Derivation of an observer model adapted to irregular signals based on convolution channels

Ivan Diaz, Craig K. Abbey, Pontus A. S. Timberg, Miguel P. Eckstein, Francis R. Verdun, Cyril Castella, and François O. Bochud.

Abstract— **Anthropomorphic model observers are mathematical algorithms which are applied to images with the ultimate goal of predicting human signal detection and classification accuracy across varieties of backgrounds, image acquisitions and display conditions. A limitation of current channelized model observers is their inability to handle irregularly-shaped signals, which are common in clinical images, without a high number of directional channels. Here, we derive a new linear model observer based on convolution channels which we refer to as the "Filtered Channel observer" (FCO), as an extension of the channelized Hotelling observer (CHO) and the nonprewhitening with an eye filter (NPWE) observer. In analogy to the CHO, this linear model observer can take the form of a single template with an external noise term. To compare with human observers, we tested signals with irregular and asymmetrical shapes spanning the size of lesions down to those of microcalfications in 4-AFC breast tomosynthesis detection tasks, with three different contrasts for each case. Whereas humans uniformly outperformed conventional CHOs, the FCO observer outperformed humans for every signal with only one exception. Additive internal noise in the models allowed us to degrade model performance and match human performance. We could not match all the human performances with a model with a single internal noise component for all signal shape, size and contrast conditions. This suggests that either the internal noise might vary across signals or that the model cannot entirely capture the human detection strategy. However, the FCO model offers an efficient way to apprehend human observer performance for a non-symmetric signal.**

Index Terms—**Image quality assessment, model observers, optimization**

I. INTRODUCTION

IMAGE quality in medical imaging has been defined at the technical level based on image parameters, such as contrast, power spectrum, modulation transfer function, etc. An alternative approach has been to frame image quality in terms of the task for which the image is intended [1][2]. For example, in mammography an image of "high" quality would be one which maximizes the physician's ability to detect the signal (e.g. tumors, microcalcifications) under the condition of an acceptable dose delivered to the patient. Modern medical imaging systems are complex and have many different parameters which are optimized often only subjectively. The large set of possible system parameters is generally too extensive to fully test within a reasonable time frame and budget using psychophysical observer performance studies.

To address this issue, there has been a considerable effort to develop mathematical anthropomorphic model observers in order to mimic the response of a human observer. These model observers allow researchers to test higher numbers of parameter configurations in an efficient manner assuming they can accurately predict human performance. The ultimate goal would be to find a standard model observer which is recognized to be reliable in mimicking human performance in detection tasks for many different types of patient backgrounds and pathologies. The construction of such a model is still very much an open question.

To date, there have been a variety of model observers proposed for signal detection which vary in usefulness depending on the type of signal, the statistics of the background, the computational time, and the framework used, such as forced choice or Receiver Operating Characteristic (ROC) curve analysis. One very successful class of model observers are known as Channelized Hotelling observers (CHOs) [2,3], which are an extension of the Hotelling observer that is tractable when the target profile and background statistics are presumed known or can be estimated [4]. CHO models use channels as a set of linear responses, which are then combined to form a decision variable, using Fisher-Hotelling weights. The channels themselves are generally based on circularly symmetric profiles such as Difference of Gaussian (DOG) functions [5, 6] or Laguerre-Gauss weighted polynomials, although more general channel

models have been suggested [7].

The performance of these models depends heavily on the choice of channels. Previous studies have shown that a model observer with circularly symmetric functions tends to underperform with respect to human observers when signals are not circularly symmetric [8]. This is problematic considering the fact that signals typically encountered in practice are not circularly symmetric. The problem can, in principle, be overcome using asymmetric channel functions [9]. However, this requires many channels and this makes the estimation and inversion of the covariance matrix much more challenging.

Another widely used model observer is the Non-Prewhitening observer with an eye filter [10], traditionally denoted NPWE, which adds a contrast sensitivity function to a simple matched filter. This observer uses the signal profile (assumed to be known) modulated by a model of the contrast sensitivity of the human visual system as a detection template. The NPWE model is considered to be non-prewhitening because it makes no use of image pixel covariance statistics (i.e. prewhitening) to form the weights of its decision variable. It extends naturally to irregularly shaped signals since the signal profile is used to generate the detection template. The NPWE observer is however not adequate in all situations because it has been repeatedly demonstrated that human observers do adapt their detection strategies to covariance between pixels [11] in some cases. This means that human observers are performing some sort of prewhitening operation. Furthermore, Abbey and Eckstein [12] have shown evidence that models of visual detection must contain some mechanism for incorporating multiple filters if they are to be used with varying magnitudes of image noise.

In this context, there is a need for a model observer that: 1) can be applied to the detection and classification of irregular shaped signals; 2) can use background statistics to adjust the weighting of different spatial frequencies (i.e., prewhitening); and 3) is computationally tractable. The present article introduces a model that fulfills these requirements. We derive a filtered channel model observer (FCO) which includes the CHO and the NPWE observers as special cases, and can detect irregular non-circularly symmetric signals. The proposed observer uses multiple filters in a way that matches the computations of the CHO models, but it creates each channel filter using bandpass filtering of an arbitrarily shaped signal that is similar to the NPWE model.

We derive the channel model from the perspective that a channel represents a whole class of linear receptive fields at different spatial locations that are tuned to particular spatial frequencies. Extending the model to be selective to both frequency and orientation selectivity, common in models of receptive fields of simple cells in the early visual system [13]-[15], is straightforward, although we do not find it necessary to pursue orientation selectivity here. We then show how the responses from these many features can be combined into a decision variable in a form that mimics the CHO computation.

As an initial validation of the FCO model, we evaluate its performance in a standard task detecting a rotationally symmetric Gaussian "bump" target embedded in white Gaussian noise. This allows us to verify that it is consistent with human-observer performance in traditional studies. We then evaluate an irregularly shaped target embedded in noise that models image variability in digital breast tomosynthesis.

II. THEORY

Let $g(x, y)$ represent the image intensity as a function of horizontal and vertical pixel locations, $x = 0, \ldots, N_x - 1$ and $y = 0, \ldots, N_y - 1$. We will consider a generic simple detection task, in which the signal-absent images consist of a noise field:

$$g(x, y) = n(x, y), \qquad (1)$$

and the signal-present images consist of a fixed (nonrandom) signal profile s added to the noise:

$$g(x, y) = n(x, y) + s(x, y). \qquad (2)$$

In terms of discrete spatial frequency variables (u, v) the 2D Fourier transform is defined as

$$\hat{g}(u, v) = \frac{1}{N_x N_y} \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} g(x, y) e^{-2\pi i \left(\frac{ux}{N_x} + \frac{vy}{N_y} \right)}. \qquad (3)$$

We use the caret symbol (^) to indicate the Fourier Transform and an overbar (‾) to indicate the complex conjugate of a function. We will also make extensive use of Parseval's relation, which equates inner products in the spatial domain and Fourier domains, for a normalized Fourier transform. For $f(x, y)$ and $g(x, y)$ functions with Fourier transforms $\hat{f}(u, v)$ and $\hat{g}(u, v)$, then by Parseval's relation,

$$\sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} \overline{f(x, y)} g(x, y) = N_x N_y \sum_{u=0}^{N_x-1} \sum_{v=0}^{N_y-1} \overline{\hat{f}(u, v)} \hat{g}(u, v). \qquad (4)$$

We note that discrete Fourier transforms generally assume periodic boundary conditions, which results in non-physical "wrap-around" effects in correlations and convolutions. We will assume that image sizes are large enough that boundary effects can be neglected.

A. The standard "channelized" approach

The standard CHO uses a single receptive field, sometimes called a channel template, to represent the channel. For a given set of channels, indexed by k ($k = 0, \cdots, K-1$), the standard channelized model specifies the channels as a set of spatial (or potentially spatiotemporal) channel templates $t_k(x, y)$. The channel responses are given by a noisy inner product of the template with the image,

$$\lambda_k = \sum_{y=0}^{N_y-1} \sum_{x=0}^{N_x-1} t_k(x,y) g(x,y) + \varepsilon_k, \qquad (5)$$

where ε_k represents a random perturbation of internal noise in the response. When used to model the human observer, ε_k is typically specified to be a zero-mean Gaussian random variable with variance σ_k^2. Internal noise is generally presumed to be independent across channels. When used to approximate the original Hotelling observer, internal noise is not used (i.e. $\sigma_k^2 = 0$).

The individual channel responses are combined into a decision variable by a weighted sum with weights α_k,

$$\lambda = \sum_{k=1}^{K} \alpha_k \lambda_k = \mathbf{\alpha}^T \mathbf{\lambda} \qquad (6)$$

where the second equality comes from defining K-element column vectors, $\mathbf{\alpha}$ and $\mathbf{\lambda}$, composed of the α_k and λ_k terms respectively, and the superscript T indicates the transpose operation. For the standard CHO, the channel weights are determined by the mean response to the target and the covariance matrix of the channel responses. The mean response s_k of the k-th channel to the signal is given by

$$s_k = \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} t_k(x,y) s(x,y) \qquad (7)$$

In general, we will assemble s_k into a K-element vector, $\mathbf{s}$. The noisy $K \times K$ covariance matrix of the channel responses, $\mathbf{\Sigma}_\lambda$, is related to the continuously defined image covariance matrix, $\Sigma_g(x,y;x',y')$, by the sum

$$[\mathbf{\Sigma}_\lambda]_{k,k'} = \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} t_k(x,y) \sum_{x'=0}^{N_x-1} \sum_{y'=0}^{N_y-1} t_{k'}(x,y) \Sigma_g(x,y;x',y') + \delta_{k,k'} \sigma_k^2 \qquad (8)$$

Note that the Kronecker delta, $\delta_{k,k'}$, in this equation means the internal noise variance is only added to the diagonal elements of the covariance matrix. When image variability can be approximated as stationary with discrete power spectrum $N(u,v)$, this expression simplifies to the following by means of the Wiener–Khinchin theorem:

$$[\mathbf{\Sigma}_\lambda]_{k,k'} = \sum_{u=0}^{N_x-1} \sum_{v=0}^{N_y-1} \overline{\hat{t}_k(u,v)} \hat{t}_{k'}(u,v) N(u,v) + \delta_{k,k'} \sigma_k^2 \qquad (9)$$

When the statistics of the images are not well known, N or $\mathbf{\Sigma}_\lambda$ may be estimated from samples.

The channelized Hotelling observer, defined by the channel templates and internal noise variances, specifies the channel weights according to

$$\mathbf{\alpha} = \mathbf{\Sigma}_\lambda^{-1} \mathbf{s} \qquad (10)$$

With this scheme for weighting and using Gaussian assumptions for the statistical distribution of the images, the detectability index for detecting $s(x,y)$ is

$$d' = \sqrt{\mathbf{s} \mathbf{\Sigma}_\lambda^{-1} \mathbf{s}} . \qquad (11)$$

It is also worth noting that Abbey and Bochud showed an equivalent implementation of the channelized Hotelling observer which formulates a single template [16],

$$t_{\text{CHO}}(x,y) = \sum_{k=0}^{K-1} \alpha_k t_k(x,y) \qquad (12)$$

and a single internal noise component, ε, with variance

$$\sigma_\varepsilon^2 = \sum_{k=0}^{K-1} \sigma_k^2 \alpha_k^2 . \qquad (13)$$

The channelized Hotelling observer can thus be thought of as a way to synthesize a linear template with a limited ability to adapt to the statistics of the images. We will show that the model developed in this paper, the FCO observer, takes a similar form, but with a broader concept of a channel.

B. Filtered channel response model.

Let us now assume that a channel consists of a large set of linear feature responses tuned to a specific spatial frequency profile, with a feature centered at every location (x,y) in the image. The responses associated with channel k, analogous to Equation (5), are given in the array $r_k(x,y)$, which can be now thought of as a convolution of the image with a channel profile, $c_k(x,y)$. If we also consider internal noise in the responses as Gaussian white noise added to each response, $\gamma_k(x,y)$, the response function can be defined as

$$r_k(x,y) = \sum_{x'=0}^{N_x-1} \sum_{y'=0}^{N_y-1} c_k(x-x', y-y') g(x',y') + \gamma_k(x,y) \qquad (14)$$

In order to perform simple detection tasks, the response functions must be transformed into a scalar decision variable, λ. As in Equation (6), we assume that this is accomplished by a linear summation process using weighting functions w_k,

$$\lambda = \sum_{k=0}^{K-1} \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} w_k(x,y) r_k(x,y) \qquad (15)$$

We will describe the form of these weighting functions next, but we note at this point that the channel model is determined by c_k and w_k, and by the statistical properties of the internal noise fields.

C. Channel weights

We will now make a strong assumption about the channel weights that define our model. We propose that each channel weighting function, $w_k(x,y)$, is the product of a spatial summation component, $m_k(x,y)$, and a channel-dependent weight, α_k, that is independent of location,

$$w_k(x,y) = \alpha_k m_k(x,y). \qquad (16)$$

Our assumption is that prewhitening operates between channels not within them. This means that the α_k values are determined in part by noise properties of the images, whereas the $m_k(x,y)$ are not,

We define $m_k(x,y)$ as the convolution of the signal and the channel profile, defined most conveniently in the Fourier domain as

$$\hat{m}_k(u,v) = \hat{c}_k(u,v)\hat{s}(u,v) \qquad (17)$$

Using Parseval's relation we can rewrite Equation (15) as,

$$\lambda = N_x N_y \sum_{k=0}^{K-1} \alpha_k \sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1} \overline{\hat{s}(u,v)}\left[\left|\hat{c}_k(u,v)\right|^2 \hat{g}(u,v) + \overline{\hat{c}_k(u,v)}\hat{\gamma}_k(u,v)\right] \qquad (18)$$

Now let us separate each element of the sum into components related to the image and the internal noise. Equation (17) suggests that the image related component for each channel can be written,

$$\sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1} \overline{\hat{s}(u,v)}\left|\hat{c}_k(u,v)\right|^2 \hat{g}(u,v) \equiv \sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1} \overline{\hat{t}_k(u,v)}\hat{g}(u,v) \qquad (19)$$

where $\hat{t}_k(u,v) = \left|c_k(u,v)\right|^2 s(u,v)$ is a single template rather than the spatial array of feature responses specified in Equation 14. Parseval's relation also allows this component to be computed in the spatial domain by the inner product

$$\sum_{x=0}^{N_x-1}\sum_{y=0}^{N_y-1} t_k(x,y)g(x,y) \qquad (20)$$

The computation in Equation 15 is now directly analogous to the Equation 5, ignoring internal noise for the moment. The difference is that the profile of $t_k(x,y)$ is determined by modulating the frequencies of the signal with the power spectrum of the channel profile.

The internal noise component in Equation 18 is dependent on the statistical properties of $\gamma_k(x,y)$. If we assume γ_k is

an array of white Gaussian noise with variance $\sigma^2_{\gamma,k}$, then the internal noise component variance is given by

$$\sigma_k^2 = \sigma^2_{\gamma,k}\sum_{x=0}^{N_x-1}\sum_{y=0}^{N_y-1} w_k^2(x,y) = \sigma^2_{\gamma,k} N_x N_y \sum_{u=0}^{N_x-1}\sum_{v=0}^{N_y-1}\left|\hat{c}_k(u,v)\right|^2\left|\hat{s}(u,v)\right|^2 \qquad (21)$$

Equations 19 and 21 show that the channel-dependence of the spatial template is completely captured by power spectrum of the channel profile, $\left|\hat{c}_k(u,v)\right|^2$. It is also worth noting that the model is completely analogous to the standard Channelized model in Equation 5, that is, each scalar response is given by a noisy inner product of the template with the image. We have shown how a set of linear channel responses within each channel can be equivalently implemented as a single noisy linear template under the assumption that the channel features are weighted based on their response to the signal profile.

We propose that the final step of determining the cross-channel weights, α_k, in Equation 16 also be directly analogous to the standard template model, with channel weights given in the fashion of Equation 10. In Figure 1, we outline the procedure necessary to obtain the model observer templates for a standard CHO and show the difference with the FCO observer in Figure 2.
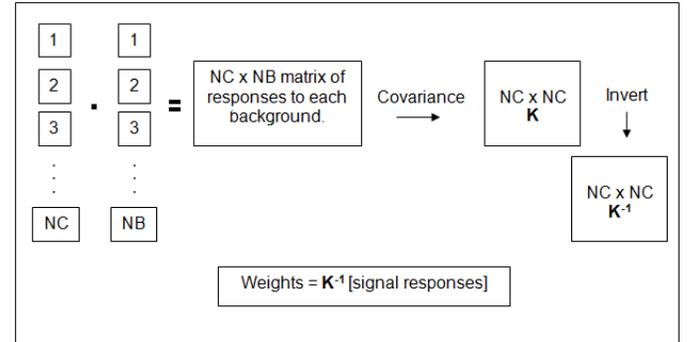

Figure 1 Graphic outline of the procedure to calculate a 'standard' channelized Hotelling observer. NC and NB stand for the number of channels and number of background images, respectively. K is the covariance matrix of the backgrounds.
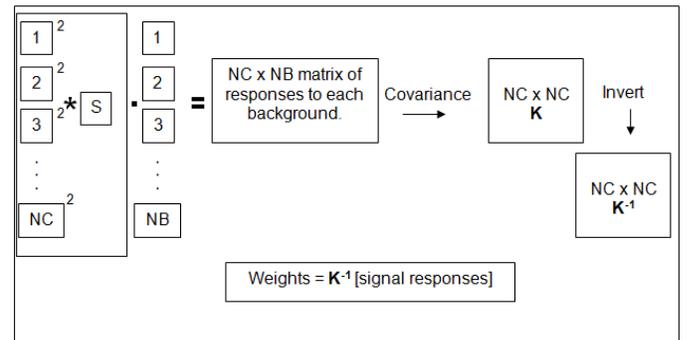

Figure 2 Graphic outline of the procedure to calculate the Filtered Channel observer (FCO). The only difference with the standard channelized Hotelling observer is that the NC channels are convolved with the signal, S, prior to taking the dot product.

For the CHO, an NCxNB matrix containing the responses of each of the NC channels to each of the NB backgrounds is created. The covariance of this matrix is calculated to yield an NCxNC matrix which is then inverted. The weights are then given by the product of this inverted matrix by an NCx1 vector of responses of the channels to the signal. The final template is the weighted sum of the original NC channels.

In the FCO observer, the channels are originally convolved with the signal and it is this set of images that is weighted. Thus the main difference is that the final template depends on the signal itself and is the reason why the overall performance is potentially higher than the standard CHOs for asymmetric signals.

D. Relationship to other model observers

Now we consider two special cases in order to show the relationship of this model with respect to other standard models. If we take the signal to be an impulse at a point (x_s, y_s) (i.e., $s(x, y) = \delta_{x_s,x}\delta_{y_s,y}$), the image component of Equation 19 is reduced to

$$\lambda_k = N_x N_y \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} \left|\hat{c}_k(u,v)\right|^2 \hat{g}(u,v) \tag{22}$$

This is simply the definition of the 'standard' channelized Hotelling observer in the Fourier domain. Thus the standard CHO results when the FCO observer is tuned to a delta function as a target.

A second special case occurs when we have only one channel ($k = 0$) in Equation 18. In this situation, we may assume that $\alpha_0 = 1$ without loss of generality, and the equation describes the convolution of a single filtered signal with the image:

$$\lambda = N_x N_y \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} s(u,v)\left|\hat{c}_0(u,v)\right|^2 \hat{g}(u,v). \tag{23}$$

When $\hat{c}_0(u,v)$ is taken to be the profile of the human contrast sensitivity function, the above equation describes the well-known Eye-filtered nonprewhitening observer model (NPWE) proposed by Burgess [10]. Thus the NPWE model results from the FCO model when it is allowed only a single channel with a somewhat different interpretation.

III. COMPARISON WITH HUMAN OBSERVERS

A. Radially-symmetric signal.

Before testing the FCO observer on an irregular signal, we evaluated its performance for a very simple case of a 2D Gaussian "bump" signal embedded in Gaussian white noise. The images had a mean background intensity of 128 gray levels, and the noise had a standard deviation of 28 gray levels. The image was 400x400 pixels and the signal had a FWHM of 19 pixels with an amplitude of 6 gray levels. Six observers then performed a 2-alternative forced choice (2-AFC) Signal-Known-Exactly detection task with 600 trials.

Five different model observers were then calculated two CHO observers, two FCO observers, and one NPWE. For the calculation of the FCO and CHO observers, two difference-of-

Gaussians (DOG) models were chosen. These were chosen to consider dense and sparse channels. The first contained 12 channels concentrated over low spatial frequencies (DOG-L) while the second contained 10 channels and was spread from very low spatial frequencies to approximately half the Nyquist frequency (DOG-H), as shown in Fig 3. Table 1 shows a summary of the results. As expected for a signal of this spatial frequency, the NPWE observer performed the worst, but still at a level higher than the humans. Figure 4 shows the normalized radial profile of the signal as well as the profile of the CHO and FCO observers for one set of DOG channels.

Table 1. Model observer performance for a 2-AFC task with a Gaussian, radially symmetric signal. The CHO and FCO implementations performed equally high. Uncertainties are standard errors with $n = 10$.

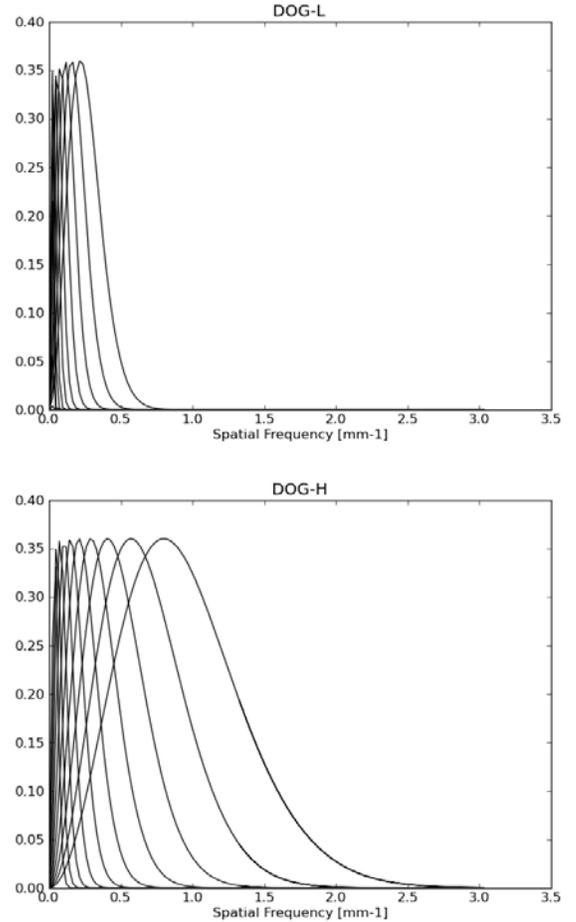| Observer	Percent Correct
Humans	80 ± 2
NPWE	90.0 ± 0.3
DOG-L CHO	98.0 ± 0.2
DOG-L FCO	98.2 ± 0.1
DOG-H CHO	98.0 ± 0.1
DOG-H FCO	98.2 ± 0.2

Fig. 3. Difference-of-Gaussian channels used for the calculation of the CHO and FCO observer. The top set of channels, designated "DOG-L" (for "low") has 12 channels clustered in the low spatial frequencies. The bottom set of channels, designated "DOG-H" (for "high") has 10 channels which span a higher range of spatial frequencies. On both figures the Nyquist frequency corresponds to 3.5 mm-1.
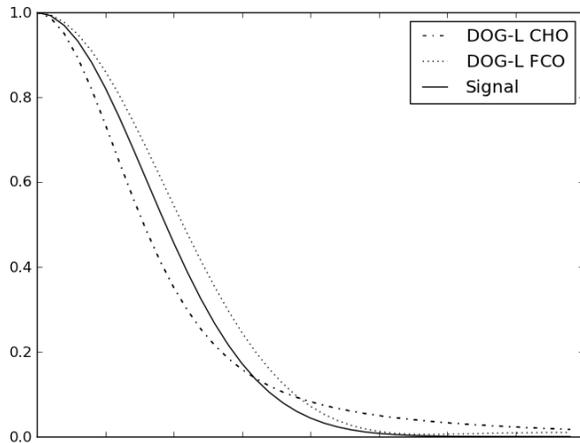
Figure 4 Normalized radial profile of the signal as well as the profile of the CHO and FCO observers calculated with DOG-L channels (spatial domain).

B. Non-symmetric signal

A natural consequence of the spatial weights described in Eq 17, is the ability to generate an irregularly shaped template in response to an irregularly shaped target. This should generate better performance in detecting such targets. Previous studies [8] have illustrated the limitations of rotationally symmetric detection templates. To verify the performance of the FCO approach for irregularly shaped nonsymmetrical targets, we ran it on the same simulated 4-AFC detection task used previously to evaluate performance in digital breast tomosynthesis.

The paradigm used in this study was the Signal Known Exactly but Varying (SKEV) task (varying in shape and size). In each trial four images are shown to the observer, next to the signal itself as a reference. Only one of these images contains the signal, and the observer is instructed to choose which of the four contains the signal. As an example, model observer templates are shown in Figure 5 for one the irregular targets used.

The observers used were a DOG-L and DOG-H FCO observer with the same channels as those of Figure 3, and an NPWE observer. Human observer performance results were reported in a previous study [8]. A Siemens prototype breast tomosynthesis unit was used to acquire images from 30 patients. During the acquisition, 25 projection images, evenly distributed over an angular span of 50 degrees, were collected. Only every other projection image was used for reconstruction because the original human trials were also meant to be compared to digital mammography, and this omission kept the corresponding average glandular doses (AGD) at the same level (i.e an AGD of 0.8 mGy for a 50 mm standard breast). For each patient, signal-absent areas were chosen from the reconstructed slices, yielding about 600 background images. Realistic lesions were then projected randomly into 60

backgrounds per trial, under the assumption of a constant attenuation coefficient difference, $\Delta\mu$. It is important to note that this is not a simple linear addition of a signal directly onto the final 3D image. Rather, the additional attenuation of the simulated lesion modified each projection before reconstruction. However, for the calculation of the model observers the signal was treated as additive. Four different lesion sizes were used, with average diameters of 0.2 mm, 1 mm, 8 mm and 25 mm. Three different contrast levels were used for each size. The two smaller sizes are meant to mimic calcifications, which are relevant in the early detection of breast cancer [17]. The contrast levels were chosen in order to give a percent correct between 70-95% in the 4-AFC experiments.

IV. RESULTS

Figure 6 shows the performance of the human and model observers. With the exception of the NPWE observer, all other model observers were calculated with the extra convolution step. These model observers matched or outperformed the human observers in 4-AFC trials for all but one signal size. The sole exception was that of a set of DOG-L channel observer searching for the 200µm microcalcifications. This is due to the fact that for this model all of the channels are clustered in the low frequencies while the microcalfication signals, which spanned just a few pixels, require channels in higher frequencies than are absent in the DOG-L set.
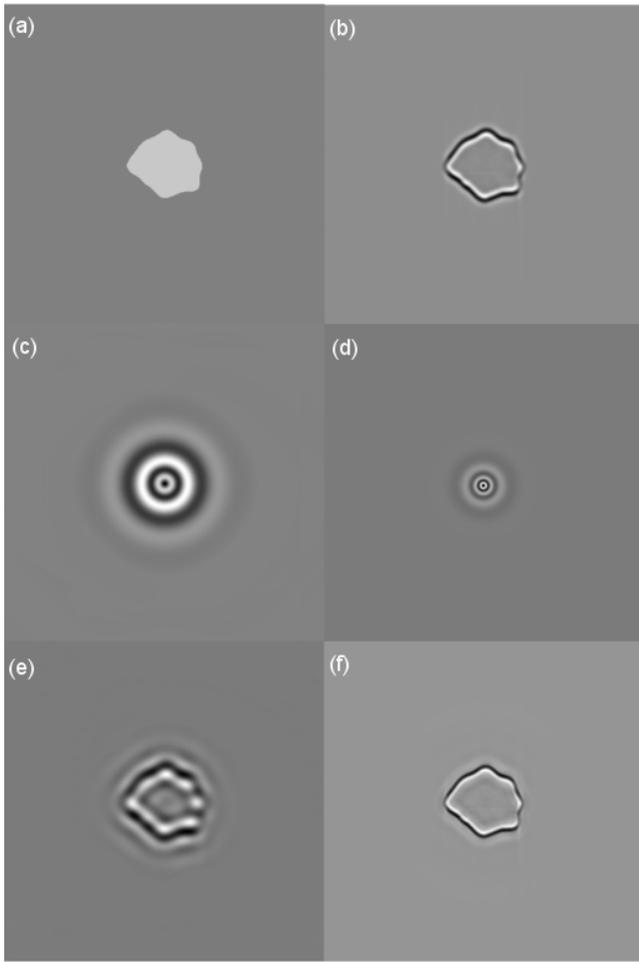
Fig. 5. Examples of model observer templates in 2D representation for an irregular signal. (a) Original signal, (b) NPWE, (c) DOG-L CHO, (d) DOG-H CHO, (e) DOG-L FCO, (f) DOG-H FCO

The high performance of the DOG-FCO observers (DOG-L and DOG-H) of the model observers was degraded by the addition of internal noise in order to match the performance of the human observers. The DOG-FCO observers had the same DOG-L channels as those of Figure 3. To assess the amount of internal noise needed to match the humans, we minimized the chi-squared residuals between the humans and model observers by using different values of the external noise proportionality constant, p_n. The method of minimizing the chi-squared residuals worked well when minimizing across different contrasts for each one of the four available mass sizes individually. However, simultaneous minimization across all sizes and contrasts pooled together resulted in poor fits. Thus we have chosen to only optimize across different contrasts for each signal size, leading to 4 different internal noise proportionality constants for each set of channels chosen (see Table 2).
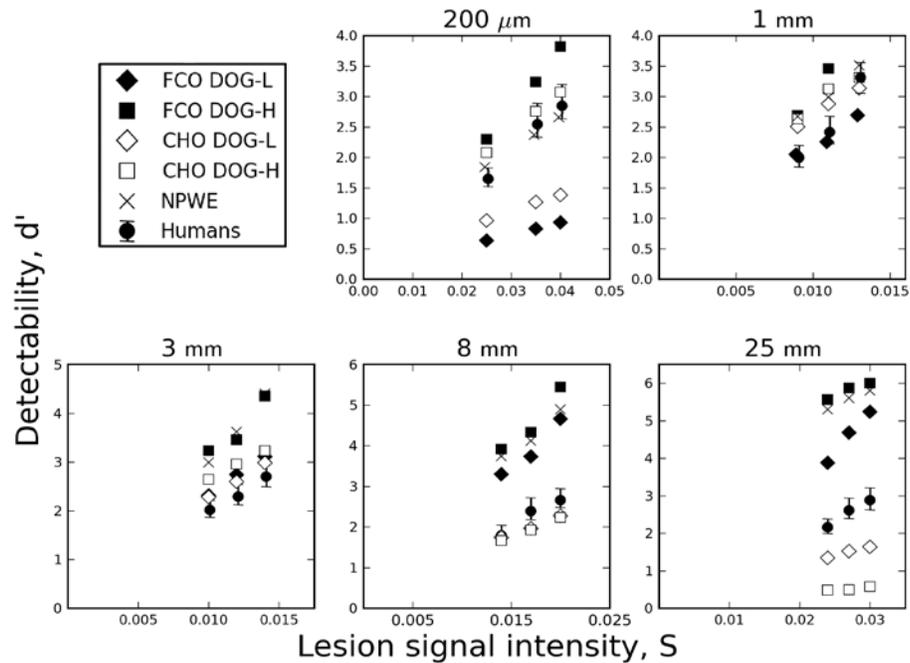

Fig. 6. Model observer and human performance. Two sets of Difference-of-Gaussians channels were used for the FCO observer, one with channels bunched up in the low frequencies (DOG-L), and one spread over higher frequencies (DOG-H). The corresponding CHO observers were also calculated and the NPWE observer is shown for comparison.

Table 2. Amount of internal noise necessary to match human performance for FCO observer. Internal noise is proportional by the constant to the variance in the external noise. Noise was only added when the model observer performance was higher than the human performance.

Channel set	p_n				
	200 μm	1 mm	3 mm	8 mm	25 mm
NPWE	-	0.8	1.4	4.4	7.1
DOG-L FCO	-	0.0	0.4	2.2	3.9
DOG-H FCO	0.8	0.9	1.7	4.6	7.7

V. Discussion

We proposed a viable model observer for irregularly shaped signals. This model can be fine-tuned to match the performance of human observers.

The derivation of the FCO is based on the notion that a channel is composed of a collection of linear feature responses tuned to specific spatial frequencies. This broader conception of a channel is more consistent with its use in vision science [13]-[15] and communication theory [18] than what has been used for the standard CHO, which equates the channel with a single filter response. As shown in Section 2.D, this occurs for the FCO only in the special case of a signal defined as a delta function. The fundamental assumption of the model is that the relative weigh of each filter within a channel is based on how well it responds to the target, and a limited form of prewhitening occurs between channels. This also makes the FCO particularly well adapted for tasks with irregular signals compared to previously developed model observers that are based on rotationally symmetric channel templates. Many of the CHOs used in the literature have been implemented for spherical and otherwise very simple symmetric signals. This is an oversimplification of what a human observer will encounter in a clinical setting and moving to asymmetrical signals of different sizes is a logical step in establishing a more realistic model of the human observer process.

We expect channelized models to be better at mimicking human observers under certain circumstances since there is evidence for spatial frequency-selective channels in the human visual system. Choosing a set of channels with frequency peaks from the very low to very high frequency domain will ensure that the template calculation procedure will optimize the channel weights in order to maximize the performance similar to the way humans perform detection tasks. The entire concept of model observers is, of course, a great simplification of the complex and not fully understood process of decision-making in humans, but we reiterate that the ultimate goal is to find a model which is predictive of human observers. This requires minimizing the complexity of the model while taking into account the relevant components of the visual system.

The FCO is a generalization of both the CHO and NPWE models. As shown in the derivation, the FCO reduces to a standard CHO when the signal is a delta function and to an NPWE observer when there is only one channel that corresponds to the contrast sensitivity function. The combination of the NPWE observer with the channel approach provides a more robust model and keeps the computational advantages when compared to the original non-channelized Hotelling observer.

In a previous study [5] performed with conventional CHOs, a set of channels which were clustered in the lowest frequencies was used. This led to good results for the larger signals, but performed very poorly for the smallest signals. This is clear from the fact that in order to detect small signals, one needs to recruit the high frequency channels which this model was lacking. These results were replicated in this study with the set of low frequency channels. We are generally looking for performance from models that is the equal of or better than the human observers. Performing worse than humans is immediately a cause for rejection from consideration as a candidate for a viable model to be used in a larger scale study since it does not allow for the addition of internal noise, a known characteristic of human performance.

However, for the other sets of channels (DOG-H) we were able to achieve a performance superior to that of the human observers. This result alone is significant when compared to the poor performance obtained by every single symmetric CHO considered in the previous study. The advantage of the high detectability rates that were obtained with the observers considered in this study is that they could be brought down to the performance of human observers. There is hope that internal noise and other inefficiencies may explain the difference. In this study we had a data set which included several different signal sizes and contrast levels and we first tried to estimate the internal noise by minimizing the chi-squared residuals between the human and model observer performance across all conditions. Unfortunately, the agreement was significantly poorer than when we did it for each size individually. This is a problem since it was part of our rationale for not recommending the NPWE observer. However, despite being similar in this sense to the NPWE observer, we point out that this observer takes more into account the workings of the human visual system. While the NPWE only requires a function with a peak corresponding to the human contrast sensitivity function, this model can incorporate any number of channels which might be more suitable to the way humans respond to visual stimuli.

Thus, we conclude that at this point it is best to match the human observer performance by using an internal noise which is different for each size. The key assumption is that having a lower proportionality factor necessary for matching human performance is indicative of a detection strategy on the part of the model observer that is closer to that of a human observer with human internal noise. A single model observer with a single proportionality constant for the addition of internal noise would be the goal, and there are still many combinations of channels and basis functions that could be tried in the future with this procedure to obtain that goal.

As mentioned earlier, breast tomosynthesis has a very large number of acquisition parameters and if these models are to be

used in automated optimization trials the computational aspects must be addressed. The code itself is procedural and straightforward to implement. The bottleneck in the generation of the templates comes when the covariance matrix is calculated from the available background images. All the templates can be calculated, 100 4-AFC experiments with 60 trials run, and internal noise optimized for each signal size in approximately 15 minutes with a standard desktop computer. This was the performance obtained when using all available backgrounds. Of course, since these observers are outperforming humans, the proportionality constant is iteratively changed to match the human observers. This process uses the variances from the responses to the backgrounds, and only needs to be calculated once. A second implementation was done with each of the 60 signals and was calculated using a leave-three-out calculation of the templates. This implies choosing three backgrounds at random. The templates are calculated from the remaining 597 backgrounds and the three original backgrounds were used as the three incorrect choices in one 4-AFC trial. This implementation takes considerably longer, on the order of 3 hours but the performance is equivalent.

A final comment is that this model uses global noise statistics to make its decision. However humans have been shown to adapt to local orientation properties in nonstationary backgrounds, which were used in this study [19]. Thus, one limitation of this model is that it is not able to adapt to anisotropic noise in the orientation domain. However, this can be remedied with orientation selective channels rather than the isotropic channels used here.

VI. CONCLUSION

We have derived the filtered channel observer model under the assumption that each channel consists of many receptive fields tuned to specific spatial frequencies, and that prewhitening occurs between channels and not within a channel. We have shown through the derivation that, in a similar fashion to the channelized Hotelling observer, the FCO observer can be implemented as a single template with an internal noise component. This renders the performance calculation very simple.

This model immediately improved the performance of the model observers with respect to the humans compared to a standard CHO with symmetric channels for the detection of a nonsymmetrical lesion embedded in image backgrounds. Model observer performance can be subsequently degraded to the level of human observers by the addition of internal noise. However, it is not clear it this point how to determine the amount of internal noise to be added in a prospective manner. This topic warrants further study. We believe that evaluating the FCO over a broader range of target sizes, shapes, and will aid in refining the model observers which are to be used in automating the optimization of breast tomosynthesis acquisition parameters, and other imaging systems.

REFERENCES

[1] Barrett, Harrison H. "Objective assessment of image quality: effects of quantum noise and object variability." *JOSA A* 7, no. 7 (1990): 1266-1278.
[2] K.J. Myers and H.H. Barrett, "Addition of a channel mechanism to the ideal-observer model," J. Opt. Soc. Am. A **4**, 2447-2457 (1987).
[3] B. Gallas, and H.H. Barrett, "Validating the use of channels to estimate the ideal linear observer," J. Opt. Soc. Am. A **20**, 1725-1738 (2003)
[4] 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. Volume 1. Physics and psychophysics*, J. Beutel, H. L. Kundel, R. L. Van Metter, eds. (SPIE press, 2000), pp. 593-628.
[5] C.K. Abbey, H.H. Barrett, "Linear Iterative Reconstruction Algorithms: Study of Observer Performance," in *Proc 14th Int Conf on Information Processing in Medical Imaging,* Y. Bizais, R. Di Paola, eds. (Kluwer Academic, 1995)
[6] S. Daly, "The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity," *Digital Images and Human Vision.* A.B. Watson, ed. (MIT Press, 1993)
[7] M.P. Eckstein, C.K. Abbey, J.S. Whiting, "Human Vs Model Observers in Anatomic Backgrounds," Proc SPIE 3340 (1998).
[8] C. Castella, M. Ruschin, M. Eckstein, C.K Abbey, K. Kinkel, F. Verdun, and A. Tinberg, "Masses detection in breast tomosynthesis and digital mammography: a model observer study, " Proc. SPIE (2009).
[9] M.P. Eckstein, A.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. Am. A **14**, 2406-2419 (1997)
[10] A. E. Burgess, "Statistically defined backgrounds: Performance of a modified non-prewhitening observer," J. Opt. Soc. Am. A **11**, 1237-1242 (1994).
[11] J. P. Rolland and H. H. Barrett, "Effect of random background inhomogeneity on observer detection performance," J. Opt. Soc. Am. A **9**, 649-658 (1992)
[12] C.K. Abbey and M.P. Eckstein, "Frequency tuning of perceptual templates changes with noise magnitude," J. Opt. Soc. Am. A **26**, B72-B83 (2009)
[13] Wilson, Hugh R., David K. McFarlane, and Gregory C. Phillips. "Spatial frequency tuning of orientation selective units estimated by oblique masking." *Vision research* 23.9 (1983): 873-882.
[14] De Valois, Russell L., Duane G. Albrecht, and Lisa G. Thorell. "Spatial frequency selectivity of cells in macaque visual cortex." *Vision research* 22.5 (1982): 545-559.
[15] Field, David J. "Relations between the statistics of natural images and the response properties of cortical cells." *J. Opt. Soc. Am. A* 4.12 (1987): 2379-2394.
[16] C. K. Abbey and F. O. Bochud, "Modeling Visual Detection Tasks in Correlated Image Noise," *Handbook of medical imaging. Volume 1. Physics and psychophysics*, J. Beutel, H. L. Kundel, R. L. Van Metter, eds. (SPIE press, 2000), pp. 646-650.
[17] W. Huda, K. M. Odgen, E. M. Scalzetti, D. R. Dance, and E. A. Bertrand, "How Do Lesion Size and Random Noise Affect Detection Performance in Digital Mammography?," *Acad. Radiol.* Vol 13, 2006.
[18] Gabor, D. (1946). Theory of communication. Part 1: The analysis of information. *Electrical Engineers-Part III: Radio and Communication Engineering, Journal of the Institution of, 93*(26), 429-441.
[19] Y. Zhang, C.K. Abbey, and M.P. Eckstein. "Adaptive detection mechanisms in globally statistically nonstationary-oriented noise," J. Opt. Soc. Am. A **23**, 1549-1558 (2006)