NEQ and Task in Dual-Energy Imaging: From Cascaded Systems Analysis to Human Observer Performance

Samuel Richard, a Jeffrey H. Siewerdsen, a,b and Daniel J. Tward b

aDepartment of Medical Biophysics, University of Toronto, Ontario, Canada M5G 2M9;
 bOntario Cancer Institute, Princess Margaret Hospital, Ontario, Canada M5G 2M9;

ABSTRACT

The relationship between theoretical descriptions of imaging performance (Fourier-based cascaded systems analysis) and the performance of real human observers was investigated for various detection and discrimination tasks. Dual-energy (DE) imaging provided a useful basis for investigating this relationship, because it presents a host of acquisition and processing parameters that can significantly affect signal and noise transfer characteristics and, correspondingly, human observer performance. The detectability index was computed theoretically using:

1) cascaded systems analysis of the modulation transfer function (MTF), and noise-power spectrum (NPS) for DE imaging; 2) a Fourier description of imaging task; and 3.) integration of MTF, NPS, and task function according to various observer models, including Fisher-Hotelling and non-prewhitening with and without an eye filter and internal noise. Three idealized tasks were considered: sphere detection, shape discrimination (sphere vs. disk), and texture discrimination (uniform vs. textured disk). Using images of phantoms acquired on a prototype DE imaging system, human observer performance was assessed in multiple-alternative forced choice (MAFC) tests, giving an estimate of area under the ROC curve \((A_Z)\). The degree to which the theoretical detectability index correlated with human observer performance was investigated, and results agreed well over a broad range of imaging conditions, depending on the choice of observer model. Results demonstrated that optimal DE image acquisition and decomposition parameters depend significantly on the imaging task. These studies provide important initial validation that the detectability index derived theoretically by Fourier-based cascaded systems analysis correlates well with actual human observer performance and represents a meaningful metric for system optimization.

Keywords: NEQ, MTF, NPS, MAFC, dual-energy imaging, flat-panel detector, observer performance, detectability index.

1. INTRODUCTION

The development of novel imaging systems requires an understanding of factors affecting the ability to perform a given imaging task; however, there is generally a gap between physical metrics that describe the performance of medical imaging systems (e.g., MTF, NPS, and NEQ) and the performance of actual human observers. There has been considerable research in the development of model observers, giving analytical models that describe human observer performance to varying degrees, depending on the characteristics of the image and the imaging task. The work described below seeks to relate physical signal and noise characteristics (MTF and NPS) with human observer performance by deriving the MTF and NPS from first principles (cascaded systems analysis), combining with a Fourier description of imaging task according to various observer models, and correlating the resulting detectability index with measurements of human observer performance.

Dual-energy (DE) imaging provided a useful platform for the research, involving both acquisition and decomposition parameters (specifically, dose allocation and noise reduction parameters, respectively) that significantly affect the MTF and NPS. The imaging chain was modeled by means of cascaded systems analysis. Three imaging tasks were considered – detection, shape discrimination, and texture discrimination – each modeled...
theoretically as Fourier domain templates. The detectability index was computed using the theoretically derived MTF, NPS, and imaging task, considering four model observers. A DE imaging phantom was constructed to present structures corresponding to each imaging task, and the resulting DE images were used in MAFC tests of human observer performance. The extent to which theoretically derived detectability index agreed with actual human observer performance was evaluated across a broad range of acquisition and decomposition parameters.

2. MODEL OBSERVERS AND IMAGING TASK

Four model observers were investigated in this study: the Fisher-Hotelling (FH) observer; the FH observer with an eye filter and internal noise (FHE); the non-prewhitening (NPW) observer; and the NPW observer with an eye filter and internal noise (NPWE). A brief description of each is presented below, with notation based on that of Burgess et al. and additional details found in previous work. Each of the terms appearing in Eqs. (1-5) were computed theoretically, providing an approach to observer performance that is based on first principles of spatial-frequency-dependent signal and noise transfer characteristics yet begins to bridge the gap to image quality as assessed by real human observers.

2.1 The Fisher-Hotelling Observer (FH)

The Fisher-Hotelling observer is defined as follows:

\[
d^2_{FH} = \int \int \frac{MTF^2(u,v)W^2_{Task}(u,v)}{NPS(u,v)} \, du \, dv,
\]

where \(u\) and \(v\) are spatial frequency coordinates and the integration is performed over the Nyquist region. \(MTF(u,v)\) denotes the modulation transfer function, \(NPS(u,v)\) denotes the normalized noise-power spectrum (i.e., the NPS divided by the square of the mean signal), and \(W_{Task}(u,v)\) denotes the task function (which describes the spatial frequencies of interest for a given imaging task, as detailed in Section 2.5). In this form, the detectability index is seen to be a sum over the weighted combination of noise equivalent quanta (NEQ) and task function. Such represents a simple observer model for which the observer matches the of the task function to the signal and compensates for noise correlations by the prewhitening of the noise.

2.2 The Fisher-Hotelling Observer with Eye Filter (FHE)

The FH observer was modified by Burgess to include an eye filter and internal observer noise, as follows:

\[
d^2_{FHE} = \int \int \frac{MTF^2(u,v)W^2_{Task}(u,v)E^2(u,v)}{NPS(u,v)E^2(u,v) + N_i} \, du \, dv,
\]

where \(E(u,v)\) is the eye filter and \(N_i\) denotes internal noise. The eye filter assumed in this work is the same as the one used by Burgess and is modeled on the contrast sensitivity function of the human visual system. The filter consists of a ramp filter multiplied by a Gaussian filter: \(f^n \exp(-cf^2)\), with a value of \(n=1.3\) and \(c=3\) shown to provide good agreement for a viewing distance of 50 cm. The internal noise was assumed to be uncorrelated (white) and set to an empirically determined fraction (0.02) of the zero-frequency quantum noise for a viewing distance of 100 cm. Internal noise therefore scales with viewing distance, \(D\) (units of cm), according to \((D/100)^2\), such that: \(N_i = 0.02 \times NNPS(0, 0) \times (D/100)^2\).

2.3 The Non-Prewhitening Observer (NPW)

The non-prewhitening observer (NPW) differs from the FH observer in that the observer does not compensate for correlations in the noise (i.e., the noise is not prewhitened). Correspondingly, as initially proposed by Wagner, the detectability index for the NPW observer is:

\[
d^2_{NPW} = \frac{\left[ \int \int MTF^2(u,v)W^2_{Task}(u,v) \, du \, dv \right]^2}{\int \int MTF^2(u,v)W^2_{Task}NPS(u,v) \, du \, dv},
\]

which is generally lower than that of the FH observer but is identical under conditions of white noise (NNPS equal to a constant). This model has been shown to agree well with human performance under white noise and high-pass filtered images such as CT images.
2.4 The Non-Prewhitening Observer with Eye Filter (NPWE)

The NPW observer may be similarly modified, as shown by Burgess,\textsuperscript{10} to include an eye filter and internal observer noise:

\[
d_{NPWE}^2 = \frac{\left[ \int \int \text{MTF}^2(u,v)W_{\text{Task}}^2(u,v)E^2(u,v) dudv \right]^2}{\int \int \text{MTF}^2(u,v)W_{\text{Task}}^2(u,v)\text{NNPS}(u,v)E^4(u,v) + \text{MTF}^2(u,v)W_{\text{Task}}^2(u,v)N_i dudv}
\]

(4)

The same eye filter and internal noise values as described above for the FHE observer were used for the NPWE observer.

2.5 Imaging Tasks

The work below investigates somewhat more complex task functions than the simple (delta-function or gaussian sphere) detection tasks previously reported.\textsuperscript{8} Imaging task functions were defined as the Fourier difference between two hypotheses (e.g., signal absent vs. signal present):\textsuperscript{9}

\[
W_{\text{Task}} = |F[h_1(x,y) - h_2(x,y)]|
\]

(5)

where \(h_1(x,y)\) and \(h_2(x,y)\) are the spatial domain representations of the two hypotheses, and \(F\) is the Fourier transform operator. Below, hypothesis 1 and hypothesis 2 are denoted "normal" and "abnormal", respectively. The three tasks considered in this work are illustrated in Fig. 1. The sphere detection task consists of sphere-absent and sphere-present hypotheses and primarily weighs low frequencies. The shape discrimination task considers hypotheses that the stimulus is a sphere or a cylinder and weighs higher frequencies associated with the differentiation of a smooth versus a sharp edge. The texture discrimination task considers hypotheses that the stimulus is a smooth or textured cylinder, giving a task function with narrow frequency bands associated with the width of the crenelations in the textured disk. While these tasks are highly idealized, they are analogous to clinically relevant tasks — e.g., detection of a small soft-tissue nodule, discrimination between a vessel and a nodule, and discrimination lesion texture (lobulated, speculated, etc.), respectively. Moreover, they present spatial-frequency spectra that are richly varied and are hypothesized to impart distinct differences in detectability index under conditions of widely varying spatial-frequency-dependent MTF and NNPS.

![Figure 1. Hypotheses and task functions for the three tasks investigated](image)

2.6 Comparison of \(d'\) with Human Observer Performance

The detectability index can be related to the area under the ROC curve, denoted \(A_Z\), under the assumption that the distributions of normal and abnormal cases are Gaussian and homoscedastic (equal variance):\textsuperscript{15}

\[
A_Z = \frac{1}{2} \left( 1 + \frac{2}{\sqrt{\pi}} \int_0^{d'} e^{x^2} dx \right)
\]

\[
= \frac{1}{2} \left( 1 + \text{erf} \left( \frac{d'}{\sqrt{2}} \right) \right).
\]

(6)
In the experiments below, MAFC tests were used to evaluate human observer performance, recognizing that while MAFC tests are less efficient than ROC tests (requiring more images to achieve a given level of statistical error), they are well suited to phantom studies involving a large number of images and are well tolerated by observers. The resulting figure of merit for human observer performance in MAFC tests is the proportion correct, denoted \( P_C \). For an MAFC test with \( M \) choices, \( M \) is related to the detectability index by:

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

where \( \Phi(x) \) is the cumulative Gaussian distribution function, \( \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-y^2/2) \, dy \). Note that for \( M = 2 \) (a 2AFC test), the proportion correct is simply equal to \( A_Z \) (under the assumption of an ideal ROC test in which observer response does not vary over the course of the test). All results below are reported in terms of \( A_Z \), with theoretically derived \( d' \) [Eqs. (1-5)] converted to \( A_Z \) via Eq. (6) and with experimentally determined \( P_C \) (from MAFC tests) converted to \( A_Z \) by numerical inversion of Eq. (7) (a reverse look-up table between \( P_C \) and \( A_Z \) computed numerically).

3. DUAL-ENERGY (DE) IMAGING PERFORMANCE

3.1 DE Image Decomposition Algorithms

3.1.1 DE image decomposition: A general form

A general algorithm for linear DE image decomposition involves the natural log of the low- and high-energy projections \([I_L(x, y)\) and \(I_H(x, y)\), respectively] filtered individually and combined to yield a DE image:

\[
I_{DE}(x, y) = I_L(x, y) * h_L(x, y) + I_H(x, y) * h_H(x, y),
\]

where \(*h_L(x, y)\) and \(*h_H(x, y)\) denote convolution with the low- and high-energy filters. As shown below, this equation is a general form of well-known log-subtraction decomposition techniques, reducing to familiar forms as special cases determined by the selection of \(h_L(x, y)\) and \(h_H(x, y)\). This form is also convenient in the derivation of \(MTF(u, v)\) and \(NPS(u, v)\) for DE images. Three linear decomposition algorithms were investigated in this work, described below. Because non-linear decomposition algorithms (e.g., noise clipping) are less amenable to cascaded systems analysis, such were not included in this work; however, the image quality associated with linear and non-linear decomposition have been shown to be comparable.

3.1.2 Standard log subtraction (SLS)

Perhaps the most familiar DE image decomposition algorithm involves a weighted subtraction of (log) low- and high-energy projections, referred to below as standard log subtraction (SLS). For example, the soft-tissue image is given by the high-energy projection minus the low-energy projection weighted by a tissue cancellation parameter, \(w\):

\[
I_{SLS}(x, y) = -wI_L(x, y) + I_H(x, y) = I_L [-w\delta(x, y)] + I_H [\delta(x, y)],
\]

where \(\delta(x, y)\) denotes a delta function. The tissue cancellation parameter is ideally given by the ratio of the effective linear attenuation coefficient of the canceled material at low- and high-energy: \(\mu_H/\mu_L\). Equation 9 yields a soft-tissue-only decomposition with positive contrast (white soft-tissue on black background). The complementary bone image with positive contrast (white bones) is given by multiplication of the right-hand side of Eq. 9 by \(-1\), with \(w\) adjusted to cancel soft-tissue. The results below pertain to the soft-tissue image, with similar analysis for the bone image currently underway.

3.1.3 Simple smoothing of the high-energy image (SSH)

The greater contributor of quantum noise in DE images is the high-energy projection. Recognizing this, Johns and Yaffe suggested application of a low-pass filter to the high-energy projection prior to subtraction, referred to below as simple smoothing of the high-energy image (SSH):

\[
I_{SSH}(x, y) = -wI_L(x, y) + I_H(x, y) * h_{LPF}(x, y) = I_L [-w\delta(x, y)] + I_H [h_{LPF}(x, y)],
\]

where \(h_{LPF}\) denotes the low-pass filter.
3.1.4 Anti-correlated noise reduction (ACNR)

A third algorithm developed by Kalendar,\textsuperscript{18} Ergun,\textsuperscript{19} and McCollough\textsuperscript{20} takes advantage of the fact that quantum noise in the soft-tissue image and bone-only image is anti-correlated. Referred to below as anti-correlated noise reduction (ACNR), the algorithm applies a high-pass filter to the complementary image, (viz., the bone image for a soft-tissue decomposition or the soft-tissue image for a bone decomposition) leaving only the quantum noise (and residual edge artifacts). The resulting high-pass-filtered complementary image is then added to the original (SLS) DE image:

\[
I_{ACNR}(x,y) = I_{SLS}(x,y) + I_{SLS}(x,y) * h_{HPF}(x,y) \\
= I_L * [w_n w_c h_{HPF}(x,y) - w \delta(x,y)] + I_H * [\delta(x,y) - w_n h_{HPF}(x,y)],
\]

where \(I_{SLS}\) denotes the complementary image, \(w_c\) denotes the tissue cancellation parameter for the complementary image, \(h_{HPF}\) denotes the high-pass filter, and \(w_n\) is a weighting parameter determined qualitatively or quantitatively to minimize quantum noise while minimizing edge artifacts.

3.1.5 Decomposition parameters and filters

<table>
<thead>
<tr>
<th>Spatial domain</th>
<th>(h_L(x,y))</th>
<th>(h_H(x,y))</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLS</td>
<td>(-w \delta(x,y))</td>
<td>(\delta(x,y))</td>
</tr>
<tr>
<td>SSH</td>
<td>(-w \delta(x,y))</td>
<td>(h_{LPF}(x,y))</td>
</tr>
<tr>
<td>ACNR</td>
<td>(w_n w_c h_{HPF}(x,y) - w \delta(x,y))</td>
<td>(\delta(x,y) - w_n h_{HPF}(x,y))</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fourier domain</th>
<th>(H_L(x,y))</th>
<th>(H_H(x,y))</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLS</td>
<td>(-w)</td>
<td>(1)</td>
</tr>
<tr>
<td>SSH</td>
<td>(-w)</td>
<td>(H_{LPF}(u,v))</td>
</tr>
<tr>
<td>ACNR</td>
<td>(w_n w_c h_{HPF}(u,v) - w)</td>
<td>(1 - w_n h_{HPF}(u,v))</td>
</tr>
</tbody>
</table>

Table 1. Summary of low- and high-energy convolution filters, represented in the spatial and Fourier domain, for the three DE decomposition algorithms (SLS, SSH, ACNR)

Table 1 summarizes the spatial and Fourier domain representation of the filters applied to the low- and high-energy projections for each of the decomposition algorithms derived from Eqs 9, 10, and 11. The low- and high-pass filters were chosen as Gaussian filters:

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

and

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

characterized by the parameters, \(d_{LPF}\) and \(d_{HPF}\), respectively.

3.2 DE Image Acquisition System

Figure 2 depicts a prototype DE imaging system developed in our laboratory\textsuperscript{21} along with a phantom designed to present three distinct imaging tasks. The system employs a Trixell Pixium 4600 flat-panel detector (250 mg/cm\(^2\) CsI:Tl x-ray converter; 3000x3000 pixel matrix; 143 \(\mu\)m pixel pitch; \(\sim\)0.8 fill factor). The system space also incorporates a 10:1 antiscatter bucky grid. The low and high kVp were fixed at 60 and 120 kVp in these studies (denoted [60/120] kVp). Total filtration for the low-energy beam was 2.5 mm Al, with 0.6 mm Ag + 2.5 mm Al added for the high-energy beam. These settings of kVp pair and filtration have been shown previously to be optimal for soft-tissue DE imaging at total dose equivalent to that of a single PA chest radiograph\textsuperscript{21}.

Table 2 summarizes the parameters employed in DE image decompositions used for the human observer study and for the theoretical modeling of detectability index. The parameters are consistent with previous studies involving anthropomorphic phantoms and patient subjects in which expert radiologists determined tissue cancellation parameters qualitatively to present soft-tissue images of acceptable diagnostic quality.
3.3 Cascaded Systems Analysis for DE Imaging

Cascaded systems analysis (CSA) provides a powerful framework for understanding and modeling imaging system performance. CSA conceptualizes the imaging chain as a series of discrete stages from which imaging performance metrics such as the MTF and NPS can be derived analytically.\textsuperscript{4,5} Previously, CSA was extended to DE imaging and DE noise reduction algorithms\textsuperscript{6–8} and is applied in this work to model the DE imaging system described in Sec. 3.2 and to compute the detectability index for the various model observers. A brief overview of the derivation and results for the DE MTF and DE NNPS is provided below.

3.3.1 DE modulation transfer function (MTF)

The DE MTF is given by the sum of the low- and high-energy filtered projection $MTF(u, v)$ weighted by the respective signal in the low- and high-energy projection, with weights denoted: $k_L$ and $k_H$.\textsuperscript{8} The DE MTF is normalized to unity at zero frequency (i.e., the DC component of the DE image is independent of decomposition algorithm). Assuming that the detector MTF is equal in low- and high-energy projections – a convenient, but not essential, assumption made simply to give a more concise analytical form. Considering polyethylene (simulating soft tissue) at [60/120] kVp, the value of $k_{rel}$ was determined theoretically and verified empirically to be equal to $\sim 1.8$\textsuperscript{8}.

Figure 3a shows theoretical calculations of the DE MTF computed using Eq. 14 and parameters found in Tables 1 and 2. The DE MTF depends significantly on the decomposition algorithms and parameters. For example, $MTF_{SSH}$ (i.e., the MTF for the SSH image) drops to 0 at 1 $mm^{-1}$, implying a significant degradation in performance of tasks around that frequency (e.g., the texture discrimination task in Fig. 1). The results also
illustrate that $MTF_{ACNR}$ (for the soft-tissue image) is typically greater than $MTF_{SLS}$ and $MTF_{SSH}$ (even though the quantum noise is decreased for ACNR as shown in Fig. 3b).

![Figure 3](image_url)

Figure 3. (a) MTF and (b) NNPS for three DE image decomposition algorithms. Each assumes a soft-tissue image decomposition from projections at [60/120] kVp.

### 3.3.2 DE image noise-power spectrum

Considering the combination of statistically independent log images $[I_L(x, y)$ and $I_H(x, y)]$, the total NNPS is equal to the sum of the individual NNPS. It follows that the total NNPS arising from such projections after the application of a filter $[H_L(u, v)$ or $H_H(u, v)]$ is related to the individual “unfiltered” NNPS multiplied by the filter squared, giving:

$$NNPS_{DE}(u, v) = NNPS_L(u, v)H_L^2(u, v) + NNPS_H(u, v)H_H^2(u, v), \quad (15)$$

where $H_L(u, v)$ and $H_H(u, v)$ are the filters discussed in Section 3.1.5.

Figure 3b shows theoretical calculations of the NNPS. As expected, both the SSH and ACNR noise reduction algorithms reduce the NNPS compared to the SLS algorithm. Empirical validation for these calculations was shown previously, and results are consistent with measurements reported by Warp and Dobbins. For the soft-tissue image, the $NNPS_{ACNR}$ is typically the lowest of the three. As ACNR also imparts the highest MTF, as shown in Fig. 3(a), the algorithm carries the all-too-rare combination of reduced noise and improved spatial resolution.

### 4. HUMAN OBSERVER STUDY

#### 4.1 Experimental Conditions:

DE images were acquired and decomposed across a range of conditions expected to significantly affect the magnitude and spatial-frequency dependence of the NEQ. As described below, the experiments considered a single acquisition parameter (dose allocation) and two decomposition parameters ($d_{LPF}$ and $d_{HPF}$). By varying each in a manner that significantly affects MTF and NNPS (Fig. 3) and testing human observer performance in tasks believed to involve distinct spatial frequencies [the phantom of Fig. 1], the correspondence of theoretically derived detectability index and empirically measured human observer performance could be quantitatively gauged.

#### 4.1.1 Dose allocation

Dose allocation, denoted $A$, is defined as the fraction of the total entrance surface dose (ESD) delivered by the low-energy image (i.e., $A = ESD_L/(ESD_L + ESD_H)$, where $ESD_L$ and $ESD_H$ are the ESD for the low- and high-energy image, respectively. The total ESD (i.e., $ESD_L + ESD_H$) was held fixed at 0.1 mGy, equivalent to the dose of a single PA chest x-ray exam. Dose allocation has been shown to be an important technique factor in system optimization, with a significant influence on image quality at a fixed patient dose. Allocation was varied from 0 to 1 through variation of the mAs in low- and high-energy projections such that total ESD was fixed (0.1 mGy).
4.1.2 Decomposition parameters
The low-and high-pass filter parameters in the SSH and ACNR algorithms ($d_{LPF}$ and $d_{HPF}$) are similarly expected to have significant effect on detectability. For example, variation of $d_{LPF}$ changes the spatial frequency at which the “zero” of $MTF_{SSH}$ occurs [Fig. 3(a)]. Each parameter was varied across a range 0 to 1, with theoretical and human observer performance in shape and texture discrimination tasks investigated for the SSH and ACNR algorithms.

4.2 Multiple-Alternative Forced Choice (MAFC) Tests
In an MAFC study, an observer chooses which of $M$ images contains a specified stimulus, where only one of the $M$ images contains the stimulus (“abnormal”), and the remaining $M - 1$ images present the “normal”. The fraction of correct responses, termed the proportion correct and denoted $P_C$, gives a figure of merit for task performance which can be related to the area under the ROC curve as discussed in Section 2.6. To choose an appropriate value of $M$ for the studies herein, the dependence of $A_Z$ on $M$ was analyzed from Eqs. 6 and 7, with the aim of obtaining $A_Z \approx 0.8$ in the observer studies, yielding a value of $M = 9$ for all studies below.

4.2.1 Task phantom
The phantom in Fig. 2 contains polyethylene objects (simulating soft-tissue lung nodules) within a 10 cm acrylic slab, simulating the mean thickness (attenuation) of an adult chest. The polyethylene stimuli were precisely machined to ensure that that the objects corresponded closely to the idealized hypothesis functions. For the detection task, the stimulus was a 3 mm diameter sphere. For the shape discrimination task, the stimulus was a 6 mm diameter sphere, and the “normal” was a cylinder machined to present total signal power equal to that of the stimulus (viz., the height and diameter of the cylinder were 87% of the sphere diameter, identified by numerical simulation and verified in measurements). Taking care to present equal signal power between both hypotheses in discrimination tasks was necessary to remove possible bias associated with the displayed contrast of the object. For the texture discrimination task, two cylinders were presented, each of diameter 11 mm. A crenelated cylinder represented the stimulus (i.e., a cylinder machined with concentric grooves in the surface), with minimum and maximum crenelation height of 12 mm and 16 mm, respectively. A smooth cylinder of height 14 mm represented the “normal,” again ensuring equal signal power between the two hypotheses.

Images of the phantom were cropped to 300x300 pixel regions about each object for presentation in MAFC tests. Each region was detrended by subtraction of a plane fit to remove possible bias associated with different shading artifacts between normal and abnormal cases. Furthermore, the location of the object in each region was slightly randomized from the center to remove possible bias based on location. Example images for each task presented in 9AFC tests are shown in Fig. 4

Figure 4. Example images presented in 9AFC human observer tests: (a) sphere detection, (b) shape discrimination, and (c) texture discrimination. For purposes of illustration, the stimulus image is in the bottom left corner.
4.2.2 Human observer studies

Six physicists were used as observers, considered sufficiently expert for purposes of a simple phantom experiment. The three tests, corresponding to the three tasks, were conducted in random order, with images corresponding to variable dose allocation and decomposition filter presented randomly within each test. Observers were instructed as to the “stimulus” and “normal” hypotheses of each task according to the illustrations in Fig. 1. For each case, 9 images were presented, 8 representing “normal” and 1 representing the stimulus (“abnormal”), and observers selected the stimulus image by mouse click. Each observer was trained on ∼70 cases before each test, with training data completely distinct from the actual test data. Normal and abnormal cases were randomly selected from fifteen images acquired at each combination of allocation and decomposition parameter, ensuring that no set of 9 images were repeated twice for the same observer. Each repeat, including the training set, was therefore statistically independent.

Images were displayed in a 3x3 matrix as shown in Fig. 4 on a diagnostic workstation (Dell Precision-380, 3GHz Pentium 4 with dual-head monochrome Barco displays, 1530 x 2048 resolution, 8-bit grayscale). Each image was 300x300 pixels in size (900 x 900 pixels for the 9AFC, 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 total of 1150 cases were presented to each observer. Viewing time was unrestricted, and short breaks were allowed between each test. An average time of ∼2.5 h was required to complete all three tests, with an average of ∼8 s per case.

4.3 Theory and Measurements

4.3.1 Dose allocation

Figure 5 plots $A_Z$ as a function of dose allocation for the three tasks, showing measurements (points) and theoretical results (lines). Error bars depict the 95% confidence interval across the 6 observers. The results illustrate the previously observed dependence of image quality on allocation, with fairly broad optima in the range A ∼0.3-0.5. The SSH algorithm is seen to perform worst in all tasks (recognizing that such is true only for the soft-tissue decompositions considered here; previous studies showed that SSH actually performs best for bone images). The ACNR algorithm, on the other hand, yielded the highest values of $A_Z$ for all tasks – so high, in fact, that it is difficult to differentiate the results, as the stimuli were too conspicuous ($A_Z > ∼0.9$).

All four observer models generally follow the expected trends exhibited by the measurements. As expected, the FH observer shows the highest performance, almost identical to NPW for the sphere detection task, since such weighs low-spatial frequencies for which there is little correlation and, therefore, little difference between prewhitening and non-prewhitening observer models. For the shape and texture discrimination tasks [Fig. 5(d-i)], the FH and NPW observer perform differently due to weighting of higher spatial frequencies in the task function. Observer models incorporating the “eye” filter (i.e., FHE and NPWE) exhibit closer correspondence to measurements than the simpler (FH and NPW) models. Overall, the NPWE observer model is seen to give the highest correspondence to the measurements for all decomposition algorithms and imaging tasks.

4.3.2 SSH and ACNR filtration parameter

Figure 6 plots calculations and measurements of $A_Z$ as a function of the low- and high-pass filter parameters used in the SSH and ACNR decomposition algorithms, respectively. Figures 6(a-b) show results for the shape discrimination task, where NPWE and FHE are seen to provide the best agreement between theory and measurement. Similar results are seen for the texture discrimination task [Fig.6(c-d)]. These results illustrate the benefits of including an eye filter model, likely associated with the notion that human observers do not use the DC component of the image but rely more on mid-spatial-frequencies.

Perhaps the most striking evidence that the theoretical model combining NEQ and imaging task in calculation of detectability index does indeed describe the performance of real human observers is seen in Fig. 6(c) – the texture discrimination task under the SSH decomposition algorithm. In this case, it was shown in Fig. 3(a) that the algorithm effectively zeroes the MTF (and therefore the NEQ) at a given spatial frequency, depending on the choice of $d_{LPF}$. In Fig. 6(c) we see that if $d_{LPF}$ is chosen such that the “zeroing” of the NEQ coincides
Figure 5. Theoretical calculations and experimental measurements of $A_Z$ for the three tasks as a function of dose allocation for the SLS, SSH, and ACNR DE decomposition algorithms. Lines are the four model observers. Points show 9AFC observer study results and the error bars depict the 95% confidence interval across the 6 human observers.

with the spatial frequencies of interest in the imaging task (i.e., the frequency of crenelations, as shown in Fig. 1), then detectability plummets. Such is predicted theoretically and borne out experimentally, demonstrating that cascaded systems analysis based on first principles of Fourier signal and noise transfer characteristics can provide a practical, flexible means of describing actual human observer performance.
Figure 6. Theoretical calculations and experimental measurements of $A_Z$ for the two discrimination tasks as a function of decomposition parameter ($d_{LPF}$ and $d_{HPF}$). Two noise reduction algorithms are shown: SSH (left column) and ACNR (right column). Cases, symbols, and error analysis are as in Fig. 5

5. CONCLUSION

The development of advanced imaging modalities benefits tremendously from an improved, quantitative understanding of detector performance (MTF, NPS, and NEQ), imaging task (described here in terms of Fourier task functions), and observer performance (either calculated theoretically by way of detectability index or measured using real human observers). In this work, a theoretical model grounded in first principles of Fourier signal and noise transfer characteristics was presented for DE imaging systems and combined with various imaging tasks as a means of predicting the performance of real human observers. The results clearly demonstrate how imaging task, acquisition parameters, and decomposition algorithms affect imaging performance, but more importantly, the work provides initial validation of cascaded systems analysis as a means of predicting real observer performance. Reasonable agreement was found between theory and measurement for model observers incorporating a simple “eye” and internal observer noise, with NPWE demonstrating the best overall correspondence to measurements for various algorithms and imaging tasks. Such work, aiming to relate fundamental detector performance to human observer performance, helps to bridge the gap between simple, practical Fourier-based imaging performance metrics (e.g., NEQ) and task-dependent performance of real observers.

ACKNOWLEDGMENTS

The authors extend their gratitude to 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 M. J.
Daly, N. A. Shkumat, 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 Institutes of Health Grant No. R01-CA-127944-01.

REFERENCES