ROC curves predicted by a model of visual search

D P Chakraborty

Department of Radiology, University of Pittsburgh, 3520 5th Avenue, Suite 300,
Pittsburgh, PA 15261, USA

E-mail: dpc10@pitt.edu

Abstract

In imaging tasks where the observer is uncertain whether lesions are present,
and where they could be present, the image is searched for lesions. In the
free-response paradigm, which closely reflects this task, the observer provides
data in the form of a variable number of mark-rating pairs per image. In a
companion paper a statistical model of visual search has been proposed that
has parameters characterizing the perceived lesion signal-to-noise ratio, the
ability of the observer to avoid marking non-lesion locations, and the ability of
the observer to find lesions. The aim of this work is to relate the search model
parameters to receiver operating characteristic (ROC) curves that would result
if the observer reported the rating of the most suspicious finding on an image as
the overall rating. Also presented are the probability density functions (pdfs)
of the underlying latent decision variables corresponding to the highest rating
for normal and abnormal images. The search-model-predicted ROC curves
are 'proper' in the sense of never crossing the chance diagonal and the slope
is monotonically changing. They also have the interesting property of not
allowing the observer to move the operating point continuously from the origin
to (1, 1). For certain choices of parameters the operating points are predicted to
be clustered near the initial steep region of the curve, as has been observed by
other investigators. The pdfs are non-Gaussians, markedly so for the abnormal
images and for certain choices of parameter values, and provide an explanation
for the well-known observation that experimental ROC data generally imply
a wider pdf for abnormal images than for normal images. Some features
of search-model-predicted ROC curves and pdfs resemble those predicted by
the contaminated binormal model, but there are significant differences. The
search model appears to provide physical explanations for several aspects of
experimental ROC curves.

Introduction

The receiver operating characteristic (ROC) paradigm is widely used in the evaluation of medical imaging systems (Metz 1978, 1986, 1989). The method consists of asking observers to rate images for suspicion of abnormality, for example, higher ratings denoting a higher degree of suspicion that the image is abnormal. The images are either normal or abnormal, but this 'truth' information is known only to the experimenter. The numerical ratings data are usually modelled by two Gaussian distributions whose separation and ratio of variances form the basic parameters of the analytical model (Dorfman and Alf 1969). Additional cutoff parameters are needed to model the ratings data, but these do not affect the overall performance or figure of merit of the observer. The ROC curve is defined as the plot of true positive fraction versus the false positive fraction. The figure of merit is usually chosen to be the area under the theoretical ROC curve.

Images in an ROC study are viewed under conditions of uncertainty regarding their true normal/abnormal status and a single rating is recorded for each image. Additional information, such as the location of a suspected abnormality or that the image has multiple suspicious regions, cannot be analysed in the ROC framework. In imaging tasks where the observer is uncertain whether lesions are present and where they could be present the observer must search the images for lesions. In the free-response paradigm, which closely reflects the search task, the observer provides data in the form of a variable number of mark-rating pairs per image (Bunch *et al* 1978). A mark is a physical location or region in the image that was deemed to be worth reporting and the rating is the associated degree of suspicion. By adopting a proximity criterion or acceptance radius (Chakraborty and Berbaum 2004) it is possible to classify each mark as a 'true positive' (near a lesion) or a 'false positive' (far from a lesion). To avoid confusion with the common usage of these same terms in ROC analysis where no localization is required, I prefer to use the terms 'lesion localization' and 'non-lesion localization', respectively. The process of classifying the marks as lesion or non-lesion localizations is referred to as *scoring* the marks.

In a companion paper (Chakraborty 2006) a method of modelling the search/free-response task was described. The search model has parameters characterizing the perceived lesion signal-to-noise ratio, the ability of the observer to avoid marking non-lesion locations, and the ability of the observer to find lesions. A figure of merit for quantifying performance of the observer under search conditions was defined as the probability that the rating of the highest rated (or most suspicious) region on an abnormal image is greater than the corresponding rating on a normal image. If one can make the assumption, as is sometimes done (Swensson 1996, Swensson *et al* 2001), that had the observer been asked for a single rating in a search task then the observer would provide the rating of the highest rated region on the image, then it should be possible to predict the ROC curve for this hypothetical observer. In other words, it should be possible to relate the search model parameters to ROC curves. The purpose of this paper is to investigate this relationship. Since ROC curves imply underlying distributions of latent decision variables for normal and abnormal images, it is of interest to determine the probability density functions (pdfs) of these distributions and their dependence on search model parameters. Some of the ROC and pdf curve predictions of the search model have similarities to another model of observer performance, and examining their relationship is another aim of this study.

Theory

Introduction to the search model

The search model is a mathematical parameterization of an existing descriptive model of visual search (Kundel and Nodine 1983, 2004, Nodine and Kundel 1987). A unique aspect of the descriptive model is that all locations on an image do not receive equal units of attention. Instead the observer reduces the potentially large number of locations on an image to a smaller number of *sites* requiring more detailed examination and at each of which a *decision* is made whether or not to report the site as a possible lesion candidate. I use the term 'locations were hit' as shorthand for 'locations where decisions were made'. Eye movement measurements on radiologists form the basis of the descriptive search model (Duchowski 2002). The mechanism by which the observer accomplishes the reduction to a smaller number of sites is not well understood and is not relevant to the mathematical parameterization. The sites corresponding to non-lesion locations are termed *noise* sites and those corresponding to lesion locations are termed *signal* sites. The numbers of noise and signal sites on an image are denoted by n and u respectively.

Search model assumptions:

(1) The number of noise sites n on an image follows the Poisson distribution with parameter λ. The number of signal sites u on an abnormal image with s lesions follows the binomial distribution with trial size s and success probability ν. In other words ν is the probability that a lesion is hit, i.e., cognitively evaluated. These assumptions have been used in a prior model of free-response data (Edwards *et al* 2002). Every abnormal image has the same number of lesions (this limitation can be removed in an analogous manner to that indicated in appendix 1, equation (15), of the previous manuscript). The number of noise sites n and the number of signal sites u on an image are statistically independent.

(2) Let $N(\mu, 1)$ denote the Gaussian distribution with mean μ and unit variance. A decision variable sample z is realized at each decision site. The z-sample from a noise site is realized from a Gaussian distribution of unit variance and zero mean, i.e., $z \sim N(0, 1)$. The z-sample from a signal site is realized from a Gaussian distribution of unit variance and mean μ, i.e., $z \sim N(\mu, 1)$. All z-samples on an image are statistically independent. The reason for assuming equal widths for the noise and signal distributions is given in the discussion section.

(3) The observer adopts R ordered cutoff parameters ζ_i ($i = 1, 2, \ldots, R$) where R is the number of rating bins employed in the free-response study. The cutoff vector $\vec{\zeta}$ is defined as $\vec{\zeta} \equiv (\zeta_0, \zeta_1, \zeta_2, \ldots, \zeta_R, \zeta_{R+1})$ and by definition $\zeta_0 -\infty$ and $\zeta_{R+1} = +\infty$. If $\zeta_i < z < \zeta_{i+1}$ then the corresponding decision site is marked and rated in bin i, and if $z < \zeta_1$ then the site is not marked.

(4) The location of the mark is at the precise centre of the decision site that exceeded a cutoff. An infinitely precise scoring criterion, i.e., an infinitesimally small acceptance radius is adopted. Consequently, there is no confusing a mark made as a consequence of a signal site z-sample exceeding the cutoff as one made as a consequence of a noise site z-sample exceeding the cutoff, and vice versa. Any mark made as a consequence of a sample $z \sim N(0, 1)$ that satisfies $\zeta_i < z < \zeta_{i+1}$ will be scored as a non-lesion mark and assigned the rating 'i', and likewise any mark made as a consequence of a sample $z \sim N(\mu, 1)$ that satisfies $\zeta_i < z < \zeta_{i+1}$ will be scored as a lesion mark and assigned the rating 'i'.

(5) When required to give a single summary rating to an image that has at least one decision site the observer gives the rating z_h of the highest rated site on the image (henceforth
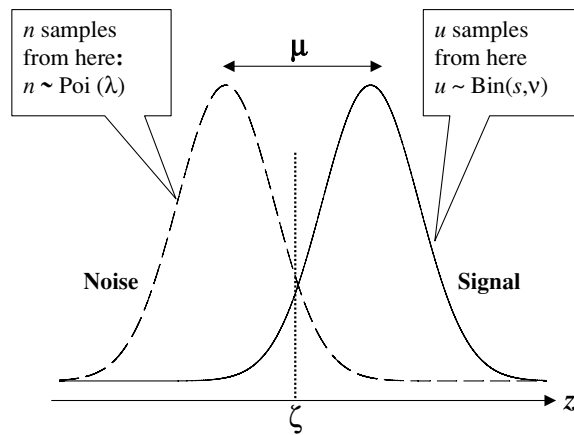
Figure 1. The search model for a single rating free-response study. The basic parameters of the model are μ, λ and ν, and s is the number of lesions per abnormal image. The two unit variance Gaussian distributions labelled Noise and Signal represent the pdfs of the z-samples from noise sites and signal sites, respectively. The number of noise sites n and the number of signal sites u are modelled by a Poisson and a Binomial distribution, respectively. The total number of decision sites per image in $n + u$. Each decision site yields a z-sample from the Noise or Signal distribution, for a noise site or a signal site, respectively. When a z-sample exceeds ζ, the observer's threshold, the observer marks the corresponding location. Noise site z-samples exceeding ζ are recorded as non-lesion localizations and corresponding signal site z-samples are recorded as lesion localizations.

abbreviated to 'highest rating'). On an abnormal image this could be the rating of a noise or a signal site. On a normal image this is necessarily the rating of a noise site.

(6) When required to give a single summary rating to an image that has no decision sites the observer gives lowest possible rating. For example, if the allowed rating scale is $1, 2, \ldots,$ 100, corresponding to $R = 100$, the observer assigns the '1' rating to such images.

The left and right Gaussian distributions in figure 1 represent the pdfs corresponding to $N(0, 1)$ and $N(\mu, 1)$, respectively. The horizontal axis is the observer's internal confidence that a decision site represents a lesion, i.e., the z-sample. According to the search model one realizes n noise site z-samples from $N(0, 1)$ and u z-samples from $N(\mu, 1)$. The integers n and u are image dependent and represent the number of z-samples corresponding to noise sites and signal sites, respectively. Those noise site z-samples that exceed the observer's lowest cutoff are marked by the observer and are scored by the experimenter as non-lesion localizations, and likewise the signal site z-samples that exceed the lowest cutoff are marked and scored as lesion localizations. The rating assigned to these marks follows the rule specified in assumption (3). The number (f) of non-lesion localizations on an image cannot exceed n (because the noise site z-samples that fall below the lowest cutoff do not result in marks; only when the lowest cutoff equals negative infinity does f equal n). For the same reason t cannot exceed u, and obviously u cannot exceed s, the number of lesions in the images.

ROC curve predicted by the search model

Definitions. The unit variance Gaussian probability density function corresponding to $N(\mu, 1)$ and the corresponding distribution function are defined by

$$\phi(z|\mu) = \frac{1}{\sqrt{2\pi}}\, e^{-(z-\mu)^2/2} \qquad \Phi(z|\mu) = \int_{-\infty}^{z} dy\, \phi(y|\mu). \tag{1}$$

The Poisson and binomial density functions are defined by

$$\text{Poi}(n \mid \lambda) = \frac{\lambda^n}{n!}\, \mathrm{e}^{-\lambda} \qquad \text{Bin}(u|s, \nu) = \binom{s}{u} \nu^u (1 - \nu)^{s-u}. \tag{2}$$

In these expressions μ, λ and ν are the parameters of the search model and s is the number of lesions per abnormal image.

The ROC curve

The ROC curve is the plot of true positive fraction versus false positive fraction. Observer generated (or experimental) operating points are obtained by a well-known procedure involving cumulating the counts in the ratings bins. For example, consider an R rating ROC study, where the numerical rating ranges from 1 to R, and assume that higher ratings correspond to greater confidence that the image is abnormal. The ordinate (true positive fraction) of the lowest point, i.e., closest to the origin and corresponding to the highest confidence, is obtained by dividing the count in the Rth abnormal bin by the number of abnormal images. The ordinate of the next higher point is obtained by cumulating the counts in bins $R - 1$ and R, and so on. A similar procedure is used for the abscissa (false positive fraction) of the experimental operating points, except that one divides by the number of normal images. In this way successive experimental ROC operating points are generated. If none of the bins are cumulated one gets the (0, 0) point and if all the bins are cumulated one gets the (1, 1) point. I use the symbol y for the true positive fraction and x for the false positive fraction. To predict a continuous ROC curve one regards x and y as functions of a continuous variable ζ. The latter is the cutoff used by the observer to render positive (abnormal) decisions. Specifically, if the decision variable for the image exceeds ζ then the observer classifies the image as abnormal. According to assumption (5) the decision variable for the image is the highest rating z_h for the image. The true positive fraction is the probability that z_h on an abnormal image exceeds ζ and the false positive fraction is the probability that z_h on a normal image exceeds ζ:

$$x(\zeta) = \text{Prob}(z_h > \zeta | N) \qquad y(\zeta) = \text{Prob}(z_h > \zeta | A), \tag{3}$$

where N and A denote normal and abnormal images, respectively. Varying the cutoff parameter ζ from ∞ to $-\infty$ generates points on the theoretical ROC curve that range from (0, 0) to (1, 1).

Limiting point on the ROC curve

Before getting into details I describe a distinctive feature of all search-model-predicted ROC curves. It will be shown that the full range of the ROC data space, namely $0 \leqslant x(\zeta) \leqslant 1$ and $0 \leqslant y(\zeta) \leqslant 1$, is *not* continuously accessible to the observer. In fact $0 \leqslant x(\zeta) \leqslant x_{\max}$ and $0 \leqslant y(\zeta) \leqslant y_{\max}$ where $x_{\max}$ and $y_{\max}$ are less than unity. By 'not continuously accessible' I mean that for any *finite* value of the cutoff ζ, no matter how small, some images will not be classified as abnormal. As the observer varies the cutoff continuously from $+\infty$ to $-\infty$, the ROC point (x, y) will move continuously from (0, 0) to $(x_{\max}, y_{\max})$, and thereafter there will be a discontinuous jump to the point (1, 1) which is obtained when counts in all bins are cumulated. This behaviour is distinct from traditional ROC curves where the entire section of the curve extending from (0, 0) to (1, 1) is continuously accessible to the observer via appropriate choice of the cutoff. The reason for this behaviour is that some images generate no hits (no decision sites) and therefore do not yield a z-sample. By assumption (6) such images are assigned to the lowest rating bin, i.e., '1', and only when this bin is included in the cumulation procedure does the point (1, 1) result. The coordinates of the operating point

resulting from cumulating the counts at or above the next higher bin, i.e., '2' and above, will yield an ROC point (x, y) with $x \leqslant x_{\max}$ and $y \leqslant y_{\max}$. (As an aside, in multi-rating ROC studies involving several bins, how closely the observer approaches the limiting point $(x_{\max}, y_{\max})$ is not related to the number of bins, but to the position of the lowest cutoff, i.e., ζ_1. As ζ_1 is lowered (i.e., the observer is encouraged to be more aggressive) the uppermost operating point approaches $(x_{\max}, y_{\max})$. How closely $(x_{\max}, y_{\max})$ approaches $(1, 1)$ depends on λ: as λ increases $(x_{\max}, y_{\max})$ approaches $(1, 1)$, see equation (4).)

The limiting point $(x_{\max}, y_{\max})$ can be calculated as follows. Consider first the normal cases and the calculation of $x_{\max}$. Basically one needs the probability that a normal image has at least one hit. Such an image will generate a finite z_h and with an appropriately low cutoff the image will be rated '2' or above. The probability of zero noise sites is $\text{Poi}(0 \mid \lambda) = \exp(-\lambda)$. Therefore the probability of at least one hit on a normal image is $1 - \exp(-\lambda)$ which is $x_{\max}$. Likewise, an abnormal image has no hits if it has zero noise sites, the probability of which is $\text{Poi}(0 \mid \lambda)$, and it has zero signal sites, the probability of which is $\text{Bin}(0 \mid s, \nu)$. Therefore the probability of zero hits on an abnormal image is the product of these two probabilities, namely $\text{Poi}(0 \mid \lambda) \, \text{Bin}(0 \mid s, \nu)$ and the probability that there is at least one hit is $1 - \text{Poi}(0 \mid \lambda) \, \text{Bin}(0 \mid s, \nu)$, which is $y_{\max}$. Summarizing,

$$
\begin{aligned}
x_{\max} &\equiv x_{\max}(\lambda) = 1 - \text{Poi}(0 \mid \lambda) = 1 - \exp(-\lambda) \\
y_{\max} &\equiv y_{\max}(\lambda, \nu, s) = 1 - \text{Poi}(0 \mid \lambda) \, \text{Bin}(0, s, \nu) = 1 - \exp(-\lambda)(1 - \nu)^s.
\end{aligned}
\tag{4}
$$

Calculation of the ROC curve

According to equation (3) to calculate the ROC curve one needs two probabilities, namely $\text{Prob}(z_h \rangle \zeta \mid N)$ and $\text{Prob}(z_h \rangle \zeta \mid A)$. Consider a normal image with n noise sites. According to assumption (2) each noise site yields a decision variable sample from $N(0, 1)$. The probability that a z-sample is smaller than ζ is $\Phi(\zeta \mid 0)$. By the independence assumption the probability that all z-samples are smaller than ζ is $[\Phi(\zeta \mid 0)]^n$. If all z-samples are smaller than ζ then the highest z_h of the samples is smaller than ζ. Therefore, the probability that z_h will exceed ζ is

$$
x(\zeta \mid n) = \text{Prob}(z_h > \zeta \mid n, N) = 1 - [\Phi(\zeta \mid 0)]^n.
\tag{5}
$$

The notation in equation (5) reflects the fact that this expression applies specifically to normal images with n noise sites. The desired x-coordinate is obtained by averaging $x(\zeta \mid n)$ over all values of n,

$$
x(\zeta \mid \lambda) = \sum_{n=0}^{\infty} \text{Poi}(n \mid \lambda) x(\zeta \mid n) = 1 - \exp\left(-\frac{\lambda}{2} + \frac{1}{2}\lambda \, \text{erf}\left(\frac{\zeta}{\sqrt{2}}\right)\right).
\tag{6}
$$

In equation (6) erf is the error function (Press *et al* 1988) that ranges from -1 to $+1$ as its argument ranges from $-\infty$ to $+\infty$. It can be easily confirmed that for $\zeta = -\infty$ equation (6) yields the same expression for $x_{\max}(\lambda)$ as equation (4).

Now consider an abnormal image with n noise sites and u signal sites. According to assumption (2) each noise site yields a decision variable sample from $N(0, 1)$ and each signal site yields a sample from $N(\mu, 1)$. Therefore the probability that z_h is larger than ζ is

$$
y(\zeta \mid \mu, n, u) = 1 - [\Phi(\zeta \mid 0)]^n [\Phi(\zeta \mid \mu)]^u.
\tag{7}
$$

As before one averages over n and u to obtain the desired ROC ordinate,

$$
y(\zeta \mid \mu, \lambda, \nu, s) = \sum_{n=0}^{\infty} \text{Poi}(n \mid \lambda) \sum_{u=0}^{s} \text{Bin}(u \mid s, \nu) \, y(\zeta \mid \mu, n, u).
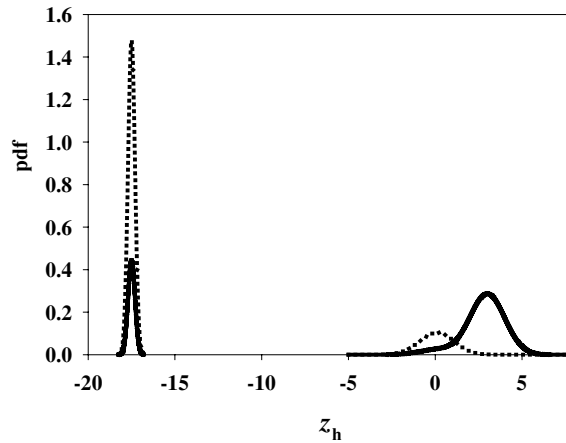\tag{8}
$$

Figure 2. The pdfs of the decision variable z_h of the highest rated site for $\mu = 3$, $\lambda = 0.3$, $\nu = 0.7$ and $s = 1$. The dotted lines correspond to the normal cases and the solid lines to the abnormal cases. The delta functions at $-\infty$ are, for convenience, shown as narrow Gaussians centred at -17.5. The two pdfs centred near 0 and 3 generate the continuously accessible portion of the ROC curve shown as the solid line in figure 3. The pdfs centred at $-\infty$ generate the inaccessible portion of the ROC curve shown as the dotted line in figure 3.

This can be evaluated using the symbolic mathematical language Maple (Maple 8.00, Waterloo Maple Inc.) yielding

$$y(\zeta|\mu, \lambda, \nu, s) = 1 - \left(1 - \frac{\nu}{2} + \frac{\nu}{2}\mathrm{erf}\left(\frac{\zeta - \mu}{\sqrt{2}}\right)\right)^s \mathrm{e}^{(-\frac{\lambda}{2} + \frac{1}{2}\lambda\, \mathrm{erf}(\frac{\zeta}{\sqrt{2}}))}. \tag{9}$$

It can be easily confirmed that for $\zeta = -\infty$ equation (9) yields the same expression for $y_{\max}(\lambda, \nu, s)$ as equation (4) (in this limit the μ dependence drops out).

The probability density functions

The probability density functions (pdfs) are given by

$$\mathrm{pdf}_N(\zeta|\lambda) = \frac{\partial}{\partial \zeta}\left[1 - x(\zeta|\lambda)\right]$$

$$\mathrm{pdf}_A(\zeta|\mu, \lambda, \nu, s) = \frac{\partial}{\partial \zeta}\left[1 - y(\zeta|\mu, \lambda, \nu, s)\right]. \tag{10}$$

Because of the images that are not hit, and for the same reason that the ROC curve does not extend continuously to $(1, 1)$, the area under these pdfs will not equal unity. In fact, it is easily seen that

$$\int_{-\infty}^{\infty} \mathrm{d}\zeta\,\mathrm{pdf}_N(\zeta|\lambda) = [1 - x(\zeta|\lambda)]|_{-\infty}^{\infty} = x_{\max}(\lambda)$$

$$\int_{-\infty}^{\infty} \mathrm{d}\zeta\,\mathrm{pdf}_A(\zeta|\mu, \lambda, \nu, s) = [1 - y(\zeta|\mu, \lambda, \nu, s)]|_{-\infty}^{\infty} = y_{\max}(\lambda, \nu, s). \tag{11}$$

To account for the missing areas one needs two delta functions at $-\infty$, one corresponding to the normal images with integrated area $1 - x_{\max}(\lambda)$, and the other corresponding to the abnormal cases with integrated area $1 - y_{\max}(\mu, \lambda, \nu, s)$. The chance level discrimination between these two delta function distributions leads to a straight line portion of the ROC curve that extends from $(x_{\max}, y_{\max})$ to $(1, 1)$. Figure 2 shows pdfs of the decision variable z_h of the
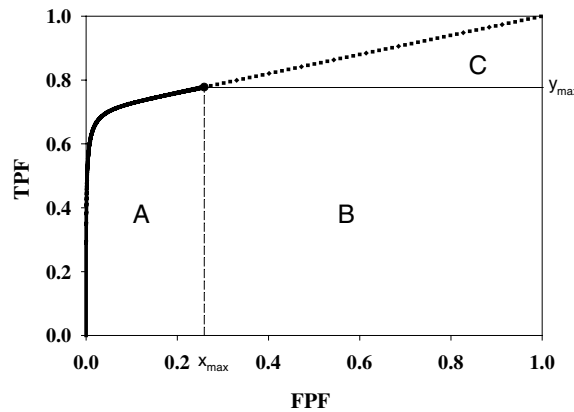
Figure 3. Geometrical interpretation of the area under the curve (AUC). The parameter values are as in figure 2. The area under the continuous section of the ROC curve, extending from $(0, 0)$ to (x_{max}, y_{max}) and which is labelled A, corresponds to AUC_1 in equation (12). The area of the rectangle labelled B is the contribution due to perfect discrimination between the abnormal image pdf in figure 2 and the delta function normal image pdf at $-\infty$. The area of the triangle labelled C is the contribution due to chance level discrimination between the two delta function pdfs at $-\infty$ in figure 2. The sum of the areas B and C corresponds to AUC_2 in equation (12).

highest rated site for $\mu = 3$, $\lambda = 0.3$, $\nu = 0.7$ and $s = 1$ for the normal and abnormal cases. The dotted lines correspond to the normal cases and the solid lines to the abnormal cases. The delta functions at $-\infty$ are, for convenience, shown as narrow Gaussians centred at -17.5. The two pdfs centred near zero and 3 generate the continuously accessible portion of the ROC curve shown as the solid line in figure 3. The pdfs centred at $-\infty$ generate the inaccessible portion of the ROC curve shown as the dotted line in figure 3. The areas under the two pdfs centred near 0 and 3 are 0.259 and 0.778, corresponding to the normal and abnormal images, respectively. The corresponding areas under the delta functions are the complements of these values. In either case the total area under a complete pdf is unity.

As an aside, the need for delta functions at negative infinity can be seen from the following argument. Let us postulate two finite width pdfs with the same shapes but different areas, centred at a common value z_h far to the left in decision space, but not at negative infinity. In the example shown in figure 2, $z_h = -17.5$. These pdfs would also yield a straight line portion to the ROC curve. However, they would be inconsistent with the search model assumptions. According to assumption (6) some images yield no decision variable samples and cannot be rated in bin 2 or higher (if they were rated 2 or higher that would imply that they did yield decision variable samples and moreover one of the samples exceeded ζ_1). Therefore, if the distributions are as postulated above then choice of a cutoff in the neighbourhood of z_h would result in some of these images being rated 2 or higher, contradicting assumption (6). The delta function pdfs at negative infinity are seen to be a consequence of the search model assumptions.

The area under the ROC curve

The total area under the ROC curve will consist of two parts, one under the continuous section and one under the straight line portion. Denoting these by AUC_1 and AUC_2 one has

$$AUC_1 = \int_{-\infty}^{\infty} d\zeta\, pdf_N(\zeta|\lambda)\, y(\zeta|\mu, \lambda, \nu, s)$$
$$AUC_2 = y_{max}(1 - x_{max}) + 0.5(1 - x_{max})(1 - y_{max}). \tag{12}$$

The area AUC_1 under the continuous portion of the ROC curve was evaluated numerically using Maple. A geometrical interpretation of the areas is shown in figure 3. The area under the continuous section of the curve labelled A corresponds to AUC_1 in equation (12). The area of the rectangle labelled B is $y_{max}(1 - x_{max})$. The area of the triangle labelled C is 0.5 $(1 - x_{max})(1 - y_{max})$. The net area under the dotted section of the ROC curve is B + C, and comparison to equation (12) shows that this area corresponds to AUC_2. The total area under the ROC curve is the sum of AUC_1 and AUC_2, namely

$$AUC = AUC_1 + AUC_2. \tag{13}$$

The quantity AUC can be regarded as a figure of merit of observer performance in a free-response study. It will depend on the parameters μ, λ and ν of the search model, and s, the number of lesions per abnormal image.

Relation to a previously proposed figure of merit

A figure of merit θ proposed in the companion paper was based on the two alternative forced choice (2AFC) paradigm. It is now shown that the two figures or merit, AUC and θ, are in fact identical. In the 2AFC paradigm the observer compares images in a normal–abnormal pair and attempts to select the abnormal image. The figure of merit θ is defined as the fraction of correct choices in this task. Four cases need to be distinguished: (a) both images have at least one hit, (b) neither image has a hit, (c) only the abnormal image has a hit and (d) only the normal image has a hit. For case (b) assume that the observer picks between the images at random so that the probability of a correct choice is 0.5. For cases (c) and (d) assume that the observer picks whichever image has a hit, so that the probability of a correct choice is one or zero depending on whether the image with a hit is abnormal or normal.

The figure of merit contribution for case (a) involves comparisons between highest rated samples from the two probability density functions, $\text{pdf}_N(\zeta|\lambda)$ and $\text{pdf}_A(\zeta|\mu, \lambda, \nu, s)$, corresponding to normal and abnormal cases, respectively. These comparisons are implicit in the first integral in equation (12), which is the average over all possible highest rating samples ζ from the normal image pdf, of the probability y that the highest rating from an abnormal image exceeds ζ. Therefore the case (a) contribution θ_a is in fact identical to the area under the continuous section of the ROC curve, i.e., $\theta_a = AUC_1$. The contribution for case (b) is 0.5 times the probability that neither image of the pair has a hit. The probability that a normal image does not have a hit is $\text{Poi}(0 \mid \lambda)$. The probability that an abnormal image does not have a hit is $\text{Poi}(0 \mid \lambda)\text{Bin}(0, s, \nu)$. The probability that neither image of the pair has a hit is the product of these probabilities, namely $[\text{Poi}(0 \mid \lambda)]^2\text{Bin}(0, s, \nu)$ and therefore the case (b) contribution to the figure of merit is

$$\theta_b = 0.5[\text{Poi}(0 \mid \lambda)]^2\text{Bin}(0, s, \nu). \tag{14}$$

In case (c) only the abnormal image has a hit. The probability that a normal image does not have a hit is $\text{Poi}(0 \mid \lambda)$. The probability that an abnormal image does have a hit was calculated previously and is $(1 - \text{Poi}(0 \mid \lambda)\text{Bin}(0, s, \nu))$. Therefore, the probability that only the abnormal image has a hit is $\text{Poi}(0 \mid \lambda)(1 - \text{Poi}(0 \mid \lambda)\text{Bin}(0, s, \nu))$ and since each of these contributions leads to a correct choice, the case (c) contribution to the figure of merit is

$$\theta_c = \text{Poi}(0 \mid \lambda)(1 - \text{Poi}(0 \mid \lambda)\text{Bin}(0, s, \nu)). \tag{15}$$

The case (d) probability is not needed, as it leads to no correct choices. The sum of the case (b) and case (c) contributions is

$$\theta_b + \theta_c = 0.5[\text{Poi}(0 \mid \lambda)]^2\text{Bin}(0, s, \nu) + \text{Poi}(0 \mid \lambda)(1 - \text{Poi}(0 \mid \lambda)\text{Bin}(0, s, \nu)). \tag{16}$$

Using equation (4) the above simplifies to

$$\theta_b + \theta_c = 0.5[1 - x_{\max}][1 - y_{\max}] + [1 - x_{\max}]y_{\max}, \tag{17}$$

which is identical to the area under the straight line section of the ROC curve, namely, AUC_2. This completes the proof of the proposition that the two figures of merits are identical.

'Proper' ROC curve

I now show that the accessible portion of the ROC curve is 'proper'. From equations (6) and (9) it is seen that one can express the ROC coordinates (x, y) as (for convenience I suppress the dependence on model parameters)

$$x(\zeta) = 1 - G(\zeta) \qquad y(\zeta) = 1 - F(\zeta)G(\zeta). \tag{18}$$

where

$$G(\zeta) = \exp\left(-\frac{\lambda}{2} + \frac{1}{2}\lambda \ \mathrm{erf}\left(\frac{\zeta}{\sqrt{2}}\right)\right), \tag{19}$$

and

$$F(\zeta) = \left(1 - \frac{\nu}{2} + \frac{\nu}{2}\mathrm{erf}\left(\frac{\zeta - \mu}{\sqrt{2}}\right)\right)^s. \tag{20}$$

These equations have exactly the same structure as Swensson's (Swensson 1996) equations (1) and (2) and the logic he used to demonstrate that ROC curves predicted by his model were 'proper' also applies to the present situation. In particular, since the error function ranges between -1 and 1 and $\nu \leqslant 1$ it follows that $F(\zeta) \leqslant 1$. Therefore $y(\zeta) \geqslant x(\zeta)$ and the ROC curve is constrained to the upper half of the ROC space, namely that lying above the 'chance' diagonal. Additionally the more general constraint shown by Swensson applies, namely the slope of the ROC curve at any operating point (x, y) cannot be less than the slope of the straight line connecting (x, y) and $(1, 1)$. This implies that the slope is monotonically changing and also rules out curves with 'hooks'.

Simulation testing of the predicted ROC curves

To test the internal consistency between the search model and the predicted ROC curves, free-response data were simulated according to the search model. Free-response data consist of counts per image in the different bins for non-lesion and lesion localizations. Consider a R-rating free-response study for which the bins are labelled $1, 2, \ldots, R$ and the cutoffs are ζ_i $(i = 1, 2, \ldots, R)$ as described in assumption (3). For specified values of the parameters λ, μ, and ν, and the number s of lesions per abnormal image the simulation proceeded as follows. For each normal image one generates a random number n $(\geqslant 0)$ from the Poisson distribution with parameter λ. Next one obtains n noise site z-samples from the Gaussian distribution $N(0, 1)$ which yields z_j $(j = 1, \ldots, n)$. Each of the z-samples is binned according to the rule in assumption (3). Specifically, if $\zeta_i < z_j < \zeta_{i+1}$ then the count in bin i is incremented by unity (all bin counts for each image are initially set equal to zeros). Note that if $z_j < \zeta_1$ then no bin count is incremented. If $n = 0$ one does not sample $N(0, 1)$ and for that image all bin-counts are zeros. As an example assume $n = 6$ for an image and that the final count vector for a 4-rating free-response study is $(2, 0, 2, 0)$. This means that of the six noise site z-samples two fell in the first bin, two fell in the third bin and two z-samples were smaller than ζ_1.

For abnormal images one has both noise and signal samples. The former are handled as for normal images, resulting in a non-lesion localization counts vector with R elements. For the signal samples one generates a random number u $(0 \leqslant u \leqslant s)$ from a binomial distribution

with trial size s and success probability ν. Next one obtains u samples from the Gaussian distribution $N(\mu, 1)$ which yields z_j $(j = 1, \ldots, u)$. Each z-sample is binned as described above except this time the counts are recorded in the lesion localization counts vector for that image, which also has R elements. If z_j is smaller than ζ_1 then the lesion localization counts vector is not incremented. If $u = 0$ one does not sample $N(\mu, 1)$ and for that image the lesion localization counts vector is zero.

Conversion of the free-response ratings to a single summary rating as required by assumption (5) proceeds as follows. For an R-rating free-response study one defines two $R + 1$ dimensional count vectors $\vec{F}$ and $\vec{T}$ with components F_k and $T_k(k = 0, 1, 2, \ldots, R)$ corresponding to the normal and abnormal images, respectively, and initializes all counts to zeros. For each normal image one determines the index j $(j = 1, 2, \ldots, R)$ of the highest bin with a non-zero entry in the (non-lesion localization) counts vector and increments F_j by unity. If the counts vector for an image has only zero elements one increments F_0. For example, if the counts vector for an image is (2, 0, 2, 0) then $j = 3$ and one increments F_3. For abnormal images one determines the index j of the highest bin with a non-zero entry when both non-lesion and lesion vectors are considered and increments T_j. For example, assuming $s = 2$ and the non-lesion and lesion counts vectors are (0, 1, 0, 0) and (0, 0, 1, 0), respectively, then $j = 3$ and one increments T_3. As another example if the corresponding vectors are (1, 1, 0, 1) and (0, 2, 0, 0) then $j = 4$ and one increments T_4. It is possible that the highest index for non-lesion and lesion vectors are identical, in which case one uses the common index j and increments T_j. If all elements in both vectors are zeros one increments T_0. The total of the counts in the vector $\vec{F}$ must equal the total number of normal images N_N, and likewise the total of the counts in the vector $\vec{T}$ must equal the total number of abnormal images N_A. Using the cumulation procedure described previously the vectors $\vec{F}$ and $\vec{T}$ determine the operating points on the ROC curve.

To generate the simulated ROC data points 20 cutoffs uniformly spaced between -2 and 8 were chosen $(R = 20)$. To minimize sampling variability a large number $(N_N = N_A = 2000)$ of images were simulated. This ensures that any deviation between the ROC operating points and the predicted ROC curves cannot be explained by sampling variability and would provide evidence that the sampling model was inconsistent with the ROC curve prediction. The random number generators in the Interactive Data Language (IDL, Research Systems Inc.) were used in this work.

Results

Table 1 summarizes μ, λ, ν, s, x_{max}, y_{max}, AUC_1, AUC_2 and AUC for the different ROC curves shown in this work. AUC_1 is the area under the accessible part of the curve (solid curve), AUC_2 is the area under the inaccessible part of the curve (dotted straight line) and AUC is the total area under the curve (AUC = AUC_1 + AUC_2). The figure of merit (AUC $\equiv \theta$) values exhibit the same dependence on search model parameters as those noted in the companion paper.

Shown in the upper panels in figures 4 through 8 are ROC curves for the choices of parameter values listed in table 1. In each case the solid curve is the continuously accessible portion of the ROC curve and the dotted portion is the inaccessible portion. The open circles are the operating points resulting from the simulation. It is seen that the open circles closely match the predicted curves. This demonstrates that the simulation model and the ROC curve prediction are internally consistent. The lower panels in these figures show the probability density functions (pdfs) for the highest rating for normal (dotted curve) and abnormal images (solid curve) respectively. The delta functions at negative infinity are not shown.
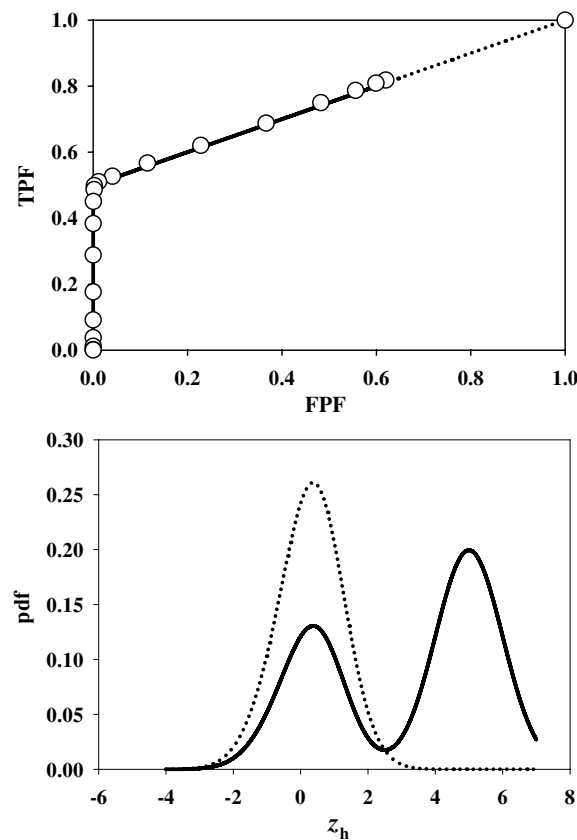
Figure 4. The ROC curve (upper panel) and pdfs (lower panel) for $\mu = 5$, $\lambda = 1$, $\nu = 0.5$ and $s = 1$. The open circles in the upper panel in this and succeeding plots are experimental ROC operating points from the simulations. The accessible portion of the ROC curve extends from $(0, 0)$ to $(0.63, 0.82)$. Note the strong bi-modality in the abnormal image pdf arising from the fact that half of the lesions are not hit. Therefore the highest decision variable for such images must have originated from a z-sample from $N(0, 1)$ yielding the peak near 0.

Table 1. This table summarizes the relevant data for the ROC curves and pdfs shown in this paper. The parameters μ, λ and ν are the basic parameters of the search model. The integer s is the number of lesions per abnormal image, assumed constant over all abnormal images. The quantities x_{max} and y_{max} are the coordinates of the end-point of the continuous section of the ROC curve. AUC_1 is the area under the continuous section of the curve shown by solid lines and AUC_2 is the area under the inaccessible section of the curve, shown by dotted lines. Their sum equals AUC, the total area under the curve, which is the figure of merit for the search task. The entries that are indicated as '1.000' are the rounded value to three decimal places; the exact value is slightly smaller than unity; similarly the entries indicated as '0.000' are actually slightly larger than zero.

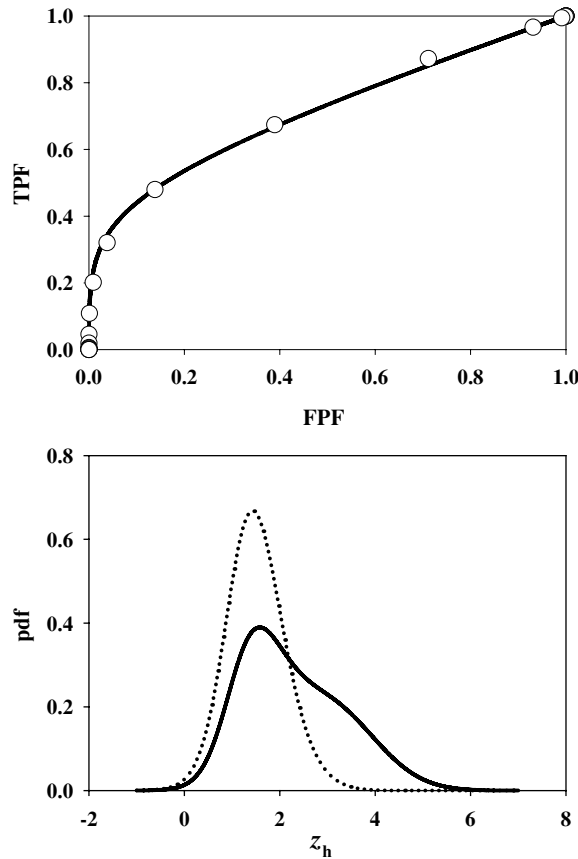| Figure	μ	λ	ν	s	x_{max}	y_{max}	AUC_1	AUC_2	AUC
4	5	1	0.5	1	0.632	0.816	0.416	0.334	0.750
5	3	10	0.5	1	1.000	1.000	0.711	0.000	0.711
6	3	3	1	1	0.950	1.000	0.914	0.050	0.964
2, 3, 7	3	0.3	0.7	1	0.259	0.778	0.188	0.658	0.847
8	3	10	0.5	2	1.000	1.000	0.831	0.000	0.831

Figure 5. The ROC curve and pdf for $\mu = 3$, $\lambda = 10$, $\nu = 0.5$ and $s = 1$. Due to the large value of λ the accessible section of the ROC curve extends almost to $(1, 1)$. Also, the highest rating on abnormal images is likely due to a noise site z-sample, yielding the large peak in the abnormal image pdf near 1.5. A slight peak is also evident near $z_h = 3$ due to the fewer times when a signal site z-sample is the highest rating.

Figure 4 corresponds to $\mu = 5$, $\lambda = 1$, $\nu = 0.5$ and $s = 1$. Note the discontinuous jump from the uppermost open circle on the continuous section of the curve to $(1, 1)$. The former point approaches (x_{max}, y_{max}) asymptotically as the lowest cutoff approaches $-\infty$. The point in question actually corresponds to $\zeta_1 = -2$, which cutoff is low enough to be smaller than almost any sample from the normal image pdf (see the lower panel), so this point is close to the limiting point. Since not all images generate decision variable samples, the limiting point is significantly below $(1, 1)$. In fact, using equation (4) and the noted parameter values, the coordinates of the limiting point are $(0.63, 0.82)$. As explained in connection with figure 3 the area under the normal image pdf is x_{max}, and that under the (bimodal) abnormal image pdf is y_{max}, where (x_{max}, y_{max}) are the coordinates of the limiting point on the accessible part of the ROC curve, i.e., $(0.63, 0.82)$ in this case.

Figure 5 shows the ROC curve and pdf for $\mu = 3$, $\lambda = 10$, $\nu = 0.5$ and $s = 1$. Due to the large value of λ the accessible section of the ROC curve extends almost to $(1, 1)$. Also, the highest rating on abnormal images is likely due to a noise site z-sample, yielding the large peak in the abnormal image pdf near 1.5 (the mean of the highest of $\sim$10 samples from $N(0, 1)$ is about equal to this value). A slight peak is also evident near $z_h = 3$ due to the fewer
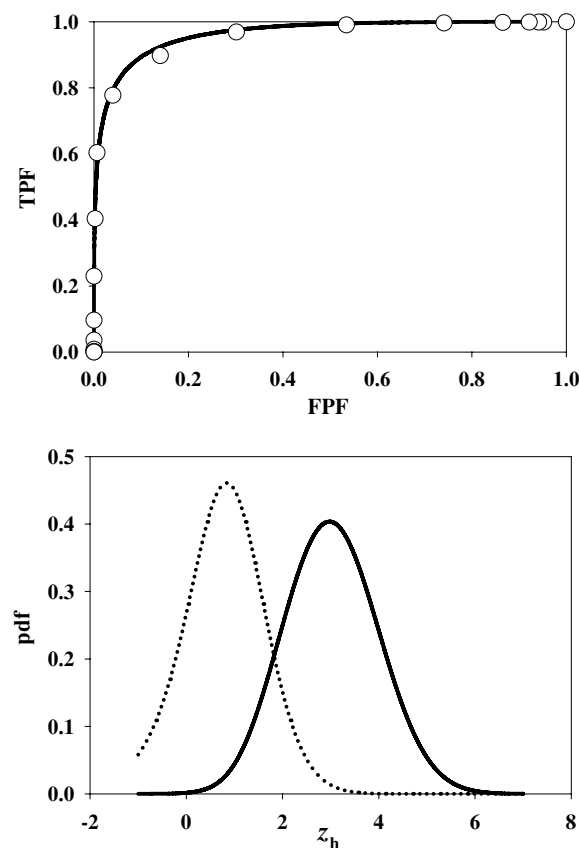
Figure 6. The ROC curve predicted by the search model for $\mu = 3$, $\lambda = 3$, $\nu = 1.0$ and $s = 1$. This example most resembles a conventional ROC curve although strictly the accessible portion of the curve does not extend to (1, 1) and the pdfs are not exactly Gaussians.

times when the signal site z-sample is the highest rating. The shapes of the *normal* image pdfs shown in figures 4 through 8 are close to Gaussian but strictly none of them are Gaussian. A subtle but visible non-Gaussian tail is evident in figure 5, which represents the largest λ value in the examples shown ($\lambda = 10$). With the exception of figure 6 the *abnormal* image pdfs are all significantly non-Gaussian. The upper panel of figure 6 shows the ROC curve predicted by the search model for $\mu = 3$, $\lambda = 3$, $\nu = 1.0$ and $s = 1$. This example most resembles a conventional ROC curve, although the accessible portion of the curve does not extend exactly to (1, 1). The pdf for the abnormal images (lower panel, solid curve) is close to Gaussian. This is related to the certainty of a lesion hit ($\nu = 1$), the fact that $s = 1$ (if $s > 1$ the highest sample from $N(3, 1)$ would have a non-Gaussian distribution), and the low probability of a noise site z-sample exceeding the signal site sample (since $\mu = 3$ is relatively large). However the pdf is not exactly Gaussian. Figure 7 (upper panel) shows the ROC curve for $\mu = 3$, $\lambda = 0.3$, $\nu = 0.7$ and $s = 1$. Due to the small value of λ, this example shows an unusually small accessible portion of the ROC curve which extends to (0.26, 0.78). Note the relatively small area under the normal image pdf (0.26) as compared to the abnormal image pdf (0.78); the missing areas are in the delta functions. Figure 8 (upper panel) shows the ROC curve for $\mu = 3$, $\lambda = 10$, $\nu = 0.5$ and $s = 2$. Excepting for the number of lesions ($s = 2$ versus $s = 1$) the
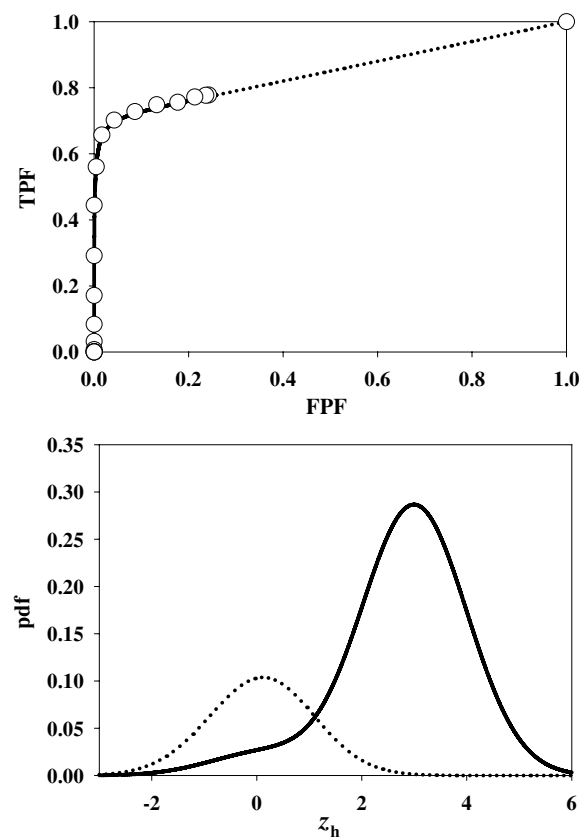
Figure 7. The ROC curve predicted by the search model for $\mu = 3$, $\lambda = 0.3$, $\nu = 0.7$ and $s = 1$. Due to the small value of λ, this example shows an unusually small accessible portion of the ROC curve which extends to (0.26, 0.78). The areas under the normal (abnormal) image pdfs are these values, namely 0.26 and 0.78, respectively. Note the clustering of the operating points near the initial vertical section of the ROC curve.

parameter values are identical to those shown in figure 5. Note the larger AUC, as compared to figure 5, demonstrating the expected increase with s.

Some of the abnormal image pdfs in the figures show evidence of bimodality. This has to do with the fact that on abnormal images sometimes the signal site z-sample from $N(\mu, 1)$ is the highest rating, and sometimes the noise site z-sample from $N(0, 1)$ is the highest rating. Bimodality is most evident in figure 4, which corresponds to $\nu = 0.5$, when this effect is expected to be maximum. Bimodal pdfs are also predicted by the contaminated binormal model (CBM) (Dorfman and Berbaum 2000).

Discussion

The search model has parameters characterizing the perceived lesion signal-to-noise ratio (μ), the ability of the observer to avoid cognitively evaluating non-lesion locations (λ), and the ability of the observer to find (i.e., cognitively evaluate) lesions (ν). An expert observer (or conversely an easy search task) would be characterized by large μ and/or small λ and/or large
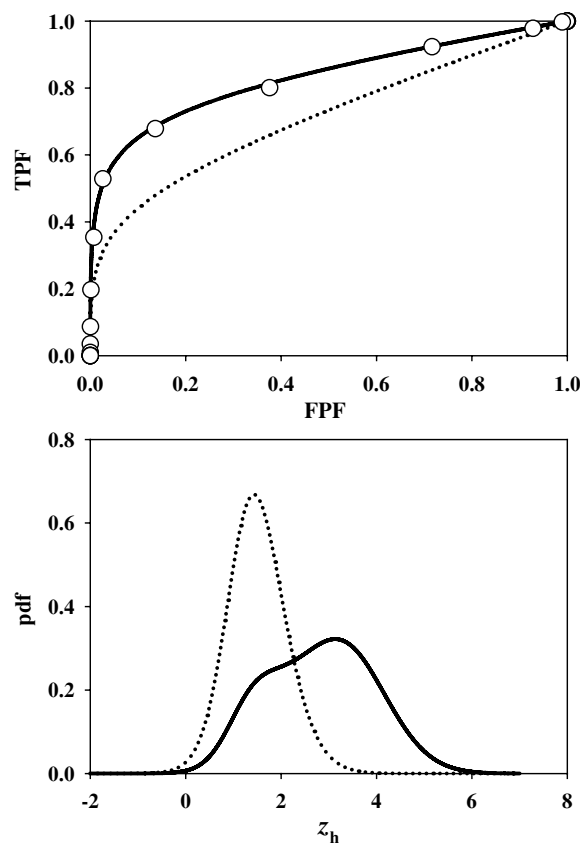
Figure 8. The ROC curve predicted by the search model for $\mu = 3$, $\lambda = 10$, $\nu = 0.5$ and $s = 2$. Excepting for the number of lesions, the parameter values are identical to those shown in figure 5, which was for $s = 1$. The dotted curve corresponding to figure 5 is shown in the upper panel for convenience.

ν. Such an observer would yield a larger figure of merit, AUC or θ, than an observer with the opposite characteristics. The net performance depends on the interplay of the parameters μ, λ and ν. For example, the expert observer (AUC = 0.964) whose ROC curve is shown in figure 6 (upper panel) is perfect at finding lesions ($\nu = 1$) but tends to cognitively evaluate about 3 non-lesion locations per image ($\lambda = 3$). It is important to note that $\nu = 1$ does not imply that all lesions are marked by this observer. It simply means that all of the lesions were hit, i.e., they were considered for marking. A lesion is actually marked only if the corresponding decision variable sample from $N(\mu, 1)$ exceeds the lowest cutoff. An unusually low z-sample and/or a strict criterion (ζ_1 large) would both result in an unmarked lesion, but the figure of merit, which is averaged over all z-samples and allows for the criterion effect, would be unaffected. Similarly, $\lambda = 10$ does not mean that the observer marks about 10 non-lesion locations per image. All search model parameters are characteristic of the observer and the task but can be influenced by the experimenter. For example, by asking the observer to evaluate normal regions that might not otherwise have been considered, one can potentially drive λ up (and make more of the ROC space accessible). The magnitude of this effect is expected to depend on the type of sites that the observer is asked to look at—if these are

obviously irrelevant sites (e.g., outside the anatomic area) then the observer will easily reject them and λ will be unaffected. However, if the sites resemble lesions then one expects λ to be adversely affected. Likewise μ and ν can also be influenced by reading conditions, e.g., poorly displayed images or a viewing time restriction.

The search-model-predicted ROC curves depend on s, the number of lesions per abnormal image. All other parameters being equal, images with $s = 2$ will yield a higher ROC curve than images with $s = 1$. An example is shown in figure 8. Note the higher figure of merit in this case, AUC $= 0.831$ versus AUC $= 0.711$ for $s = 1$. The increase of the figure of merit with s is expected since with more lesions in the image the observer is expected to hit more lesions but this should not be interpreted as an expertise effect. If it were possible to estimate the parameters μ, λ and ν from the free-response data obtained on an image set with $s = 2$, it would be possible, via equations (12) and (13), to predict the performance for $s = 1$, and the latter quantity should be interpreted as the true performance of the observer.

The ROC curves shown in figures 4 through 8 are 'proper' in the sense that the slope is monotonically changing, the curve always lies above the chance diagonal and the slope of any point on the ROC curve is never smaller than the slope of the line connecting that point to (1, 1). A number of alternative procedures are available for fitting 'proper' ROC curves, namely the likelihood ratio based model (Metz and Pan 1999), the contaminated binormal model, or CBM (Dorfman *et al* 2000), the bigamma model (Dorfman *et al* 1997) and the localization receiver operating characteristic model (Swensson 1996, Swensson *et al* 2001).

An interesting consequence of the search model is the inaccessible portion of the ROC curve. Any attempt by the experimenter to force the observer into the inaccessible portion is not expected to succeed (this statement is only true to the extent that λ is unaffected by the forcing). The inaccessible portion is particularly pronounced for observers characterized by small values of λ. In such cases the operating points are clustered near the initial near-vertical section of the ROC curve. There is evidence (Dorfman *et al* 2000) that observers sometimes provide data like this and that it is difficult to get them to generate appreciable numbers of false positives. This type of data clustering presents degeneracy problems for binormal model based analysis of such data. Instructions to the observers to 'spread their ratings' (Metz 1989) or to use 'continuous' ratings (Metz *et al* 1998) do not always seem to work. According to the search model the images that produce no decision variable samples will always be classified in the lowest bin no matter how lax the criterion. To the experimenter such observers will appear to be not heeding the advice to spread their ratings. Note that the CBM also provides an alternative explanation for the data clustering and initial near-vertical section of the ROC curve.

Some reasons for the inaccessible portion of the ROC curve, and for the need for the delta functions, have already been given. Basically these are both due to the fact that some images generate no hits. Since these concepts are fundamentally different from all existing ROC models, we elaborate further on the need for the delta functions and the inaccessible portion. One could argue that when the observer sees nothing to report then he starts guessing, and indeed this would enable the observer to move along the dotted portion of the curve in figure 3. This argument implies that the observer knows when his threshold is at negative infinity, at which point he turns on the guessing mechanism (the observer who always guesses would move along the chance diagonal connecting (0, 0) and (1, 1)). In my opinion this is unreasonable to expect. It is more likely that the observer will turn on the guessing mechanism at a low but finite value of the threshold. Different choices of the switching threshold would result in different ROC curves (the limiting point in figure 3 would move along the continuous portion and consequently the slope of the dotted line connecting it to (1, 1) would change). The existence of two thresholds, one for moving along the non-guessing portion and one for

switching to the guessing mode would require abandoning the concept of a universal ROC curve. To preserve this concept one needs the delta functions at negative infinity and the inaccessible portion of the ROC curve.

One may wonder why the widths of the signal and noise distribution in figure 1 are assumed to be the same. The conventional ROC model assumes that the widths in general are different. One reason is model parsimony—it is undesirable to introduce a parameter that may not be needed. Introducing a wider width for the signal distribution in figure 1 would also destroy the 'proper' ROC curve characteristic of the equal variance model. Another reason is that the search model, as it stands, provides an explanation for the well-known observation (Green and Swets 1966) that most binormal model fitted ROC curves imply a larger width for the abnormal image pdf relative to the normal image pdf. This is evident in all the pdfs shown in this work and is due to two effects. (a) Multiple samples occurring on normal images yield a narrower highest rating pdf (Fisher and Tippett 1928) which effect is more pronounced when λ is large, see figure 5 for an example. (b) For abnormal images a broadening of the pdf occurs since sometimes the noise site, sampled from $N(0, 1)$, yields the highest rating and sometimes the signal site, sampled from $N(\mu, 1)$, yields the highest rating. This effect is most pronounced when $\nu = 0.5$, see figure 4. It should be noted that the CBM also explains the observation that most binormal model fitted ROC curves imply a larger width for the abnormal image pdf.

Several similarities between search-model predicted ROC curves and those predicted by the CBM have already been noted. CBM is intended for ROC data, i.e., without localization, whereas the search model describes localization studies. Because of this difference comparisons only become possible when one considers ROC curve predictions of the two models. A detailed comparison of the two models is presented in the appendix. Comparisons with other ROC models are outside the scope of this work.

The ROC curves shown in the paper are logical predictions of the search model assumptions. Whether or not they fit actual operating points of real observers cannot be established until one has a method for estimating the parameters of the search model from observer search data. Currently I do not have the estimation capability. A similar situation applied to the case of ROC curves predicted by the likelihood ratio decision variable which had the interesting characteristic of being 'proper'. Only recently was a procedure developed to estimate likelihood ratio based ROC curves from ratings data (Metz and Pan 1999). I make no claim that search-model-predicted ROC curves will yield demonstrably better fits to ROC data. If the limiting point is far enough into the right-hand portion of ROC space, the extension of the curve by the straight line is minimal, and such cases can probably be well-fitted by any competing continuous model. When the limiting point moves towards the left-hand portion of ROC space, then the data are not expected to be well-fitted by the binormal model, but can probably be fitted by CBM. Strictly speaking the existence of a limiting point is inconsistent with all continuous ROC models, and especially so for small λ. However, due to the relatively small number of cases in most ROC studies, it will probably be difficult to distinguish between search model and continuous model predicted ROC curves (Hanley 1988).

The present search model builds on work by several other authors who have attempted to advance beyond the simple ROC paradigm. Some of these works have already been commented on. Swensson authored one of the earliest search models for medical imaging (Swensson 1980) when he described a two-stage process that is similar in concept to the present search model. His work on the location receiver operating characteristic (LROC) paradigm and a curve fitting procedure that simultaneously fits LROC data and yields proper ROC curves (Swensson 1996, Swensson *et al* 2001) is noteworthy insofar as it too attempts to address the localization issue. The issue of satisfaction of search is another attempt to go beyond the simple ROC paradigm (Berbaum *et al* 1990). Finally, it should be noted that

there are several tasks in imaging that go beyond the search task considered in this work. As an example, in mammography one is interested in finding (i.e., detecting) lesions and once a lesion has been found one seeks to classify the lesion as benign or malignant. The search model applies to the detection task and not to the classification task. Current ROC concepts and analyses continue to be applicable to the classification task.

Acknowledgments

Appendix

Relation of the search model to contaminated binormal model (CBM)

Some similarities between contaminated binormal model (CBM) and search-model-predicted ROC curves have already been noted. Here a more detailed comparison of the two models is presented. CBM is intended for ROC data, i.e., without localization, whereas the search model describes localization studies. CBM models a single decision variable sample per image whereas the search model allows multiple samples per image. In CBM there is an α parameter ($0 \leqslant \alpha \leqslant 1$) defined as the proportion of abnormal cases where the abnormality is visible. In CBM each normal image yields one sample from $N(0, 1)$ and each abnormal image yields *one* sample from either $N(\mu, 1)$, with probability α, or from $N(0, 1)$, with probability $1 - \alpha$, i.e., the abnormal image pdf is bimodal. In the search model each normal image yields n samples from $N(0, 1)$ where $n \geqslant 0$, and each abnormal image yields in addition u samples from $N(\mu, 1)$ where $u \geqslant 0$. CBM strictly applies to one lesion (or abnormality) per abnormal image ($s = 1$), whereas the search model allows multiple lesions. If CBM is applied to more than one lesion per abnormal image, the strictly Gaussian bimodal pdf model is not expected to hold, since lesion visibility will not be binary (visible or not visible).

In CBM the single z-sample determines the ROC rating whereas in the search model the highest of $n+u$ samples, z_h, determines the ROC rating. Although the pdfs shown in figures 4 and 7 appear similar to some of the plots in the CBM paper, there are fundamental differences. In CBM the normal image pdf and the bimodal abnormal image pdf are constructed from strictly Gaussian shaped functions. In the search model the pdfs are always strictly non-Gaussians, in some cases obviously so, e.g., figures 5 and 8. In CBM the total area under the normal image pdf is unity, and the total area under the two components of the abnormal image pdf is also equal to unity. This is not true of the search-model-predicted pdfs, where the corresponding quantities, excluding the delta functions, are x_{max} and y_{max}. See figure 7 for a pronounced example of the different areas.

Both models predict a straight line portion to the ROC curve that ends at (1, 1) but in CBM this portion is continuously accessible to the observer, whereas in the search model it is not. The reason is that in CBM a decision variable sample always occurs whereas in the search model there are images where no samples occur. The ν parameter of the search model (the fraction of lesions that are hit) is related to the CBM α parameter (the proportion of abnormal cases where the abnormality is visible). Both approach the limits 0 and 1 as μ approaches 0 and ∞, respectively, but the two are not identical. The reason is that CBM does not use localization information whereas the search model does. Therefore the terms 'visible' and 'hit' are not equivalent. The mathematical relationship between α and ν is outside the scope

of this paper, but as an illustration of the difference consider the following example. Suppose CBM is applied to a localization study and consider an image with $n = u = 1$ and where the signal site is *more* suspicious than the noise site. In CBM this would yield a sample from $N(\mu, 1)$. Therefore the ensemble of such images would tend to increase the estimate of α. Now consider an image with $n = u = 1$ but where the signal is *less* suspicious than the noise. In CBM this would yield a sample from $N(0, 1)$. Therefore the ensemble of such images would tend to decrease the estimate of α. In contrast to the search model, the estimate of ν would be indifferent to the two cases (since $u = 1$ in either case).

References

Berbaum K S *et al* 1990 *Invest. Radiol.* **25** 133–40
Bunch P C, Hamilton J F, Sanderson G K and Simmons A H 1978 *J. Appl. Photogr. Eng.* **4** 166–71
Chakraborty D P 2006 *Phys. Med. Biol.* **51** 3449–62
Chakraborty D P and Berbaum K S 2004 *Med. Phys.* **31** 2313–30
Dorfman D D and Alf E Jr 1969 *J. Math. Psychol.* **6** 487–96
Dorfman D D and Berbaum K S 2000 *Acad. Radiol.* **7** 427–37
Dorfman D D, Berbaum K S and Brandser E A 2000 *Acad. Radiol.* **7** 420–6
Dorfman D D, Berbaum K S, Metz C E, Lenth R V, Hanley J A and Abu Dagga H 1997 *Acad. Radiol.* **4** 138–49
Duchowski A T 2002 *Eye Tracking Methodology: Theory and Practice* (Clemson, SC: Clemson University)
Edwards D C, Kupinski M A, Metz C E and Nishikawa R M 2002 *Med. Phys.* **29** 2861–70
Fisher R A and Tippett L H C 1928 *Proc. Cambridge Phil. Soc.* **24** 180–90
Green D M and Swets J A 1966 *Signal Detection Theory and Psychophysics* (New York: Wiley)
Hanley J A 1988 *Med. Decis. Making* **8** 197–203
Kundel H L and Nodine C F 1983 *Radiology* **146** 363–8
Kundel H L and Nodine C F 2004 *Proc. SPIE* **5372** 110–5
Metz C E 1978 *Semin. Nucl. Med.* **VIII** 283–98
Metz C E 1986 *Invest. Radiol.* **21** 720–33
Metz C E 1989 *Invest. Radiol.* **24** 234–45
Metz C E, Herman B A and Shen J-H 1998 *Stat. Med.* **17** 1033–53
Metz C E and Pan X 1999 *J. Math. Psychol.* **43** 1–33
Nodine C F and Kundel H L 1987 *Radiographics* **7** 1241–50
Press W H, Flannery B P, Teukolsky S A and Vetterling W T 1988 *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge: Cambridge University Press)
Swensson R G 1980 *Percept. Psychophys.* **27** 11–6
Swensson R G 1996 *Med. Phys.* **23** 1709–25
Swensson R G, King J L and Gur D 2001 *Med. Phys.* **28** 1597–609