



Published in final edited form as:

IEEE Trans Nucl Sci. 2016 June ; 63(3): 1426–1434. doi:10.1109/TNS.2016.2542042.

Task Equivalence for Model and Human-Observer Comparisons in SPECT Localization Studies

Anando Sen,

Department of Biomedical Informatics, Columbia University, New York City, NY, USA

Faraz Kalantari, and

Department of Radiation Oncology, University of Texas Southwestern, Dallas, TX, USA

Howard C. Gifford

Department of Biomedical Engineering, University of Houston, Houston, TX, USA

Howard C. Gifford: hgifford@uh.edu

Abstract

While mathematical model observers are intended for efficient assessment of medical imaging systems, their findings should be relevant for human observers as the primary clinical end users. We have investigated whether pursuing equivalence between the model and human-observer tasks can help ensure this goal. A localization ROC (LROC) study tested prostate lesion detection in simulated In-111 SPECT imaging with anthropomorphic phantoms. The test images were 2D slices extracted from reconstructed volumes. The iterative OSEM reconstruction method was used with Gaussian postsMOOTHING. Variations in the number of iterations and the level of postfiltering defined the test strategies in the study. Human-observer performance was compared with that of a visual-search (VS) observer, a scanning channelized Hotelling observer, and a scanning nonprewhitening (CNPW) observer. These model observers were applied with precise information about the target regions of interest (ROIs). ROI knowledge was a study variable for the human observers. In one study format, the humans read the SPECT image alone. With a dual-modality format, the SPECT image was presented alongside an anatomical image slice extracted from the density map of the phantom. Performance was scored by area under the LROC curve. The human observers performed significantly better with the dual-modality format, and correlation with the model observers was also improved. Given the human-observer data from the SPECT study format, the Pearson correlation coefficients for the model observers were 0.58 (VS), -0.12 (CH), and -0.23 (CNPW). The respective coefficients based on the human-observer data from the dual-modality study were 0.72, 0.27, and -0.11 . These results point towards the continued development of the VS observer for enhancing task equivalence in model-observer studies.

Index Terms

lesion detection; model observers; visual search; image quality assessment; SPECT; dual-modality imaging

I. Introduction

The prospect of better noninvasive clinical diagnoses is a major motivation for tomographic medical imaging research. This motivation is put into practice through task-based assessments. The task could be lesion detection or estimation of physical quantities at a suspected lesion site. The task is performed by an observer, which in a clinical setting is a radiologist. Within this framework, image quality is defined by how well observers can perform a specified task with a given set of images [1]. For example, diagnostic accuracy as measured in human-observer studies provides a basic measure of imaging system performance [2]. Image quality may be improved by optimizing the hardware or acquisition protocols.

However, conducting large-scale studies with radiologists for developmental research is highly impractical. This situation has led to the development of mathematical model observers. Much of the model-observer research has investigated statistical ideal observers for system optimization, but the range of applicable tasks for these models is limited. Also, design evaluations based on model observers should generally be relevant for human observers as the chief end users of clinical imaging systems, which is not assured with ideal observers. We have focused instead on developing model observers for tasks that better approximate clinical tasks of interest, under the hypothesis that emphasizing task equivalence can lead to reliable human-observer models for clinically realistic search tasks that are impractical with existing model observers.

Of those existing models, the channelized Hotelling (CH) observer has been most widely used as a human-observer substitute [3]. The CH observer operates with extensive prior information in the form of image ensemble statistics, with the level of information defined by several well-known paradigms. Detection tasks involving a known lesion profile at a fixed location are signal-known-exactly (SKE) in nature. For these tasks, the observer has only to judge whether the lesion is present or not. The observer's knowledge about the anatomical background can be either precise [background-known-exactly (BKE)] or statistical [background-known-statistically (BKS)].

These background paradigms also dictate the types of noise that will affect model observer performance in a given task, with quantum and anatomical noise being the main processes for much of radiological imaging. With BKE tasks, quantum noise alone affects observer performance. More vital from a task equivalence standpoint are BKS tasks where anatomical noise in the form of normal background structure can contribute false positives or obscure actual lesions.

The CH observer has information about the relevant noise processes as provided through ensemble class covariance matrices. Scanning forms of the CH observer that have been applied for search tasks analyze each location within a particular region of interest (ROI) as a potential lesion site [4], [5]. For these inherently signal-known-statistically (SKS) tasks, the scanning CH observer requires covariance matrices at every ROI location. This can be a computationally expensive undertaking, particularly when applied for tasks with reconstructed images [6]. Downgrading to computationally simpler models like the scanning

channelized nonprewhitening (CNPW) observer [7] or nonsearch observers can raise important interpretational issues if the original and downgraded tasks are too dissimilar.

We have recently proposed [8] a computationally efficient way of accounting for the search effects of anatomical noise that is motivated by how radiologists read images [9]. Under this two-stage visual-search (VS) paradigm, an initial search identifies suspicious lesion locations and subsequent analysis is made at just those candidate locations. With a basic form of VS model observer for nuclear medicine, a search for hot “blobs” as candidate locations is degraded by image texture (quantum and anatomical noise combined) [8]. The observer forgoes much of the statistical information provided to the scanning CH observer and which humans generally do not have.

For this paper, we compared the VS and scanning CH observers as potential human-observer surrogates in localization ROC (LROC) studies with SPECT images. A scanning CNPW observer was also included to provide a benchmark for BKE task performance. The application was prostate and pelvic lymph-node lesion detection based on simulations with In-111-labeled Prostatecint [10]. Prostatecint images have long been known to be most useful when combined with the anatomical information garnered by SPECT-CT [11]. The model observers had precise knowledge of the ROIs, and another study objective was to test whether their performance might better describe the performance of human observer with SPECT images alone or with a dual-modality study format that provided both functional and anatomical imagery.

II. Background

A. Detection-Localization Tasks

We focus on search tasks conducted with 2D slices extracted from reconstructed image volumes. An $N \times N$ test image in our study is denoted by vector \mathbf{f} , with pixel values f_i , $i = 1, \dots, N^2$. Image \mathbf{f} might contain a single lesion or otherwise be lesion-free. We let Ω represent the region of interest (ROI) as represented by the pixel indices of possible lesion locations. Note that Ω may vary with slice.

When Ω contains J locations, a model observer faces $J+1$ detection hypotheses. Under the lesion-absent hypothesis

$$H_0: \mathbf{f} = \mathbf{b} + \mathbf{n}, \quad (1)$$

the image consists of a background \mathbf{b} and zero-mean quantum noise \mathbf{n} . For each of the remaining hypotheses

$$H_j: \mathbf{f} = \mathbf{b} + \mathbf{s}_j + \mathbf{n}, \quad j = 1, \dots, J, \quad (2)$$

the image also has a lesion \mathbf{s}_j in the j th location. In this work, we consider the special case of a location-invariant lesion profile \mathbf{s} and let \mathbf{s}_j denote the lesion shifted to the j th location.

B. Image Statistics

Image statistics are required to construct the model observers and we shall use a bracket notation to denote mean quantities. The quantum-noise mean of the lesion-absent cases is $\langle \mathbf{f} \rangle_{\mathbf{n}|\mathbf{b},0} = \mathbf{b}$, where the bracket subscript indicates an average over vector \mathbf{n} for fixed \mathbf{b} when hypothesis H_0 is true. The corresponding mean when hypothesis H_j holds is $\langle \mathbf{f} \rangle_{\mathbf{n}|\mathbf{b},j} = \mathbf{b} + \mathbf{s}_j$. When both quantum and anatomical noise are averaged, the respective means are $\langle \mathbf{f} \rangle_{\mathbf{n},\mathbf{b}|0} = \bar{\mathbf{b}}$ and $\langle \mathbf{f} \rangle_{\mathbf{n},\mathbf{b}|j} = \bar{\mathbf{b}} + \mathbf{s}_j$.

In computing noise covariance matrices for the test images, only lesion-absent images were used based on the assumption that the presence of a low-contrast lesion would not substantially affect these calculations. With this weak-signal approximation, the covariance matrix is

$$\mathbf{K} = \langle (\mathbf{f} - \langle \mathbf{f} \rangle_{\mathbf{n},\mathbf{b}|0})(\mathbf{f} - \langle \mathbf{f} \rangle_{\mathbf{n},\mathbf{b}|0})^t \rangle_{\mathbf{n},\mathbf{b}|0} \quad (3)$$

$$= \langle (\mathbf{f} - \bar{\mathbf{b}})(\mathbf{f} - \bar{\mathbf{b}})^t \rangle_{\mathbf{n},\mathbf{b}|0}. \quad (4)$$

Matrix \mathbf{K} may be decomposed into the sum [12]

$$\mathbf{K} = \mathbf{K}_{\text{quant}} + \mathbf{K}_{\text{anat}}, \quad (5)$$

with a quantum noise term

$$\mathbf{K}_{\text{quant}} = \langle (\mathbf{f} - \mathbf{b})(\mathbf{f} - \mathbf{b})^t \rangle_{\mathbf{n},\mathbf{b}|0}. \quad (6)$$

and an anatomical noise contribution

$$\mathbf{K}_{\text{anat}} = \langle (\mathbf{b} - \bar{\mathbf{b}})(\mathbf{b} - \bar{\mathbf{b}})^t \rangle_{\mathbf{b}}. \quad (7)$$

C. LROC Analysis for Model Observers

Observer performance with single-target search tasks may be evaluated with LROC methodology. For a given test image in an LROC study, the observer provides the most suspicious location (denoted herein by pixel index r) along with a rating λ that reflects the observer's confidence that the image is either lesion-present or lesion-absent. An image is classified as lesion-present at threshold λ_t if $\lambda > \lambda_t$. With a lesion-present image, the localization is scored as correct if it lies within a fixed radius of correct localization R_{CL} about the true location. The LROC curve is a plot of the true-positive fraction conditioned on correct localization against the false-positive fraction as λ_t varies. The area under the curve (A_L) serves as a performance figure of merit.

D. Scanning Observers for Detection-Localization Tasks

A scanning observer generates the required LROC image data for \mathbf{f} by first computing a perception measurement λ_j for each location j within Ω . The rating and localization for the image are then obtained according to the rules

$$\lambda = \max_{j \in \Omega} \lambda_j \quad (8)$$

$$r = \operatorname{argmax}_{j \in \Omega} \lambda_j. \quad (9)$$

The general form for the observer rating considered in our work is

$$\lambda_j = (\mathbf{w}_j^{\text{obs}})^t [\mathbf{f} - \mathbf{c}_j], \quad (10)$$

where $\mathbf{w}_j^{\text{obs}}$ is a location-specific scan template for a given model observer and the superscript t denotes the transpose.

The translation vector \mathbf{c}_j in Eq. 10 establishes the origin for the rating at a given location. This term is derived from the conditional means of \mathbf{f} under the hypotheses H_0 and H_j . For a BKE task that involves only quantum noise,

$$\mathbf{c}_j = \frac{\langle \mathbf{f} \rangle_{\mathbf{n}|\mathbf{b},0} + \langle \mathbf{f} \rangle_{\mathbf{n}|\mathbf{b},j}}{2} = \mathbf{b} + \frac{\bar{\mathbf{s}}_j}{2}. \quad (11)$$

For large J , the computation of \mathbf{c}_j at every location would be intensive were the lesion profile not treated as shift-invariant. Note that subtracting \mathbf{c}_j from \mathbf{f} in Eq. 10 creates a lesion detection task that is largely free of anatomical noise. With BKS tasks, both quantum and anatomical noise are averaged, and \mathbf{b} is replaced by $\bar{\mathbf{b}}$ in Eq. 11. In that case, the subtraction in Eq. 10 generally retains some anatomical noise.

1) Scanning Channelized Hotelling Observer—A CH observer performs detection tasks with the data obtained by processing \mathbf{f} through a set of C spatial frequency channels. In our studies, these channels provide a basic model for the human visual system. With $C \ll N^2$, the channels also serve to greatly reduce the dimensions of the model-observer computation. We let \mathbf{U} be the $N^2 \times C$ matrix whose columns are the location-invariant spatial responses generated as the inverse Fourier transforms of the channels. The location-specific template for the scanning CH observer then takes the form

$$\mathbf{w}_j^{CH} = \mathbf{U}_j \mathbf{K}_j^{-1} \mathbf{U}_j^t \bar{\mathbf{s}}_j. \quad (12)$$

The subscript on \mathbf{U} denotes a shift of the channel responses to the j th location, while \mathbf{K}_j is the $C \times C$ channelized covariance matrix obtained for location j from the formula

$$\mathbf{K}_j = \langle \mathbf{U}_j^t (\mathbf{f} - \bar{\mathbf{b}}) (\mathbf{f} - \bar{\mathbf{b}})^t \mathbf{U}_j \rangle_{n, \mathbf{b}|0}. \quad (13)$$

An equivalent expression for this matrix is $\mathbf{K}_j = \mathbf{U}_j^t \mathbf{K} \mathbf{U}_j$, and one may refer to Eq. 5 to express \mathbf{K}_j in terms of quantum and anatomical covariance contributions. For BKS tasks, both contributions are relevant and the CH perception measurement is given by

$$\lambda_j = \bar{\mathbf{s}}_j^t \mathbf{U}_j \mathbf{K}_j^{-1} \mathbf{U}_j^t \left[\mathbf{f} - \left(\bar{\mathbf{b}} + \frac{\bar{\mathbf{s}}_j}{2} \right) \right]. \quad (14)$$

The process of inverting \mathbf{K}_j is simplified by the use of a relative handful of channels, but having to compute a covariance matrix at each lesion location makes application of the scanning CH observer cumbersome for high-resolution or volumetric imaging applications. As explained in Sec. V, this is particularly true when there is a large degree of anatomical variation among the phantoms.

2) Scanning Channelized Nonprewhitening Observer—The CNPW observer template

$$\mathbf{w}_j^{CNPW} = \mathbf{U}_j \mathbf{U}_j^t \bar{\mathbf{s}}_j \quad (15)$$

is obtained by replacing the CH covariance term in Eq. 12 by the identity matrix. Omitting the noise covariances makes the template shift-invariant, whereby the CNPW perception measurements for BKE tasks may be efficiently computed from the discrete 2D cross-correlation

$$\mathbf{\Lambda} = \mathbf{w}^{CNPW} \star \star [\mathbf{f} - \mathbf{b}]. \quad (16)$$

In this equation, $\mathbf{\Lambda}$ is a measurement vector with entries λ_j for $j \in \Omega$, \mathbf{w}^{CNPW} is the centered observer template, and the double stars indicate a two-dimensional cross-correlation between $N \times N$ images. Since both \mathbf{w}^{CNPW} and $\bar{\mathbf{s}}$ are shift-invariant, the translation image for this observer simplifies to $\mathbf{c}_j = \mathbf{b}$, which is independent of location. Given this background subtraction in Eq. 16, the BKE task is lesion detection in reconstructed quantum noise.

As the scanning CNPW observer does not have covariance matrices, adjusting Eq. 16 for BKS tasks solely requires substituting $\bar{\mathbf{b}}$ for \mathbf{b} . However, an important problem for BKS tasks is the nonspecificity of the CNPW template—the anatomical structure that remains after subtracting $\bar{\mathbf{b}}$ from the test image can create pathological false positives. VS observers provide a means of addressing this lack of specificity for the more complicated tasks while largely maintaining the computational efficiency of the CNPW observer.

E. Visual-Search Observers

As related in our introduction, the VS observer performs a detection-localization task using sequential search and analysis. Prior knowledge of the lesion profiles in the study makes this an SKS task. An initial feature-driven search for suspicious locations reduces the ROI Ω to a much smaller set Ω' . The feature could be greyscale intensity (as in [8]), in which case the test image \mathbf{f} is first segmented into hot blobs. The local maximum within each blob is considered a focal point, and those focal points within Ω constitute Ω' . Other search features may be defined by incorporating lesion profiles or profile gradients [13] as matched filters to generate a correlation map for segmentation. A matched filter effectively integrates the saliency of a given feature in the image against the task-dependent knowledge of the observer.

The subsequent analysis of Ω' is performed by a statistical discriminant. In our previous work (and this study as well), the CNPW discriminant of Eq. 16 has been applied. Note that while the initial search does not include information about the mean image background, the discriminant analysis applies background subtraction. We thus categorize the CNPW-based VS observer task as quasi-BKE. The number of candidate locations in Ω' will be less than those in Ω and for lesion-present images may exclude true lesions due to masking effects, a situation not faced with the scanning CNPW observer. For this reason, the performances of the scanning CNPW observer and CNPW-based VS observer may differ.

III. Methods

A. The XCAT phantom

The extended cardiac torso (XCAT) phantom developed by Segars *et al.* [14] was used as our object. The XCAT phantom strikes a balance between mathematical and pixel-based phantoms by modeling the organ boundaries through nonuniform rational B-splines. Five different In-111 biodistributions were created for a single XCAT phantom geometry. These distributions were formed by sampling from normal probability distributions placed on the relative uptakes for certain organs. The means and standard deviations for the normal distributions were derived from clinical studies (see Table I). The relative uptakes for the remaining organs and the background were set to 2.0. A portion of the XCAT phantom between the liver and the thighs was extracted for our study. This portion had dimensions $256 \times 256 \times 256$ and a voxel width of 0.208 cm. Representative slices from corresponding activity and attenuation phantoms are shown in Fig. 1.

B. SPECT Simulation

The SPECT simulation followed clinical In-111 imaging protocols in mimicking acquisitions with a dual-headed gamma camera and medium-energy parallel-hole collimators. An analytical projector incorporated a rotator algorithm and accounted for inhomogeneous photon attenuation and depth-dependent detector blur. Although photon scatter is an important component of In-111 SPECT imaging, it was not modeled as our primary focus for this study was on comparing the human and model observers.

For each acquisition, a total of 120 noise-free projections were generated at equal intervals on a 360° circular orbit. The projection dimensions were 256×256 , with a pixel size of 0.208 cm. Separate projections were calculated for the 171-keV and 245-keV gamma rays of In-111 and then scaled by their respective abundances and absorption coefficients before being combined. The relative abundances for these two energies are 90% and 94%, respectively, while the absorption coefficients for a sodium iodide detector with one-cm crystal thickness were taken to be 0.95 and 0.99. The projections were then rebinned to 128×128 (pixel size = 0.416 cm) to match clinical acquisitions. The initial acquisition at higher resolution was intended to control aliasing effects.

The lesion-present cases featured a single spherical soft-tissue target with a one-cm diameter. Lesion locations were selected with the help of binary masks that were created for the prostate and lymph nodes. These masks assumed the value of 1 for voxels within the organ of interest and 0 otherwise. The lesion center coordinates were chosen randomly within the support of these masks, with the constraint that any two lesion centers be separated by at least two voxels. A total of 225 lesion locations were generated, with an approximately equal distribution between the prostate and the lymph nodes. Sets of lesion projections were created, scaled for contrast and then added to the noise-free XCAT background projections. The lesion-to-prostate activities were from the set of values {6.6, 8.5, 10.5, 12.4}. These values were based on pilot study data [15] with the intention of generating average human-observer A_L values on the order of 0.75–0.80 as computed over the range of test strategies (see also Sec. III-E). Each relative activity was used with each lesion location for the image formation.

Photon-count levels from clinical data were used to guide the addition of Poisson noise to the noise-free projections. Some count-level adjustment was required due to the relatively high-count contributions of the liver in the XCAT acquisitions. We generated 15 different noise levels for the projection sets by sampling from a uniform probability distribution on total projection counts that assumed a mean of 5 million counts. The sampled count levels were between 4.1 and 5.6 million. Poisson noise based on one of these count levels was added to each projection set. All told, there were 225 noisy lesion-present sets and an equal number of noisy lesion-absent sets for each relative lesion activity.

C. Image Reconstruction

The projection sets were reconstructed using the ordered-subset expectation-maximization (OSEM) algorithm [16]. The 120 projections were divided into 15 subsets of 8 projections each, and even numbers of OSEM iterations from 2 through 8 were obtained. The reconstructions included iterative corrections for attenuation and detector blur. For attenuation correction, we used an attenuation map based on an effective energy of 210 keV. This map was computed by scaling the XCAT density map using the method described by Seo et al. [17]. The $128 \times 128 \times 128$ reconstructed volumes (voxel width of 0.416 cm) were smoothed with a 3D Gaussian postfilter. The applied blurs corresponded to integral full-widths-at-half-max (FWHM) from 0 to 4 voxels, with a blur of 0 implying no postfiltering. The test strategies in our study were represented by selected combinations of iteration number and blur FWHM as described in Sec. III-E.

After postfiltering, each reconstructed volume was converted (without thresholding) to byte format. The transverse slice through the lesion center of a lesion-present volume and the same slice from a corresponding lesion-absent volume were extracted for the observer study. Fig. 2 displays the noisy and noise-free images based on a pair of lesion locations. Corresponding slices from the XCAT attenuation phantom were extracted to provide anatomical information for the human observers in our study.

D. Model-observer Details

As described in Sec. II-E, the VS observer applies a two-step search-analysis process. For this work, the initial search considered the matched-filter image

$$\mathbf{z} = \bar{\mathbf{s}} \star \mathbf{f}. \quad (17)$$

Image \mathbf{z} was first segmented into blobs by means of a watershed algorithm [18]. The algorithm was applied to the additive inverse $-\mathbf{z}$ in order to identify regions of relatively high activity (hot blobs). Conceptually, the algorithm forms “catchment basins” (or watersheds) around local minima in $-\mathbf{z}$, building dams where the basins meet. Each basin in the segmented output represented a blob in \mathbf{z} . Determination of candidacy locations from these blobs and application of the CNPW discriminant to the search output was as described in Sec. II-E.

Each of the model observers in our study made use of a set of three 2D difference-of-Gaussian (DOG) spatial frequency channels [19]. The elements for the i th channel ($i = 0, 1, 2$) were

$$\tilde{u}_i(\boldsymbol{\xi}) = \exp \left[-\left(\frac{\|\boldsymbol{\xi}\|}{2^{i+1}\sigma_0} \right)^2 \right] - \exp \left[-\left(\frac{\|\boldsymbol{\xi}\|}{2^i\sigma_0} \right)^2 \right], \quad (18)$$

with $\boldsymbol{\xi}$ the 2D spatial frequency and $\sigma_0 = 0.015$. The discretized spatial responses for these channels formed matrix \mathbf{U} in Eqs. 12–15.

E. Observer Studies

Four human observers participated in the study. These were imaging scientists (nonradiologists) from our laboratory, three of whom had minimal prior experience reading simulated images. The observers read images from ten test strategies, representing the five postfilter FWHMs for two and six OSEM iterations. This subset of the saved iterations was selected based on pilot studies which indicated relatively little effect of iteration number on observer performance [15]. A set of 150 images per strategy consisted of 75 pairs of lesion-present/lesion-absent images. The same lesion locations were used for all the strategies. Each location was associated with a single relative lesion activity.

The 75 image pairs for a given strategy were split into sets of 25 training pairs and 50 test pairs. The training images were read immediately before the 50 test images. During the training process, feedback was provided as to the true disease status of a given image. The

SPECT and dual-modality study formats were both implemented. For the dual-modality study, each test image was displayed alongside the corresponding anatomical image derived from the XCAT attenuation map (Fig. 3). The reader should note that attenuation maps are not used clinically for this purpose. With each format, confidence ratings were collected on a four-point ordinal scale.

The model observers read the same image sets as the human observers, but were also applied for additional contrast-specific sets covering the same test strategies. These additional sets consisted of 450 images (a lesion-present/lesion-absent pair for 225 lesion locations), with all the lesions in a given set having the same relative activity. Of the 450 images, 150 were used for training and 300 for testing. The CNPW observer training for a given strategy consisted of estimating the mean reconstructed lesion profile $\bar{\mathbf{s}}$ and the mean background \mathbf{b} . In this work, $\bar{\mathbf{s}}$ was estimated from the training images, whereas \mathbf{b} was approximated from noise-free lesion-absent reconstructions associated with the training images. A separate set of 225 noisy lesion-absent images was used for estimation of the channelized covariance matrices. The model-observer rating data were maintained as floating-point values, as opposed to being converted to an ordinal scale.

Observer performance for all observers was measured as A_L . Correct lesion localizations were decided using a five-pixel R_{CL} that was determined by the process described in [20]. Estimates of A_L were obtained with a Wilcoxon nonparametric ranking method [21]. Uncertainties for the A_L estimates were calculated using the formula given in [22]. The overall human performance for a given strategy was calculated as the average A_L over the individual observers, and standard errors for these averages were also calculated. Analysis of variance (ANOVA) and posthoc comparisons were employed to test the statistical significance of the observer results.

IV. Results

A. The Human-Observer Study

Figure 4 shows how average human-observer performance varied with study format as a function of postfilter FWHM for two and six OSEM iterations. The error bars represent one standard error in the mean areas. For both iterations and all blur levels, average performance was higher with the dual-modality format than with the SPECT format. Qualitatively, these plots demonstrate predictable trends in that the nominal performance peak is at a higher blur for six iterations than for two iterations, reflecting the quantum noise amplification that occurs with increased OSEM iterations. The results of a four-way ANOVA conducted with the individual observer scores are summarized in Table II. The main effects for the analysis were blur, study format, iteration and observer. The blur, format and observer effects were statistically significant at the 5% level. A subsequent Tukey HSD multiple comparisons test attributed the latter effect to the differences in reading experience among the observers. The two-way interaction between format and observer was also significant. After controlling for blur and iteration, each observer was found to have scored higher on average in the dual-modality study than in the SPECT study, but the differences were considerably larger for the relatively inexperienced observers. A follow-up two-way ANOVA was conducted with the dual-modality results alone. Observer and reconstruction strategy were the main factors. For

this format-specific analysis, strategy alone was found to be significant (p -value = 0.01). At two OSEM iterations, observer performance with four-voxel postsmoothing was significantly lower than the performance with either zero or one-voxel postsmoothing. No significant differences were found on the six-iteration curve. The p -value for the observer effect was 0.55.

B. Model-Observer Performance

Performances of the CH and VS observers with the contrast-specific image sets are compared in Fig. 5 as a function of iterations and FWHM. Figures 5a and b relate to two and six iterations, respectively. Of the four lesion contrasts used in the study, only the two lowest ones (6.6 and 8.5) are shown as the higher contrasts produced little variation in observer performance. The uncertainties in the plotted A_L values were in the range of ± 0.01 to ± 0.03 and varied inversely with contrast.

Unlike the performance of the CNPW observer with the BKE task (not shown), which increased monotonically with FWHM [15], the VS and CH observer performances generally peaked at intermediate and iteration-dependent FWHMs due to the combined effects of quantum and anatomical noise. As with the human-observer results in Fig. 4, the optimal FWHM shifted to the right as increased numbers of iterations amplified quantum noise. Too much postsmoothing amplified anatomical noise as the lesion became lost in the background.

Figure 5 also shows that the VS observer tended to outperform the CH observer at low FWHM while performing worse at high FWHM. One reason for this lies in how image blur affects each observer's response to anatomical noise. With the VS observer, partial-volume effects for small structures [23] dictate that true lesion locations are more likely to be selected as candidate locations at low levels of blur than with high levels. Given the relatively high efficiency of the BKE CNPW candidate analysis at all FWHM values, the result is relatively lower anatomical-noise effects for decreasing FWHM. By contrast, anatomical noise for the CH observer is primarily introduced through the subtraction of $\bar{\mathbf{b}}$ (see Eq. 14). Generally, the effects of anatomical noise for a given test image will increase for larger deviations between $\bar{\mathbf{b}}$ and \mathbf{b} , which in our study occurred for lower FWHM values.

For the VS observer, the number of focal points per image determined from the watershed segmentation averaged in the 60s. The number of candidate locations per image, decided by applying the binary ROI mask to the focal points, was always less than five.

C. Model Observers versus Human Observers

While Fig. 5 pertains to contrast-specific image sets, the model observers also read the mixed-contrast image sets from the human-observer study. Figure 6 compares the performance of the model observers with the average dual-modality performance of the human observers. All three models demonstrated a sizable positive performance bias relative to the human observers. The CNPW observer performed its BKE task nearly perfectly and with negligible uncertainties for all reconstruction strategies. The differences between the CNPW and VS observers are attributable to the initial candidate search performed by the

latter. The VS and CH model observers handled anatomical noise under different task paradigms, yet performed similarly relative to the CNPW and human observers. As with the contrast-specific results, the VS observer outperformed the CH observer at low FWHM while underperforming at high FWHM. The CH and VS uncertainties ranged from ± 0.03 to ± 0.06 , consistent with the uncertainties for the individual human observers. A two-way ANOVA applied to the CH and VS results found no significant differences between the observers (p-value = 0.45) or among the strategies (p-value = 0.21).

Correlation coefficients quantified the agreement between the various model observers and the human observers. The VS observer produced the highest correlation. With the average human-observer data from the SPECT study, the VS, CH and CNPW coefficients were 0.58, -0.12, and -0.23, respectively. For the dual-modality data, the respective coefficients were 0.72, 0.27, and -0.11.

V. Discussion

Several approaches to increased task equivalence in SPECT detection-localization studies have been investigated. The dual-modality display for the human observers was intended to balance the ROI knowledge of the model observers, while the VS observer invoked a more realistic search paradigm for humans than what scanning observers provide. The VS observer also mimicked human observers to an extent in relying less on prior knowledge compared to the scanning CH observer. Section IV shows that in terms of model and human observer agreement, these approaches improved our study outcomes.

Model observers operating with prior knowledge about the image classes will generally outperform human observers, as was the case in this study. For SKE tasks, internal noise is customarily added to the models to bridge this performance gap [24]. However, the use of standard internal-noise mechanisms to compensate for ROI uncertainty in search tasks can be unreliable [7]. Our results indicate that augmenting human-observer knowledge may be the better option. Dual-modality studies in the form of registered SPECT and CT reconstructions have long been standard protocol for clinical In-111 Prostatecint imaging.

Detailed comparisons of our study with clinical imaging are restricted by several limitations, including the omission of scatter in the SPECT acquisitions, the 2D nature of our search task, and the manner in which anatomical information was presented to the human observers. The anatomical images for the dual-modality study were simply scaled extractions from the XCAT density map, presented side-by-side with the corresponding SPECT slice. These images provided sufficient anatomical information for our observer comparisons. However, simulating a true SPECT-CT display with registered images might have reduced interobserver variations with some of the test strategies. The inclusion of In-111 scatter would have changed the noise texture in the test images, potentially affecting the relative performance magnitudes of the human and model observers. Correlations between the observers should be less affected, based on results from previous SPECT studies involving scatter (e.g., [8], [25]). Lastly, the use of single-slice test images was acceptable given our primary objective of observer comparison. Methods for applying these

model observers to multislice images in other SPECT applications have been described previously [7], [8] and future studies can incorporate these methods.

We assessed observer agreement with A_L . Localization concordance is a more-stringent measure of agreement utilized in previous studies [26]. On the whole, it is reasonable to expect that the VS observer should improve concordance with humans when compared with scanning observers, as the initial hot-blob search is predicated on a basic tenet that humans understand about lesion detection for much of nuclear medicine. However, further consideration of the concordance measure awaits new extensions of the VS model. One shortcoming with the current model is its reliance on the quasi-BKE task. Also, the two-stage VS process allows considerable flexibility in modeling human-observer inefficiencies, and we have been testing separate internal-noise mechanisms for the search and analysis stages along with constraints on search sensitivity [27].

This study has offered the first direct comparison of the scanning CH and VS model observers for tomographic imaging. The latter demonstrated better correlation with the human observers, although the overall performance magnitudes were similar. With the scanning CH model, the covariance matrices aid performance by providing a partial prewhitening based on ensemble texture information. The deleterious effects of anatomical noise are largely determined by the ensemble-mean background $\bar{\mathbf{b}}$ in Eq. 14. The anatomical variations in our study were limited to changes in radiotracer uptake within a fixed phantom geometry. With structural variations between phantoms, each phantom would be accorded its own $\bar{\mathbf{b}}$ calculated from the Markov-chain Monte Carlo methods described in [6]. The manner in which $\bar{\mathbf{b}}$ is obtained by this approach can lead to ambiguity in the BKS task.

To select candidate locations, the VS observer identifies lesion-like features in the unmodified test image. This amounts to empirical testing of the image texture correlations without resorting to background statistics. Candidate locations may represent actual lesions or be due to noise blobs. The quasi-BKE task likely inflated observer performance, particularly with the highly smoothed test images where anatomical noise was greatest. Future studies will investigate discriminants that do not require background task paradigms [13].

The computational requirements of the two model observers were not similar. Implemented with IDL code on a single CPU of a Dell 7600 Linux workstation, the CH observer read a set of 100 images in two minutes. Given the training images, precalculation of the location-specific channelized covariance matrices for the image set took approximately two minutes. (The matrices were not preinverted.) The shift-variant nature of the CH template was the principal bottleneck, but the covariance calculations would be a major factor in applications with larger ROI dimensions or when there are structural variations between the phantoms. By comparison, the VS observer read the image set in under five seconds, slightly longer than what the CNPW observer required. This timing includes application of the watershed segmentation.

VI. Conclusions

We have investigated task equivalence as a means of improving the agreement between model and human observers in SPECT observer studies. Use of the dual-modality study format and the VS paradigm helped to balance the prior knowledge of our human and model observers. The VS observer also improves on scanning as a model for human search while providing a computationally efficient way of accounting for how image texture affects task performance. Our study outcomes argue for the continued development of VS observers for enhancing task equivalence in model-observer studies.

Acknowledgments

This work was supported by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) under grants R01 EB012070 and R21 EB012529. The contents are solely the responsibility of the authors and do not necessarily represent official NIBIB views.

We thank Paul Segars and Chi Liu for providing the In-111 uptake data. The dual-modality display interface for the human observers was developed jointly with Nagarohit Katta. Aixia Guo, Kheya Banerjee and Raksha Raghunathan participated in the human-observer study.

References

1. Tech. Rep. Report 54. Bethesda, MD: International Commission on Radiation Units and Measurements; 1996. Medical Imaging - The Assessment of Image Quality.
2. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Medical Decision Making*. 1991; 11:88–94. [PubMed: 1907710]
3. Barrett HH, Yao J, Rolland JP, Myers KJ. Model observers for assessment of image quality. *Proceedings of the National Academy of Sciences*. 1993; 90:9758–9765.
4. Gifford HC, Pretorius PH, King M. Comparison of human- and model-observer LROC studies. *Proceedings of SPIE*. 2003; 5034:112–122.
5. Khurd P, Gindi GR. Fast LROC analysis of Bayesian reconstructed emission tomographic images using model observers. *Physics in Medicine and Biology*. 2005; 50:1519–1532. [PubMed: 15798341]
6. He X, Caffo BS, Frey EC. Toward realistic and practical ideal observer (IO) estimation for the optimization of medical imaging systems. *IEEE Transactions on Medical Imaging*. 2008; 27:1535–1543. [PubMed: 18815105]
7. Gifford HC, King MA, Pretorius PH, Wells RG. A comparison of human and model observers in multislice LROC studies. *IEEE Transactions on Medical Imaging*. 2005; 24:160–169. [PubMed: 15707242]
8. Gifford HC. A visual-search model observer for multislice-multiview SPECT images. *Medical Physics*. 2013; 40:092505. [PubMed: 24007181]
9. Kundel HL, Nodine C, Conant EF, Weinstein SP. Holistic component of image perception in mammogram interpretation: gaze-tracking study. *Radiology*. 2007; 242:396–402. [PubMed: 17255410]
10. Taneja S. ProstaScint® Scan: Contemporary Use in Clinical Practice. *Reviews in Urology*. 2004
11. Seo Y, Franc BL, Hawkins RA, Wong KH, Hasegawa BH. Progress in SPECT/CT imaging of prostate cancer. *Technology in Cancer Research & Treatment*. 2006; 5:329–336. [PubMed: 16866563]
12. Whitaker MK, Clarkson EW, Barrett HH. Estimating random signal parameters from noisy images with nuisance parameters: linear and scanning-linear methods. *Optics Express*. 2008; 16:8150–8173. [PubMed: 18545527]
13. Lau BA, Das M, Gifford HC. Towards Visual-Search Model Observers for Mass Detection in Breast Tomosynthesis. *Proceedings of SPIE*. 2013; 8668:86680X.

14. Segars WP, Tsui BMW. MCAT to XCAT: The evolution of 4-D computerized phantoms for imaging research. *Proceedings of the IEEE*. 2009; 97:1954–1968. [PubMed: 26472880]
15. Sen A, Kalantari F, Gifford HC. Assessment of prostate cancer detection with a visual-search human model observer. *Proceedings of SPIE*. 2014; 9037:90370Q.
16. Hudson HM, Larkin RS. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Transactions on Medical Imaging*. 1994; 13:601–609. [PubMed: 18218538]
17. Seo Y, Wong KH, Hasegawa BH. Calculation and validation of the use of effective attenuation coefficient for attenuation correction in In-111 SPECT. *Medical Physics*. 2005; 32:3628–3635. [PubMed: 16475761]
18. Roerdink J, Meijster A. The watershed transform: definitions, algorithms and parallelization strategies. *Fundamenta Informaticae*. 2001; 41:187–228.
19. Abbey CK, Barrett HH. Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability. *Journal of the Optical Society of America A*. 2001; 18:473–488.
20. Wells RG, Simkin PH, Judy PF, King M, Pretorius PH, Gifford HC. Effect of filtering on the detection and localization of small Ga-67 lesions in thoracic single photon emission computed tomography images. *Medical Physics*. 1999; 26:1382–1388. [PubMed: 10435542]
21. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982; 143:29–36. [PubMed: 7063747]
22. Tang LL, Balakrishnan N. A random-sum Wilcoxon statistic and its application to analysis of ROC and LROC data. *Journal of Statistical Planning and Inference*. 2011; 141:335–344. [PubMed: 22639485]
23. Wernick, MN.; Aarsvold, JN., editors. *Emission Tomography: The Fundamentals of PET and SPECT*. Academic Press; 2004.
24. Zhang Y, Pham BT, Eckstein MP. Evaluation of internal noise methods for Hotelling observer models. *Medical Physics*. 2007; 34:3312–3322. [PubMed: 17879795]
25. Farncombe TH, Gifford HC, Narayanan MV, Pretorius PH, Frey EC, King MA. Assessment of scatter compensation strategies for (67)Ga SPECT using numerical observers and human LROC studies. *Journal of Nuclear Medicine*. 2004; 45:802–812. [PubMed: 15136630]
26. Gifford HC. Efficient Visual-Search Model Observers for PET. *British Journal of Radiology*. 2014; 87:20140017. [PubMed: 24837105]
27. Sen, A.; Kalantari, F.; Gifford, HC. Impact of anatomical noise on model observers for prostate SPECT; *IEEE Nuclear Science Symposium and Medical Imaging Conference*; 2014.

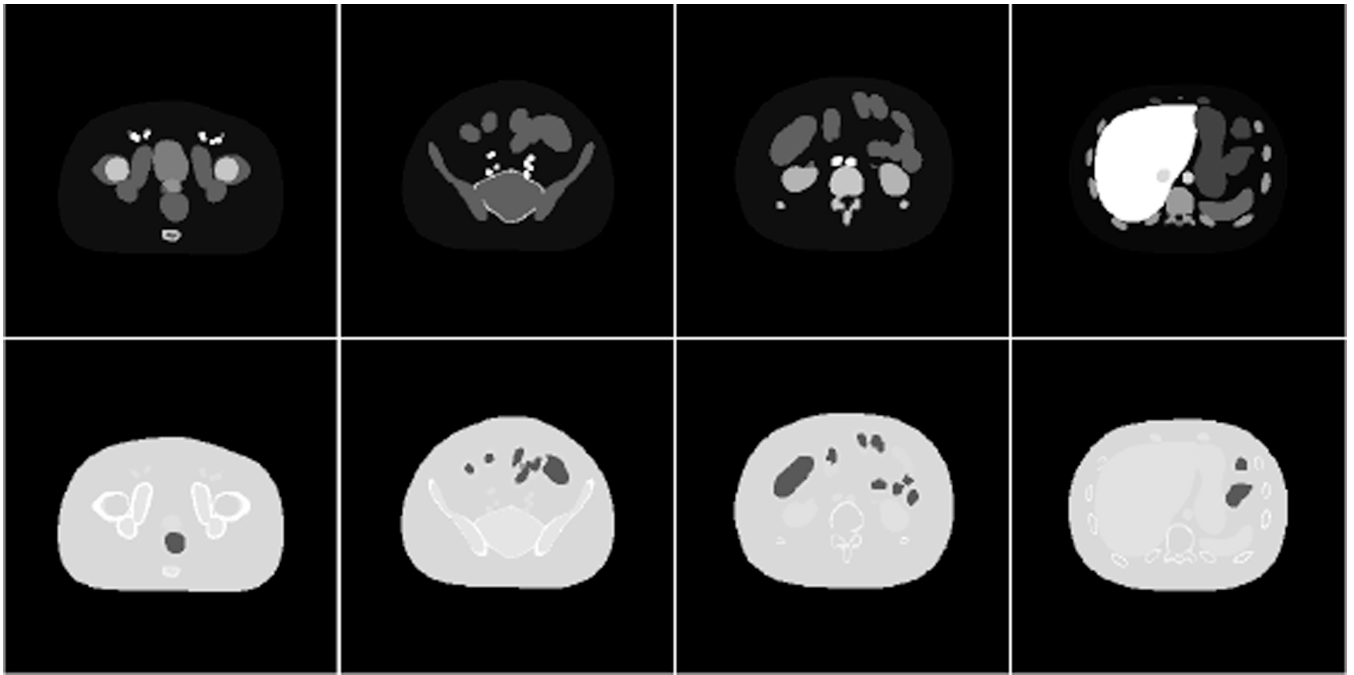


Fig. 1. Representative slices of the activity phantom with In-111 biodistribution are shown in the top row and corresponding attenuation phantom slices (171 keV) shown below it. The slice at left includes the prostate. The other slices are progressively higher in the torso, with the hot liver shown at right. The images have been contrast-enhanced for publication.

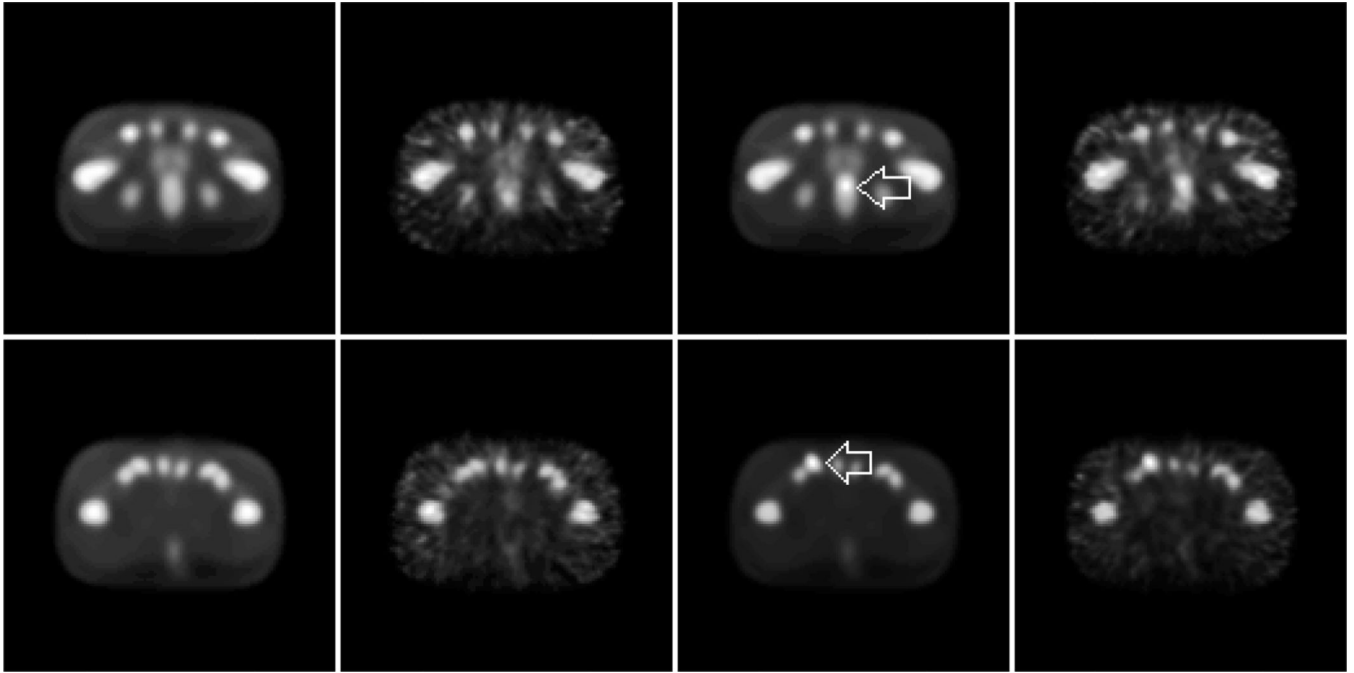


Fig. 2. Example OSEM reconstructed images generated using two iterations and a two-voxel postfilter FWHM. From left to right, each row shows the noise-free lesion-absent, noisy lesion-absent, noise-free lesion-present, and noisy lesion-present images from a different slice of the XCAT phantom. At top, the lesion is in the prostate. The bottom row shows a nodal lesion. Each lesion is indicated by an arrow in the noise-free image. The images have been contrast-enhanced for publication.

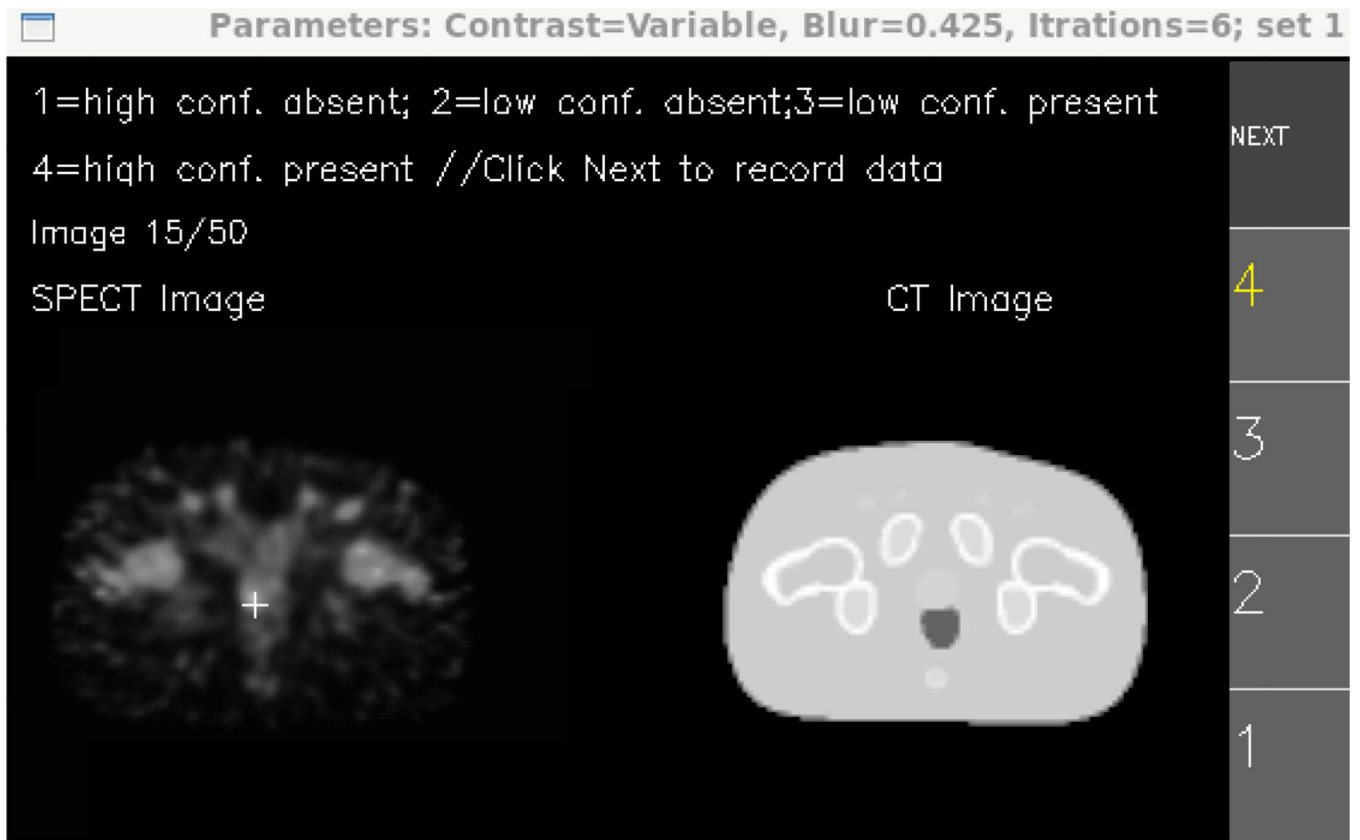


Fig. 3. The human-observer interface for the dual-modality study. Shown is the display during a training session, with the lesion location and confidence rating marked. The SPECT study interface omitted the anatomical image.

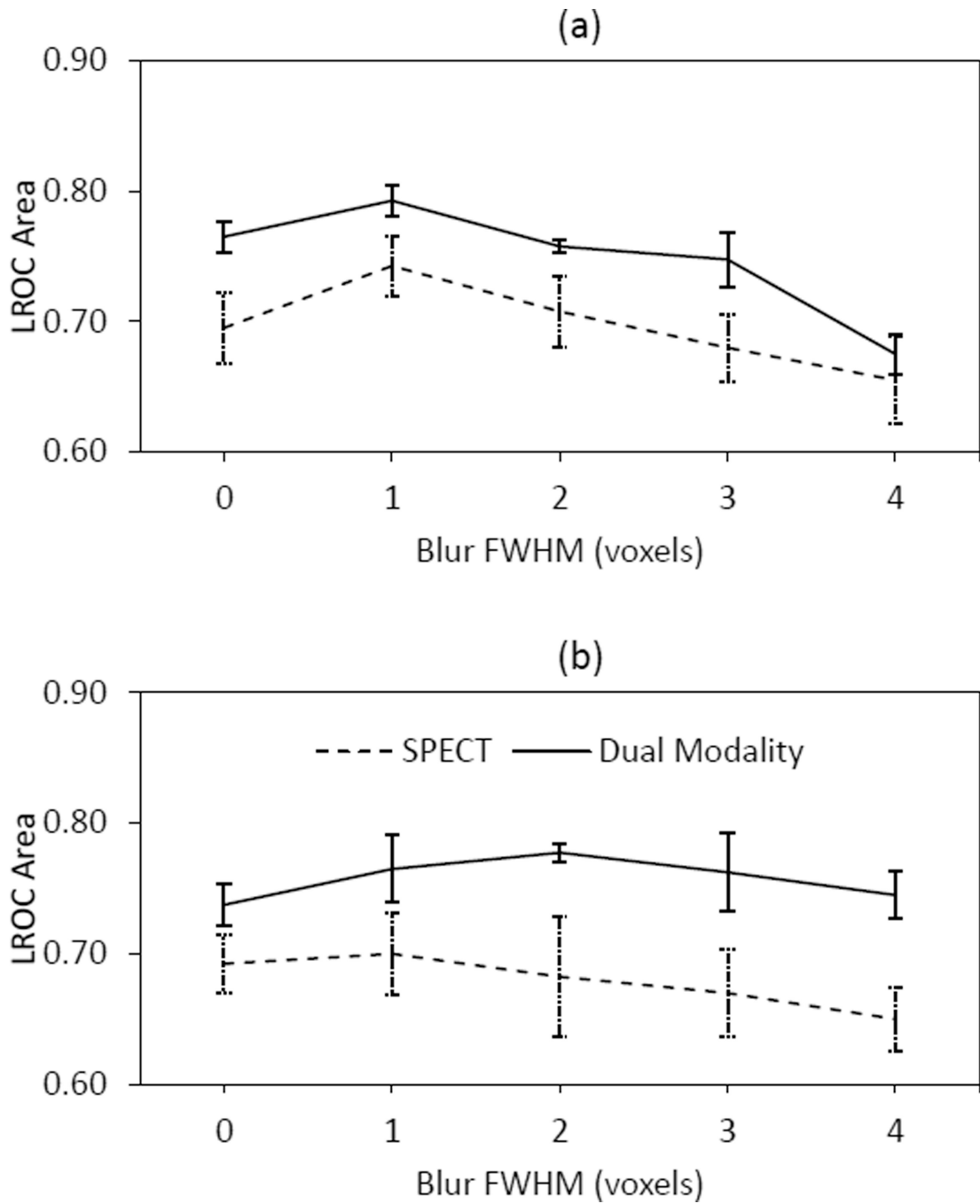


Fig. 4. Average human-observer A_L with the SPECT and dual-modality studies as a function of postfilter FWHM for (a) two and (b) six iterations.

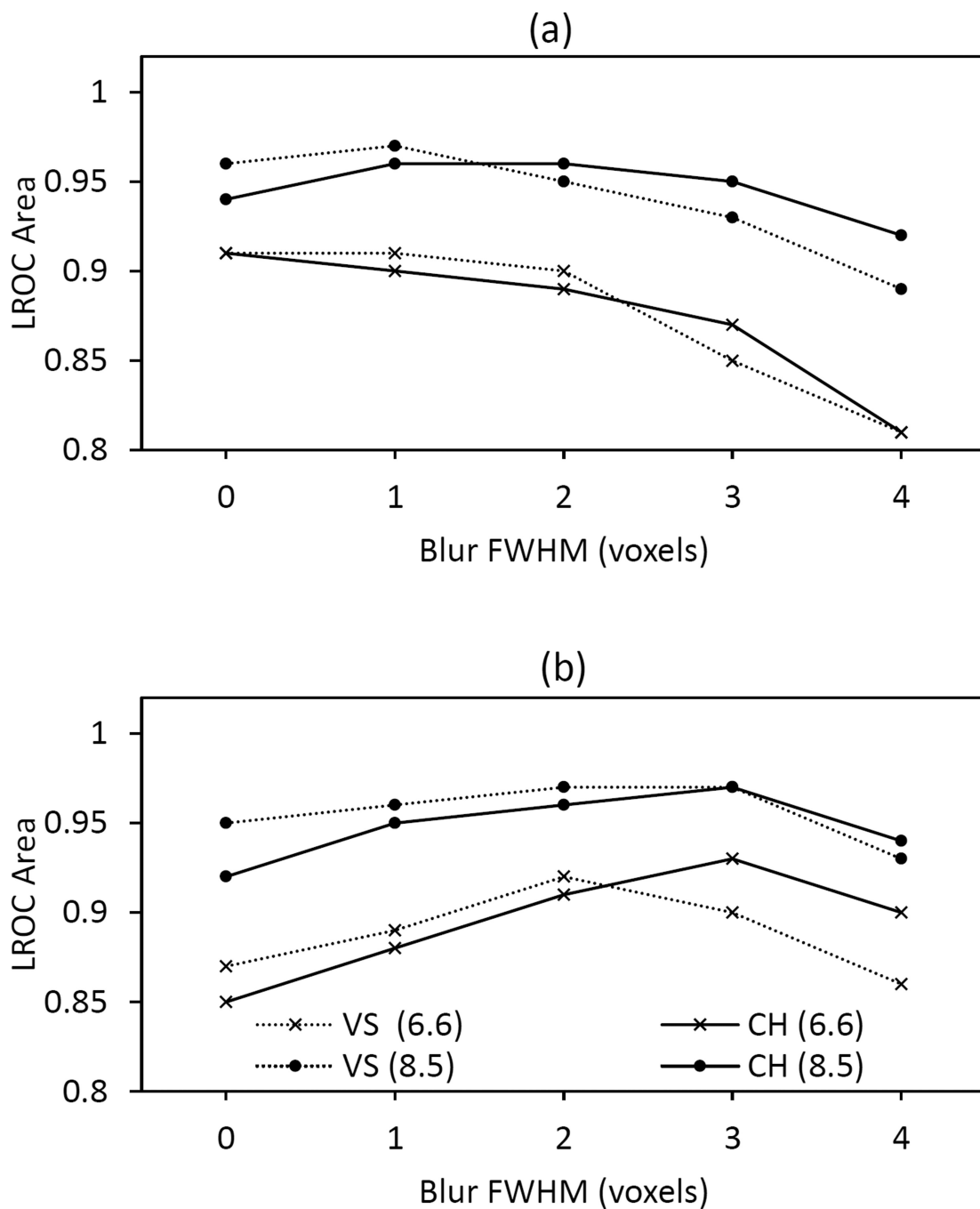


Fig. 5. Performance of the VS and CH observers with the contrast-specific image sets for a) two and b) six OSEM iterations. The lesion contrasts were 6.6 and 8.5. Each plot line shows observer performance for a given lesion contrast as a function of postfilter FWHM.

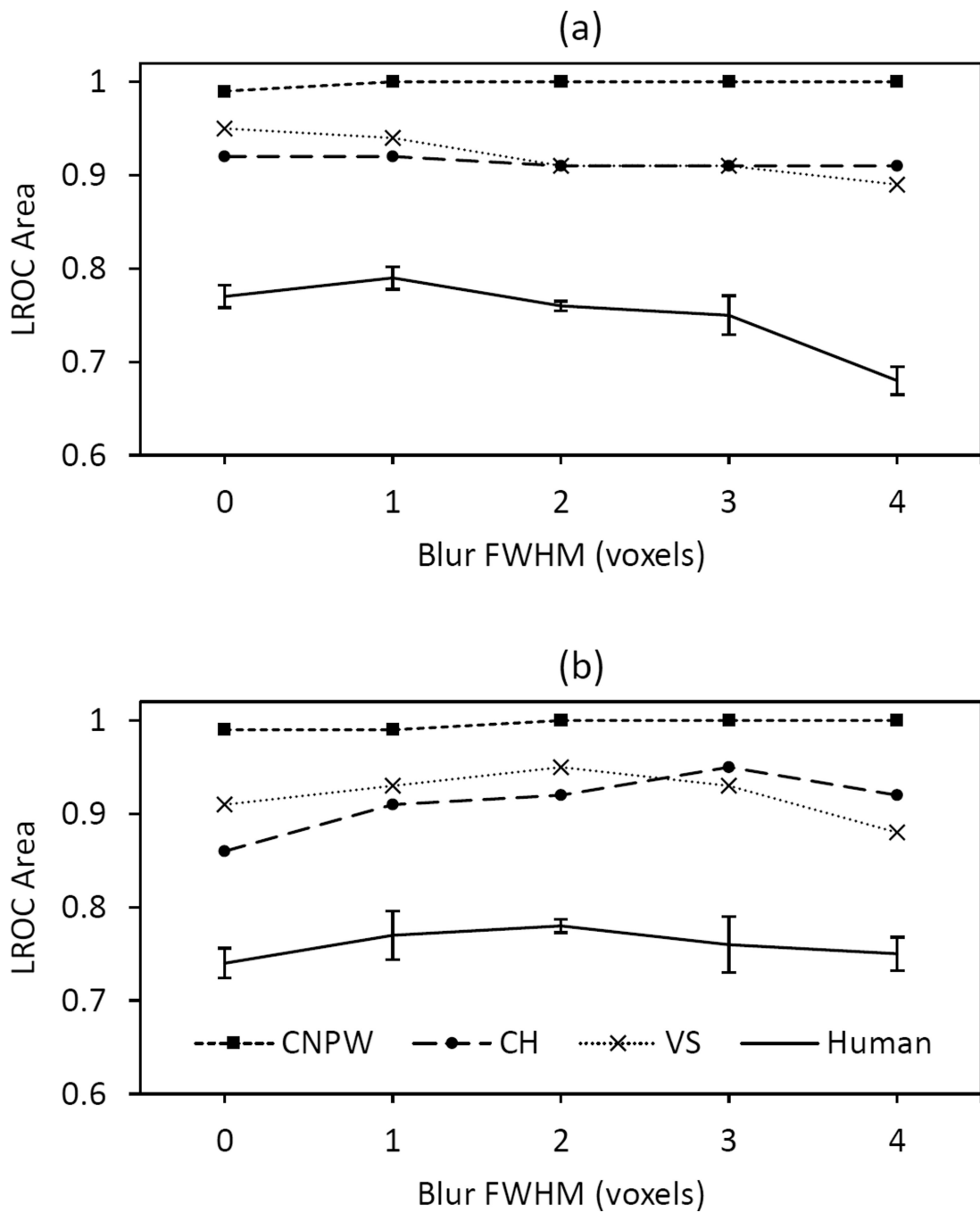


Fig. 6. Comparison of model and human-observer performance as a function of postfilter FWHM for (a) two and (b) six OSEM iterations. The human data is from the dual-modality study.

Organ mean (μ) and standard deviation (σ) uptake parameters for XCAT biodistribution sampling.

TABLE I

Parameter	Organ					
	Bladder	Blood Vessels	Bone Marrow	Liver	Lymph Nodes	Prostate
μ	3.4	9.7	6.2	10.0	3.0	4.0
σ	1.2	7.9	2.6	1.6	1.4	1.4

TABLE II

Results from the four-way ANOVA with the human observer scores. Study format, observer, OSEM iteration and postfilter blur were the main factors. Format, observer and blur effects, as well as the observer-format interaction, were significant at the $\alpha = 0.05$ level.

Factor	df	ss	F	Pr(>F)
Blur	4	0.041	10.03	8.4E-4
Format	1	0.084	82.78	9.8E-7
Iteration	1	2.4E-4	0.24	0.63
Observer	3	0.061	19.84	6.0E-5
Blur:Format	4	0.0018	0.44	0.78
Blur:Iteration	4	0.0098	2.41	0.11
Blur:Observer	12	4.1E-4	0.40	0.94
Format:Iteration	1	0.0036	3.57	0.083
Format:Observer	3	0.032	10.37	1.2E-3
Iteration:Observer	3	0.0013	0.43	0.74
Blur:Format:Iteration	4	0.0055	1.34	0.31
Blur:Format:Observer	12	0.014	1.19	0.38
Blur:Iteration:Observer	12	0.020	1.61	0.21
Format:Iteration:Observer	3	0.0025	0.82	0.50
Residual	12	0.012		