesis and CBCT, beginning with a 3D cascaded systems model for the imaging systems and validating it in comparison to human observer response for several simple imaging tasks over a wide range of imaging conditions. A reasonable level of agreement was observed between theoretical predictions and experimental results. Interesting, nontrivial trends were revealed, which suggest important design considerations for system optimization. Tradeoffs among anatomical noise, quantum noise, electronics noise, and view-sampling effects result in complex behavior (e.g., nonmonotonic dependence of d'_{slice} on θ_{tot}) that suggest optimal choices of acquisition parameters specific to the imaging tasks.

Such behavior was reasonably predicted and explained by the cascaded systems model, with the level of agreement depending on the imaging task, perhaps not surprisingly, considering the simplicity of Fourier hypothesis-testing task functions [Eqs. (15)–(25)] in comparison to the complexities of human visual perception. The case for which the model deviated most from measurement was the encapsulated sphere discrimination task under the constant-N_{proj} case for θ_{tot} above 180°, where theoretical prediction did not accurately predict the large drop observed in human observer performance. This suggests room for improvement when modeling high-frequency tasks which are more prone to deterioration by stochastic noise and view-sampling effects. Overall, results demonstrate that generalized detectability index yielded reasonable correspondence with human observer performance for a variety of simple imaging tasks over a broad range of experimental conditions in both the constant-$\Delta\theta$ and constant-N_{proj} schemes, helping to bridge the gap between Fourier-based metrics (e.g., NEQ) of system performance and observer-based characterization of image quality (e.g., ROC).

Of the simple observer models considered, the nonprewhitening models, especially the NPWEi model, yielded the best overall agreement with human observer response. The fair agreement for the NPW model may seem surprising, considering that such models have been shown previously to correspond poorly with human observer performance due to a large zero-frequency (DC) response.[58] In the generalized detectability index calculation above, inclusion of the power-law noise in the denominator introduces a large zero-frequency noise component which diminishes DC response, similar to the effect achieved by an eye filter in the NPWE or NPWEi models. Therefore, the DC effect that confounded NPW model agreement in a previous work[59] was not a significant factor in the experiments considered above, with the exception of the uniform background task. Burgess *et al.*[39] further modified the NPW model to include an eye filter and internal noise, which improved agreement with observer response significantly, consistent with the findings above. Prewhitening observer models' consistently overestimated response likely due to the fact that the discrete spheres in the cluttered background were indistinguishable from the signal and human observers were not able to completely decorrelate noise. That said, it bears reiteration that the purpose of this work was not to advance or improve any particular observer model; rather, this work aimed to determine the extent to which trends in human observer response may be predicted by first principles of cascaded systems analysis (GNEQ) combined with task functions through any of these simple observer models. More sophisticated observer models yielded through ongoing work in perception science will presumably yield further improvement, including channelized Hotelling observer models[60–62] and forms of eye filter and internal noise models.[63,64]

This study involved simple, idealized detection and dis-

crimination tasks as a starting point to assess imaging performance. Modeling of more complex and higher order tasks forms an important area of future work to better relate such work to clinical applications. Examples include "search" tasks analogous to an observer detecting and localizing suspicious lesions on a radiograph or "estimation" tasks in which an observer needs to approximate the size of a lesion. Also of interest are task functions representing multiple hypotheses, e.g., discrimination of signal-absent from signal-benign and signal-malignant hypotheses. Similarly, the model can be incorporated in analysis of multiple tasks, e.g., detection of a (low-frequency) lesion followed by detection of (high-frequency) calcifications.[65] In the context of volumetric imaging, it remains to be shown to what extent the fully 3D detectability index [Eqs. (2)–(7)] corresponds to cases in which the observer "scrolls" slices, is presented with a montage of multiple slices simultaneously or perceives a volumetric rendering all at once. Another limitation of this work is that only one reconstruction algorithm (FBP) with one reconstruction filter (a smooth cosine Hann filter) was investigated. Cascaded systems analysis is well suited to description of FBP reconstruction and extension to other methods (e.g., iterative reconstruction) would require a substantially modified approach. Accommodating various reconstruction filters within the model is straightforward, has been examined in a previous work,[45] but was not investigated directly here, since it had less influence on task performance than angular range and number of projections.

In summary, the generalized detectability index was compared to human observer performance for a variety of simple tasks over a broad range of experimental conditions. Reasonable agreement was obtained for all tasks across the tomosynthesis angular range $\theta_{tot} < 180°$. Discrepancy was observed for high-frequency tasks (e.g., small sphere detection and encapsulated sphere vs solid sphere discrimination) under the constant-N_{proj} scheme under conditions dominated by view-sampling artifacts (i.e., small N_{proj} with $\theta_{tot} > 180°$). Such discrepancies identify areas for improvement of the model and future investigation for task performance under conditions dominated by image artifact (rather than purely stochastic noise). Still, generalized detectability index derived from the 3D cascaded systems model demonstrates considerable promise in relating simple Fourier metrics to human observer performance and suggests utility as an objective function in the design and optimization of 3D imaging systems in CBCT and tomosynthesis.

ACKNOWLEDGMENTS

The first and second authors contributed equally to this work. The authors would like to extend their gratitude to Dr. Arthur E. Burgess (Harvard University, retired), Dr. Rebecca Fahrig and Dr. Sungwon Yoon (Stanford University), and Dr. Angel Pineda (California State University-Fullerton) for stimulating discussions throughout the evolution of this work. Collaboration with Dr. John Carrino and Dr. Russell Taylor (Johns Hopkins University) are gratefully acknowledged. This work was supported by the National Institutes of Health Grant No. R01-CA-112163.

[a]Author to whom correspondence should be addressed. Electronic mail: jeff.siewerdsen@jhu.edu; Telephone: 443-287-6269.

[1]D. Gur, "Tomosynthesis: Potential clinical role in breast imaging," AJR, Am. J. Roentgenol. **189**(3), 614–615 (2007).
[2]A. O'Connell, D. L. Conover, Y. Zhang, P. Seifert, W. Logan-Young, C. F. Lin, L. Sahler, and R. Ning, "Cone-beam CT for breast imaging: Radiation dose, breast coverage, and image quality," AJR, Am. J. Roentgenol. **195**(2), 496–509 (2010).
[3]J. M. Park, E. A. Franken, Jr., M. Garg, L. L. Fajardo, and L. T. Niklason, "Breast tomosynthesis: Present considerations and future applications," Radiographics **27**, S231–S240 (2007).
[4]V. Singh, C. Saunders, L. Wylie, and A. Bourke, "New diagnostic techniques for breast cancer detection," Future Oncol. **4**(4), 501–513 (2008).
[5]J. T. Dobbins III, H. P. McAdams, D. J. Godfrey, and C. M. Li, "Digital tomosynthesis of the chest," J. Thorac. Imaging **23**(2), 86–92 (2008).
[6]J. T. Dobbins III and H. P. McAdams, "Chest tomosynthesis: Technical principles and clinical update," Eur. J. Radiol. **72**(2), 244–251 (2009).
[7]A. C. Miracle and S. K. Mukherji, "Conebeam CT of the head and neck, Part 2: Clinical applications," AJNR Am. J. Neuroradiol. **30**(7), 1285–1292 (2009).
[8]M. J. Daly, J. H. Siewerdsen, D. J. Moseley, D. A. Jaffray, and J. C. Irish, "Intraoperative cone-beam CT for guidance of head and neck surgery: Assessment of dose and image quality using a C-Arm prototype," Med. Phys. **33**(10), 3767–3780 (2006).
[9]"Medical imaging—The assessment of image quality," ICRU Report No. 54 (International Commission on Radiation Units and Measurements, Bethesda, MD, 1996).
[10]X. Gong, S. J. Glick, B. Liu, A. A. Vedula, and S. Thacker, "A computer simulation study comparing lesion detection accuracy with digital mammography, breast tomosynthesis, and cone-beam CT breast imaging," Med. Phys. **33**(4), 1041–1052 (2006).
[11]A. 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).
[12]I. Reiser and R. M. Nishikawa, "Task-based assessment of breast tomosynthesis: Effect of acquisition parameters and quantum noise," Med. Phys. **37**(4), 1591–1600 (2010).
[13]S. Park, R. J. Jennings, H. Liu, A. Badano, and K. J. Myers, "A statistical, task-based evaluation method for three-dimensional x-ray breast imaging systems using variable-background phantoms," Med. Phys. **37**(12), 6253–6270 (2010).
[14]J. 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).
[15]J. H. Siewerdsen, L. E. Antonuk, Y. el-Mohri, J. Yorkston, W. Huang, and I. A. Cunningham, "Signal, noise power spectrum, and detective quantum efficiency of indirect-detection flat-panel imagers for diagnostic radiology," Med. Phys. **25**(5), 614–628 (1998).
[16]W. Zhao, W. G. Ji, and J. A. Rowlands, "Effects of characteristic x rays on the noise power spectra and detective quantum efficiency of photoconductive x-ray detectors," Med. Phys. **28**(10), 2039–2049 (2001).
[17]W. Zhao, W. G. Ji, A. Debrie, and J. A. Rowlands, "Imaging performance of amorphous selenium based flat-panel detectors for digital mammography: Characterization of a small area prototype detector," Med. Phys. **30**(2), 254–263 (2003).
[18]E. Samei and M. J. Flynn, "Physical measures of image quality in photostimulable phosphor radiographic systems," Proc. SPIE **3032**, 328–338 (1997).
[19]W. Zhao and J. A. Rowlands, "Digital radiology using active matrix readout of amorphous selenium: Theoretical analysis of detective quantum efficiency," Med. Phys. **24**(12), 1819–1833 (1997).
[20]S. Richard, J. H. Siewerdsen, D. A. Jaffray, D. J. Moseley, and B. Bakhtiar, "Generalized DQE analysis of radiographic and dual-energy imaging using flat-panel detectors," Med. Phys. **32**(5), 1397–1413 (2005).
[21]J. P. Johnson, J. Lo, T. Mertelmeier, J. S. Nafziger, P. Timberg, and E. Samei, "Visual image quality metrics for optimization of breast tomosynthesis acquisition technique," Proc. SPIE **6515**, 65150P-1–65150P-10 (2007).

[22]R. Saunders, E. Samei, C. Badea, K. Ghaghada, Y. Qi, L. W. Hedlund, and S. Mukundan, "Optimization of dual energy contrast enhanced breast tomosynthesis for improved mammographic lesion detection and diagnosis," Proc. SPIE **6913**, 69130Y-1–69130Y-11 (2008).
[23]B. Zhao, J. Zhou, Y. H. Hu, T. Mertelmeier, J. Ludwig, and W. Zhao, "Experimental validation of a three-dimensional linear system model for breast tomosynthesis," Med. Phys. **36**(1), 240–251 (2009).
[24]B. Zhao and W. Zhao, "Three-dimensional linear system analysis for breast tomosynthesis," Med. Phys. **35**(12), 5219–5232 (2008).
[25]J. H. Siewerdsen and D. A. Jaffray, "Cone-beam computed tomography with a flat-panel imager: Noise considerations for fully 3-D imaging," Proc. SPIE **3977**, 408–416 (2000).
[26]D. J. Tward and J. H. Siewerdsen, "Cascaded systems analysis of the 3D noise transfer characteristics of flat-panel cone-beam CT," Med. Phys. **35**(12), 5510–5529 (2008).
[27]D. J. Tward, J. H. Siewerdsen, R. A. Fahrig, and A. R. Pineda, "Cascaded systems analysis of the 3D NEQ for cone-beam CT and tomosynthesis," Proc. SPIE **6913**, 69131S-1–69131S-12 (2008).
[28]Y. Lin, X. Wang, W. J. Sehnert, D. H. Foos, L. Barski, and E. Samei, "Quantification of radiographic image quality based on patient anatomical contrast-to-noise ratio: A preliminary study with chest images," Proc. SPIE **7627**, 76271F-1–76271F-12 (2010).
[29]H. H. Barrett, "Objective assessment of image quality: Effects of quantum noise and object variability," J. Opt. Soc. Am. A **7**(7), 1266–1278 (1990).
[30]H. H. Barrett, J. L. Denny, R. F. Wagner, and K. J. Myers, "Objective assessment of image quality. II. Fisher information, Fourier crosstalk, and figures of merit for task performance," J. Opt. Soc. Am. A Opt. Image Sci. Vis **12**(5), 834–852 (1995).
[31]H. H. Barrett, C. K. Abbey, and E. Clarkson, "Objective assessment of image quality. III. ROC metrics, ideal observers, and likelihood-generating functions," J. Opt. Soc. Am. A Opt. Image Sci. Vis **15**(6), 1520–1535 (1998).
[32]C. E. Metz, "Basic principles of ROC analysis," Semin Nucl. Med. **8**(4), 283–298 (1978).
[33]H. H. Barrett and K. J. Myers, *Foundations of Image Science* (Wiley, New York, 2004).
[34]D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).
[35]D. J. Goodenough, K. Rossmann, and L. B. Lusted, "Radiographic applications of signal detection theory," Radiology **105**(1), 199–200 (1972).
[36]C. K. Abbey and H. H. Barrett, "Human- and model-observer performance in ramp-spectrum noise: Effects of regularization and object variability," J. Opt. Soc. Am. A Opt. Image Sci. Vis **18**(3), 473–488 (2001).
[37]M. A. Kupinski, J. W. Hoppin, E. Clarkson, and H. H. Barrett, "Ideal-observer computation in medical imaging with use of Markov-chain Monte Carlo techniques," J. Opt. Soc. Am. A Opt. Image Sci. Vis **20**(3), 430–438 (2003).
[38]H. H. Barrett, J. P. Rolland, R. F. Wagner, and K. J. Myers, "Detection and discrimination of known signals in inhomogeneous, random backgrounds," Proc. SPIE **1090**, 176–182 (1989).
[39]A. E. Burgess, X. Li, and C. K. Abbey, "Visual signal detectability with two noise components: Anomalous masking effects," J. Opt. Soc. Am. A Opt. Image Sci. Vis **14**(9), 2420–2442 (1997).
[40]A. E. Burgess, "Mammographic structure: Data preparation and spatial statistics analysis," Proc. SPIE **3661**, 642–653 (1999).
[41]A. E. Burgess, F. L. Jacobson, and P. F. Judy, "Human observer detection experiments with mammograms and power-law noise," Med. Phys. **28**(4), 419–437 (2001).
[42]J. J. Heine, S. R. Deans, R. P. Velthuizen, and L. P. Clarke, "On the statistical nature of mammograms," Med. Phys. **26**(11), 2254–2265 (1999).
[43]E. 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).
[44]E. Samei, M. J. Flynn, E. Peterson, and W. R. Eyler, "Subtle lung nodules: Influence of local anatomic variations on detection," Radiology **228**(1), 76–84 (2003).
[45]J. 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).
[46]J. 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).
[47]G. J. Gang, D. J. Tward, J. Lee, and J. H. Siewerdsen, "Anatomical background and generalized detectability in tomosynthesis and cone-beam CT," Med. Phys. **37**(5), 1948–1965 (2010).
[48]R. 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).
[49]A. E. Burgess, R. F. Wagner, R. J. Jennings, and H. B. Barlow, "Efficiency of human visual signal discrimination," Science **214**(4516), 93–94 (1981).
[50]M. Ishida, K. Doi, L. N. Loo, C. E. Metz, and J. L. Lehr, "Digital image processing: Effect on detectability of simulated low-contrast radiographic patterns," Radiology **150**(2), 569–575 (1984).
[51]J. H. Siewerdsen, I. A. Cunningham, and D. A. Jaffray, "A framework for noise-power spectrum analysis of multidimensional images," Med. Phys. **29**(11), 2655–2671 (2002).
[52]P. G. J. Barten, *Contrast Sensitivity of the Human Eye and Its Effects on Image Quality* (SPIE, Bellingham, 1999).
[53]S. Richard and J. H. Siewerdsen, "Comparison of model and human observer performance for detection and discrimination tasks using dual-energy x-ray images," Med. Phys. **35**(11), 5043–5053 (2008).
[54]M. D. Silver, "A method for including redundant data in computed tomography," Med. Phys. **27**(4), 773–774 (2000).
[55]P. Elliot, in *Signal Detection and Recognition by Human Observers: Contemporary Readings*, edited by J. A. Sweets (Wiley, New York, 1964), Appendix 1.
[56]C. K. Abbey, R. J. Zemp, J. Liu, K. K. Lindfors, and M. F. Insana, "Observer efficiency in discrimination tasks simulating malignant and benign breast lesions imaged with ultrasound," IEEE Trans. Med. Imaging **25**(2), 198–209 (2006).
[57]K. 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).
[58]A. E. Burgess, "Statistically defined backgrounds: Performance of a modified nonprewhitening observer model," J. Opt. Soc. Am. A Opt. Image Sci. Vis **11**(4), 1237–1242 (1994).
[59]J. P. Rolland and H. H. Barrett, "Effect of random background inhomogeneity on observer detection performance," J. Opt. Soc. Am. A **9**(5), 649–658 (1992).
[60]K. J. Myers and H. H. Barrett, "Addition of a channel mechanism to the ideal-observer model," J. Opt. Soc. Am. A **4**(12), 2447–2457 (1987).
[61]J. Yao and H. H. Barrett, "Predicting human performance by a channelized Hotelling observer model," Proc. SPIE **1768**, 161–168 (1992).
[62]S. Park, H. H. Barrett, E. Clarkson, M. A. Kupinski, and K. J. Myers, "Channelized-ideal observer using Laguerre-Gauss channels in detection tasks involving non-Gaussian distributed lumpy backgrounds and a Gaussian signal," J. Opt. Soc. Am. A Opt. Image Sci. Vis **24**(12), B136–B150 (2007).
[63]Z. L. Lu and B. A. Dosher, "Characterizing human perceptual inefficiencies with equivalent internal noise," J. Opt. Soc. Am. A Opt. Image Sci. Vis **16**(3), 764–778 (1999).
[64]Y. Zhang, B. T. Pham, and M. P. Eckstein, "Evaluation of internal noise methods for Hotelling observer models," Med. Phys. **34**(8), 3312–3322 (2007).
[65]S. Richard and J. H. Siewerdsen, "Optimization of dual-energy imaging systems using generalized NEQ and imaging task," Med. Phys. **34**(1), 127–139 (2007).