Channelized Hotelling and Human Observer Correlation for Lesion Detection in Hepatic SPECT Imaging

Howard C. Gifford, Michael A. King, Daniel J. de Vries, and Edward J. Soares

Department of Radiology, University of Massachusetts Medical School, Worcester; Department of Mathematics, College of the Holy Cross, Worcester; Department of Radiology, Brigham and Women's Hospital, Boston; and Department of Radiology, Harvard Medical School, Boston, Massachusetts

Mathematic "model" observers that predict human performance are of interest in medical imaging as substitutes in psychophysical studies. We have examined the correlations between human observers and several forms of the channelized Hotelling observer (CHO) for a tumor detection task with simulated SPECT liver images that were used to study the effects of scatter and scatter correction on detection. **Methods:** A receiver operating characteristic (ROC) study was devised to investigate the relative value of a scatter-subtraction strategy in SPECT imaging. The study used simulated images of the biodistribution of ^{99m}Tc-labeled FO23C5 anticarcinoembryonic antigen antibodies within the liver. Projection data for 3 separate tumor locations and 5 strategies for handling scatter were obtained using Monte Carlo software applied to an anthropomorphic phantom. The strategies were (a) perfect scatter rejection, (b) no scatter correction, (c) no scatter correction under an assumption of an elevated amount of scatter, (d) an energy-spectrum–based scatter compensation of the normal-scatter case (b), and (e) similar scatter compensation for the elevated-scatter case (c). Image reconstruction approximated current clinical procedures at the University of Massachusetts Medical School. Human performance for each combination of location and strategy was based on averaging the areas under the ROC curve for 7 individuals. A set of 15 signal-to-noise ratios (SNRs) was derived from these averages for comparison with SNRs for CHO models featuring constant-Q and difference-of-gaussian (DOG) filters. **Results:** The Spearman rank correlation coefficient was 0.92 (P = 0.000001) when comparing task performances for the average human and a constant-Q CHO using 4 square-profile channels. For the DOG version of the CHO, comparison with the average human found a coefficient of 0.84 (P = 0.00005). **Conclusion:** The significant positive correlations found between the rankings of the average human observer and the CHOs for our detection task indicate that a channelized model observer could eventually serve as a replacement for human observers. The specific CHO models we have used are best suited to screen for significant differences between strategies before a human psychophysical study.

Key Words: numeric observer; receiver operating characteristic analysis; lesion detection; scatter correction

J Nucl Med 2000; 41:514–521

Psychophysical studies with human observers are a standard means of making objective assessments of detection performance in medical imaging. However, a credible demonstration that some method of data acquisition or image processing is an improvement over another often requires an extensive series of time- and patience-consuming studies. Mathematic "model" observers that correlate with humans for clinically realistic tasks are being considered to expedite such studies (*1*).

A model observer is an equation that accepts a multivariate input (the image) and returns a scalar value (the test statistic) for comparison with a threshold value. This equation is defined by the statistical properties of the image, characteristics of the lesion, anatomic background, and system noise, which also largely determine the detection task difficulty for human observers.

Current research on model observers as predictors of human performance has centered on linear observers, which incorporate linear functions of the image pixel values. The channelized Hotelling observer (CHO) is a linear observer with the outstanding feature of a bank of band-pass prefilters that extracts a small number of image details. This number is much smaller than the number of image pixels and has a physiologic basis in the frequency-selective channels of the human visual system (*2*). The CHO has been shown to agree with humans for a variety of "signal-known-exactly" (SKE) detection tasks, for which the observer is told the tumor location and the only decision to be made for a given image is whether the tumor is actually present. Initial studies were of SKE-"background-known-exactly" (SKE-BKE) tasks, with images of signals embedded in filtered white noise (*3*). More realistic studies investigated SKE tasks with random, heterogeneous ("lumpy") backgrounds (SKE-RB) (*4*). The CHO has also correlated with human performance for ranking maximimum-likelihood expectation-maximization reconstructions as a function of iteration stopping point for both SKE-BKE and SKE-RB tasks (*5*). More recently, Burgess et al. (*6*) considered signal detection in

---

Received Nov. 24, 1998; revision accepted Jun. 21, 1999.

For correspondence or reprints contact: Howard C. Gifford, PhD, Department of Radiology, University of Massachusetts Medical School, 55 Lake Ave. N., Worcester, MA 01655.

images with quantified mixes of noise and slower-varying "structural background," and Eckstein et al. (*7*) studied lesion detection in actual clinical backgrounds with superimposed white noise.

For this study, we have investigated correlations between CHO and human performances for a detection task with simulated SPECT images, considering detection of "hot" tumors at known locations within the liver when varying the amount of scatter and applying scatter correction. Although the receiver operating characteristic (ROC) studies were of SKE-BKE format, the data acquisition and image processing were otherwise intended to approximate clinical conditions (*8*). Our goal was to develop a model observer sufficiently capable of predicting human performance in similar tasks to facilitate psychophysical experiments that are under way at our laboratory.

MATERIALS AND METHODS

Simulation Models

The ROC study used simulated images of the biodistribution of FO23C5 anticarcinoembryonic antigen antibodies labeled with ^{99m}Tc (*9*). The SPECT imaging system was modeled on a Prism 3000 (Picker International, Inc., Cleveland, OH) with low-energy, ultra-high-resolution, parallel-hole collimators, a circular radius of rotation of 21.5 cm, and an energy resolution of 9.4% full width at half maximum at 140 keV. A 20% photopeak window centered on 140 keV was subdivided into 15% upper and 5% lower windows for scatter-compensation purposes.

Projection data consisted of 128 × 128 images with a pixel width of 0.36 cm. The counts were obtained with the SIMIND Monte Carlo software (*10*), which permitted classification of the detected photons as either primary or scattered. The data were synthesized in 2 steps. Projections of the background antibody distribution were obtained from the Zubal anthropomorphic phantom (*11*). Projection sets of a 2.5-cm-diameter spherical tumor within the liver were simulated separately. Separate ROC studies were performed for each of 3 tumor sites, with the intention of exploring a range of task difficulties for the observers. These locations were determined on the basis of a preliminary ROC study (*8*). For each location, the tumor projections were scaled to produce a tumor-to-liver contrast ratio of 13% and then added to the projections of the background distribution to form tumor-present datasets. Independent Poisson noise realizations of a desired count level were then created.

Scatter-Correction Strategies

Scatter in SPECT data is attributed to photons whose points of emission are not consistent with the spatial information recorded by the detector. Scatter adds a low-frequency component to the data, with the dual effect of lowering both the relative magnitude of the Poisson noise and the contrast in the reconstructed images (*12*).

Scatter-correction algorithms are designed to improve image contrast, but do so at a cost of increased noise correlation and magnitude in the images. The dual-photopeak window (DPW) method is a subtraction method of scatter compensation that is representative of the class of subtraction scatter-correction methods. DPW estimates the scatter-to-total photon count ratio $R_{scatter}$ of the projection images on a pixel-by-pixel basis from the power-law expression:

$$R_{scatter} = a\,R_{low}^{b} + c, \qquad \text{Eq. 1}$$

where R_{low} is the ratio of the counts in the lower subwindow to the counts in the total photopeak window, and a, b, and c are coefficients determined through system calibration (*13,14*). The result is a noisy estimate of the scatter, which must be low-pass filtered before subtraction from the projection data. This filtering increases spatial correlations in the projections.

For each of the 3 tumor locations, 5 datasets were created by applying the following scatter-compensation strategies: (a) Primary: The ideal of perfect scatter rejection, implemented by considering only the primary photons in the data. This represents the best-case performance for scatter subtraction. Primary datasets consisted of 8.77 × 10^6 counts. (b) Scatter: A no-correction strategy, using the primary plus scattered photons with a scatter-to-primary count ratio or scatter fraction (SF) ranging from 0.4 to 0.5. This SF is typical for ^{99m}Tc. A scatter dataset consisted of 12.5 × 10^6 counts. (c) High-scatter: A no-correction strategy, using data with SFs raised to between 1.0 and 1.2. This simulated imaging with a radionuclide that produces elevated levels of scatter. In both no-correction strategies, the number of primary counts was held constant. For a high-scatter dataset, 21.82 × 10^6 counts were used. (d) DPW: DPW correction applied to the scatter dataset. (e) High-DPW: DPW correction applied to the high-scatter dataset. We refer to the primary, scatter, and high-scatter strategies as the non-DPW strategies, and to the DPW and high-DPW strategies as the DPW strategies.

Image Reconstruction

The 2-dimensional slices through the tumor centers were chosen for filtered backprojection reconstruction with a ramp filter. We applied 2-dimensional, fourth-order Butterworth prefiltering with a cutoff frequency of 0.25/cm to each dataset as apodization for the ramp filter (*8*). Postprocessing included multiplicative Chang attenuation correction (*15*) using a uniform attenuation map based on an elliptical approximation to the body outline, along with the zeroing of negative pixel values. The 128 × 128 images were bilinearly interpolated to 384 × 384 and then cropped to the central 256 × 256 pixels (0.12-cm pixel width). As a final step, gray-scale mapping was used that scaled the mean liver pixel value to mid gray-scale level. The tumor placements within the liver, numbered 1, 2, and 3, are shown in the noise-free images of Figure 1.

Human ROC Studies

For each location–strategy combination, a set of 20 training images and a set of 80 study images were prepared. Both sets were equally divided between tumor-present and tumor-absent images. Images from all 3 locations were then combined to produce a training set of 60 images and a study set of 240 images for each strategy. These were read by each observer in 2 separate sessions featuring 30 training images and 120 study images. With 5 strategies, the ROC study thus consisted of 10 such sessions. For each observer, the sessions were scheduled in a different, nonrandom order. The reading order of the images within a session was randomized for each observer. Seven members of the medical physics research group at the University of Massachusetts Medical School participated as observers.

An observer's performance for a particular location and strategy was computed as an area under the ROC curve, A_z (*16*). Overall human performance as a function of location and strategy was

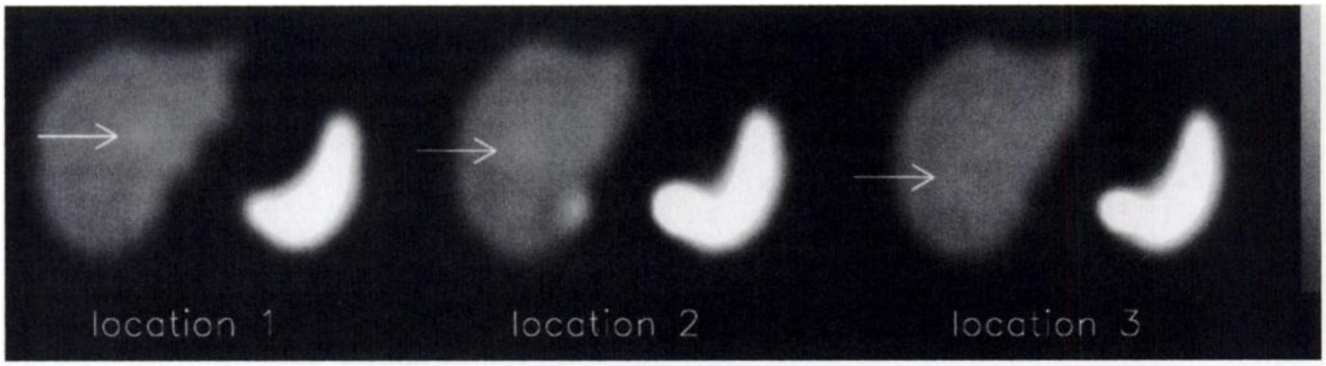

FIGURE 1. In noise-free images, arrows indicate tumor locations (1, 2, and 3, left to right). ROC study was run for each location.

measured as the average area $\overline{A}_z$ among observers. For comparison with the model observers, a signal-to-noise ratio (SNR) was determined for each $\overline{A}_z$ by the relation (*17*):

$$SNR_{human} = 2\,\mathrm{erf}^{-1}(2\overline{A}_z - 1), \qquad \text{Eq. 2}$$

where $\mathrm{erf}^{-1}(\cdot)$ denotes the inverse of the error function:

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt. \qquad \text{Eq. 3}$$

Model Observers

A model observer is a mathematic equation that computes a scalar test statistic on the basis of the pixel values of an image and renders a decision by comparing the statistic to a threshold value λ_c. Assume that vector $\mathbf{f}$ has as elements f_j the pixel values of an arbitrary 256 × 256 image in the ROC study. A linear model observer computes the statistic $\lambda(\mathbf{f})$ from the formula:

$$\lambda(\mathbf{f}) = \sum_{j=1}^{256^2} w_j f_j, \qquad \text{Eq. 4}$$

where w_j is the importance the observer assigns to the j^{th} pixel value. Equation 4 can also be viewed as the scalar product $\mathbf{w}^\dagger \mathbf{f}$ of the image and an observer template image $\mathbf{w}$ with w_j as the j^{th} pixel value. The term $\mathbf{w}^\dagger$ represents the transpose of $\mathbf{w}$.

The SKE-BKE task for a single tumor location describes a binary hypothesis test for which we assume that $\mathbf{f}$ is drawn at random from either a lesion-absent ensemble of images H_0 or a lesion-present ensemble H_1. (The probability $p(H_i)$ that $\mathbf{f}$ comes from the i^{th} ensemble is fixed by the task definition. Our study images were evenly divided between lesion-present and lesion-absent, so $p(H_0)$ and $p(H_1)$ were both 0.5.) If $\lambda(\mathbf{f})$ exceeds the decision threshold λ_c, H_1 is selected. Otherwise, the choice is H_0. The function notation $\lambda(\mathbf{f})$ emphasizes that the test statistic is dependent on the particular test image and is therefore a random variable. Assume that $p(\lambda|H_i)$ is the conditional probability distribution of $\lambda(\mathbf{f})$ for the i^{th} ensemble. For the model observer to be effective, the separation between the distributions $p(\lambda|H_0)$ and $p(\lambda|H_1)$ should in some sense be large. When this is the case, λ_c can be set so that the number of incorrect decisions will be small (Fig. 2). One measure of this separation is the SNR (*18*):

$$SNR = \left[\frac{[\langle\lambda\rangle_1 - \langle\lambda\rangle_0]^2}{p(H_1)\,\mathrm{var}(\lambda)_1 + p(H_0)\,\mathrm{var}(\lambda)_0}\right]^{1/2}, \qquad \text{Eq. 5}$$

where $\langle\lambda\rangle_i$ and $\mathrm{var}(\lambda)_i$ are the conditional mean and variance, respectively, of λ for $\mathbf{f} \in H_i$.

CHO Models

As implied in our description of model observers, an observer template for the scalar product of equation 4 is determined by the statistical properties of the ensembles H_0 and H_1. We let $\bar{\mathbf{f}}_i$ be the statistical mean image of ensemble H_i, and take K_i to be the $256^2 \times 256^2$ noise covariance matrix for H_i, as defined by the equation:

$$K_i = \langle[\mathbf{f} - \bar{\mathbf{f}}_i][\mathbf{f} - \bar{\mathbf{f}}_i]^\dagger\rangle_{\mathbf{f}\in H_i}. \qquad \text{Eq. 6}$$

The notation $\langle\cdot\rangle_{\mathbf{f}\in H_i}$ indicates an average over the images in H_i.

The CHO is related to a pair of well-known linear discriminants from signal detection theory. One of these, the nonprewhitening (NPW) observer, performs tumor detection by adopting the mean reconstructed tumor:

$$\mathbf{w}_{npw} = \bar{\mathbf{f}}_1 - \bar{\mathbf{f}}_0, \qquad \text{Eq. 7}$$

as an observer template (*18*). This is the optimal observer (in the sense of maximizing the SNR of equation 5) for SKE-BKE tasks when the images have uncorrelated Gaussian noise.

Another standard observer is the prewhitening (PW) matched filter observer, which precedes the matched filter step with a prewhitening operation. Prewhitening seeks to decorrelate noise in an image through a multiplication by the inverse of the noise covariance matrix (*18*). The PW template is:

$$\mathbf{w}_{pw} = \frac{1}{2}\,[K_0 + K_1]^{-1}\,[\bar{\mathbf{f}}_1 - \bar{\mathbf{f}}_0]. \qquad \text{Eq. 8}$$

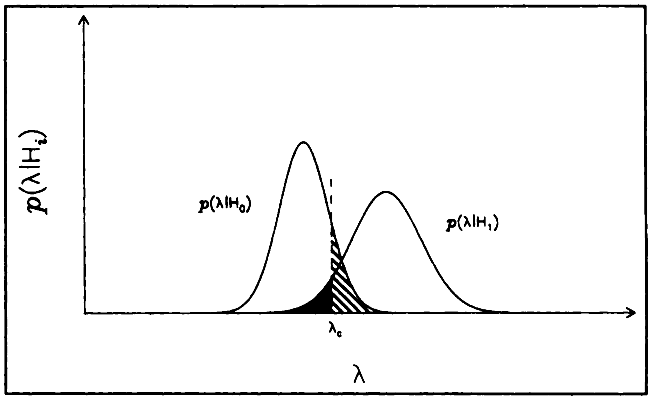

FIGURE 2. Examples of conditional probability distributions $p(\lambda|H_0)$ and $p(\lambda|H_1)$ and decision threshold λ_c. With good separation of distributions, threshold can be set so that probability of false-positive response (shaded part of $p(\lambda|H_0)$) or false-negative response (shaded part of $p(\lambda|H_1)$) will be small.

For SKE-BKE tasks with correlated Gaussian noise, the PW observer is the optimal observer.

However, neither the NPW observer nor the PW observer has been as consistent as the CHO in predicting human detection performance (*19,20*). The CHO can be divided into 3 components: the channels, a prewhitening operator, and a matched filter, as shown in Figure 3. For an SKE task, a channel can be represented by the 256 × 256 vector $\mathbf{u}_n$ ($n = 1, \cdots, N$) that describes the channel's impulse response centered on the tumor location. The scalar product of $\mathbf{u}_n$ and $\mathbf{f}$ is the nth element of an N-dimensional channel output vector. We write this output vector as $U\mathbf{f}$, where U is the $N \times 256^2$ matrix with the nth row containing the elements of $\mathbf{u}_n$.

The CHO prewhitening is aimed at decorrelating the noise in the channels (as opposed to the image pixels). Therefore, the CHO prewhitening operator is an N × N matrix defined by U and the ensemble noise covariance matrices K_0 and K_1. The composite channel noise covariance matrix is:

$$K_{chan} = \frac{1}{2}\, U[K_0 + K_1]U^{\dagger}, \qquad \text{Eq. 9}$$

and the prewhitening operator is its inverse, K_{chan}^{-1}. As $N \ll 256^2$, this inversion is much less demanding than the inversion of $K_0 + K_1$ required of the PW observer. In addition, for tasks in which the ensemble statistics are not available, so that covariance matrices must be estimated using sample statistics, the number of images required for an inverse to exist is on the order of the number of elements in the covariance matrix (*21*).

Finally, the CHO matched filter uses the ensemble mean difference $\bar{\mathbf{f}}_1 - \bar{\mathbf{f}}_0$ after it has been processed through the channels. In terms of these components, the CHO template is:

$$\mathbf{w}_{cho} = U^{\dagger} K_{chan}^{-1} U(\bar{\mathbf{f}}_1 - \bar{\mathbf{f}}_0), \qquad \text{Eq. 10}$$

and the CHO SNR is:

$$SNR_{cho} = [(\bar{\mathbf{f}}_1 - \bar{\mathbf{f}}_0)^{\dagger} U^{\dagger} K_{chan}^{-1} U(\bar{\mathbf{f}}_1 - \bar{\mathbf{f}}_0)]^{1/2}. \qquad \text{Eq. 11}$$

CHO Application

CHO models are distinguished by the type of band-pass filter used. Discretizations of the filters correspond to the Fourier transforms $\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_N$ of the channel impulse responses. We tested 2 forms, profiles of which are shown in Figure 4.

A constant-Q model features filters with central frequency magnitudes that are in a constant ratio Q. With nonoverlapping, rotationally symmetric, concentric filters having square profiles, the nth constant-Q filter takes the value 1 in the range of frequencies $[f_c Q^{n-1}/256, f_c Q^n/256)$ (pixel^{-1}) and 0 elsewhere (*3*). For the results presented here, this model was applied with 4 channels, a filter-width ratio of Q = 2, and a low-frequency cutoff of f_c = 6.5/256 pixel^{-1}. This parameter selection was made through trial and error from tests using CHOs with 3–5 channels, Q between 1.7 and 2.3, and f_c in the range of 2/256 pixel^{-1} to 20/256 pixel^{-1}. Burgess et al. (*6*) have suggested averaging results over an interval of cutoffs to model shifts in an observer's viewing distance during a study, but this procedure was not applied here.

Our second model observer, a difference-of-gaussian (DOG) model, used 3 channels. This CHO was used by Abbey et al. (*5*) in a study of iterative reconstruction stopping points. The filters were derived by taking the differences of pairs of a set of 4 2-dimensional gaussians with 0 means and an SD of $(2d\sqrt{\pi})^{-1}$ pixel^{-1},

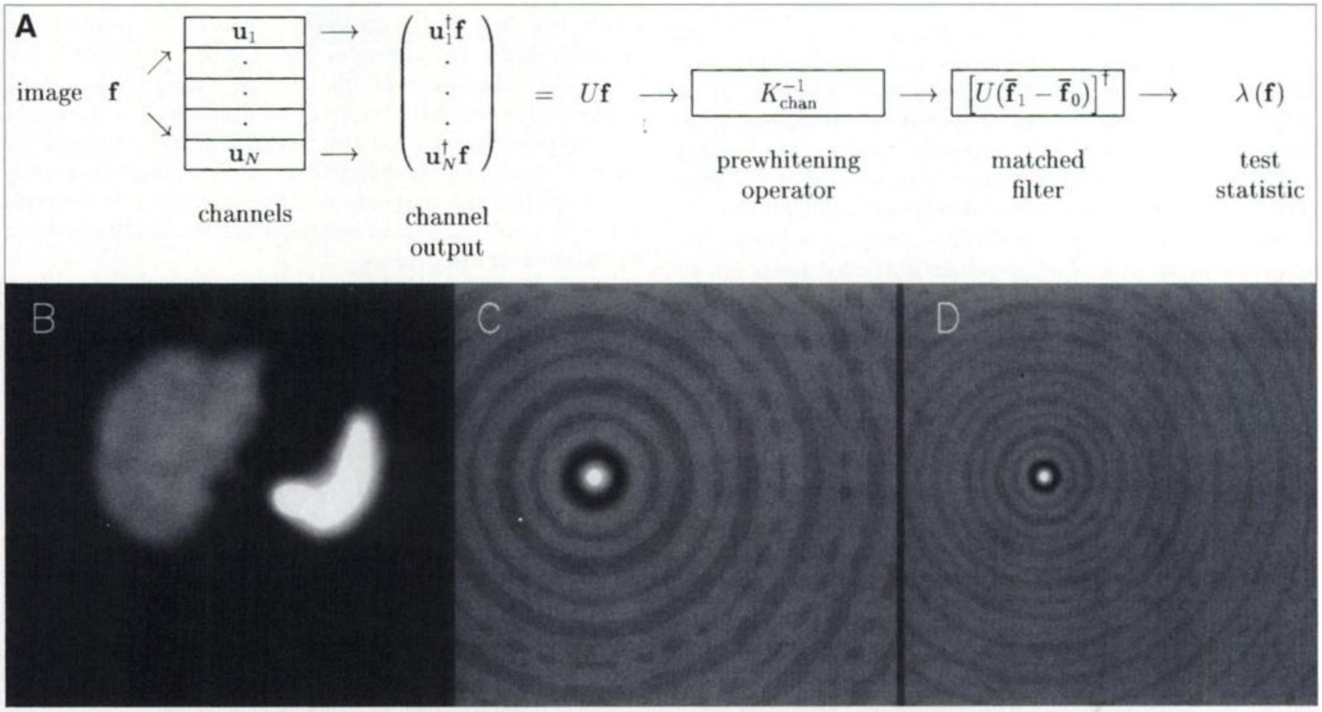

FIGURE 3. (A) Flow chart of CHO operation. (B) Image processed by N frequency-selective channels, with tumor at location 3. (C) Impulse response for low-frequency channel template. (D) Impulse response for higher-frequency channel. Both impulse responses are centered on location 3. Contrast of channel rings in C and D was exaggerated for purpose of display. Prewhitening and matched filtering of output vector **Uf** produced test statistic.

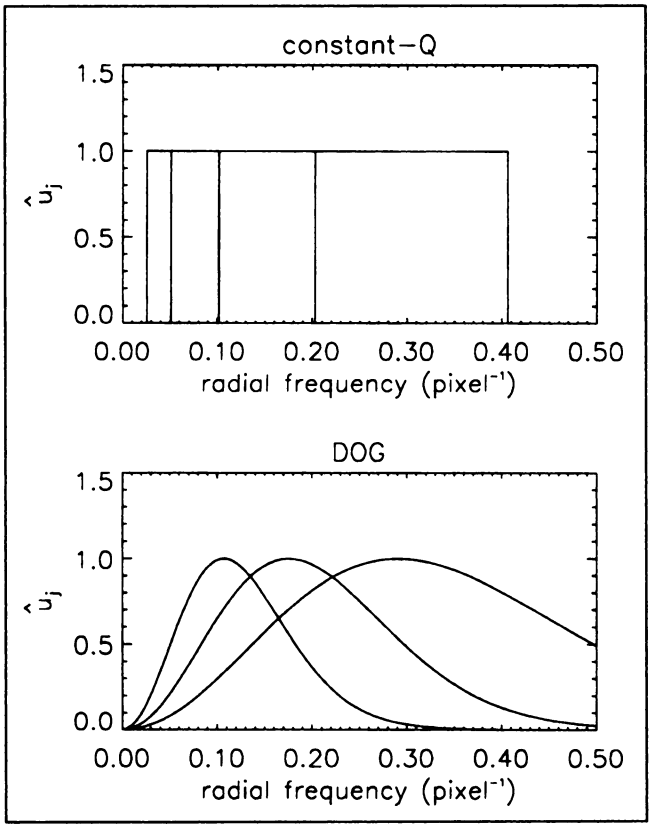

FIGURE 4. Profiles of constant-Q (top) and DOG (bottom) filter banks.

where d = 0.573, 0.995, 1.592, and 2.653. The DOG channels are also rotationally symmetric but have substantial overlap.

The first- and second-order ensemble statistics necessary for the prewhitening and matched-filter processes were estimated from the ROC images. This introduced an uncertainty in the SNRs of the model observers. Whereas human observer performance is based on only the study sets of 80 images per strategy and location, sample statistics for the CHOs used both the training and study images (totaling 100 images per strategy and location). This is a fairly small set of images on which to base the estimates, and having an equal ratio of lesion-present and lesion-absent images helped to ensure that the uncertainties in the estimates would be uniform. However, more clinically realistic probabilities $p(\lambda|H_0)$ and $p(\lambda|H_1)$ could be chosen to alter this ratio.

Finally, note that if our reconstruction process had not included the negative-pixel truncation, the ensemble means $\bar{\mathbf{f}}_0$ and $\bar{\mathbf{f}}_1$ of Equation 10 would be the reconstructions of the noise-free tumor-absent and tumor-present projection datasets, respectively, and an analytic expression for the ensemble covariance matrices would be available (*18*). The fact that the truncation operation invalidates this approach to deriving the ensemble statistics was not recognized in a previous comparison (*12*) of the CHO to the human ROC data used in this work, and as a consequence, an incorrect CHO template was applied.

Statistical Significance and Correlation Tests

Differences in an observer's performance as the result of strategy were tested at each lesion location using 2-way analysis of variance and a multiple-comparison *t* test for paired data (*22*). For the average human observer, the data were the individual observers' values of A_z. For the CHOs, the data consisted of the differences in $\lambda(\mathbf{f})$ computed from lesion-present–lesion-absent image pairs. In comparing pairs of strategies, we used a significance level of 0.05 to define statistically significant differences in observer performance.

Correlations between the average human and model observers were measured with the Spearman rank correlation test (*22*), which yields a linear correlation coefficient ρ_s of the observers' SNR rankings of the 15 location–strategy combinations. These rankings are susceptible to uncertainties in the SNRs as a result of variations among the human observers and the use of sample statistics for the model observers. The standard errors in the average human SNRs were determined from the standard errors in the values of $\overline{A}_z$ (*17*). For the CHO, SDs in the values of SNR_{cho} were found from a propagation-of-error analysis (*23*) applied to Equation 5. This calculation required estimates of the test-statistic quantities $\langle\lambda_i\rangle$ and $var(\lambda)_i$ ($i = 1, 2$), which were acquired by applying $\mathbf{w}_{cho}$ to the study images.

The sensitivity of the rank correlation coefficients to the uncertainties in SNR was evaluated through a Monte Carlo analysis. We created 2 sets of 15 independent gaussian random variables. The SNRs and uncertainties for the average human observer were assigned as means and SDs for the variables in 1 of these sets, whereas those for the CHO were assigned to the variables in the other set. By generating multiple realizations of these 30 variables and conducting repeated (simulated) Spearman tests, we were able to calculate an average rank correlation coefficient that accounted for the observer uncertainties.

RESULTS

Table 1 lists the observer SNRs for the 15 location–strategy combinations, along with SEs for the average

TABLE 1
SNR Values and Uncertainties for the Human, Constant-Q, and DOG Observers

	Location 1			Location 2			Location 3		
Strategy	Human	Constant-Q	DOG	Human	Constant-Q	DOG	Human	Constant-Q	DOG
Primary	1.50 ± 0.08	1.32 ± 0.22	1.79 ± 0.24	1.42 ± 0.05	1.20 ± 0.22	1.58 ± 0.23	1.65 ± 0.12	1.68 ± 0.23	2.19 ± 0.25
Scatter	1.18 ± 0.10	1.03 ± 0.21	1.45 ± 0.22	0.98 ± 0.09	0.92 ± 0.21	1.35 ± 0.22	1.52 ± 0.08	1.46 ± 0.23	1.70 ± 0.23
High-scatter	0.73 ± 0.08	0.70 ± 0.21	1.03 ± 0.21	0.93 ± 0.08	0.70 ± 0.21	1.24 ± 0.22	1.16 ± 0.09	0.99 ± 0.21	1.22 ± 0.22
DPW	1.28 ± 0.06	1.04 ± 0.21	1.42 ± 0.22	1.23 ± 0.06	0.85 ± 0.21	1.17 ± 0.22	1.47 ± 0.09	1.23 ± 0.22	1.68 ± 0.23
High-DPW	1.10 ± 0.04	0.77 ± 0.21	1.18 ± 0.22	1.09 ± 0.12	0.79 ± 0.21	1.12 ± 0.22	1.14 ± 0.05	1.14 ± 0.22	1.41 ± 0.22

For humans, data are SNR ± 1 SE; for CHOs, data are SNR ± 1 SD.

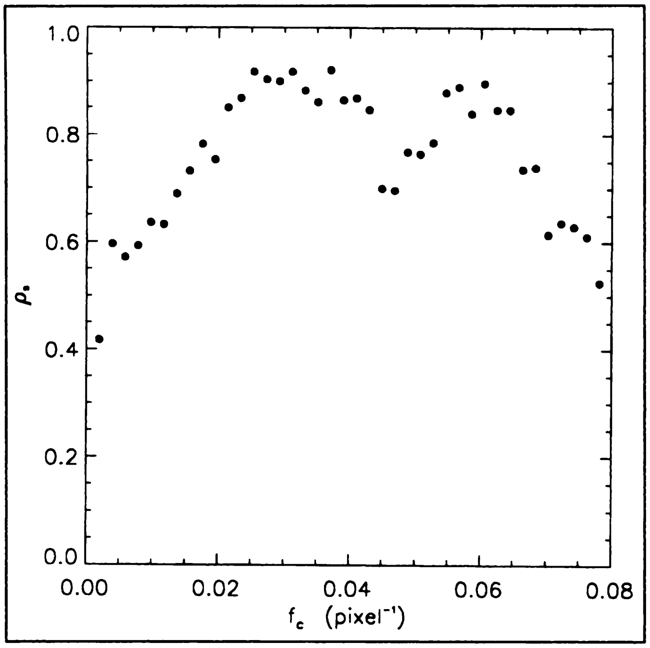

FIGURE 5. Spearman rank correlation coefficient ρ_s as function of constant-Q CHO cutoff frequency f_c.

human observer and propagation-of-error estimates of the SDs for the CHOs. For every strategy, each observer performed better with the tumor at location 3 than at other locations. This location was closest to the body's outer surface (Fig. 1), leading to less attenuation and scatter in the projection data than was the case with the other locations.

The human SNRs were generally higher than the constant-Q observer and lower than the DOG observer, but the SNRs for the constant-Q observer can be improved by lowering the cutoff frequency f_c. The parameter Q and the number of channels had smaller effects. Our final choices for the constant-Q parameters were based on outcomes of the Spearman rank correlation test. We found little difference in the CHO's ability to correlate with the human observer for values of Q between 1.7 and 2.3. This is consistent with the determination by Myers and Barrett (*3*). There was also little difference among 3, 4, and 5 channels. The correlation was more sensitive to f_c. A band between 5.5/256 pixel^{-1} and 10.5/256 pixel^{-1} provided fairly consistent results for the 4-channel model with Q = 2 (Fig. 5).

Pairs of strategies that differed significantly at a given location for at least 1 observer are indicated in Table 2. For these pairs, Table 2 also shows how close to significance the differences were for the other observers. For both the human and constant-Q observers, the SNR differences between the primary and high-scatter strategies were significant at all locations. For the DOG observer, this was the case at locations 1 and 3. In addition, the differences between the primary and high-DPW strategies were significant for the CHOs and either significant or nearly so for the human. In 5 of 6 significant differences for the human that did not involve the primary strategy, no significance was found for either CHO. The human observer also supplied the only instance (between high-scatter and high-DPW at location 1) in which the DPW scatter compensation provided statistically significant improvement over an uncorrected strategy.

Overall, more significant differences between strategy pairs were found for the human observer than for the CHOs. The constant-Q observer was shown to have 8 significant differences, and, for 7 of these, the human observer showed either significant or nearly significant differences. The exception was the primary and DPW strategy pair at location 3, where the significance level for the human was 0.59. However, significant differences for the human were less likely to be good indicators of significance levels for the constant-Q observer.

The ranking of the location–strategy combinations by the constant-Q observer correlated better with the human observer than did the DOG ranking. The rank correlation coefficient for the human and constant-Q observers comparison was ρ_s = 0.92, which allowed rejection of the null

TABLE 2
Significance Levels Between Pairs of Strategies by Observer and Location

Strategy	Location	Scatter	High-scatter	DPW	High-DPW
Primary	1	*	($\mathbf{<10^{-4}}$, $\mathbf{<10^{-2}}$, $\mathbf{<10^{-4}}$)	*	(0.1, $\mathbf{<10^{-2}}$, $\mathbf{<10^{-2}}$)
	2	(**0.03**, 0.27, 0.51)	(**0.01**, **0.01**, 0.15)	*	(0.14, **0.05**, **0.03**)
	3	(0.37, 0.86, $\mathbf{<10^{-2}}$)	($\mathbf{<10^{-3}}$, $\mathbf{<10^{-4}}$, $\mathbf{<10^{-4}}$)	(0.59, $\mathbf{<10^{-2}}$, $\mathbf{<10^{-3}}$)	($\mathbf{<10^{-3}}$, $\mathbf{<10^{-3}}$, $\mathbf{<10^{-4}}$)
Scatter	1		($\mathbf{<10^{-2}}$, 0.47, 0.23)	*	*
	2		*	*	*
	3		(**0.01**, **0.01**, **0.05**)	*	(**0.01**, 0.15, 0.4)
High-scatter	1			($\mathbf{<10^{-2}}$, 0.44, 0.3)	(**0.01**, 1.0, 0.97)
	2			*	*
	3			*	*
DPW	1		*		*
	2				*
	3				(**0.03**, 0.99, 0.48)

*No significant differences were found at that location.

Significance levels are given as (human, constant-Q, difference-of-gaussian filter). Boldface indicates level was significant.

hypothesis of no positive correlation at the $P = 0.000001$ level. The coefficient for the human and DOG observers comparison was $\rho_s = 0.84$ ($P = 0.00004$).

Figure 6 presents a scatter plot of the SNRs for the constant-Q and human observers. The hollow symbols for a given tumor location represent the non-DPW strategies (i.e., primary, scatter, and high-scatter), and the solid symbols represent the DPW scatter-compensation strategies (i.e., DPW and high-DPW). It is evident from this plot that our rank correlation coefficients could be very sensitive to the uncertainties in the SNRs. To test this, we applied the Monte Carlo technique that interpreted the SNRs and uncertainties of Table 1 as the means and SDs of gaussian random variables. The result from 1 million trials was an average rank correlation coefficient of 0.66 (SD = 0.13) for the human and constant-Q observers and 0.64 (SD = 0.14) for the human and DOG observers.

DISCUSSION

The results of this study show a significant positive correlation between the rankings of the average human observer and the constant-Q CHO for the considered SKE-BKE detection task. The correlation with the human observer was not perfect, but a true optimization of the CHO parameters was not an option, because efficient techniques for doing so are not currently available. As used to produce the data in Table 1, the constant-Q CHO is very similar to those used by other researchers (*3,6*). We found that changes in the value of Q and the number of channels did not have a large effect on the ability of the CHO to predict human performance. Given the good performance of the DOG model as well, these results suggest a fair degree of stability in the channelized model for correlating with human observers.

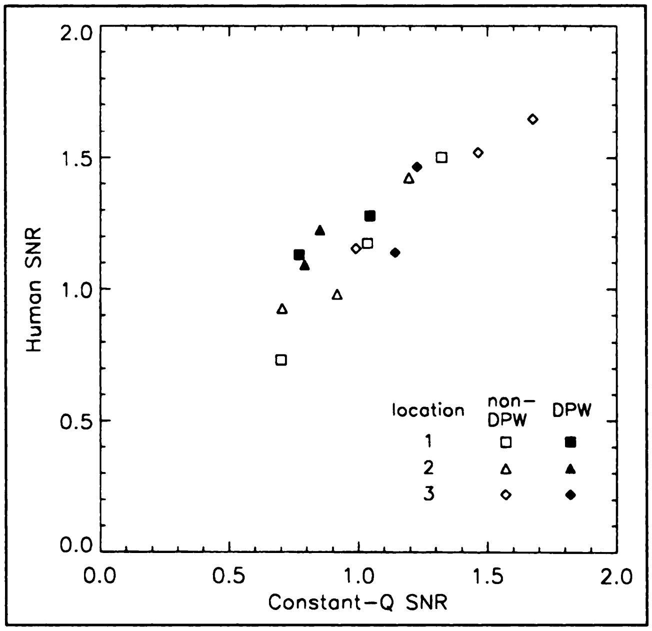

FIGURE 6. Scatter plot of average human and constant-Q CHO SNRs. Strategies that implemented DPW correction are shaded in black. Uncertainties are listed in Table 1.

Other model observer comparisons to the human observer data used in this work have been reported (*12,13,24*). The NPW observer was shown to correlate well with the average human observer for this detection task (*13*) and for related tasks (*24*). The CHOs applied here compare favorably with the NPW observer for our SKE-BKE detection task, because the rank correlation coefficient for the NPW observer was reported as 0.69 ($P = 0.005$) (*13*). The NPW observer also did a good job of predicting human performance for the non-DPW strategies, with a rank correlation coefficient of $\rho_s = 0.95$ ($P = 0.001$) when only these 9 combinations of lesion location and strategy were considered. This was also the case for the CHOs: the constant-Q observer produced a rank correlation coefficient of $\rho_s = 1.00$, and the DOG observer yielded $\rho_s = 0.93$ ($P = 0.0003$) when only the non-DPW strategies were ranked. This indicates that the CHO prewhitening operation is not essential for modeling human lesion detection in images with variable amounts of scatter. On the other hand, there was an increase in overall rank correlation coefficient in going from the NPW observer to the constant-Q and DOG observers. This is attributable to the CHO's superior performance for the DPW strategies and is evidence of a human ability to perform some prewhitening to overcome noise correlations created by the scatter subtraction method.

Although the CHOs showed improved correlation with the average human for the DPW strategies, there were some notable differences in how the CHOs and human observers reacted to DPW scatter correction. At the levels of scatter used in the scatter and DPW strategies, the human performed better (albeit not significantly so) with the corrected images, whereas the CHOs showed either a degradation or no improvement in performance. Yet at the elevated scatter levels used in the high-scatter and high-DPW strategies, all observers tended toward improved performance with the corrected images. This apparent contradiction for the CHOs is explained by the trade-off between the advantage of DPW contrast enhancement versus the problems caused by the increased noise correlation in the scatter-corrected images. This trade-off is a function of scatter level. For the normal scatter level, the reduced image contrast resulting from scatter was not debilitating for the CHOs, so that the DPW contrast enhancement had little effect, whereas the added noise correlation led to a net loss in performance. With elevated scatter, poor contrast in the uncorrected images was a much greater impediment for the CHOs, and the contrast enhancement provided by DPW overcame the disadvantage of increased noise correlation.

These conflicting trends for the normal and elevated levels of scatter show that despite the good correlation exhibited between the CHOs and human observers, the CHOs do not predict human performance accurately enough to serve as replacements in our psychophysical studies. But the outcome of our significance testing does suggest that the constant-Q observer could be used to screen for significant differences between strategies before a psychophysical study with human observers. However, in psychophysical studies comparing image processing or acquisition strategies for which the true differences in human detection performance may be small, proving that statistically significant differences exist can require many more than the 100 images per each strategy we have used (*23*), and the benefit of the CHO may be negligible. This points to the problem of using sample statistics as estimates of the ensemble quantities in the CHO template and suggests that efforts should be made to find an expression for the ensemble statistics if at all possible. Once the images have been reconstructed, however, application of the CHO template can be very efficient. The time required to generate the set of 15 SNRs and uncertainties given in Table 1 for either CHO was on the order of 40 s using a 400-MHz DEC Alpha workstation (Compaq Computer Corp., Houston, TX).

This study is the starting point for model observer research involving detection tasks that more closely approximate clinical tasks with SPECT imaging. The most substantial change will be the use of localization ROC (LROC) studies, in which tumor locations are not furnished to the observers, in place of the SKE-BKE ROC studies (25). Other sources of task variability to be considered are multiple lesion sizes and fluctuations in the anatomic background. Such alterations in the tasks present the challenge of modeling the human's detection processes but also of developing an effective model that is computationally efficient. Our first steps in this direction have entailed comparisons of human LROC and CHO SKE-ROC performance to see whether correlation persists.

CONCLUSION

The results of this study show a significant positive correlation between the average rankings of the human observers and the CHOs for the considered SKE-BKE detection task. Although this provides further evidence that a channelized observer might serve as a replacement for human observers in psychophysical studies, the role of the CHOs used in this study would be limited to providing an initial screening for significant differences between strategies before the human study. More research is required to determine the validity of this role for other detection tasks.

ACKNOWLEDGMENTS

The authors have benefited from discussions with C.K. Abbey, H.H. Barrett, A.E. Burgess, P.F. Judy, C.E. Metz, and R.G. Wells. This work was supported by the National Cancer Institute (NCI) under grant number CA-42165. Its contents are solely the responsibility of the authors and do not necessarily reflect the official views of NCI.

REFERENCES

1. Barrett HH, Yao J, Rolland JP, Myers KJ. Model observers for assessment of image quality. *Proc Natl Acad Sci USA.* 1993;90:9758–9765.
2. Burgess AE. From light to optic nerve: optimization of the front-end of visual systems. *Proc SPIE.* 1998;3340:2–13.
3. Myers KJ, Barrett HH. Addition of a channel mechanism to the ideal-observer model. *J Opt Soc Am A.* 1987;4:2447–2457.
4. Yao J, Barrett HH. Predicting human performance by a channelized Hotelling observer model. *Proc SPIE.* 1992;1768:161–168.
5. Abbey CK, Barrett HH, Wilson DW. Observer signal-to-noise ratio for the ML-EM algorithm. *Proc SPIE.* 1996;2712:47–58.
6. Burgess AE, Li X, Abbey CK. Nodule detection in two-component noise: toward patient structure. *Proc SPIE.* 1997;3036:2420–2442.
7. Eckstein MP, Abbey CK, Whiting JS. Human vs. model observers in anatomical backgrounds. *Proc SPIE.* 1998;3340:16–26.
8. de Vries DJ, King MA, Soares EJ, Tsui BMW, Metz CE. Evaluation of the effect of scatter correction on lesion detection in hepatic SPECT imaging. *IEEE Trans Nucl Sci.* 1997;44:1733–1740.
9. Hnatowich DJ, Mardirossian G, Ruscowski M, et al. Pharmacokinetics of the FO23C5 anti-CEA antibody fragment labeled with Tc-99m and In-111: a comparison in patients. *Nucl Med Commun.* 1993;14:52–63.
10. Ljungberg M, Strand SE. A Monte Carlo program for the simulation of scintillation camera characteristics. *Comput Meth Prog Biomed.* 1989;29:257–272.
11. Zubal IG, Harrell CR, Smith EO, Rattner Z, Gindi G, Hoffer PB. Computerized three-dimensional segmented human anatomy. *Med Phys.* 1994;21:299–302.
12. King MA, de Vries DJ, Soares EJ. Comparison of the channelized Hotelling and human observers for lesion detection in hepatic SPECT imaging. *Proc SPIE.* 1997;3036:14–20.
13. de Vries DJ. *Development and Evaluation of Scatter Subtraction for SPECT Imaging* [dissertation]. Worcester, MA: Worcester Polytechnic Institute; 1997.
14. de Vries DJ, King MA. Window selection for dual photopeak window scatter correction in Tc-99m imaging. *IEEE Trans Nucl Sci.* 1994;41:2771–2778.
15. Chang LT. A method for attenuation correction in radionuclide computed tomography. *IEEE Trans Nucl Sci.* 1978;25:638–642.
16. Metz CE. ROC methodology in radiological imaging. *Invest Radiol.* 1986;21:720–733.
17. Burgess AE. Comparison of receiver operating characteristic and forced choice observer performance measurement methods. *Med Phys.* 1995;22:643–655.
18. Barrett HH. Objective assessment of image quality: effects of quantum noise and object variability. *J Opt Soc Am A.* 1990;7:1266–1278.
19. Burgess AE. Comparison of non-prewhitening and Hotelling observer models. *Proc SPIE.* 1995;2436:1–8.
20. Burgess AE, Li X, Abbey CK. Visual signal detectability with two noise components: anomalous masking effects. *J Opt Soc Am.* 1997;14:2420–2442.
21. Barrett HH, Abbey CK, Gallas B. Stabilized estimates of Hotelling-observer detection performance in patient-structured noise. *Proc SPIE.* 1998;3340:27–43.
22. Pollard JH. *A Handbook of Numerical and Statistical Techniques.* Cambridge, UK: Cambridge University Press; 1977.
23. Abbey CK, Barrett HH, Eckstein MP. Practical issues and methodology in assessment of image quality using model observers. *Proc SPIE.* 1997;3032:182–194.
24. de Vries DJ, King MA, Soares EJ, Tsui BMW, Metz CE. Evaluation of scatter subtraction for hepatic SPECT imaging: comparison of human detection with numerical models for detection and quantitation. *J Nucl Med.* 1999;40:1011–1023.
25. Swensson RG. Unified measurement of observer performance in detecting and localizing target objects on images. *Med Phys.* 1996;23:1709–1725.