Published in final edited form as:

Phys Med Biol. 2009 May 7; 54(9): 2651–2666. doi:10.1088/0031-9155/54/9/004.

Efficiency of the human observer for lesion detection and localization in emission tomography

Bin Liu¹, Lili Zhou¹, Santosh Kulkarni², and Gene Gindi^{1,2}

Gene Gindi: gindi@mil.sunysb.edu

¹ Department of Radiology, Stony Brook University, NY 11790, USA (Phone: 631-444-2539)

² Department of Electrical & Computer Engineering, Stony Brook University, NY 11790, USA

Abstract

For the medically relevant task of joint detection and localization of a signal (lesion) in an emission computed tomographic (ECT) images, it is of interest to measure the efficiency, defined as the relative task performance of a human observer vs that of an ideal observer. Low efficiency implies that improvements in reconstruction algorithms may be possible. Calculation of ideal observer performance for ECT is highly computationally complex. We can, however, compute ideal observer performance exactly using a simplified “filtered-noise” model of ECT. This model results in images whose correlation structure, due to quantum noise, background variability and regularization, is similar to that of real ECT reconstructed images. A two-alternative forced choice test is used to obtain the performance of the human observers. We compare the efficiency of our joint detection-localization task with that of a corresponding signal-known-exactly (SKE) detection task. For the joint task, efficiency is low when the search tolerance is stringent. Efficiency for the joint task rises with signal intensity but is flat for the SKE task. For both tasks, efficiency peaks at a midrange level of regularization corresponding to a particular noise-resolution tradeoff.

1. Introduction

A principled approach to assess image quality in emission computed tomography (ECT) is to use a scalar task-based figure of merit (FOM) (Barrett and Myers 2004). These FOMs can be used to compare or optimize imaging systems or reconstruction algorithms. A commonly used task is the detection of a signal (a lesion) in a noisy complex image. In ECT, Poisson (photon) noise is present in the sinogram, but an important additional source of noise is the statistical variation in the spatial distribution of uptake of radionuclide due to anatomical and other effects in the underlying object (Barrett and Myers 2004). We shall term this form of noise background variability (BV).

Task FOMs can be evaluated using human or mathematical model observers. Model observers can be designed to emulate human performance (Barrett *et al.* 1993), but a different sort of model observer, the Bayesian ideal observer is used here. These ideal observers are mathematical (model) observers that deliver the best possible performance for the task at hand. They have knowledge of all relevant probability densities related to BV, photon noise, and any other sources of uncertainty. For example, for a signal present/absent detection problem, the Bayesian ideal observer maximizes a common FOM, the area under the receiver operating characteristic (ROC) curve, AROC (Barrett and Myers 2004). A human observer is not ideal, and its performance will be less than that of the ideal observer. A measure of their relative performance is called efficiency. A plethora of studies, for example (Tanner and Birdsall 1958, Burgess *et al.* 1981, 1982, 1997, Park *et al.* 2005, 2007), have explored human efficiency for different tasks and different forms of BV. If the ideal observer performance is much greater than that of the human, there is an implication

that further processing of the image might well be used to improve human performance (Rolland *et al.* 1991, Abbey *et al.* 2006). Thus it is of interest to obtain the knowledge of human efficiency relative to the ideal observer.

So far we have focused on detection tasks, but in this paper we are interested in the medically more realistic task of joint detection and localization of a signal in a noisy ECT image. This differs from the more conventional pure detection task of a known signal at a known location embedded in a noisy ECT image. Indeed we shall compare efficiency for these two tasks. Recently we have formulated an ideal observer for this joint detection-localization task (Khurd and Gindi 2005). This observer is optimal in that it maximizes the area under the localization ROC (LROC) curve (Swensson 1996).

We shall make use of the following acronyms: the “SKS (signal-known-statistically) task” shall specifically refer to the joint detection and localization of a signal in a noisy background. Here the signal is known exactly except for location. The “SKE (signal-known-exactly) task” shall specifically refer to the detection of a known signal in a noisy background. Here, the signal form and location are known exactly. We note that both the SKS and SKE tasks differ from a multiple-alternative forced choice (MAFC) task in which an observer is presented with an image containing one signal of known form but unknown location. The MAFC observer knows that the signal will be in any one of M precisely specified locations and his task is to try to correctly identify the signal location.

We also mention a different approach to image quality evaluation as espoused in the scan-statistics literature (Swensson 1996, Popescu and Lewitt 2006). This approach typically uses detection or joint detection and localization FOMs. A review of the scan-statistics methodology and an excellent resource of scan-statistics (and related) methods in imaging can be found in (Popescu and Lewitt 2006). In this approach, one eschews the often hard-to-obtain knowledge of probability densities related to any sources of uncertainty needed by ideal observers. Instead one takes a more practical approach, directly applying empirical (non-ideal) observers to samples of images to obtain histogram estimates of probabilities that can then be used to derive image-quality FOMs. In (Popescu and Lewitt 2006), PET versus time-of-flight PET are compared under a task context of detection of a signal at an unknown location.

We are interested in ECT, but the calculation of ideal observer performance for realistic SPECT or PET systems is highly computationally complex. To address this computational problem, we make a major simplification by using a “filtered-noise” model proposed in (Abbey and Barrett 2001). In (Abbey and Barrett 2001), the reconstructed ECT image is modelled as a stationary noise field with a possible added signal. The correlation structure of the noise field is designed to approximate the effects of: (1) propagating photon noise from a sinogram into the reconstruction via an FBP algorithm (Riederer *et al.* 1978); (2) propagating a form of BV (Rolland and Barrett 1992) into the reconstruction and (3) capturing the effects of any smoothing operations on the reconstruction. This model was used in (Abbey and Barrett 2001) not for efficiency studies, but to examine the effects of regularization, BV and photon noise on the performance of both human and human-emulating (non-ideal) model observers. These studies were also used to improve the predictive power of human-emulating observers.

The rest of the paper is organized as follows: In Section 2, we introduce the mathematical background. Experimental methods are presented in Section 3 and experimental results are given in Section 4. We conclude in Section 5 with a conclusion and discussion.

2. Mathematical Background

Scalars and scalar-valued functions are denoted by unbolded letters. Vectors are indicated by bolded lowercase letters. Although we use 2-D images, we use lexicographically ordered vectors to represent images, with each element containing the intensity of an image pixel. Bolded uppercase letters are used to represent matrices.

2.1. Joint detection-localization task

We now focus on the mathematical description of the SKS task. A test image $\mathbf{y}$ is either signal-absent or includes a signal that is present at one of J locations, and the objective of an observer is to determine which class the test image $\mathbf{y}$ belongs to. This task is a $(J + 1)$ -class classification problem (also known as multiple hypothesis-testing problem) with classes (hypotheses) expressed as:

$$\begin{aligned} H_0: \mathbf{y} &= \mathbf{b} + \mathbf{n}, \\ H_j: \mathbf{y} &= \mathbf{s}_j + \mathbf{b} + \mathbf{n}, \quad (j=1, 2, \dots, J), \end{aligned} \quad (1)$$

where $\mathbf{n}$ is an $M \times 1$ vector of noise with mean zero. For class H_0 , the test image $\mathbf{y}$ comprises a $M \times 1$ background image $\mathbf{b}$ plus zero-mean noise $\mathbf{n}$, but has no signal. For class H_j , $\mathbf{y}$ also contains an $M \times 1$ signal (lesion) vector $\mathbf{s}_j$ present at the j^{th} location. The form of the signal is assumed to be known and independent of j . Note that when $J = 1$, we obtain the SKE task.

We assume that an observer performs the SKE task by computing a scalar test statistic t based on $\mathbf{y}$. The scalar test statistic is compared to a threshold τ to decide signal-present if $t(\mathbf{y}) \geq \tau$ and signal absent if $t(\mathbf{y}) < \tau$. For each threshold τ , one calculates $P_{TP}(\tau)$ indicating the probability of deciding signal-present when a signal is actually present, and $P_{FP}(\tau)$ indicating the probability of deciding signal-present when a signal is actually not present. An ROC curve can be generated by plotting $P_{TP}(\tau)$ as a function of $P_{FP}(\tau)$ as the threshold τ is swept. Figure 1 shows a hypothetical ROC curve. The area under the ROC curve, AROC, is a well-known FOM for the SKE detection problem.

In order to introduce location uncertainty into the detection problem, we consider the localization ROC (LROC) curve (Swensson 1996). In addition to reporting whether or not the image contains a signal, the observer also needs to localize the signal within a search tolerance in a signal-present image. As shown in Figure 1, the LROC curve plots the probability of correct joint detection and localization $P_{CL}(\tau)$ as a function of $P_{FP}(\tau)$. As with AROC, the area under the LROC curve (ALROC) is a FOM for the joint detection and localization problem. The details of generating a $t(\mathbf{y})$ and a location estimate, as well as quantities $P_{CL}(\tau)$ and $P_{FP}(\tau)$ for this task are given below. Note that $AROC \geq 0.5$ (Barrett and Myers 2004) whereas ALROC can be < 0.5 (Swensson 1996).

2.2. Ideal observer for a joint detection and localization task

This ideal observer (Khurd and Gindi 2005) searches for the signal in the test image $\mathbf{y}$ and reports whether the signal is present, and if so, reports one of L candidate locations. We decide the signal is correctly localized if the reported location is reasonably close to the true signal location, i.e., within a circular tolerance region surrounding the true location.

Let P_j , $j = 0, 1, \dots, J$, denote the prior probabilities of each class and $p(\mathbf{y}|H_j)$, $j = 0, 1, \dots, J$, denote the data pdfs conditioned on each hypothesis. By defining $T(l)$, $l = 1, \dots, L$, as a tolerance region centered at location l , we decide that the signal is correctly detected and localized if the signal is actually present, the observer reports the l^{th} location, and the true

signal location is within $T(l)$. From (Khurd and Gindi 2005), the ideal observer that maximizes the ALROC for this SKS task is

$$\begin{aligned} t_{SKS}(\mathbf{y}) &= \max_{l \in \{1, \dots, L\}} \sum_{j \in T(l)} P_j LR(\mathbf{y}, H_j) \\ l(\mathbf{y}) &= \arg \max_{l \in \{1, \dots, L\}} \sum_{j \in T(l)} P_j LR(\mathbf{y}, H_j) \\ \text{Report signal in } T(l(\mathbf{y})) &\text{ if } t_{SKS}(\mathbf{y}) > \tau, \text{ else decide } H_0 \end{aligned} \quad (2)$$

where the likelihood ratio LR is defined as

$$LR(\mathbf{y}, H_j) \equiv \frac{p(\mathbf{y}|H_j)}{p(\mathbf{y}|H_0)}, \quad j=1, \dots, J \quad (3)$$

For the SKE task, we are given a signal with *known* location j . The ideal observer $t_{SKE} = LR(\mathbf{y}, H_j)$ simply computes the likelihood ratio for this signal. We decide signal present if $t_{SKE} \geq \tau$.

2.3. Human observer for a joint detection and localization task

We use a two-alternative forced-choice (2AFC) test to evaluate the performance of a human observer. In a 2AFC test appropriate for the SKS task, an observer is shown many pairs of test images. Each pair, as seen in Figure 2, comprises a signal-absent image $\mathbf{y}$ and a signal-present image $\mathbf{y}'$. The observer is then forced to choose which image contains the signal and report the location of the signal. We assume the observer forms two internal responses (test statistics), $\lambda^+ = \lambda(\mathbf{y}')$ and $\lambda^- = \lambda(\mathbf{y})$, and chooses the image with larger response as the signal-present image. A candidate signal location l is also reported by the observer. Therefore, the observer correctly detects the signal-present image if $\lambda^+ \geq \lambda^-$ and correctly localizes the signal if the true signal location is within $T(l)$. The proportion of correctly detected and localized images, P_C , is computed after showing many pairs of images to the human observer. It can be shown that, for a human observer experiment, ALROC equals to P_C in the limit of a large number of image samples. The detailed derivation is given in Appendix A. An alternative derivation can be found in (Clarkson 2007). Unlike the SKS case, the 2AFC test for the SKE case is well-established (Barrett and Myers 2004) and discussed in Section 3.3. In Section 3.3, we also give further methodological details of our 2AFC-SKS implementation.

3. Methods

3.1. Filtered noise model

Consider a general ECT system where the sinogram $\mathbf{g}$ is given by

$$\mathbf{g} = \mathcal{H}\mathbf{f} + \mathbf{n}_p \quad (4)$$

where $\mathbf{n}_p$ is the Poisson noise, and $\mathcal{H}$ is a system matrix, which for this case would be a digital version of the Radon transform. The object $\mathbf{f}$ comprises a random background and may include a signal $\mathbf{s}_j'$. The reconstruction, $\hat{\mathbf{f}}$, can be expressed as

$$\hat{\mathbf{f}} = \Theta \mathbf{g} = \Theta (\mathcal{H} \mathbf{f} + \mathbf{n}_p)$$

where we assume the reconstruction operator Θ is an FBP operator that may include regularization.

Take the object $\mathbf{f} = \mathbf{s}'_j + \mathbf{b}' + \mathbf{n}_b$, where $\mathbf{s}'_j$ is the signal present in the object domain ($\mathbf{s}'_j = 0$ for a signal-absent object), $\mathbf{b}'$ is the *deterministic* part of the background in the object domain and $\mathbf{n}_b$ is zero mean BV in the object. Then the reconstructed image $\hat{\mathbf{f}}$ can be written as

$$\hat{\mathbf{f}} = (\Theta \mathcal{H} \mathbf{s}'_j + \Theta \mathcal{H} \mathbf{b}') + (\Theta \mathcal{H} \mathbf{n}_b + \Theta \mathbf{n}_p) \quad (5)$$

where the first two terms are deterministic and the latter two terms random. The first term, $\Theta \mathcal{H} \mathbf{s}'_j$, is the noiseless reconstructed signal and the second term, $\Theta \mathcal{H} \mathbf{b}'$, is the noiseless reconstructed deterministic background. The third term, $\Theta \mathcal{H} \mathbf{n}_b$, is the object background variability propagated into the reconstruction, and the fourth term, $\Theta \mathbf{n}_p$, is the quantum noise propagated into the reconstruction. Therefore, the correlation structure of the reconstructed image $\hat{\mathbf{f}}$ is due to both background variability and quantum noise.

The ECT model (5) can be approximated by the filtered-noise model introduced in (Abbey and Barrett 2001). This noise model is used to emulate the tomographic reconstruction process wherein the correlation structure of the reconstructed images is a combination of propagated quantum noise and BV modulated by some form of regularization. By “emulate”, we mean that no actual reconstruction is done; instead, a noisy image is generated whose noise structure is designed to be that of a reconstruction.

We now describe the process of generating filtered noise. The filtered noise $\mathbf{n}_f$ can be generated by the process $\mathbf{n}_f = \mathbf{F}^H \mathbf{\Lambda} \mathbf{F} \mathbf{n}_0$, where $\mathbf{n}_0$ is zero-mean i.i.d. Gaussian noise with unit variance, $\mathbf{F}$ is the DFT matrix, and $\mathbf{\Lambda}$ is a diagonal matrix that defines the discrete transfer function of the noise generation process. Note that $\mathbf{n}_f$ follows a multivariate Gaussian distribution. Following the notation in (Abbey and Barrett 2001), the discrete transfer function $\mathbf{\Lambda}$ is defined as

$$\mathbf{\Lambda} \equiv \mathbf{B}(\mathbf{N}_q + \mathbf{N}_a)^{1/2}.$$

Here the background noise-power spectrum (NPS) is denoted by $\mathbf{N}_a$ which represents the fluctuations in the reconstructed images due to BV. The quantum NPS, $\mathbf{N}_q$, represents the fluctuations in the reconstructed images from quantum noise in the sinogram. The regularization imposed by the reconstruction algorithm is incorporated through $\mathbf{B}$. Given these definitions, the random terms in the reconstructed image, $\Theta \mathcal{H} \mathbf{n}_b + \Theta \mathbf{n}_p$, can be modelled by $\mathbf{n}_f$. The deterministic signal and background terms in the reconstructed image, $\Theta \mathcal{H} \mathbf{s}'_j$ and $\Theta \mathcal{H} \mathbf{b}'$, can be modelled by $\mathbf{s}'_f$ and $\mathbf{b}_f$, respectively, where

$$\mathbf{s}'_f = \mathbf{F}^H \mathbf{B} \mathbf{F} \mathbf{s}'_j, \quad \mathbf{b}_f = \mathbf{F}^H \mathbf{B} \mathbf{F} \mathbf{b}',$$

also include the effects of regularization.

Given the filtered noise model, we can rewrite (1) as

$$\begin{aligned} H_0: \mathbf{y} &= \mathbf{b}_f + \mathbf{n}_f, \\ H_j: \mathbf{y} &= \mathbf{s}_f^j + \mathbf{b}_f + \mathbf{n}_f, \quad j=1, 2, \dots, J. \end{aligned} \quad (6)$$

and since $\mathbf{n}_f$ is multi-variate Gaussian, the corresponding likelihood functions can be written as

$$\begin{aligned} p(\mathbf{y}|H_0) &= \frac{1}{(2\pi)^{M/2} |\mathbf{K}_{\mathbf{n}_f}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{b}_f)^T \mathbf{K}_{\mathbf{n}_f}^{-1} (\mathbf{y} - \mathbf{b}_f)\right), \\ p(\mathbf{y}|H_j) &= \frac{1}{(2\pi)^{M/2} |\mathbf{K}_{\mathbf{n}_f}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{b}_f - \mathbf{s}_f^j)^T \mathbf{K}_{\mathbf{n}_f}^{-1} (\mathbf{y} - \mathbf{b}_f - \mathbf{s}_f^j)\right), \end{aligned} \quad (7)$$

where $\mathbf{K}_{\mathbf{n}_f}$ is the covariance matrix of the filtered noise $\mathbf{n}_f$ and is derived as

$$\mathbf{K}_{\mathbf{n}_f} = \mathbf{F}^H \mathbf{\Lambda} \mathbf{F} (\mathbf{F}^H \mathbf{\Lambda} \mathbf{F})^H = \mathbf{F}^H \mathbf{B}^2 (\mathbf{N}_a + \mathbf{N}_q) \mathbf{F}. \quad (8)$$

For the components in the discrete transfer function $\mathbf{\Lambda}$, we use the same functional forms as in (Abbey and Barrett 2001). For fluctuations due to background variability, they recommend the use of inverse power law noise, whose isotropic NPS, $\mathbf{N}_a$, is expressed as

$$[\mathbf{N}_a]_{kk} = \frac{W_a}{1 + \left(\frac{\rho_k}{\rho_a}\right)^\beta}$$

where $[\mathbf{X}]_{kk}$ denotes the k^{th} diagonal element of matrix $\mathbf{X}$ and ρ_k is the radial frequency. The parameter W_a controls the magnitude of the background fluctuations in the images and the exponent β controls the rate of falloff of the NPS. The parameter ρ_a is a constant that acts approximately as a frequency-axis scaling parameter. To emulate an FBP-like reconstruction, the functional form of the quantum NPS, $\mathbf{N}_q$, is ramplike in frequency space (Riederer *et al.* 1978). Abbey and Barrett (2001) use the form

$$[\mathbf{N}_q]_{kk} = \begin{cases} W_q \rho_k & \text{if } \rho_k \geq \rho_q, \\ W_q \rho_q & \text{if } \rho_k < \rho_q \end{cases},$$

where ρ_q is a constant that imposes a small DC component near the origin for normalization purposes, and W_q controls the slope of the ramp. Abbey and Barrett (2001) use an isotropic Butterworth filter as the apodizing filter with a functional form given by

$$[\mathbf{B}]_{kk} = \frac{1}{1 + \left(\frac{\rho_k}{\rho_c}\right)^{2\nu}}$$

where ρ_c is the cutoff point of the filter and the order ν of the Butterworth filter determines how fast the filter falls off near the cutoff point. Note that $\mathbf{B}$ controls the noise-resolution tradeoff in the reconstruction.

3.2. Calculating the performance of the ideal observer

For the SKS task, without loss of generality, we assume the signal is uniformly distributed at all possible locations. We use a circle as the tolerance region and represent tolerance by its

radius r_{tol} . The signal $\mathbf{s}_j$ is a Gaussian blob of peak intensity a_G and fixed width σ_G . The background $\mathbf{b}'$ is assumed to be uniform.

To calculate ALROC in a simulation experiment, we generate $N^+ = 5000$ signal-present images and $N^- = 4000$ signal-absent images. Unlike the case for real ECT, for the filtered-noise model, the likelihood ratio $LR_f(\mathbf{y}, H_j)$ can be written exactly in a simple closed-form expression using (3) and (7) to obtain,

$$LR_f(\mathbf{y}, H_j) = \exp\left((\mathbf{y} - \mathbf{b}_f)^T \mathbf{K}_{\mathbf{n}_f}^{-1} \mathbf{s}_f^j - \frac{1}{2} (\mathbf{s}_f^j)^T \mathbf{K}_{\mathbf{n}_f}^{-1} \mathbf{s}_f^j\right). \quad (9)$$

Using (2) and (9), we compute the corresponding observer responses t^+ for each signal-present image and t^- for each signal-absent image and we also obtain the observer-reported location l for each signal-present image. If the reported location for one signal-present image leads to an incorrect localization, i.e., the actual signal location $j \notin T(l)$, we simply discard the corresponding t^+ . We then determine a set of N_τ thresholds from the range of the values of t^+ and t^- . For each threshold τ , we can obtain $P_{CL}(\tau)$ and $P_{FP}(\tau)$ by computing the fraction of the observer responses that exceed the threshold, i.e.,

$$P_{CL}(\tau) = \frac{\text{Number of correctly localized } t^+ \text{ that exceeds } \tau}{N^+} \\ P_{FP}(\tau) = \frac{\text{Number of } t^- \text{ that exceeds } \tau}{N^-}$$

Note that only those t^+ with correct localization are counted in computing $P_{CL}(\tau)$. The LROC curve can then be plotted and ALROC can then be calculated by trapezoidal integration. For the SKE case, the process is the same in computing AROC but no localization is involved, and we use $t_{SKE} = LR_f(\mathbf{y}, H_j)$ instead of (2).

3.3. Calculating the performance of the human observer

As seen in Figure 2, each observer in a 2AFC test for an SKS task is shown a pair of test images on a Sony Multiscan200ES CRT monitor. The 128×128 test image is magnified to 256×256 pixels using bilinear interpolation so that the image subtends a reasonable solid angle as viewed by the human observer. The human is free to adjust her viewing position. The image containing the signal is randomly determined to be either on the left or the right side with equal probability. A signal located at the center of an empty background is also shown to the observer. The observer is forced to choose with a mouse-click which of the two images contains the signal and where in that image the signal is located. The observer is presented with 100 image pairs for training and 300 pairs for testing. In the training session, the correctness of the observer's answer is immediately reported. If the observer chooses the correct signal-present image and the distance between the true signal location (indicated by a "+" in Figure 2) and the observer reported-location (indicated by a "x" in Figure 2) is shorter than the radius of tolerance region, then the signal is deemed to be correctly detected and

localized. In the case in Figure 2, this pair of images is counted as one correct detection and localization. In the testing session, no feedback is provided to the observer except the final report of P_C . While Figure 2 illustrates the 2AFC procedure, Figure 3 better illustrates the qualitative nature of the images viewed by the human observers.

For the 2AFC test for the SKE case, the training and testing procedures are similar to that of the SKS case. The only difference is that the signal location is pointed out by a crosshair. The observer is forced only to choose which side he thinks the signal is on. For all SKE and SKS human experiments, performance was averaged over four observers. Error bars in performance represent 68% confidence intervals.

A photometer from Quantum Instruments Inc. model PMLX was used to calibrate the monitor. For each grey level [0–255], the photometer reading of luminance in (cd/m^2) was recorded. The plot of luminance vs. grey value was nearly flat from [0–50] and monotonically rising (approximately quadratically) from [50–255]. The histogram of grey values of all displayed images fit within this monotonic region.

3.4. Efficiency definitions

A commonly used definition of efficiency for the SKE case is the squared ratio of the detectability index of the human observer, d_h , and the ideal observer, d_i (Tanner and Birdsall 1958), i.e., $e_1 = (d_h/d_i)^2$. A common expression for d (Burgess *et al.* 1981) is defined as $d \equiv 2 \operatorname{erf}^{-1}[2(\operatorname{AROC}) - 1]$. However, this definition of d is appropriate only for a detection task due to the fact that it requires $\operatorname{AROC} \geq 0.5$ lest d becomes negative. For SKS, we cannot simply substitute ALROC in the above definition of d since ALROC can be < 0.5 . Thus for an SKS task, we use $d \equiv 2 \operatorname{erf}^{-1}(\operatorname{ALROC})$ which solves these problems.

We introduce a second definition of human efficiency which is simply the ratio of the AUC (area under curve) of the human observer and the ideal observer, i.e.,

$$e_2 = \frac{AUC_{human}}{AUC_{ideal}}, \quad (10)$$

where AUC equals AROC or ALROC as appropriate.

4. Results

Several experiments were implemented to compare the SKE and SKS performance of ideal and human observers for this filtered-noise ECT model. It is of interest (Park *et al.* 2005) to perform psychometric studies to see how efficiency varies with signal and background quantities, and in this study we have focused on the variation of efficiency with two important variables: signal amplitude and level of regularization. The two parameters varied were a_G to control signal intensity, and cutoff frequency ρ_c to control regularization. The level of regularization controls the noise-resolution tradeoff in the image, as seen in Figure 3. Further insight can be gained by examining quantities in the frequency domain, as seen in Figure 4. Examination of the spectra in Figure 4(b) shows a correspondence with the anecdotal realizations in Figure 3 (i.e., lower ρ_c implies blurrier image). Figure 5 illustrates the information loss in the signal due to regularization. Other parameters were fixed as follows: $W_a = 284\,949$, $\rho_a = 0.0156$, $\beta = 3$, $W_q = 6145.3$, $\rho_q = 0.0078$, $v = 4$, and $\sigma_G^2 = 5$. The deterministic background image $\mathbf{b}'$ was a constant with pixel value = 5. Thus the parameters controlling background variability and quantum noise were fixed.

One of our goals is to compare SKE and SKS performance. However, the results for SKS experiments will vary as the tolerance radius r_{tol} changes. For example, with ρ_c and a_G fixed, increasing r_{tol} increases P_{CL} , which increases ALROC. To study the effects of r_{tol} for the SKS task, we conducted ideal and human observer experiments. Figure 6(a) shows performance as a function of r_{tol} for both ideal and human observers. As seen in Figure 6(a), the performance for both the ideal observer and the human observer increase quickly in the small radius region and then become relatively flat in the large radius region. This phenomenon can be explained as follows: When the signal-present image is correctly detected, it is difficult for the observer to precisely locate the center of the signal, but it is much easier to estimate a rough location of the signal center, i.e., within a small region about the center the signal. Beyond this small region, the signal is grossly mis-localized and increasing r_{tol} provides only a small benefit in performance. Figure 6(b) presents the efficiency curves as a function of r_{tol} for both of our definitions of efficiency. The efficiency of the human observer increases quickly in the small-radius region and becomes relatively flat in the large-radius region. Thus, the human observer is even worse than the ideal observer in precisely localizing the center of the signal. Note that from Figure 6, a natural tolerance with which to evaluate the effects of signal amplitude and regularization is the curve shoulder at $r_{tol} = 3$. We thus use $r_{tol} = 3$ in our subsequent experiments.

With the tolerance issue addressed, we can now compare SKE and SKS performance for the two psychometric experiments. In Figure 7(a), we vary the signal intensity a_G from 45 to 85 and plot the SKS performance curves for both ideal and human observers. The relative efficiency curves are plotted in Figure 7(b). As expected, the performances of both ideal and human observers increase monotonically as the signal intensity increases. However, the efficiency curves have a slight dip at an intermediate intensity.

The corresponding SKE results are also shown in Figure 7, where the intensity range is lower in order to get ideal observer performance equivalent to that of the SKS case (ALROC for the highest 3 intensities matches AROC for intensities $a_G = 11, 24, 37$). As seen in Figure 7(a), we again observe the monotonic (with intensity) trend in performance. However, in Figure 7(b) we observe relatively flat efficiency curves in the selected intensity region. The efficiency curve for SKS rises over its intensity range since the human is better able (relative to the ideal observer) to take advantage of the increased localization ability afforded by brighter signals. For the corresponding SKE intensity range, no such rise is seen because no localization is involved in the SKE task.

To study the effects of regularization on performance, we vary the cutoff frequency ρ_c of the Butterworth filter. For the SKS case, as shown in Figure 8(a), the cutoff frequency has no effect on the performance of the ideal observer (the slight variation in the ideal observer curve is due to finite sample effects). Indeed, it is readily shown mathematically that ALROC does not vary with ρ_c . Clearly the cutoff has much more effect on the human observer. The performance of the human observer is degraded for both the over-smoothed case (low cutoff frequency) and under-smoothed case (high cutoff frequency) and the performance degradation is more severe for the over-smoothed case. This phenomenon is also present for both definitions of efficiency, as seen in Figure 8(b). The explanation is that for high ρ_c , we retain signal information but also include much of the quantum noise, thus degrading performance, but for low ρ_c , we lose signal information, also degrading performance, in this case even more so than the degradation for high ρ_c . Indeed, this explanation is manifested on the frequency plots in Figure 4 and Figure 5. Since the ideal observer performance is flat with ρ_c , efficiency shows the same trend as human performance.

The corresponding SKE results are also shown in Figure 8. The range of cutoff frequencies is the same as that of the SKS case and again, the ideal observer performance is independent of ρ_c . The signal intensity a_G is adjusted to attain the same ideal observer performance as for the SKS case (as seen in Figure 8(a)). Human observer performance follows the same trend as for the SKS task.

Examination of Figures 8(a) and 8(b) shows that SKE efficiency apparently exceeds that of SKS for either definition of efficiency. However, comparison of efficiencies for these two cases is not meaningful due to the different nature of the SKE and SKS tasks.

5. Conclusions and Discussions

We have, for the first time, calculated the efficiency of a human observer relative to the ideal observer for a joint detection and localization task. The effects of task, tolerance radius, signal intensity and cutoff frequency on human efficiency were analyzed in the context of a simplified filtered-noise model for ECT. For the SKS task, results showed that both the ideal and human observers performed poorly in localizing the precise signal center but performed well in localizing the rough location of the signal center. Psychometric tests with varying signal intensity showed for the SKS task a rising trend of efficiency with intensity. For the SKE task, this trend was approximately flat. Psychometric tests with varying cutoff frequency showed a similar trend for both tasks: efficiency peaked at a mid-range cutoff corresponding to a particular noise-resolution tradeoff.

While there is an extensive literature on human efficiency in visual tasks, the most relevant to our work is that in (Park *et al.* 2005, 2007). They studied human efficiency for a planar pinhole emission imaging system that delivered images containing a form of “lumpy background” BV, uncorrelated photon noise, and smoothing (regularization) due to the finite size of the pinhole. Ideal observer performance and human 2AFC tests were done for an SKE detection task and an SKS task in which the observer was required to only detect - but not localize - the signal. For the psychometric intensity tests in (Park *et al.* 2005), performance for the SKS-detection-only task was better than that for SKE.

It is also relevant to mention the work of Gifford *et al.* (2003, 2005) who addressed the joint detection-localization task for SPECT. However, they did not consider ideal observers and instead focused on model observers designed to emulate the performance of humans.

We have addressed task performance in ECT, but used the filtered-noise model approximation to allow exact calculation of ideal observer performance. Clearly this study needs to be extended toward more realistic ECT. If BV is excluded, then we can indeed calculate ideal observer performance exactly for realistic SPECT (Liu *et al.* 2008). If we include Gaussian BV, we can calculate approximate ideal observer performance for realistic SPECT (Zhou *et al.* 2008). In future work, we shall address computational problems in calculating good approximations for ideal observer performance for more realistic ECT models.

Acknowledgments

This work was supported by NIH NIBIB 02629.

References

- Abbey CK, Barrett HH. Human- and model-observer performance in ramp-spectrum noise: effect of regularization and object variability. *J Opt Soc Am A* 2001;18(3):473–488.

- Abbey CK, Zemp RJ, Liu J, Lindfors KK, Insasa MF. Observer efficiency in discrimination tasks simulating malignant and benign breast lesions imaged with ultrasound. *IEEE Trans Medical Imaging* 2006;25(2):198–209.
- Barrett, HH.; Myers, KJ. *Foundations of Image Science*. Wiley Interscience; 2004.
- Barrett, HH.; Yao, J.; Rolland, JP.; Myers, KJ. Model observers for assessment of image quality. *Proc. of the National Academy of Sciences*; 1993. p. 9758–9765.
- Burgess AE, Li X, Abbey CK. Visual signal detectability with two noise components: anomalous masking effects. *J Opt Soc Am A* 1997;14(9):2420–2442.
- Burgess AE, Wagner RF, Jennings RJ. Statistical efficiency: a measure of human visual signal detection performance. *J Appl Photogr Eng* 1982;8:76–78.
- Burgess AE, Wagner RF, Jennings RJ, Barlow HB. Efficiency of human visual signal discrimination. *Science* 1981;214:93–94. [PubMed: 7280685]
- Clarkson E. Estimation receiver operating characteristic curve and ideal observers for combined detection/estimation tasks. *J Opt Soc Am A* 2007;24(12):91–98.
- Gifford HC, King MA, Pretorius PH, Wells RG. A comparison of human and model observers in multislice LROC studies. *IEEE Trans Medical Imaging* 2005;24(2):160–169.
- Gifford, HC.; Pretorius, PH.; King, MA. Comparison of human- and model-observer LROC studies. *Proc. SPIE*; 2003. p. 112–122.
- Khurd P, Gindi G. Decision strategies that maximize the area under the LROC curve. *IEEE Trans Medical Imaging* 2005;24(12):1626–1636.
- Liu, B.; Zhou, L.; Kulkarni, S.; Gindi, G. Effects of background complexity on LROC analysis of SPECT. *Conf. Rec. IEEE Nucl. Sci. Symp. Med. Imaging*; 2008. p. M06–479
- Park S, Clarkson E, Kupinski MA, Barrett HH. Efficiency of the human observer detecting random signals in random backgrounds. *J Opt Soc Am A* 2005;22(1):3–16.
- Park S, Gallas BD, Badano A, Petrick NA, Myers KJ. Efficiency of the human observer for detecting a Gaussian signal at a known location in non-Gaussian distributed lumpy backgrounds. *J Opt Soc Am A* 2007;24(4):911–921.
- Popescu LM, Lewitt RM. Small nodule detectability evaluation using a generalized scan statistic model. *Phys Med Biol* 2006;51(23):6225–6244. [PubMed: 17110782]
- Riederer SJ, Pelc NJ, Chesler DA. The noise power spectrum in computed X-ray tomography. *Phys Med Biol* 1978;23(3):446–454. [PubMed: 674361]
- Rolland JP, Barrett HH. Effect of random background inhomogeneity on observer detection performance. *J Opt Soc Am A* 1992;9(5):649–658. [PubMed: 1588452]
- Rolland JP, Barrett HH, Seeley GW. Ideal versus human observer for long-tailed point spread functions: does deconvolution help? *Phys Med Biol* 1991;36(8):1091–1109. [PubMed: 1924544]
- Swensson RG. Unified measurement of observer performance in detecting and localizing target objects on images. *Med Phys* 1996;23(10):1709–1725. [PubMed: 8946368]
- Tanner WP, Birdsall TG. Definitions of d' and η as psychophysical measures. *J Acoust Soc Am* 1958;30(10):922–928.
- Zhou, L.; Gindi, G. Collimator optimization in SPECT based on an LROC ideal observer. *Conf. Rec. IEEE Nucl. Sci. Symp. Med. Imaging*; 2008. p. M15–2

Appendix A. Equivalence between the ALROC of a human observer and the probability of correct detection and localization in a 2AFC test

The probability of correct detection and localization in a 2AFC test can be written as $P_C = P_C = P(\lambda^+ \geq \lambda^-) P(\text{correct localization})$

$$P_C = P(\lambda^+ \geq \lambda^-) P(\text{correct localization})$$

$$= \sum_{j=1}^J P_j \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(j \in T(l) | H_j) \text{step}(\lambda^+ - \lambda^-) p(\lambda^+ | H_j) p(\lambda^-) d\lambda^+ d\lambda^- \quad (\text{A.1})$$

where

$$\text{step}(x) = \begin{cases} 1, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Since both λ^+ and λ^- are random variables through their dependence on the test images $\mathbf{y}'$ and $\mathbf{y}$ respectively, we have

$$\begin{aligned} p(\lambda^+|H_j) &= \int_{\mathbf{y}'} p(\mathbf{y}'|H_j) \delta(\lambda^+ - \lambda(\mathbf{y}')) d\mathbf{y}', j=1, \dots, J \\ p(\lambda^-) &= \int_{\mathbf{y}} p(\mathbf{y}|H_0) \delta(\lambda^- - \lambda(\mathbf{y})) d\mathbf{y} \end{aligned}$$

and (A.1) becomes

$$P_c = \sum_{j=1}^J P_j \int_{-\infty}^{\infty} d\lambda^+ \int_{-\infty}^{\infty} d\lambda^- \int_{\mathbf{y}'} \int_{\mathbf{y}} P(j \in T(l(\mathbf{y}'))|H_j) \text{step}(\lambda^+ - \lambda^-) \times p(\mathbf{y}'|H_j) \delta(\lambda^+ - \lambda(\mathbf{y}')) p(\mathbf{y}|H_0) \delta(\lambda^- - \lambda(\mathbf{y})) d\mathbf{y}' d\mathbf{y} \quad (\text{A.2})$$

Taking the integrals over λ^+ and λ^- , we have

$$P_c = \sum_{j=1}^J P_j \int_{\mathbf{y}'} \int_{\mathbf{y}} p(\mathbf{y}|H_0) p(\mathbf{y}'|H_j) P(j \in T(l(\mathbf{y}'))|H_j) \text{step}(\lambda(\mathbf{y}') - \lambda(\mathbf{y})) d\mathbf{y}' d\mathbf{y} \quad (\text{A.3})$$

For any observer with test statistic $t(\mathbf{y})$, the ALROC is given by

$$ALROC = \int_0^1 P_{CL}(\tau) dP_{FP}(\tau) \quad (\text{A.4})$$

where

$$P_{FP}(\tau) = P(t > \tau | H_0) = \int_{\tau}^{\infty} p(t|H_0) dt. \quad (\text{A.5})$$

Since $P_{FP}(\tau)$ is a monotonic function of τ , we can change the variable of integration in (A.4) from $P_{FP}(\tau)$ to τ and rewrite the expression of ALROC as

$$ALROC = - \int_{\infty}^{-\infty} P_{CL}(\tau) \frac{d}{d\tau} P_{FP}(\tau) d\tau \quad (\text{A.6})$$

where the minus sign and the change of integration limits are due to the fact that $P_{FP}(\tau)$ moves from 1 to 0 as we sweep τ from $-\infty$ to ∞ .

From (A.5), we have

$$\frac{d}{d\tau}P_{FP}(\tau) = -p(\tau|H_0),$$

and we can further rewrite ALROC in an alternative expression as follows,

$$\begin{aligned} ALROC &= \int_{-\infty}^{\infty} P_{CL}(y)p(y|H_0)dy \\ &= \int_{-\infty}^{\infty} \sum_{j=1}^J P_j P(x > y|H_j) P(j \in T(l)|H_j) p(y|H_0) dy \\ &= \sum_{j=1}^J P_j \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x|H_j) \text{step}(x-y) P(j \in T(l)|H_j) p(y|H_0) dx dy \end{aligned} \quad (A.7)$$

where we replace τ by a new variable y as τ will only be used to represent the threshold.

Again, both x and y are random variables through their dependence on the test images and we have

$$\begin{aligned} p(x|H_j) &= \int_{\mathbf{y}'} p(\mathbf{y}'|H_j) \delta(x - t(\mathbf{y}')) d\mathbf{y}', \quad j=1, \dots, J \\ p(y|H_0) &= \int_{\mathbf{y}} p(\mathbf{y}|H_0) \delta(y - t(\mathbf{y})) d\mathbf{y} \end{aligned} \quad (A.8)$$

From (A.7) and (A.8), the ALROC can be written as

$$\begin{aligned} ALROC &= \sum_{j=1}^J P_j \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \int_{\mathbf{y}'} \int_{\mathbf{y}} P(j \in T(l(\mathbf{y}'))|H_j) \text{step}(x-y) \times p(\mathbf{y}'|H_j) \delta(x - t(\mathbf{y}')) p(\mathbf{y}|H_0) \delta(y - t(\mathbf{y})) d\mathbf{y}' dy \\ &= \sum_{j=1}^J P_j \int_{\mathbf{y}} \int_{\mathbf{y}'} p(\mathbf{y}|H_0) p(\mathbf{y}'|H_j) P(j \in T(l(\mathbf{y}'))|H_j) \text{step}(t(\mathbf{y}') - t(\mathbf{y})) d\mathbf{y} d\mathbf{y}' \end{aligned}$$

which is the same as P_C if the test statistics $\lambda(\cdot)$ and $t(\cdot)$ are the same.

Figure 1.
Typical ROC and LROC curves.

Figure 2.

Typical display as seen by a human observer during the training session of an SKS 2AFC test. The true signal and guessed signal locations appear in the right image. (The circular tolerance region is shown for clarity but does not appear during the training session of the 2AFC test.) A third panel (not shown) displays the signal centered in an empty background.

Figure 3.

Sample signal-present test images with different cutoff frequencies. The signal is pointed out by an arrow. From left to right: $\rho_c = 0.06, 0.11, 0.165, 0.30, 0.45$. These figures are typical of those seen by observers. In the first, third and fifth panels the signal is not hard to detect and localize, but in the second and fourth panels, it is difficult to detect and localize.

Figure 4.

Effects of regularization on noise in the frequency domain: (a) Noise power spectrum before regularization: the spectrum comprises the background and quantum NPS; (b) Discrete transfer function Λ with different cutoff frequencies ρ_c .

Figure 5.

Effects of ρ_c on signal spectrum: (a) Signal spectrum (before regularization); (b) Signal spectrum with different cutoff frequencies (after regularization). Note that the curves for $\rho_c = 0.30$ and $\rho_c = 0.45$ overlap.

(a)

(b)

Figure 6.

Experimental results investigating tolerance effects. We use a Gaussian blob signal with $a_G = 65$ and regularization parameter $\rho_c = 0.165$. (a) ALROC versus radius of tolerance region for both the ideal and human observers; (b) Efficiency of the human observer relative to the ideal observer. Note the shoulder at $r_{tol} = 3$.

Figure 7. Experimental results with $p_c = 0.165$ for various signal intensities and for SKE and SKS tasks. (a) Ideal and human observer performance; (b) Human efficiency relative to the ideal observer.

Figure 8.

Experimental results for various cutoff frequencies and for both SKS and SKE tasks: (a) Ideal and human observer performance; (b) Human efficiency relative to the ideal observer for the SKS case; (c) Human efficiency relative to the ideal observer for the SKE case.