

The Effect of Nonlinear Human Visual System Components on Performance of a Channelized Hotelling Observer in Structured Backgrounds

Yani Zhang*, *Member, IEEE*, Binh T. Pham, and Miguel P. Eckstein

Abstract—Linear model observers based on statistical decision theory have been used successfully to predict human visual detection of aperiodic signals in a variety of noisy backgrounds. However, some models have included nonlinearities such as a transducer or nonlinear decision rules to handle intrinsic uncertainty. In addition, masking models used to predict human visual detection of signals superimposed on one of two identical backgrounds (masks) usually include a number of nonlinear components in the channels that reflect properties of the firing of cells in the primary visual cortex (V1). The effect of these nonlinearities on the ability of linear model observers to predict human signal detection in real patient structured backgrounds is unknown. We evaluate the effect of including different nonlinear human visual system components into a linear channelized Hotelling observer (CHO) using a signal known exactly but variable (SKEV) task. In particular, we evaluate whether the rank order of two compression algorithms (JPEG versus JPEG 2000) and two compression encoder settings (JPEG 2000 default versus JPEG 2000 optimized) based on model observer signal detection performance in X-ray coronary angiograms is altered by inclusion of nonlinear components. The results show: 1) the simpler linear CHO model observer outperforms CHO model with the nonlinear components; 2) the rank order of model observer performance for the compression algorithms/parameters does not change when the nonlinear components are included. For the present task and images, the results suggest that the addition of the nonlinearities to a channelized Hotelling model may add complexity to the model observers without great impact on rank order evaluation of image processing and/or acquisition algorithms.

Index Terms—Image compression, model observer, nonlinear component, signal detection.

I. INTRODUCTION

MEDICAL images are typically visually scrutinized by physicians searching and classifying morphological features of interest that might predict the state of disease. In this context, medical image quality has been objectively defined in terms of physicians' performance in clinically relevant visual tasks. Thus, evaluating and optimizing medical image quality entails conducting psychophysical studies measuring

human performance in these visual detection or classification tasks. Because conducting such psychophysical studies with physicians is time consuming there has been a rationale to develop model observers (also known as numerical observers) that can accurately predict human performance across a range of conditions. Advantages of using these model observers are that they can assist in evaluating, rank ordering, and optimizing image quality in a time efficient manner and also allow for investigation of a range of conditions that is simply not feasible using psychophysical studies. There is now a large literature of successful use of linear model observers to predict human visual detection performance in noise with varying power spectra [1]–[8]. Model observers have also been used to optimize single photon emission computed tomography, lens aperture, and image compression of coronary angiograms [7]–[11].

Linear model observers involve processing input image/images and making a decision about the presence/absence, location and/or class of a signal. When the images are computer generated to be statistically stationary and Gaussian then performance of the linear models can be computed using closed-form expressions rather than running computer simulations [12]. With patient anatomic backgrounds, investigators can also attempt to approximate performance of the model by using Fourier methods and assumptions about stationarity of the background and Gaussian distribution of the model's decision variable [13] or conduct sample driven computer simulations without making stationarity assumptions [14]–[16]. Specific linear models used include: nonprewhitening matched filter with an eye filter (NPWE), Hotelling observer (HO), channelized Hotelling observer (CHO), and Laguerre Gaussian Hotelling observer (LGHO). These models typically include some human visual system constraints such as a contrast sensitivity function (i.e., an eye filter) or a set of spatial frequency tuned channel mechanism, and decision rules that make use of knowledge about signal and noise. These linear models can be reduced to the correlation between a model template and the image data at different possible signal locations. One notable property about most implementations of these model observers is that they do not include any nonlinear components. Efforts to include a nonlinear component into single-template linear models have included intrinsic uncertainty via a maximum rule [17], [18] and a nonlinear transducer [19]. In addition, there has been another family of models developed to predict the detectability of a target added to one of two identical backgrounds, known as masks. Pattern masking refers to the performance degradation (or contrast threshold elevation) in human visual signal

Manuscript received December 1, 2005; revised June 13, 2006. *Asterisk indicates corresponding author.*

*Y. Zhang is with the Vision and Image Understanding Laboratory, Department of Psychology, University of California, Santa Barbara, CA 93106 USA (e-mail: yanizh@yahoo.com)

B. T. Pham and M. P. Eckstein are with the Vision and Image Understanding Laboratory, Department of Psychology, University of California, Santa Barbara, CA 93106 USA.

Digital Object Identifier 10.1109/TMI.2006.880681

detection when the signal appears added to a deterministic background. Masks could differ from the target in spatial frequency, orientation, and/or other pattern dimensions. Note, that linear model observers used in medical imaging do not predict any effect on performance of two identical backgrounds unless one assumes that the internal noise depends on the background. Masking models usually start with the processing of the image by a set of spatial frequency and orientation tuned channels and include a number of nonlinear components. These nonlinear components include an accelerating nonlinear relationship between the input and the channel response, referred to as the excitatory component, and a divisive suppressive of channel response by other channels, referred to as the inhibitory component. These components reflect the nonlinear properties of the firing of cells in the primary visual cortex (V1) and are thought to give rise to pattern masking [20]–[24]. Daly and Lubin [23], [24] described a multiple spatial frequency channel model based on the cortex transform with within-channel masking. Heeger [25] proposed a model where the response of each neuron had an accelerating nonlinearity and also divisively inhibited by a pool of responses of other neurons. Foley [20] investigated masking of one spatial frequency component by another and found interactions that can be explained by assuming that the contrast gain of a channel is reduced by activity in other channels. More recently, Watson and Solomon [22] implemented an image driven multiple channel model incorporating lateral interactions among channels in an effort to account for the between frequency masking. This implementation included a division of the excitatory response by an inhibitory response which was a linear combination of responses within a neighborhood in space, frequency, orientation, and phase [21], [26]. The effect of including the nonlinear components on the ability of channel models to predict human performance in structured backgrounds has not been investigated.

The goal of this paper was to study the effect of including a number of these nonlinearities on model observer performance detecting a simulated signal embedded in structured backgrounds extracted from patient X-ray coronary angiograms. In particular, we evaluated whether the rank order of image compression algorithms based on model observer performance was altered by inclusion of the nonlinear components. We also compared the ability of different models to quantitatively predict human performance as a function of compression ratio.¹ The algorithms to be considered were: 1) Joint Photographers Expert Group (JPEG), a still-image compression standard based on discrete cosine transform (DCT) [27]; 2) JPEG 2000, a still-image compression algorithm based on discrete wavelet transform (DWT) [28]; 3) Optimized JPEG 2000, a set of JPEG 2000 encoder parameters optimized via a genetic algorithm resulting in improved linear model and human observer performance for a signal detection task [8].

To evaluate a model observer that was general and flexible to allow for inclusion of nonlinearities, we selected the CHO [29] with a Gabor channel mechanism [30], [31]. The nonlinear components examined include: 1) maximum rule among U detec-

tors/templates due to intrinsic uncertainty about the spatial location of the target; 2) excitatory transducer; 3) inhibitory channel interaction. Many of the studied nonlinear components have free parameters. For the parameters we investigated a range of values found in previous studies to predict human performance [20], [22]. In addition to the nonlinear components, we investigated a number of linear model components based on human visual optics and sampling used by Geisler and Davila [32]. We also explored implementations of models using input images in gray level values, contrast, and/or luminance, and their effect on the rank order of the compression algorithms.

The task was detection of a simulated lesion within one of four simulated arteries (see locations indicated by the fiducial marks in Fig. 1 embedded in real structured X-ray coronary angiographic backgrounds. In the present study a signal known exactly but variable (SKEV) task was used. In the SKEV task the signal varied in shape and size across trials but was known exactly to the observer. In addition to the theoretical model calculations, a human observer psychophysical study was conducted and compared with the obtained model observer performance under different compression schemes/ratios.

II. GENERAL FRAMEWORK FOR CHANNELIZED HOTELLING MODEL OBSERVERS

Fig. 1 shows a general framework to study the effect of different model components on the channelized Hotelling model observer using X-ray coronary angiograms [7], [15]. Common elements to all models to be investigated include: different input format including gray value, luminance, and contrast; front end optics and receptor sampling; the processing of the image by a set of spatial frequency and orientation channels; the combination (linear or nonlinear) of responses across channels to a single decision variable for each of the four possible signal locations; and the use of the maximum rule among the four decision variables to choose the signal location. The particular models to be compared include: 1) linear model; 2) inclusion of intrinsic spatial uncertainty; 3) inclusion of excitatory channel transducer; 4) inclusion of excitatory transducer and inhibitory channel interactions. In the following sections, we present the theory for some of the model components to be evaluated.

A. Front End

Human visual optics and sampling are components that are included in many masking and basic vision models [32] but often omitted in many linear model observers used for detection in noise.² The optics of the eye can be described by a point-spread function, $\text{PSF}(x, y)$ [33], in a weighted sum of two Gaussians format

$$\text{PSF}(x, y) = aG_1(x, y) + (1 - a)G_2(x, y) \quad (1)$$

where $G_i(x, y) = (\sqrt{2\pi}s_i)^{-2} \exp[-(x^2 + y^2)/2s_i^2]$. The parameters were selected as reported by Geisler and Davila [32], $a = 0.583$, $s_1 = 0.443$ arcmin, and $s_2 = 2.04$ arcmin.

¹Nonlinear models were not fit to the data because their large number of adjustable parameters would lead to over fitting the human data. Instead model parameters were selected from previous studies.

²Although one might argue that the nonprewhitening matched filter with an eye filter model (NPWE) includes a composite of early optics, sampling and neural sensitivity.

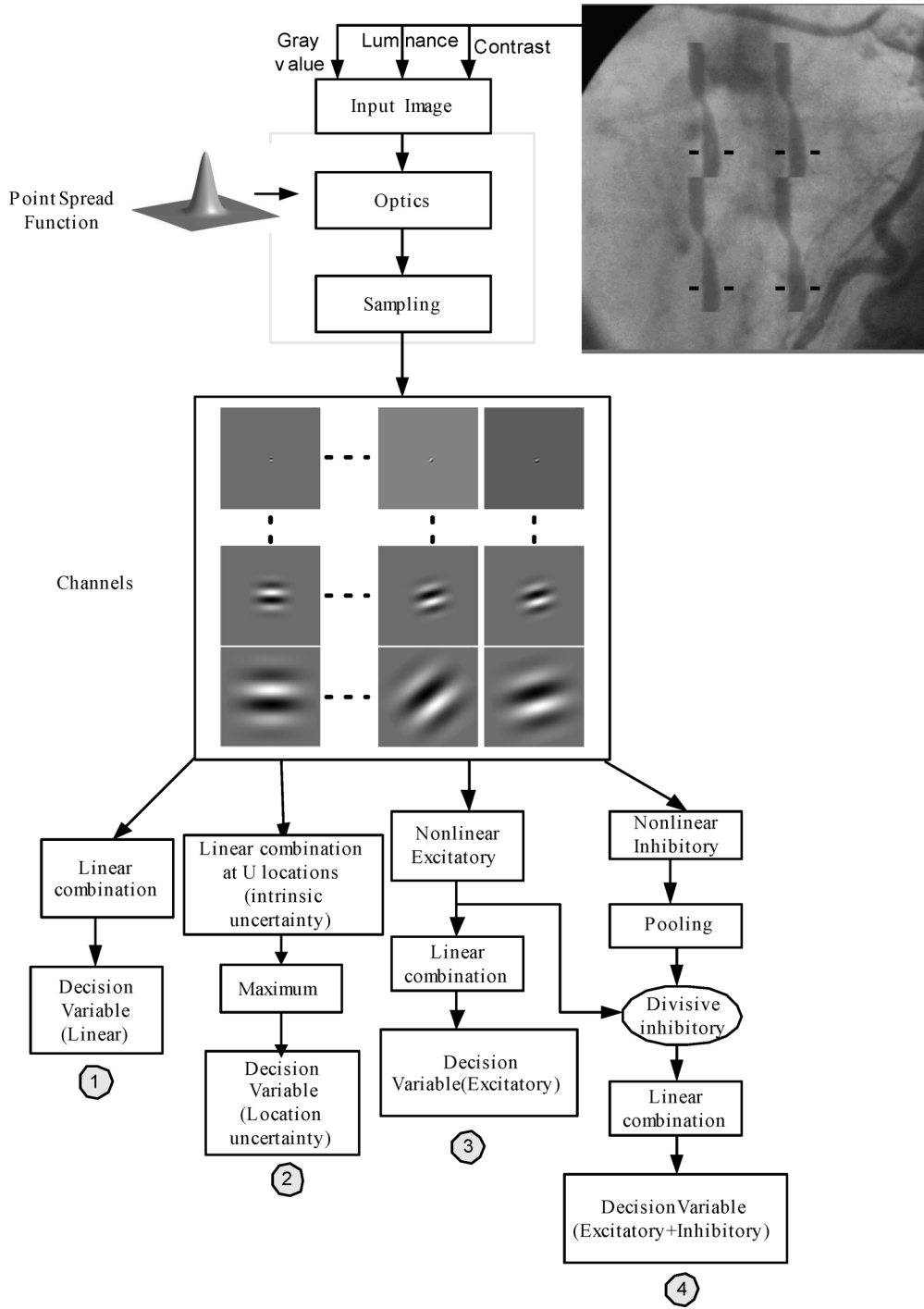


Fig. 1. Structure of the channelized Hotelling model observer with the front end: 1) linear components only; 2) location uncertainty; 3) nonlinear components-excitatory; 4) nonlinear components-excitatory plus inhibitory.

Two crucial stages of sampling take place in the retina. One is via photoreceptors (rods and cones) that initially transform light into voltage and electrochemical signals. The other is via the retinal ganglion cells which sample the outputs of the photoreceptors (after being processed by the horizontal, bipolar, and amacrine cells). Photoreceptors are assumed to be of a circular shape with a diameter of 0.6 mm at the widest portion of the inner segment. An additional assumption is that all quanta entering the inner segment follow into the layers of photopigment inside a photoreceptor [32]. As a consequence, each photore-

ceptor performs a small amount of spatial integration. This integration can be modeled by convolving the image with a circular mask $R(x, y)$ with a diameter 0.6 mm

$$R(x, y) = \frac{1}{\pi(\text{rad})^2}, \quad \text{for all } \sqrt{x^2 + y^2} \leq \text{rad}. \quad (2)$$

For our particular implementation with an observer viewing distance of 400 mm (with a resolution of 0.3 mm per pixel), a pixel

has a visual angle of 2.58 arcmin. Thus, rad in (2) is less than one pixel and sampling will have no effect on the image.

B. Implementation of the Channelized Hotelling Model Observer

The use of spatial-frequency-selective channels in the human visual system has long been part of the field of vision science [34]–[36], and has often been incorporated into models of visual detection [24]. The CHO was developed as a method to incorporate elements from human visual system into models of signal detection for medical images [29], [37]–[39]. The CHO model has been successfully used as a predictor of human performance in a number of detection experiments [6], [8], [40]. The existence of a channel mechanism is supported by psychophysical and physiological evidence [41]. In the CHO model, entire frequency content of an image within a given frequency band or channel is reduced to a single output value, there is a potentially large loss of information as an image is processed by a sparse channel mechanism. CHO model uses a weighted sum of the linear channels to derive a linear template. Weights to each channel take into account signal information transmitted by the channel and the variance of the channel responses as well as the covariance between different channel responses. A key step to implement CHO is choosing channels. There are a number of different channelized Hotelling models in the literature including square channels [29], difference of Gaussians [6], difference of Mesa filters [5], and Gabor channels [7], [31]. Gabor function has been used to model the spatial summation properties (of the receptive fields) of simple cells in the visual cortex [42]. Further, Jones, and Palmer [43] found that two-dimensional (2-D) Gabor filters fitted the 2-D spatial and spectral response profile of each simple cell in the least-squared error sense. Unlike most other channel models used with the CHO model, the Gabor functions are not rotationally symmetric and have different channels tuned to different orientations which can be used to explain the orientation and spatial frequency selectivity of simple cells. Here, we used a set of Gabor channels: 80 channels with five spatial frequencies (central frequencies, 16, 8, 4, 2, and 1 cycle per degree), eight orientations (equally spaced in orientation), and two phases (odd, 0; and even, $\pi/2$). Spatial frequency bandwidth of the channels was approximately 1 octave. Gabor channels are given by

$$C_{\text{cho}} = \exp\left(-\frac{4\ln 2(x^2 + y^2)}{w_s^2}\right) \times \cos[2\pi f_c(x \cos \theta + y \sin \theta) + \beta] \quad (3)$$

where f_c is the spatial frequency, θ is the orientation, w_s is the full width at half height, and β is the phase. w_s is related to the standard deviation σ of the Gaussian functions by: $w_s = 2.35 \sigma$. Fourier transform of a Gabor channel is a Gaussian function centered at the central frequency f_c and with a half-height full width given by $w_f = 0.8825/w_s$.

All of the investigated models (see 1–4 in Fig. 1) involve taking the dot product of the individual channels and the image

data. However, they differ on the subsequent step to manipulate this dot product to obtain a channel response.

1) *Linear CHO*: For the linear CHO model, the decision variable, $\lambda_{L,i,j}$ for each possible signal location i ($i = 1, \dots, 4$) and each known signal j ($j = 1, \dots, 184$) is calculated as a linear combination of the response of each channel

$$\lambda_{L,i,j} = \sum_{k=1}^{80} W_{L,k,j} \mathbf{R}_{L,k,i} \quad (4)$$

where the subscript L denotes that the decision variable is obtained from a linear operation. The channel response $\mathbf{R}_{L,k,i} = \mathbf{C}_{k,\text{cho}}^T \mathbf{g}_i$ is the dot product of transpose of the k th channel, $\mathbf{C}_{k,\text{cho}}$, with the image data \mathbf{g}_i at the i th location. Vector $\mathbf{W}_{L,j}$ is the channel weights for the j th signal with each element represents the optimal linear weights³ for each channel, and defined by [3], [29], [38]

$$\mathbf{W}_{L,j} = \mathbf{K}_{L,v}^{-1} [\langle \mathbf{R}_{L,j,c/s} \rangle - \langle \mathbf{R}_{L,c/b} \rangle] \quad (5)$$

where vector $\langle \mathbf{R}_{L,j,c/s} \rangle$ contains the mean of the j th signal plus background as seen by the channels. For each of the 184 signals, 400 additional samples were created as a training set to estimate $\langle \mathbf{R}_{L,j,c/s} \rangle$. The vector $\langle \mathbf{R}_{L,c/b} \rangle$ is the mean of the background as seen by the channels. Estimation of $\langle \mathbf{R}_{L,c/b} \rangle$ is based on 2700 samples. $\mathbf{K}_{L,v}$ is an $N \times N$ matrix describing the covariance of the output of the channels to the images

$$\mathbf{K}_{L,v} = \text{Dev}_{L,c/b} \text{Dev}_{L,c/b}^T / (2700 - 1)$$

with

$$\text{Dev}_{L,c/b} = [\mathbf{R}_{L,c/b,1} - \langle \mathbf{R}_{L,c/b} \rangle, \dots, \mathbf{R}_{L,c/b,2700} - \langle \mathbf{R}_{L,c/b} \rangle] \quad (6)$$

where $\text{Dev}_{L,c/b}$ is a matrix with each column as the channel response of each noise sample. For our particular implementation the covariance matrix is an 80×80 matrix. Thus, a total of 3100 samples are computed for the CHO model to estimate the best linear weights for each signal. For a particular signal j , a corresponding $\mathbf{W}_{L,j}$ is obtained.

Given that the model is fully linear, the CHO model can be implemented by obtaining a single template (in the pixel domain) calculated as the linear combination of the channels. The single template corresponding to the displayed signal is then applied to each of the possible signal locations. The response can be expressed in matrix formulation as [40]

$$\lambda_{L,i,j} = \boldsymbol{\omega}_j^T \mathbf{g}_i \quad (7)$$

The subscript j on $\boldsymbol{\omega}$ refers to the j th template, and the subscript i for \mathbf{g} refers to data at the i th location. The linear template

³For non-Gaussian noise processes, there might be a channelized template that performs better than the channelized Hotelling template.

that can be obtained from the channel weights and the channel profiles is calculated as [7]

$$\omega_j = \sum_{k=1}^{80} W_{L,k,j} \mathbf{C}_{k,\text{cho}}. \quad (8)$$

2) *Intrinsic Spatial Uncertainty*: Intrinsic uncertainty refers to observers' inability to optimally use all information provided by the experimenter about the signal [44], [45]. For example, even though the experimenter provides the observer a copy of the signal and indicates the possible signal locations with high-contrast fiducial marks, the observer might not be able to use this information optimally and still have some uncertainty about the location, spatial frequency, and orientation of the signal. The observer is, therefore, assumed to monitor not a single template but multiple templates tuned to different parameters s/h is uncertain about [17], [44]–[46]. In the context of spatial uncertainty, the observer is assumed to monitor the response of templates centered at different spatial locations. For an M alternative forced choice, the observer monitors U templates/locations per alternative and uses the maximum template response for each alternative to make the decision. In the context of a channel model with 80 channels, one might implement spatial uncertainty by assuming that the observer monitors $U \times 80$ channels per location. Alternatively, one might assume that the observer synthesizes a single template from the channels [see (9)] and, therefore, monitors U template responses per alternative. We chose the second implementation for simplicity. In addition, previous studies [17] have assumed that U responses to be statistically independent which allows for calculating the effect of uncertainty from closed-form expressions rather than computer simulations. However, if the considered locations are overlapping and the statistically independent internal noise does not dominate the external noise, this assumption is not likely. Thus, in the present paper we calculated the effect of uncertainty by using Monte Carlo simulations. The decision variable on each trial was calculated as follows:

$$\begin{aligned} \lambda_{L,i,j} &= \max(\lambda_{L,i,j,u}) \\ &= \max \left(\sum_{k=1}^{80} W_{L,k,j} \mathbf{C}_{k,\text{cho}}^T \mathbf{g}_{i,u}, u = 1, \dots, U \right) \end{aligned} \quad (9)$$

where $\mathbf{g}_{i,u}$ is the image data at the i th location with the u spatial locations. We shifted the pixel in which the template was centered on prior to the dot product to implement the location uncertainty. Fig. 2 (right) illustrates an example where the known center of the signal is at location 13 (a 5×5 uncertainty area, 0.3010 degree of visual angle horizontally and vertically assuming a viewing distance of 400 mm) The model's decision variable for this location was calculated as the maximum of the 25 responses corresponding to different alignments of the model template and the image data.

3) *Inclusion of the Excitatory (E) Nonlinearity Into the Linear Channelized Hotelling Observer*: Excitatory nonlinearity refers to the accelerating response of channels to the visual input typically implemented with an exponent and used

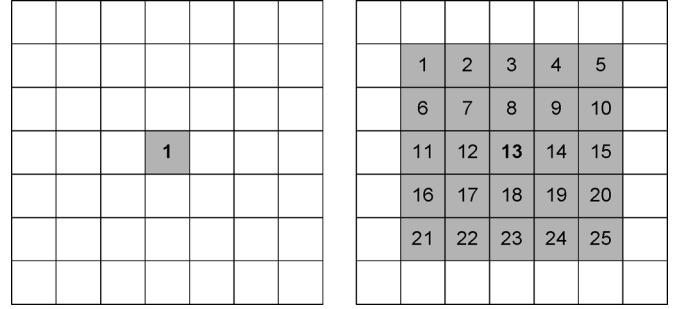


Fig. 2. Intrinsic spatial uncertainty implemented by obtaining the dot product of the model template and the image data with different spatial alignments. Shaded areas indicate the uncertainty areas represented by the image pixels aligned with the center of the template. Left: No location uncertainty. Right: Spatial uncertainty: template is centered at 25 different possible locations.

to model neuronal response. This section incorporates an excitatory nonlinearity in the CHO model. In the excitatory (E) path, the response undergoes a power-law nonlinearity with an exponent of p

$$E = \text{sign}(r) \text{abs}(r)^p \quad (10)$$

where r represents the channel response, i.e., the dot product of channel C_{cho} with the data. In the model by Foley [20], p is typically a value between 2 and 3, while in Teo–Heeger's model [21] it has a constant value of 2. In Watson and Solomon [22], the nonlinearity is applied to the unsigned response magnitude, to which the sign is then reattached. In our implementation, we selected to use Watson and Solomon's scheme but with different parameters.

Following the standard CHO scheme, channel weights for this model are estimated as

$$\mathbf{W}_{E,j} = \mathbf{K}_{E,v}^{-1} [\langle \mathbf{R}_{E,j,c/s} \rangle - \langle \mathbf{R}_{E,c/b} \rangle]. \quad (11)$$

The vector $\langle \mathbf{R}_{E,j,c/s} \rangle$ is estimated as the mean of p powered signal plus backgrounds seen through the channels. The vector $\langle \mathbf{R}_{E,c/b} \rangle$ is estimated as the mean of p powered backgrounds alone seen through the channels. $\mathbf{K}_{E,v}$ is the covariance of the p powered output of the channels to the images. The decision variable for each location is obtained as

$$\lambda_{E,i,j} = \sum_{k=1}^{80} W_{E,k,j} \mathbf{R}_{E,k,i} \quad (12)$$

where $\mathbf{R}_{E,k,i} = \text{sign}(\mathbf{C}_{k,\text{cho}}^T \mathbf{g}_i) \text{abs}(\mathbf{C}_{k,\text{cho}}^T \mathbf{g}_i)^p$ is the p powered channel response of the data. The optimality of the weights depends on whether the channel response distributions (across images) are Gaussian after the nonlinearity. This is most likely not true. For a consistent comparison of linear and nonlinear combination rules, we use the same weighted linear sum of channel outputs for both linear and nonlinear observers. For the nonlinear models, the nonlinearity was contained in the channel output.

4) *Inclusion of the Excitatory and Inhibitory (EI) Nonlinearity Into the Linear Channelized Hotelling Observer:* Inhibitory nonlinearity refers to the suppressive response of channels to the visual input. This model incorporates an excitatory nonlinearity that is divisively inhibited by a pool of responses from other neurons/channels [20]–[22]. Divisive inhibition has been considered an essential feature to describe the characteristic of human striate cortical neurons. In Teo and Heeger’s implementation [21], which is based on a direct model of neural cell responses [25], pooling is limited to the orientation only. The model presented by Watson and Solomon [22] allows for an easy integration of many kinds of channel interactions and spatial pooling.

The excitatory and inhibitory (EI) nonlinearity is expressed as

$$\begin{aligned} \text{EI} &= \frac{\text{sign}(r)\text{abs}(r)^p}{b^q + \text{sign}(r)\text{abs}(r)^q * H} \\ &= \frac{\text{sign}(r)\text{abs}(r)^p}{b^q + \text{sign}(r)\text{abs}(r)^q * \left(\frac{1}{2\pi\sigma_H} e^{-\frac{x^2+y^2}{2\sigma_H^2}} \right)} \end{aligned} \quad (13)$$

where $\text{sign}(r)\text{abs}(r)^p$ represents the excitatory nonlinearity, $\text{sign}(r)\text{abs}(r)^q$ represents the inhibitory nonlinearity, $\text{sign}(r)\text{abs}(r)^q * H$ represents the pooling in the inhibitory path, and b is a positive constant which defines the saturation point and also prevents division by zero. The standard deviation of the convolution kernel H , σ_H is also a parameter to be evaluated. The excitatory term in the numerator consists of a simple power-law nonlinearity with exponent p . The inhibitory term in the denominator controls the gain of the excitatory path. It also includes a nonlinearity with a different exponent q . Responses are pooled over different channels in the inhibitory term through convolution with a Gaussian pooling kernel. We used the Watson and Solomon [22] model because it applied image based Gabor channels and thus made it particularly appropriate to our version of the channelized Hotelling model that also uses Gabor channels. Other models such as Daly’s scheme used Mesa filters [24]. The Foley [20] model was not image based but started with the responses of channels.

In this study, we divided 80 channel responses into two groups according to the phase. Thus, in each group we had five spatial frequencies and eight orientations. The scalar responses of each group can be represented by a 5×8 channel response matrix. To achieve the inhibition across similar orientations, we used the circular convolution for orientation, in which the borders of the two operands are implicitly connected at their edges. The output of the inhibitory stage has the same dimensions as that of the excitatory stage. The pooling operation in the inhibitory path linearly combines signals over the frequency and orientation. After pooling, we obtained the postexcitatory inhibitory responses through a divisive operation.

The vector of weights \mathbf{W}_{EI} for the individual channels is then estimated by

$$\mathbf{W}_{\text{EI},j} = \mathbf{K}_{\text{EI},v}^{-1} [\langle \mathbf{R}_{\text{EI},j,c/s} \rangle - \langle \mathbf{R}_{\text{EI},c/b} \rangle]. \quad (14)$$

The responses undergo an EI nonlinearity. The vector $\langle \mathbf{R}_{\text{EI},j,c/s} \rangle$ is estimated as the mean of EI signal plus backgrounds filtered by the channels. The vector $\langle \mathbf{R}_{\text{EI},c/b} \rangle$ is estimated as the mean of EI backgrounds filtered by the channels. The term $\mathbf{K}_{\text{EI},v}$ represents the covariance of the EI output of the channels to the images. The decision variable is calculated through

$$\lambda_{\text{EI},i,j} = \sum_{k=1}^{80} W_{\text{EI},k,j} \mathbf{R}_{\text{EI},k,i} \quad (15)$$

where $\mathbf{R}_{\text{EI},k,i}$ is the k th EI channel response of the data.

C. Decision Rule and Performance Calculations

1) *Proportion Correct:* The most general way to compute model observer performance in an MAFC task is to calculate the probability of a correct outcome on each trial. A correct outcome occurs when the j th decision variable to the signal location exceeds the maximum decision variable to the noise only locations. One obtains an estimate of this probability from samples by calculating the decision variable for different locations and tallying the proportion of trials where the model correctly identifies the signal location. This can be mathematically expressed as

$$\begin{aligned} O_t &= \text{step}(\lambda_{s,j,t} - \max(\lambda_{n,j,t})) \\ &= \begin{cases} 1, & \text{if } \lambda_{s,j,t} \geq \max(\lambda_{n,j,t}) \\ 0, & \text{if } \lambda_{s,j,t} < \max(\lambda_{n,j,t}) \end{cases} \\ P_c &= \frac{1}{T_r} \sum_{t=1}^{T_r} O_t \end{aligned} \quad (16)$$

where T_r is the total number of trials, $\lambda_{s,j,t}$ is the decision variable on trial t to the j th signal present location, $\lambda_{n,j,t}$ is the decision variable on trial t to the j th signal absent locations, and the $\max(\cdot)$ function takes the maximum among the $M - 1$ decision variables to the noise locations.

2) *Index of Detectability From Decision Variable:* In many tasks, the signal appears in one of M orthogonal locations where each location is within an independent background sample. Thus, the decision variable can be assumed to be statistically independent across all possible signal locations. Another common assumption is that the template responses obey Gaussian distributions. Under these assumptions, one can represent the model observer responses to the signal location and to the background only locations as being sampled from two univariate Gaussian distributions. Performance can then be defined in terms of the distance in standard-deviation units between the signal-plus-background and the background-only distribution. This metric is known as the index of detectability [47]

$$d' = \frac{\langle \lambda_s \rangle - \langle \lambda_n \rangle}{\sigma_\lambda} \quad (17)$$

where $\langle \lambda_s \rangle$ is the mean model response to the signal plus background locations, $\langle \lambda_n \rangle$ is the mean model response to the back-

ground only locations. For the equal variance case, $\sigma_\lambda = \sigma_{\lambda,s} = \sigma_{\lambda,n}$ where $\sigma_{\lambda,s}$ is the standard deviation for the signal present location, and $\sigma_{\lambda,n}$ is the standard deviation of the noise only locations. For the unequal variance scenario, the denominator in (17) is expressed as $\sigma_\lambda = \sqrt{(1/2)\sigma_{\lambda,s}^2 + (1/2)\sigma_{\lambda,n}^2}$. Assuming Gaussian independent decision variable for each location and equal variance for the signal plus background and background only distributions, a proportion correct measure of performance can be obtained from d' using the following relationship [48]:

$$Pc(d', M) = \int_{-\infty}^{+\infty} \phi(x - d') [\Phi(x)]^{M-1} dx \quad (18)$$

where $\phi(x) = (1/\sqrt{2\pi})e^{-x^2/2}$, $\Phi(x)$ is the cumulative Gaussian distribution function, $\Phi(x) = \int_{-\infty}^x \phi(y)dy$, and M is the number of possible signal locations in the experiment. An index of detectability can be calculated from proportion correct using a precalculated table from (18). In the following, we refer to the index of detectability transformed from proportion correct as d_{maf_c} to differentiate it from the index of detectability calculated from the decision variable [(17): d']. When the decision variables are indeed Gaussian distributed, equal variance, and independent, the index of detectability calculated from the decision variable d' will be equal to that calculated from transformations of Pc [via a table created with (18)], d_{maf_c} . However, when the model responses depart from Gaussian and independent these equalities will fail. The assumptions need to be verified when the backgrounds are samples from real medical images. However, when the observer responses are non-Gaussian, d_{maf_c} [from Pc by (18)] is an acceptable figure of merit and similar to the monotonic transformation of area under the receive operating characteristic (ROC) curve to d_{auc} .

III. METHODS

A. Preparation of the Test Image

Test images were created by using clinical medical image backgrounds with computer-generated signals inserted. The use of real backgrounds gave the images a realism that was difficult to achieve with computer simulations. These clinical digital coronary angiographic backgrounds were acquired at 30 frames/s with a 7-in image intensifier field size (Advantx/DXC, General Electric Medical Systems) and digitized with a linear analog amplification achieve a 512×512 pixel matrix. The procedure to create the simulated artery attempted to mimic the X-ray coronary angiograms generation process [49]. The simulated arteries were obtained by tracing the contours of a real artery, and generating 3-D right circular cylinders with varying diameters (see Appendix for the details about the simulated artery model). Signals were projected ellipsoids with the vertical axis ranging from 3 to 25 pixels (0.128–1.067 degrees of visual angle) and the horizontal axis ranging from 3 to 10 pixels (0.128–0.422 degrees of visual angle). This resulted in a total of 184 possible signals. The choice of 184 signals was arbitrary to reflect some variability for signal size/shape within ranges of realistic filling defect sizes. Four simulated arteries were created

and projected 64 pixels apart in width and 100 pixels in height as a group into 512×512 pixel real patient digital X-ray coronary angiogram backgrounds. A total of 541 backgrounds were extracted from 50 different image sequences of 17 different patients. A signal was randomly inserted into one of the four simulated artery segments. Then we compressed/uncompressed the 512×512 images and extracted 256×256 images centered around each group of inserted arteries from the 512×512 images. The final test set consisted of a total of 900 256×256 pixel images. All signals had the same peak contrast. The average signal contrast calculated from all of the 900 test images was 0.17.

B. Human Psychophysical Study

A human psychophysical study was conducted to compare the CHO performance with human observer performance for an SKEV task. For each studied compression algorithm (JPEG, JPEG 2000) and each JPEG 2000 encoder setting (default, optimal), we conducted the experiment at five compression ratios 7:1, 10:1, 15:1, 20:1, and 30:1. We also included the uncompressed condition (compression ratio 1:1) in the experiment.

The display equipment used was M17LMAX monochrome monitor with a maximum resolution of 1664×1280 pixels (Image Systems, Minnetonka, MN). Before each experiment session, the monitor was calibrated according to the Digital Imaging and Communications in Medicine (DICOM) standard.⁴ The observer's task was to select the location where s/he thought the signal appeared from four possible locations. Experiments were conducted in a darkened room with a viewing distance of 400 mm. On each trial, an image was randomly sampled from the 900 test image database and displayed. A higher-contrast copy of the signal was shown at the bottom of the test image to inform the observer about the size and shape of the signal. There was no time restriction for the observer to make a decision. Observers were given feedback about their decisions (correct/incorrect). An experimental session consisted of 100 trials and lasted approximately 10 min. All observers participated in a training session. In the training stage, we used five sessions of 100 trials with images selected from the uncompressed condition and five sessions with images randomly selected from the compressed condition (1000 trials in total). After training, each observer ran nine sessions per compression condition and thus had a total of 900 trials per experimental condition. Human performance Pc was measured by calculating the proportion of trials in which the observer correctly localized the signal. An index of detectability d_{maf_c} was then transformed from Pc by (18).

IV. RESULTS

A. Variability in Model Observer Performance Due to Case Selection

Model observer performance data has uncertainties from the following two components: 1) the particular training set of images used to fix the model parameters; 2) the test set of images used to determine the model observer's performance. We used bootstrapping [50] procedures to determine the variability in the

⁴<http://medical.nema.org/>

TABLE I
 AVERAGED BEST FITTED β ACROSS 11 COMPRESSION CONDITIONS AND FOUR SIGNALS

	Linear	Excitatory	Excitatory+Inhibitory
Signal-present	2.64	2.61	2.69
Background (noise)	2.19	2.50	5.50

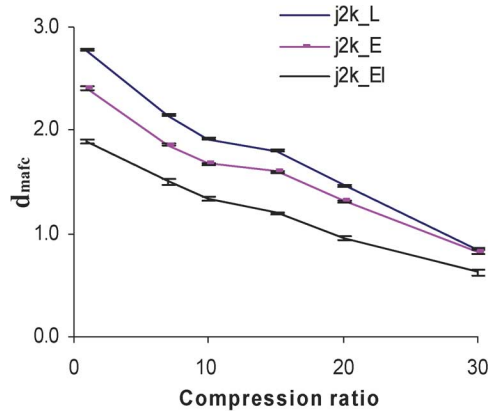
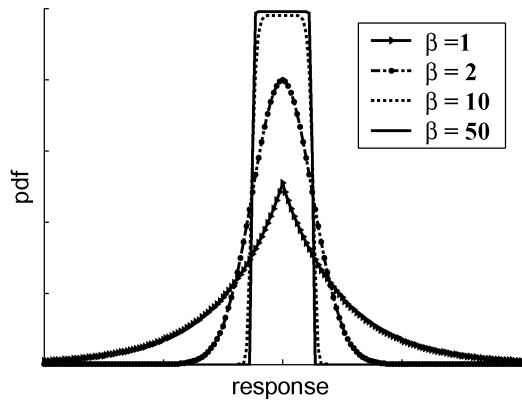


Fig. 3. CHO model observer performance with error bars for JPEG 2000. Error bars were calculated from ten samples of bootstrapping resampled training set and test set.


 Fig. 4. Different pdfs defined by (19) for different values of β .

model data from these two contributions [51]. The bootstrap resampling method was used to create both the training set for the calculation of model observer templates and the test set for the calculation of the model performance. We created 10 resample sets for the test. The variability in the models' performance (P_c) ranged (across compression) from 0.007 to 0.017 for the linear model, from 0.009 to 0.017 for the excitatory model, and from 0.008 to 0.040 for the excitatory and inhibitory model. Parameters used for the nonlinearity are the same as those used in Section IV-B for the evaluation of statistical properties of the decision variable. Error bars in Fig. 3 are the standard errors of the mean across ten bootstrap resample groups. We will not show error bars for model observer performance in the following results because error bars are much smaller than graph symbols.

B. Statistical Properties of the Decision Variable

Prior to presenting the results on the effect of compression we investigated the statistical properties of the decision variable to determine whether using of d' via calculation of the mean and standard deviation of the decision variable (17) is inappropriate due to either unequal variance, departures from normality, and/or response correlations. In this section for simplicity, in all of the experiment we use the combinations of p (excitatory) and q (inhibitory) parameters: $p = q = 2$, $b = 0.5$, $\sigma_H = 0.8$. [p , q , b , and σ_H are the parameters expressed in (10) and (13)].

1) *Equal Variance Assumption for Signal Plus Background and Background Only Decision Variables*: First we tested whether the standard deviations of the model's decision variable are the same for the signal-present (signal plus background) and background only locations. We used four signals randomly selected from the 184 signal pool and evaluated 11 compression conditions of the JPEG 2000 algorithm. For each studied signal, the test set used for calculating the signal-present locations included 400 samples. The background only location set included 2700 samples. The average ratios of $\sigma_{\lambda,n}/\sigma_{\lambda,s}$ across four signals and 11 compression conditions were 0.4303 ± 0.0630 for the linear model, 0.4454 ± 0.0703 for the excitatory model and 23.88 ± 10.2902 for the excitatory and inhibitory model. The ratios were significantly different from 1 ($p < 0.05$) and thus suggest that the equal variance assumption is violated.

2) *Normality of Decision Variable Distributions*: To verify whether the template responses were normal (Gaussian) distributed, we fitted the probability distribution function (pdf) given by

$$p(\lambda) = \alpha \exp\left(-\frac{1}{2} \left(\frac{|\lambda|}{\sigma}\right)^\beta\right) \quad (19)$$

where α is a normalizing parameter and β defines the shape of the distribution. For $\beta = 2$, the equation reduces to the usual Gaussian distribution. Values of β smaller than 2 result in a pdf that is more spread than the Gaussian and values of β greater than 2 result in a pdf that is more compact than the Gaussian, as shown in Fig. 4. Table I shows the best-fit β values to the distribution of decision variables for signal plus background and background only locations for all three models. All of the distributions are more compact than the Gaussian ($\beta > 2$). The distribution that is closest in shape to Gaussian is the decision variable for background only locations under the linear model.

3) *Interlocation Decision Variable Correlations*: Finally, we examined whether the decision variable across locations was

TABLE II
STATISTICS OF THE CORRELATION COEFFICIENTS ACROSS 11 COMPRESSION
CONDITIONS AND SIX SIGNALS

	Linear	Excitatory	Excitatory + Inhibitory
Mean	0.0117	0.0119	0.003
Std	0.0396	0.038	0.033
Max	0.1276	0.1368	0.1819
Min	-0.1437	-0.0914	-0.1491

correlated for different models. The noise response is the response obtained from the backgrounds only locations. We separated the models' noise responses obtained from the four possible locations as N1, N2, N3, and N4. We then calculated the correlation coefficients between (N1, N2), (N1, N3), (N1, N4), (N2, N3), (N2, N4), and (N3, N4). The following statistics on the interlocation correlation of the decision variable were obtained from 396 samples (6 signals \times 11 compression conditions \times 6 combinations). Data in Table II show that the noise responses from the four possible signal locations are weakly correlated and the nonlinearity does not change the correlation coefficient dramatically.

In summary, although the decision variable correlations across locations are not significantly different than zero, the decision variable's equal variance and normal distribution assumptions are violated. To evaluate the impact of these departures from the equal variance and normality on the figure of merit to calculate the index of detectability, we randomly selected 12 signals and generated a set of 900 test images for each signal. We then calculated the model observer detectability using (17) with nonequal variance (d') and (18) (d_{maf_c}). If the decision variables were indeed Gaussian and equal variance then the two figures of merit would result in the same number and their ratio (d_{maf_c}/d') would be unity. The average ratios of d_{maf_c}/d' for our test images across 12 signals were 0.98 ± 0.04 for the linear model, 1.40 ± 0.07 for the excitatory model and 3.96 ± 3.22 for the excitatory and inhibitory model. Thus, the results indicate that it is inappropriate to estimate performance using the index of detectability based on means and a single standard deviation of the response variable for the present anatomic backgrounds and/or model nonlinearities. In the following sections, the performance was measured by d_{maf_c} which was calculated from Pc [via a precalculated table using (18)].

C. Effect of Front End (Optics) Using Gray Level, Luminance and Contrast

The original images were expressed in gray values ranging from 0 (black) to 255 (white). We transformed the digital gray values of the input images to luminance based on the nonlinear relationship of the display described by the Digital Imaging and Communications in Medicine standard (DICOM).⁵ However, the human visual system's sensitivity to variation in luminance

depends on the local mean luminance, so we also used the contrast rather than absolute values of luminance. Local contrast $C(x, y)$ at point (x, y) is defined as

$$C(x, y) = \frac{L(x, y) - L_{\text{mean}}}{L_{\text{mean}}} \quad (20)$$

$L(x, y)$ is the luminance value and L_{mean} is the mean luminance calculated from the 92×92 pixel local region (which is the spatial extent of the largest Gabor) centered at the possible signal location. Conversion to local contrast is also a nonlinear transformation due to the inhomogeneity of the local luminance values of the X-ray coronary backgrounds. Fig. 5 shows the CHO performance obtained with and without the optics preprocessing for both the JPEG and JPEG 2000 compression algorithms at compression ratios 1:1, 7:1, 10:1, 15:1, 20:1, and 30:1. The left graph is the performance of the CHO model using images represented in gray values. The middle graph is the model's performance using luminance. The right graph is the performance obtained using images expressed in contrast (20). The curves with solid markers refer to the performance without front end optics (L) and the curves with empty markers refer to the performance after inclusion of the front end optics component (OL). Results show that the performance rank order of different compression algorithms is the same for all three different input image units (gray value, luminance, and contrast): JPEG (triangles) results in higher performance than JPEG 2000 (squares). In addition, results also show that including the optics front end into the model produces small changes on model observer performance for our display conditions (i.e., viewing distance). The maximum performance difference is 1.5%. In the following sections, model observer results were obtained using images transformed to contrast and including the front end optics.

D. Location Uncertainty

Fig. 6 shows the effect of location uncertainty on model observer performance. We introduced the location uncertainty by calculating the model observer response using (9) with the alignment of center of the template and image data shifted around a local area. The size of the local uncertainty area (in pixels) varied from 1×1 (perfect registration between the template and the correct signal location without uncertainty), 3×3 , 5×5 , 7×7 , to 9×9 . These areas correspond to 0.0430, 0.1290, 0.2150, 0.3010, and 0.3869 degrees of visual angle (for 40 cm viewing distance) horizontally and vertically. Fig. 6 shows that the model observer (linear CHO) performance decreases with the increase of the uncertainty area for JPEG 2000 (J2k) compression algorithm with default encoder setting. However, the rank order of the performance with regard to compression ratios is preserved across levels of location uncertainty: the performance decreases while compression ratio increases. JPEG algorithm shows the same trend as JPEG 2000, so we only presented the results with JPEG 2000.

E. Excitatory Nonlinearity With Different Parameters

Fig. 7 shows the CHO model performance with the excitatory nonlinearity for JPEG 2000 compression algorithm with power-law parameters: $p = 2, 2.5, 3,$ and 3.5 . The performance

⁵<http://medical.nema.org/>

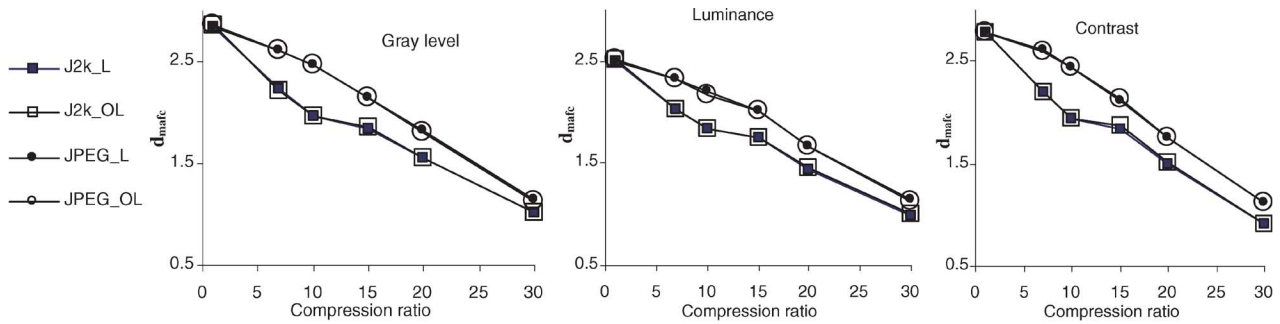


Fig. 5. Linear CHO performance as a function of compression ratio before (solid symbols, L) and after the “optics” (empty symbols, OL) for both the JPEG and JPEG 2000. Left: Performance obtained from gray-value. Middle: Performance obtained from luminance. Right: Performance obtained from local contrast.

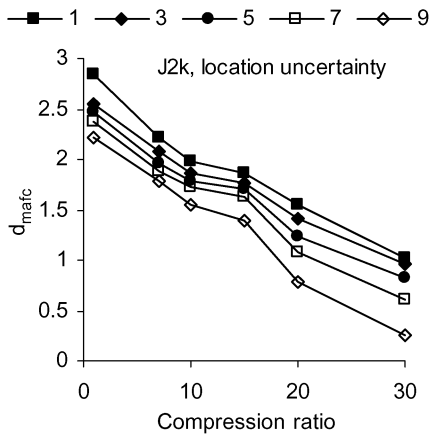


Fig. 6. Location uncertainty results. CHO model observer performance as a function of compression ratios (JPEG2000) with the model observer’s template shifted around areas of 1 × 1, 3 × 3, 5 × 5, 7 × 7, and 9 × 9 pixels.

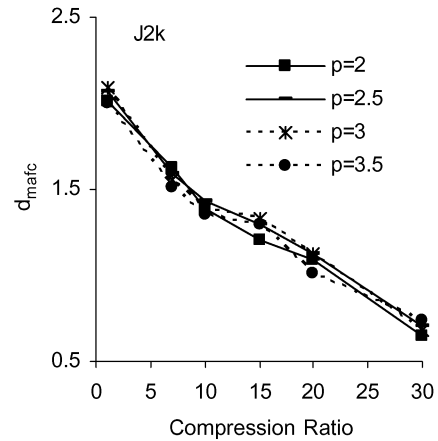


Fig. 8. CHO model performance as a function of JPEG 2000 compression ratios. Inhibitory component with constant $q = 2$, different p . $p = 2, 2.5, 3$, and 3.5 .

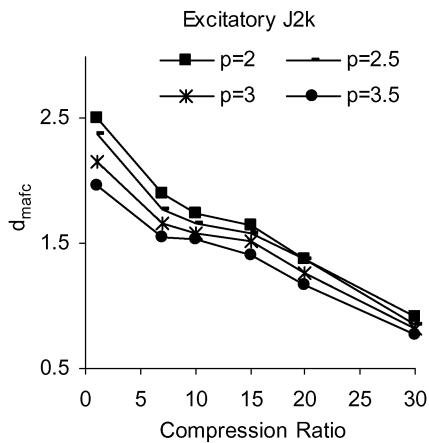


Fig. 7. CHO model performance with different excitatory parameters using JPEG 2000 as a function of compression ratio for different excitatory parameters, $p = 2, 2.5, 3, 3.5$.

slowly decreases with the increasing value of p . The performance rank order with regard to compression ratios remains unchanged after inclusion of the excitatory nonlinearity into the channel mechanism. Fig. 7 shows that the absolute performance of excitatory model observer is lower than that of the linear model observer (see right graph of Fig. 5 for comparison).

F. Inhibitory Nonlinearity With Different Parameters

As shown in (13), the model parameters p and q control the slopes of the accelerating and compressive (decelerating) parts of the nonlinearity, typically $p > q$ and $q \approx 2$ [52]. The saturation constant b is added to prevent division by zero. In Teo and Heeger’s implementation [21], the exponents of both the excitatory and inhibitory nonlinearity are fixed at $p = q = 2$ so as to be able to work with local energy measurements. In the following section, we extensively investigate the influence of different excitatory and inhibitory parameters on the channelized Hotelling model observer performance.

1) *Excitatory and Inhibitory Parameters: p, q* : First, we fixed the saturation constant ($b = 0.5$), convolution kernel standard deviation ($\sigma_H = 0.8$), and $q = 2$ while letting p vary from 2 to 3.5. Results in Fig. 8 show that performance varies little with the different values of p for JPEG 2000 algorithm. We also tested the following relationships of p (excitatory) and q (inhibitory) with parameters: $p = 3, q = 0.6p, q = 0.8p$, and $q = p$. Results in Fig. 9 show that there is little difference among these three groups of parameters. Comparisons to Fig. 5 (right) show that the linear model observer outperforms the models with excitatory plus inhibitory nonlinear components.

2) *Saturation Constant b* : The saturation constant [see (13)] controls the deceleration of the model response. Smaller saturation constant values make the model response more com-

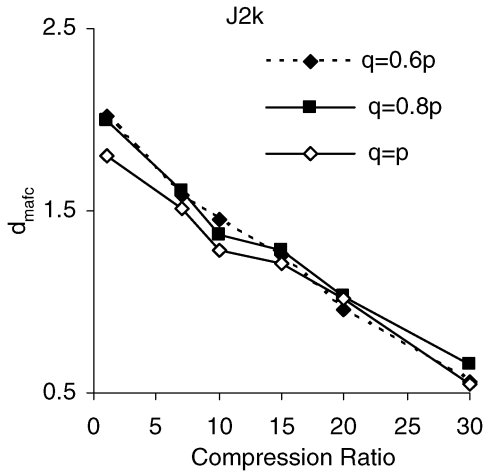


Fig. 9. CHO model performance as a function of JPEG 2000 compression ratios. Inhibitory component with constant $p = 3$. $q = 0.6p$, $q = 0.8p$, $q = p$.

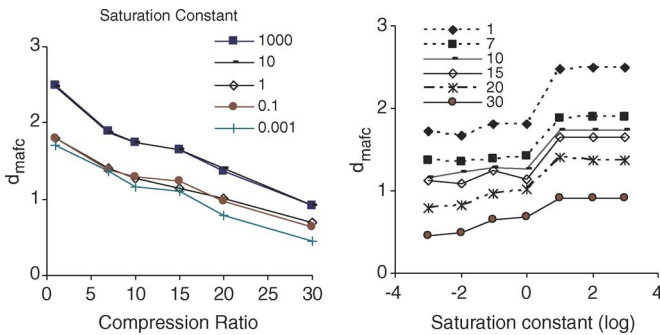


Fig. 10. Saturation constant (JPEG 2000) with $p = q = 2$, and $\sigma_H = 0.8$. Left graph: CHO performance as a function of compression ratio for different saturation constant values. Right graph: CHO performance as a function of saturation constant values (in log scale) for different compression ratios.

pressive. Fig. 10 (left) shows that the model observer performance increases with the increasing of the saturation constant (ranging from 0.001 to 1000). The right graph of Fig. 10 shows the model observer performance at different compression ratios as a function of saturation constant (log scale). For our specific test, model observer performance saturated when the saturation constant was 10 or larger for the following parameters: $p = q = 2$, and $\sigma_H = 0.8$.

3) *Inhibitory Convolution Kernel Size*: Convolution kernel size determines the bandwidth of the inhibitory mechanism. We varied the convolution parameter σ_H [see (13)] from 0.4 to 1.2. The left graph of Fig. 11 shows that the model performance did not vary greatly with the variation of convolution parameter σ_H . We also implemented inhibitory pooling only across orientations (without the frequency pooling). The right graph of Fig. 11 shows the model observer performance with orientation inhibition only. For the tested three groups of parameters ($p = 2, q = 0.6p$; $p = 2, q = 0.8p$; $p = 2, q = p$), the CHO performance resulted in small differences.

G. Human Observer Results

1) *Nonlinear Model Observer Performance versus Human Observer Performance*: In this section, we compare the model

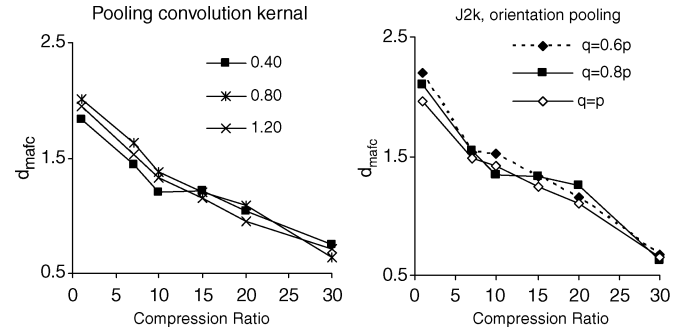


Fig. 11. Left: Inhibitory with different convolution Gaussian kernel standard deviations (σ_H varies from 0.4 to 1.2). Right: Inhibitory on orientation only with parameters $p = 2$, $q = 0.6p$, $q = 0.8p$, and $q = p$.

observer performance with the average human observer (AO) performance. The average human observer performance was obtained from four observers. Error bars in figures are the standard error of the mean across nine experimental sessions. The left column of Fig. 12 gives the comparison between two different algorithms JPEG and JPEG 2000 and the right column presents the comparison between two different encoder settings of JPEG 2000: default and optimal [8]. The first two rows show the CHO performance with divisive inhibitory nonlinearity with two sets of nonlinear parameters: 1) $p = q = 2$, $b = 0.5$, and $\sigma_H = 0.8$; 2) $p = 2$, $q = 0.8p$, $b = 0.5$, and $\sigma_H = 0.8$. The model observer performance correctly predicts the human observer performance rank order for two different compression algorithms: JPEG over JPEG 2000, and two different JPEG 2000 encoder settings: optimized over default. In addition, with the second group of parameters (the right graph on second row), the model observer performance approximated very well the human observer performance for the different encoder settings of JPEG 2000. The third row presents the CHO model performance with excitatory nonlinearity ($p = 2$). The fourth row shows the linear CHO performance. Again, the linear CHO performance predicts the human performance rank order very well although the absolute value of the model observer performance is higher than the absolute performance of averaged human observer. Typically internal noise is used to degrade linear models to the performance levels of human observers [6], [17], [53].

2) *Nonlinearity versus Internal Noise*: Fig. 13 compares two CHO models with different sources of inefficiency [nonlinearities (NL) versus internal noise [54] (IN)] to human performance. Here, internal noise was implemented by adding a random internal noise term to the scalar decision variables with the internal noise's standard deviation proportional to the decision variable's standard deviation [7], [54]. The proportionality constant was iteratively adjusted to match model and human performance at the uncompressed condition. The selected proportionality constant was then used to add internal noise to the model observer's performance at all compression conditions. For the default JPEG 2000 encoder setting, the mean square errors between the performance predicted by the model and the average human observer were 0.18 and 0.29 for the NL model and IN model respectively. For the optimized JPEG 2000 encoder setting, the corresponding mean square errors were 0.07 (NL model) and 0.14 (IN model).

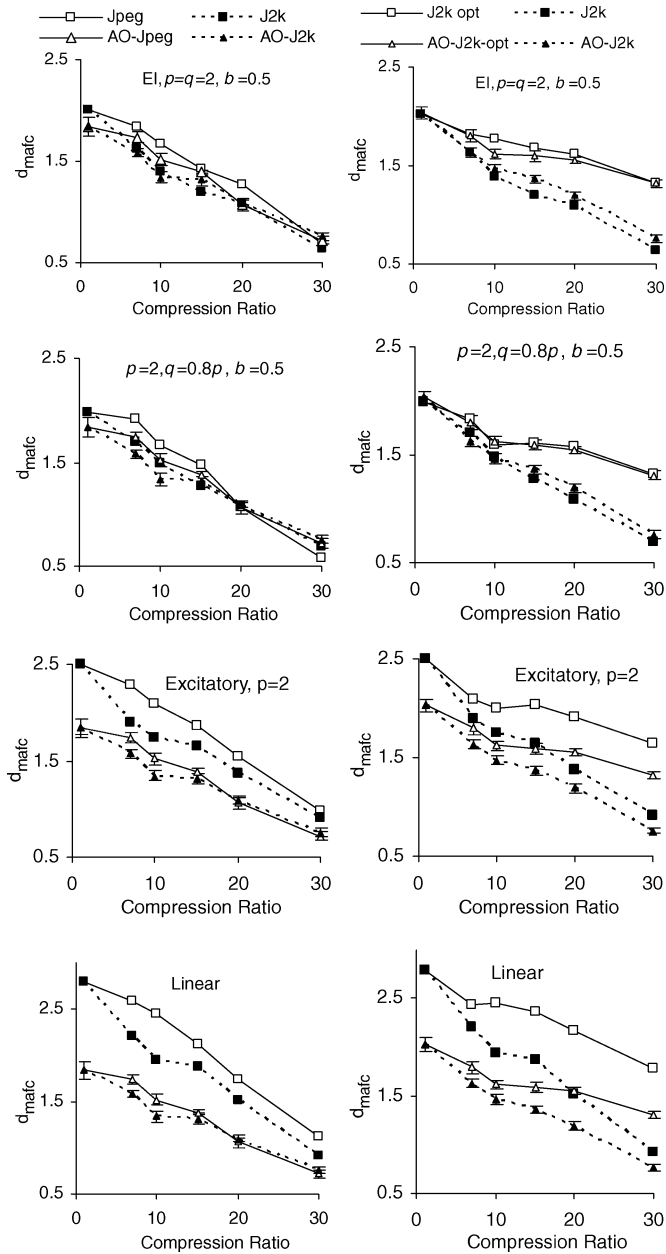


Fig. 12. Model and average human observer (AO) performance comparison. Left column: JPEG versus JPEG 2000 (J2k). Right column: JPEG 2000 default (J2k) versus optimal (J2k_opt). First row: Excitatory inhibitory (EI) nonlinearity with parameter $p = q = 2$, $b = 0.5$, $\sigma_H = 0.8$. Second row: EI nonlinearity with parameter $p = 2$, $q = 1.8$, $b = 0.5$, $\sigma_H = 0.8$. Third row: Excitatory with $p = 2$. Fourth row: Linear model observer performance.

V. SUMMARY AND DISCUSSION

A. Linear versus Nonlinear Models

Although linear model observers have been successfully used to predict visual detection and discrimination in noisy backgrounds, it has been correctly argued that they fall short of a complete account of early vision. Teo and Heeger [21] pointed out that one major problem with the linear model is the fact that V1 cell responses saturate at high contrast [55]. Moreover, a typical V1 neuron responds vigorously to its preferred orientation but not at all to the perpendicular orientation. According to the linear model, the response to the superimposed pair of stimuli

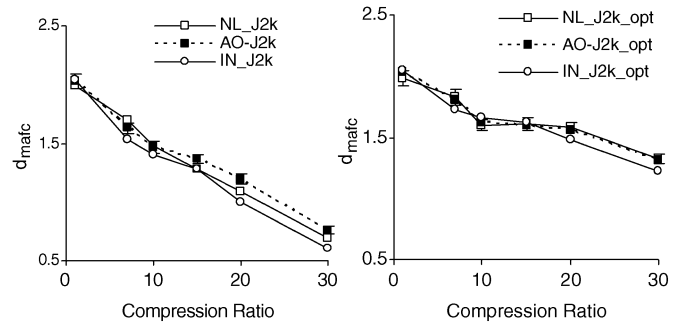


Fig. 13. Nonlinearity (NL) and internal noise (IN): JPEG 2000 (J2k) default and optimized (J2k_opt). Dotted line is the average human observer performance. Solid line with triangle symbols is the performance of model observer with internal noise. Solid line with square symbols is the model observer performance with nonlinearity ($p = 2$, $q = 0.8p$, $b = 0.5$, and $\sigma_H = 0.8$).

(preferred signal plus high-contrast perpendicular mask) should equal the response to the preferred stimulus presented alone. In fact, the response to the superimposed pair is about half that predicted [56], a phenomenon known as cross-orientation inhibition. In addition, the linear models fail to predict degradation in human performance in the presence of deterministic masks. On the other hand, many of the nonlinear models proposed to predict detection in one of two identical backgrounds (or masks [20]–[22]) are incapable of achieving performance levels comparable to humans in detection in noise tasks because they lack decision rules that make use of knowledge about the signal and the noise. It has been suggested previously that in order to account for both literatures models should include both nonlinear components and decision rules that make use of knowledge about the signal and noise [57], [58]. It seems plausible that to account for both sets of experiments (deterministic masking and noise masking), models along those proposed in Fig. 1 including both nonlinearities and decision strategies that make use of the properties of the signal and noise are needed.

However, the scope of the current paper was narrower. We investigated how the inclusion of nonlinearities affected performance in the CHO model for detection in noise. Table III summarizes the parameters we studied and their effect on the model observer performance. We use the symbol “ \uparrow ” to denote up, “ \downarrow ” to denote down, and “ \updownarrow ” to denote up and down.

The nonlinear response operator raises the excitatory input to a power, and divides it by the inhibitory input plus a constant to produce the response. The decision variable is computed by combining the difference in response to signal-plus-background and background alone across the three nonlinear operators [59]. We found that the CHO model performance degraded once the nonlinear components are included. This lower performance might be related to the fact that the distribution of decision variable departs greatly from the equal variance Gaussian assumption after the nonlinearity making suboptimal the method to derive the optimal weights from mean and a single covariance which assumes equal variance. Aside from the overall performance degradation with the inclusion of the nonlinear stages, our results showed that for the present task, both the simpler linear and nonlinear models seem to capture how human performance varied across and within compression schemes. Furthermore, from a practical point of view the

TABLE III
SUMMARY OF THE EFFECT OF THE PARAMETERS INVESTIGATED ON PERFORMANCE

Mechanism	Location uncertainty	Excitatory nonlinearity	Inhibitory nonlinearity			
Parameter	Area (from 1×1 to 9×9)	p (from 2 to 3.5)	p, q combination		Saturation Constant b (from .001 to 1000)	Convolution kernel σ_H (from 0.4 to 1.2)
			$q=2,$ p from 2 to 3.5	$p=3$ q from $0.6p$ to p		
Parameter change	↑	↑	↑	↑	↑	↑
Performance	↓	↓	↕	↕	↑	↕

decision about which compression scheme (JPEG versus JPEG 2000) preserves task performance the best would not be altered if one used the fully linear model or any of the nonlinear models investigated. This suggests that, for our image set, the nonlinearity components of the model were not a major factor. A possible future aim could be to investigate other nonlinear methods to combine the channel outputs after the nonlinear stages.

In addition, we also investigated whether including a front-end in the CHO model reflecting some of the early limitations in visual processing due to optics and receptor sampling [32] affected model performance. Previous work had included a front-end in the CHO model but did not investigate whether its presence had an effect on model performance [31]. Our result showed that for the viewing distances and signal sizes used in the present study, inclusion of the front-end results in very small differences in the CHO model performance. Note also that the models were able to predict the psychophysical data independent of whether the input image to the model was expressed in gray level, luminance, or contrast.

B. Degrading Model Performance for Comparison to Human Performance: Internal Noise versus Nonlinearity

Linear model observers typically outperform humans. Because researchers are interested in directly comparing model observer with human performance, internal noise is typically used [6], [17], [53] with model observers to degrade the model's performance to levels comparable to humans. Our previous results [8] illustrated that the inclusion of internal noise did not change the rank order conclusions about the compression conditions/algorithms. The present results showed that nonlinearities investigated in the present work might also be used to degrade model observer performance to levels comparable to human observers. Furthermore, our results indicated that the channelized Hotelling with a number of nonlinear components was a slightly better predictor of human performance than the addition of internal noise (Fig. 13).

C. Measures of Model Performance

Under the assumption that the model response is Gaussian distributed and the standard deviation of the model responses is assumed to be the same for signal-present background and

background only locations, respectively, model observer performance can be calculated directly in terms of mean and standard deviation of the decision variable [d' , (17)]. Mean and standard deviation of the decision variable can be calculated through Monte Carlo simulations (i.e., this paper) or through analytic propagation methods [19], [48]. In contrast, proportion correct (P_c) can be directly calculated from model observer performance based on the template responses without making any assumptions about equal variance and/or response correlations. For the X-ray coronary backgrounds in our study, results showed that the Gaussian equal variance decision variable assumption was violated. Furthermore, the unequal variance increased with the presence of the nonlinearities. In addition, the Gaussian decision variable assumption also fails. On the other hand, the decision variable correlations across locations were not significantly different than zero. Together, the results emphasize the inadequacy of estimating performance using the index of detectability based on means and a single standard deviation of the response variable in the presence of some nonsimulated backgrounds and/or model nonlinearities.

VI. CONCLUSION

We investigated the effect of the inclusion of an optics preprocessing stage, location uncertainty, nonlinear excitatory transducer, and divisive channel inhibition into the CHO on performance of detecting a variable signal embedded in X-ray angiographic backgrounds as a function of image compression ratios. Our results showed that for the task and images investigated, the simpler linear CHO model observer resulted in higher absolute detection performance than the nonlinear CHO model. Inclusion of the optics and/or nonlinear components did not change the rank order of model performance for two compression algorithms: JPEG always rendered higher performance than the JPEG 2000 algorithm for compression ratios less than 15:1, and for two JPEG 2000 encoder settings: the optimized encoder setting outperformed the default one. The results might suggest that the inclusion of the location uncertainty and the addition of the physiologically based channel nonlinearity to a channelized Hotelling add complexity to the model observers without great impact on automated medical image evaluation. Together these results might guide future researchers in deciding what components to include in their model observers. However, further research should test whether the current findings hold for

other imaging modalities, tasks, and/or image processing algorithms.

APPENDIX

A. Simulation of Arteries

Let image $I(x, y)$ be composed of a primary component due to the primary beam $P(x, y)$ and a secondary component due to scattered rays $S(x, y)$ then $I(x, y) = P(x, y) + S(x, y)$. $P(x, y)$ is related to the incident rays (I_0) and the densities of the projected patient (D_p) and contrast filled artery (D_a) by

$$P(x, y) = I_0 \exp(-D_p(x, y)) \exp(-D_a(x, y)). \quad (\text{A.1})$$

The simulated object is inserted so that the primary component of the image needs to be attenuated by the density of the simulated object ($\exp(-D_s(x, y))$), $P_s(x, y) = P(x, y) \exp(-D_s(x, y))$, where $P_s(x, y)$ is the primary including the simulated object. The density of the simulated object is given by $D_s = \mu t_s(x, y)$ where μ is the attenuation coefficient and $t_s(x, y)$ is the thickness of the contrast filled simulated object traversed by the X-rays as a function of the image position. The process of embedding a simulated object into a real angiographic image can be expressed by

$$I_s(x, y) = [I(x, y) - S(x, y)] \exp(-D_s(x, y)) + S(x, y) \quad (\text{A.2})$$

where $I_s(x, y)$ is the image after the simulated object is inserted. The secondary component $S(x, y)$ can be approximated according to [60]

$$S(x, y) = I(x, y) * \alpha h(x, y) + \beta \quad (\text{A.3})$$

where $\alpha = 0.483$, and $\beta = 7.69$. The convolution kernel is given by $h(x, y) = 20 \exp(-A|x - x_0|) \exp(-B|y - y_0|)$, where $A = B = 2 \log 2 / \sqrt{75}$.

The reduction in the density of the artery due to the filling defect is given by

$$D_t(x, y) = -\kappa 2.0 \sqrt{r^2 - [(x - x_0)^2 + (y - y_0)^2]} \quad (\text{A.4})$$

where x_0 and y_0 are the locations of the center of the filling defect, r is the radius, and κ is a contrast factor.

ACKNOWLEDGMENT

The authors would like to thank C. K. Abbey and J. M. Foley for valuable discussions and suggestions.

REFERENCES

- [1] A. E. Burgess, R. F. Wagner, R. J. Jennings, and H. B. Barlow, "Efficiency of human visual signal discrimination," *Science*, vol. 214, pp. 93–94, 1981.
- [2] A. E. Burgess and H. Ghandeharian, "Visual signal-detection. II. Signal-location identification," *J. Opt. Soc. Amer. A*, vol. 1, pp. 906–910, 1984.
- [3] J. Yao and H. H. Barrett, "Predicting human performance by a channelized Hotelling observer model," in *Proc. SPIE*, D. C. Wilson and J. N. Wilson, Eds., 1992, vol. 1768, pp. 161–168.
- [4] J. P. Rolland and H. H. Barrett, "Effect of random background in homogeneity on observer detection performance," *J. Opt. Soc. Amer. A*, vol. 9, pp. 649–658, May 1992.
- [5] A. E. Burgess, X. Li, and C. K. Abbey, "Visual signal detectability with two noise components: Anomalous masking effects," *J. Opt. Soc. Amer. A*, vol. 14, pp. 2420–2442, Sept. 1997.
- [6] 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. Amer. A*, vol. 18, pp. 473–488, Mar. 2001.
- [7] M. P. Eckstein, J. L. Bartroff, C. K. Abbey, J. S. Whiting, and F. O. Bochud, "Automated computer evaluation and optimization of image compression of X-ray coronary angiograms for signal known exactly detection tasks," *Opt. Express*, vol. 11, pp. 460–475, Mar. 2003.
- [8] Y. Zhang, B. T. Pham, and M. P. Eckstein, "Automated optimization of JPEG 2000 encoder options based on model observer performance for detecting variable signals in X-ray coronary angiograms," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 459–474, Apr. 2004.
- [9] L. Y. Chen and H. H. Barrett, "Task-based lens design with application to digital mammography," *J. Opt. Soc. Amer. A*, vol. 22, pp. 148–167, Jan. 2005.
- [10] H. C. Gifford, M. A. King, D. J. de Vries, and E. J. Soares, "Channelized hotelling and human observer correlation for lesion detection in hepatic SPECT imaging," *J. Nucl. Med.*, vol. 41, pp. 514–521, Mar. 2000.
- [11] K. J. Myers, J. P. Rolland, H. H. Barrett, and R. F. Wagner, "Aperture optimization for emission imaging—effect of a spatially varying background," *J. Opt. Soc. Amer. A*, vol. 7, pp. 1279–1293, July 1990.
- [12] K. J. Myers, "Ideal Observer Models of Visual Signal Detection," in *Handbook of Medical Imaging*. Bellingham, WA: SPIE Press, 2000, vol. 1, pp. 559–592.
- [13] 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*, vol. 213, pp. 727–734, Dec. 1999.
- [14] A. E. Burgess, F. L. Jacobson, and P. F. Judy, "Human observer detection experiments with mammograms and power-law noise," *Med. Phys.*, vol. 28, pp. 419–437, Apr. 2001.
- [15] Y. Zhang, B. T. Pham, and M. P. Eckstein, "Evaluation of JPEG 2000 encoder options: Human and model observer detection of variable signals in X-ray coronary angiograms," *IEEE Trans. Med. Imag.*, vol. 23, no. 5, pp. 613–632, May 2004.
- [16] S. Suryanarayanan, A. Karellas, S. Vedantham, S. M. Waldrop, and C. J. D'Orsi, "Detection of simulated lesions on data-compressed digital mammograms," *Radiology*, vol. 236, pp. 31–36, Jul. 2005.
- [17] M. P. Eckstein, A. J. Ahumada, and A. B. Watson, "Visual signal detection in structured backgrounds. II. Effects of contrast gain control, background variations, and white noise," *J. Opt. Soc. Amer. A*, vol. 14, pp. 2406–2419, Sep. 1997.
- [18] B. R. Beutter, M. P. Eckstein, and L. S. Stone, "Saccadic and perceptual performance in visual search tasks. I. Contrast detection and discrimination," *J. Opt. Soc. Amer. A*, vol. 20, pp. 1341–1355, Jul. 2003.
- [19] Z. L. Lu and B. A. Doshier, "Characterizing human perceptual inefficiencies with equivalent internal noise," *J. Opt. Soc. Amer. A*, vol. 16, pp. 764–778, Mar. 1999.
- [20] J. M. Foley, "Human luminance pattern-vision mechanisms—Masking experiments require a new model," *J. Opt. Soc. Amer. A*, vol. 11, pp. 1710–1719, Jun. 1994.
- [21] P. C. Teo and D. J. Heeger, "Perceptual image distortion," *Proc. SPIE*, B. E. Rogowitz and J. P. Allebach, Eds., vol. 2179, pp. 127–139, 1994.
- [22] A. B. Watson and J. A. Solomon, "Model of visual contrast gain control and pattern masking," *J. Opt. Soc. Amer. A*, vol. 14, pp. 2379–2391, Sep. 1997.
- [23] J. Lubin, "The use of psychophysical data and models in the analysis of display system performance," in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp. 163–178.
- [24] S. Daly, "The visible differences predictor: an algorithm for the assessment of image fidelity quality," in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp. 179–206.
- [25] D. J. Heeger, "Normalization of cell responses in cat striate cortex," *Vis. Neurosci.*, vol. 9, pp. 181–197, Aug. 1992.

- [26] A. M. Rohaly, A. J. Ahumada, and A. B. Watson, "Object detection in natural backgrounds predicted by discrimination performance and models," *Vision Res.*, vol. 37, pp. 3225–3235, Dec. 1997.
- [27] W. B. Pennebaker and J. L. Mitchell, *JPEG Still Image Data Compression Standard*. New York: Van Nostrand Reinhold, 1993.
- [28] M. Rabbani and R. Joshi, "An overview of the JPEG 2000 still image compression standard," *Signal Process. Image Commun.*, vol. 17, pp. 3–48, 2002.
- [29] K. J. Myers and H. H. Barrett, "Addition of a channel mechanism to the ideal-observer model," *J. Opt. Soc. Amer. A*, vol. 4, pp. 2447–2457, Dec. 1987.
- [30] S. Marcelja, "Mathematical description of the responses of simple cortical-cells," *J. Opt. Soc. Amer. A*, vol. 70, pp. 1297–1300, 1980.
- [31] M. P. Eckstein and J. S. Whiting, "Lesion detection in structured noise," *Acad. Radiol.*, vol. 2, pp. 249–253, Mar. 1995.
- [32] W. S. Geisler and K. D. Davila, "Ideal discriminators in spatial vision: Two-point stimuli," *J. Opt. Soc. Amer. A*, vol. 2, pp. 1483–1497, Sep. 1985.
- [33] F. W. Campbell and R. W. Gubisch, "Optical quality of the human eye," *J. Physiol.*, vol. 186, pp. 558–578, 1966.
- [34] F. W. Campbell and J. G. Robson, "Application of Fourier analysis to visibility of gratings," *J. Physiol. (Lond)*, vol. 197, pp. 551–566, 1968.
- [35] M. B. Sachs, J. Nachmias, and J. G. Robson, "Spatial-frequency channels in human vision," *J. Opt. Soc. Amer. A*, vol. 61, pp. 1176–1186, 1971.
- [36] N. Graham and J. Nachmias, "Detection of grating patterns containing two spatial frequencies—Comparison of single-channel and multiple-channels models," *Vision Res.*, vol. 11, pp. 251–259, 1971.
- [37] A. E. Burgess, "Visual signal detection with two-component noise: Low-pass spectrum effects," *J. Opt. Soc. Amer. A*, vol. 16, pp. 694–704, Mar. 1999.
- [38] M. P. Eckstein, C. K. Abbey, and F. O. Bochud, "A practical guide to model observers for visual detection in synthetic and natural noisy images," in *Handbook of Medical Imaging*. Bellingham, WA: SPIE Press, 2000, vol. 1, pp. 593–628.
- [39] M. P. Eckstein and C. K. Abbey, "Model observers for signal known statistically tasks," *Proc. SPIE*, E. A. Krupinski and D. P. Chakraborty, Eds., vol. 4324, pp. 91–102, 2001.
- [40] H. H. Barrett, J. Yao, J. P. Rolland, and K. J. Myers, "Model observers for assessment of image quality," *Proc. Nat. Acad. Sci.*, vol. 90, pp. 9758–9765, Nov. 1993.
- [41] N. Graham, *Visual Pattern Analyzers*. Cambridge, U.K.: Oxford Univ. Press, 1989.
- [42] J. G. Daugman, "Uncertainty relation for resolution in space, spatial-frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Amer. A*, vol. 2, pp. 1160–1169, 1985.
- [43] J. P. Jones and L. A. Palmer, "An evaluation of the two-dimensional Gabor filter model of simple receptive-fields in cat striate cortex," *J. Neurophysiol.*, vol. 58, pp. 1233–1258, Dec. 1987.
- [44] W. P. Tanner, "Physiological implications of psychophysical data," *Ann. NY Acad. Sci.*, vol. 89, pp. 752–765, 1961.
- [45] D. G. Pelli, "Uncertainty explains many aspects of visual contrast detection and discrimination," *J. Opt. Soc. Amer. A*, vol. 2, pp. 1508–1532, 1985.
- [46] R. M. Manjeshwar and D. L. Wilson, "Effect of inherent location uncertainty on detection of stationary targets in noisy image sequences," *J. Opt. Soc. Amer. A*, vol. 18, pp. 78–85, Jan. 2001.
- [47] D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*. New York: Wiley, 1966.
- [48] F. O. Bochud, C. K. Abbey, and M. P. Eckstein, "Visual signal detection in structured backgrounds. III. Calculation of figures of merit for model observers in statistically nonstationary backgrounds," *J. Opt. Soc. Amer. A*, vol. 17, pp. 193–205, Feb. 2000.
- [49] C. A. Morioka, M. P. Eckstein, J. L. Bartroff, J. Hausleiter, G. Aharanov, and J. S. Whiting, "Observer performance for JPEG versus wavelet image compression of X-ray coronary angiograms," *Opt. Express*, vol. 5, pp. 8–19, Jul. 1999.
- [50] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. New York: Chapman & Hall, 1993.
- [51] R. F. Wagner, H. P. Chan, B. Sahiner, N. Petrick, and J. T. Mossoba, "Components of variance in ROC analysis of CADx classifier performance: II. Applications of the bootstrap," *Proc. SPIE*, K. M. Hanson, Ed., pp. 523–532, 1999.
- [52] T. S. Meese and D. J. Holmes, "Adaptation and gain pool summation: Alternative models and masking data," *Vision Res.*, vol. 42, pp. 1113–1125, Apr. 2002.
- [53] A. E. Burgess and B. Colborne, "Visual signal-detection. IV. Observer inconsistency," *J. Opt. Soc. Amer. A*, vol. 5, pp. 617–627, Apr. 1988.
- [54] Y. Zhang, B. T. Pham, and M. P. Eckstein, "Evaluation of internal noise methods for Hotelling observers," *Proc. SPIE*, vol. 5749, pp. 162–173, 2005.
- [55] D. G. Albrecht and D. B. Hamilton, "Striate cortex of monkey and cat—Contrast response function," *J. Neurophysiol.*, vol. 48, pp. 217–237, 1982.
- [56] A. B. Bonds, "Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex," *Vis. Neurosci.*, vol. 2, pp. 41–55, 1989.
- [57] M. P. Eckstein, A. J. Ahumada, and A. B. Watson, "Image discrimination models predict visual detection in natural medical image backgrounds," *Proc. SPIE*, B. E. Rogowitz and T. N. Pappas, Eds., vol. 3016, pp. 44–56, 1997.
- [58] M. P. Eckstein, A. J. Ahumada, Jr., A. B. Watson, and J. S. Whiting, "What is degrading human visual detection performance in natural medical image backgrounds?," *Proc. SPIE*, H. L. Kundel, vol. 3036, pp. 50–64, 1997.
- [59] C. C. Chen, J. M. Foley, and D. H. Brainard, "Detection of chromoluminance patterns on chromoluminance pedestals II: Model," *Vision Res.*, vol. 40, pp. 789–803, 2000.
- [60] L. A. Love and R. A. Kruger, "Scatter estimation for a digital radiographic system using convolution filtering," *Med. Phys.*, vol. 14, pp. 178–185, Mar. 1987.