The Rose model, revisited

Arthur E. Burgess

Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115

Received June 9, 1998; accepted August 17, 1998

In 1946 and 1948, three very important papers by Albert Rose [J. Soc. Motion Pict. Eng. **47**, 273 (1946); J. Opt. Soc. Am. **38**, 196 (1948); L. Marton, ed. (Academic, New York, 1948)] were published on the role that photon fluctuations have in setting fundamental performance limits for both human vision and electronic imaging systems. The papers were important because Rose demonstrated that the performance of imaging devices can be evaluated with an absolute scale (quantum efficiency). The analysis of human visual signal detection used in these papers (developed before the formal theory of signal detectability) was based on an approach that has come to be known as the Rose model. In spite of its simplicity, the Rose model is a very good approximation of a Bayesian ideal observer for the carefully and narrowly defined conditions that Rose considered. This simple model can be used effectively for back-of-the-envelope calculations, but it needs to be used with care because of its limited range of validity. One important conclusion arising from Rose's investigations is that pixel signal-to-noise ratio is not a good figure of merit for imaging systems or components, even though it is still occasionally used as such by some researchers. In the present study, (1) aspects of signal detection theory are presented, (2) Rose's model is described and discussed, (3) pixel signal-to-noise ratio is discussed, and (4) progress on modeling human noise-limited performance is summarized. This study is intended to be a tutorial with presentation of the main ideas and provision of references to the (dispersed) technical literature.

OCIS codes: 100.2960, 110.4280, 330.1880, 330.4060, 330.7310, 030.4280, 040.1880, 110.3000, 330.5510.

1. INTRODUCTION

A. Summary

Rose[1–3] in his seminal publications proposed and described the use of (1) an absolute scale (quantum efficiency) for evaluating imaging system performance, and (2) a simple model of signal detectability by human observers. Rose used a particle-based fluctuations theory to compare realized devices with an ideal device, with the performances of both being assumed to be limited only by photon fluctuations. The images shown in Fig. 1 were used by Rose[4] in 1953 to demonstrate the maximum amount of information that can be conveyed by various known numbers of photons. Cunningham and Shaw,[5] in another paper in this feature, review the subsequent evolution of the quantum efficiency approach that culminated in spatial-frequency-dependent metrics based on modern theories of linear systems and stochastic processes. Rose's quantum efficiency is now viewed as a zero-frequency (dc) limit. This paper is concerned with the Rose signal detection model, which was based on a signal-to-noise ratio (SNR) also calculated by the particle fluctuation theory approach. Rose's work was done several years before the development of the modern theory of signal detectability[6,7] based on Bayesian probabilistic analysis. The Rose model, as it has come to be known, is useful for simple calculations but has a very narrow range of validity.

Research since 1948 has revealed that the problem of evaluation of imaging system performance and image quality is complex. A system that is optimized for one task may be suboptimal for another task. Tasks fall into two broad categories: One is classification (including detection, discrimination, and identification); the other is estimation of signal parameters. It is, in principle, possible to use a two-step process for image quality evaluation.[8] The first step is to determine the best possible task performance. This is achieved by the ideal observer, which makes optimum use of (1) the information available in the underlying physical phenomenon used to create the image, together with (2) available prior information to determine the most probable solution to the classification or estimation task. The second step is to determine task performance accuracy for the user, which could be a human or a numerical algorithm, based on realized images. The disparity between ideal and user-achieved performance gives an indication of how much room there is for improvement in the imaging system. Modeling of the task, the imaging system, and the user may give clues as to how to best improve the imaging chain. This is completely in the spirit of Rose's approach.

This paper is concerned only with the simplest classification task: detection of a signal with known parameters (size, shape, location) in additive, uncorrelated Gaussian noise on a known background. This is usually referred to as the signal-known-exactly (SKE), background-known-exactly (BKE) detection task. Rose's model is a useful approximation for SKE and BKE detection of a limited subset of signals and noise. After 50 years it is still a convenient and popular method for simple calculations and presentations of results. One purpose of this paper is to alert potential users to the limited range of validity of the model. Another is to demonstrate how it is related to more-general signal detection models. A third purpose is to caution against misuse of the Rose model in the form of a pixel SNR calculation. Readers interested in more-advanced aspects of image quality assessment are re-

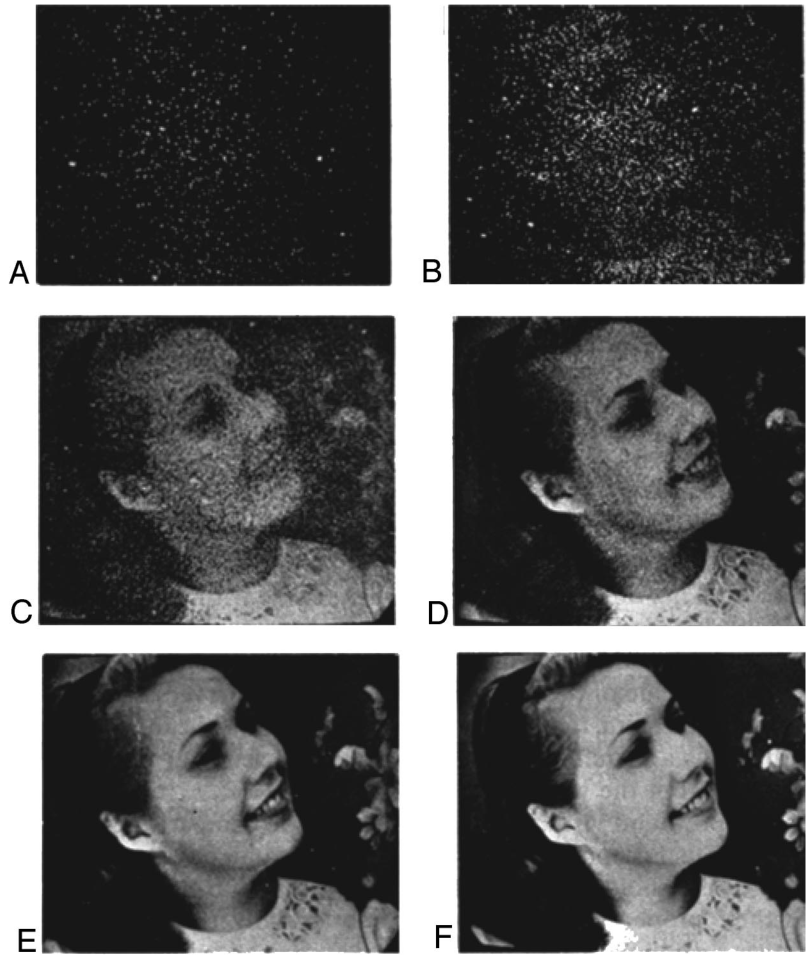

Fig. 1. Picture used by Rose,[4] of woman with flowers, to demonstrate the maximum amount of information that can be represented with varying numbers of photons. A, 3×10^3; B, 1.2×10^4; C, 9.3×10^5; D, 7.6×10^5; E, 3.6×10^5; F, 2.8×10^7. Each photon is represented as a discrete visible speck. The inherent statistical fluctuations in photon density limit one's ability to detect or identify features in the original scene.

ferred to a paper by Cunningham and Shaw in this feature[5] and to a series of papers by Barrett and co-workers.[9–11]

This paper begins with a description of the historical context of Rose's papers, gives a brief introduction of theoretical analysis of signal detection, introduces Rose's SNR model, and discusses the limited domain of validity of the Rose model. Then methods of evaluating human observer performance are described. Next, the inappropriateness of pixel SNR as a detectability measure is demonstrated. Finally, work on modeling noise-limited human performance is briefly reviewed.

B. Historical Perspective

Rose's papers on assessment of the performance of electronic devices and of the human visual system on an absolute scale came during a time of major technological change. Rose had been an important participant in the development of electronic television camera tubes at RCA since 1935. There had been a very rapid increase in camera sensitivity and a resulting need for evaluation metrics. Rose explored the consequences of the quantum nature of light as the fundamental determinant of performance. This was not a completely new idea; it had already been applied to the study of the limits of human visual system performance. In 1932, Barnes and Czerny[12] suggested that statistical fluctuations in photon arrival might limit human visual perception. In 1942, Hecht[13] estimated the minimum number of photons necessary to produce a sensation of light in the retina. de Vries[14] estimated effects of photon statistics on contrast sensitivity and acuity. Rose was the first to make complete use of a statistical approach.

Rose's work on noise effects and ultimate limits to sensitivity was concerned with just one of a number of factors limiting image quality. Early evaluations of the imaging properties of photography, television, and optical and visual systems were based on different, noncommensurable methods. O. Schade, at RCA, was developing spatial-frequency-dependent measurement techniques, measurement scales, and performance indices that could be used across the entire spectrum of imaging modalities. The results of that work are summarized by examples in his marvelous book *Image Quality: a Comparison of Photographic and Television Systems*,[15] with a foreword by Rose. Much of Schade's work has been translated into modern signal detection theory notation by Wagner.[16,17]

Rose's ideas found immediate application in the medical imaging field. In 1948, Coltman described[18] the development of a prototype electronic x-ray image intensification device and pointed out that the images would be impaired by the quantum nature of light. Sturm and Morgan[19] used Rose's concepts and observer model to calculate the expected benefits of electronic intensification in fluoroscopy. They explicitly pointed out the fundamental role that photon statistics would play in limiting the clarity with which fluoroscopic images would be visualized.

C. What is Signal-to-Noise Ratio?

Noise is a ubiquitous problem in science and engineering. Anyone attempting to define a SNR must somehow characterize the strengths of both signal and noise. The theory of signal detectability provides a definition of SNR that is not arbitrary. For pedagogical reasons, the theoretical development given here is as follows. The calculation of the optimum detection filter[20] based on linear systems theory is described in some detail. The main points of the theory of signal detectability[6] are presented to show that the optimum linear detection filter is also the optimum strategy in the probabilistic sense for special cases. Then the Rose model is described and discussed. The author has encountered other definitions of SNR over the years, based on some ratio of signal mean and noise standard deviation. There are many ways to define signal mean. Noise standard deviation is not a good measure if the noise is correlated. So SNR calculations can be very misleading if not done correctly. The key point is that attempts to minimize the effect of noise on task performance, by filtering, for example, must take into account the nature of the task. This leads to very different optimization strategies for different tasks: maximizing a SNR by matched filtering[20] for detection of a known signal, and minimizing rms error by filtration[21–23] for estimation and prediction tasks.

Consider the problem of optimizing signal detection. The basic concepts were described very nicely by Harris[24]:

> Every sensor is fundamentally limited by some form of statistical variability, i.e., noise. In the presence of noise, an object cannot be defined exactly. Where a frequency function is assumed for the noise which takes on all values from zero to infinity, it is theoretically possible for any object to result in any image. In practice many of these combinations are events of such small probability that they have no significance.... Since a given image can result from any number of objects, a decision as to which object actually produced the image can be made in statistical terms only. If the set of possible objects is known, then the probability that the image is due to each of the objects can be calculated. A decision as to which object actually produced the image should be based on comparison of these probabilities.

D. Matched Filter

North developed the theory of the prewhitening matched filter method for detecting a known signal in additive Gaussian noise. This was described in a superbly written 1943 RCA report[20] that was circulated by mimeographed copy to a generation of engineering graduate students. Eventually it was reproduced in 1963 as an Institute of Electrical and Electronics Engineers paper[20] so that it could be easily accessible—and legible. North's work was done in the context of radar, where signals and noise are functions of one variable (time), which is used here to simplify notation. Extension to two spatial dimensions is straightforward. His approach is presented in some detail because it nicely illustrates a number of important points. Note that North was a contemporary of Rose's at the RCA Laboratories and that Rose repeatedly acknowledged in his publications the benefits of discussions with North.

Consider analyzing the binary decision problem of deciding whether a signal has been received in the presence of noise. A number of constraints and assumptions must be introduced to make this a well-defined problem. North asked the question, What is the best filter to use at the radar receiver so that its output gives the best contrast between signal and noise? In other words, if the signal is present, then the filter should give a sharply peaked output, whereas, if the signal is not present, the filter should give as low an output as possible. It was assumed that the filter designer knew the temporal profile and arrival time of the signal (a short radar echo pulse of a particular shape). So the problem is to detect a known signal, $s(t)$, in a waveform, $x(t)$, with additive random noise, $n(t)$, by use of a filter with an impulse response, $h(t)$. The filter input and output functions are given by

$$\begin{aligned} \text{input:}\quad & x(t) = s(t) + n(t), \\ \text{output:}\quad & y(t) = h(t) * x(t), \quad s_0(t) = h(t) * s(t), \\ & n_0(t) = h(t) * n(t), \end{aligned} \tag{1}$$

where $*$ denotes convolution. North chose to select the optimum filter by maximizing the ratio of the signal amplitude to the noise amplitude of the output at some instant in time, based on the known arrival time of the signal. Since the noise is a random variable, it was characterized by the mean-square value, $\langle {n_0}^2(t) \rangle$, of its amplitude, where $\langle ... \rangle$ denotes an expectation value. This gave the following quadratic ratio, ρ, to characterize the relative strength of output signal and noise at a particular measurement time, t_m:

$$\rho(t_m) = \frac{{s_0}^2(t_m)}{\langle {n_0}^2(t_m) \rangle}. \tag{2}$$

Note that use of Eq. (2) does not consider or guarantee that the output signal will in any way resemble the input signal. The only concern is to maximize a scalar quantity at t_m.

To simplify the mathematics it was assumed that the noise is a stationary stochastic process.[25] Mathematical analysis of ρ is more convenient in the frequency domain by means of the signal amplitude spectrum, $S(f)$; the detector filter transfer function, $H(f)$; the noise function Fourier transform, $N(f)$; and the noise power spectrum, $P(f)$. The input and output spectra are related by

$$S_0(f) = H(f)S(f),$$

$$P_0(f) = \langle |N_0(f)|^2 \rangle = \langle |H(f)|^2 |N(f)|^2 \rangle = |H(f)|^2 P(f), \tag{3}$$

where |...| denotes an absolute value.

The filter output ratio can be evaluated in terms of the inputs. To simplify notation, the time parameter is defined so that the expected arrival time, t_m, of the signal is equal to zero. Since we are interested in a spatial filter for image applications, we do not have to worry about the causality issue, which is important for temporal filters. The filter output ratio, with frequency-domain functions, then becomes

$$\rho(t_m) = \frac{{s_0}^2(t_m)}{\langle {n_0}^2(t_m) \rangle} = \frac{\left[\int_{-\infty}^{\infty} H(f)S(f)\exp(i2\pi f t_m)\mathrm{d}f \right]^2}{\int_{-\infty}^{\infty} |H(f)|^2 P(f)\exp(i2\pi f t_m)\mathrm{d}f},$$

$$\rho(0) = \frac{\left[\int_{-\infty}^{\infty} H(f)S(f)\mathrm{d}f \right]^2}{\int_{-\infty}^{\infty} |H(f)|^2 P(f)\mathrm{d}f}. \tag{4}$$

The goal is to determine the particular filter that maximizes this ratio. We multiply and divide the integrand of the denominator by $\sqrt{P(f)}$. Since the denominator and the numerator of the ratio are scalars, we make use of the following Schwarz inequality[22]:

$$\left[\int_{-\infty}^{\infty} H(f)\sqrt{P(f)} \frac{S(f)}{\sqrt{P(f)}} \mathrm{d}f \right]^2 \leqslant \int_{-\infty}^{\infty} |H(f)|^2 |P(f)\mathrm{d}f \int_{-\infty}^{\infty} \frac{|S(f)|^2}{P(f)} \mathrm{d}f. \tag{5}$$

The range of possible output ratios is then given by the inequality

$$\rho(0)| \leqslant \frac{\int_{-\infty}^{\infty} |H(f)|^2 P(f)\mathrm{d}f \int_{-\infty}^{\infty} \frac{|S(f)|^2}{P(f)} \mathrm{d}f}{\int_{-\infty}^{\infty} |H(f)|^2 P(f)\mathrm{d}f},$$

$$\rho(0) \leqslant \int_{-\infty}^{\infty} \frac{|S(f)|^2}{P(f)} \mathrm{d}f. \tag{6}$$

The maximum value of $\rho(0)$ is obtained when the equality holds and requires that

$$H_{\max}(f) = \frac{\alpha S^*(f)}{P(f)}, \tag{7}$$

where * denotes the complex conjugate and α is an arbitrary real scalar constant. This result for $H(f)$ is the prewhitening matched filter. The output ratio with this filter is

$$\rho_{\max}(0) = \int_{-\infty}^{\infty} \frac{|S(f)|^2}{P(f)} \mathrm{d}f. \tag{8}$$

Prewhitened matched filtering can be understood by consideration of the following two-stage procedure. The first step is inverse filtering of the input by use of a filter $H_1(f)$ to remove the intrinsic correlations in the noise (i.e., prewhiten the noise). The transfer function of $H_1(f)$ is determined with the criterion that $|H_1(f)|^2 P(f)$ equal a constant. The second step is to detect the signal in the waveform with white noise. Inverse filtering of the input waveform by use of $H_1(f)$ will also change the expected signal profile, so the revised detection filter matched to the signal must be $H_2(f) = \alpha S^*(f) H_1(f)$. Assuming linear filters, we can select a single linear filter that simultaneously prewhitens and matches the revised signal with the product $H(f) = H_1(f)H_2(f)$, which yields Eq. (7).

For the special case of white noise, the matched filter is $H(f) = \alpha S^*(f)$. The corresponding optimum impulse response for detection by convolution is $h(t) = \alpha s(-t)$, a time-reversed version of the signal profile. The convolution output is evaluated at the known signal position. The equivalent result can be obtained by cross correlation with the template, $\alpha s(t)$, correctly positioned at the expected signal location. Note that, since the constant α is arbitrary and disappears in the ratio calculation, it is usually set to unity. But the implication of arbitrary α is that, at this point in the task, the decision device (observer) does not need to know the amplitude of the signal and can use any template proportional to the signal. For example, the template can be normalized to unit peak signal amplitude or unit signal energy.

We have obtained the linear filter that maximizes the ratio of signal and noise strengths as defined by Eq. (2). However, it remains to be shown that this ratio is actually the appropriate quantity to use for optimization of SKE signal detection. In Subsection 1.E it is demonstrated that, for the special case of noise with a Gaussian probability distribution, the prewhitening matched linear filter is the absolute optimum filter; that is, there is no other filter, linear or otherwise, that can improve detection efficiency. There are two additional shortcomings to the matched filter approach. First, the task of detecting an exactly known signal is too simplistic. The SKE assumption is rarely valid for real-life problems. Second, the assumption of stationary noise is rarely valid. Bayesian SDT provides the means to investigate optimum strategies for more complex tasks.

E. Likelihood Ratio Approach

The likelihood ratio approach comes from the theory of signal detectability developed by Peterson *et al.*[6] for radar applications. The theory is now more commonly referred to as signal detection theory (SDT). SDT is based on the Bayesian mathematics of testing statistical hypotheses (statistical decision theory) developed by Fisher,[26] Neyman and Pearson[27] and Wald,[28] among others. Decision theory deals with the problem of decision making under conditions of risk and uncertainty. It provides (1) a method of controlling and measuring the criterion used in making decisions about signal existence, and (2) an absolute performance measure that is independent of signal intensity and the selected decision criterion. The analysis of signal detectability is described in great detail in books by Swets,[29] Green and Swets,[30] Van Trees,[31] and Helstrom.[32] Next it is shown that SDT proves that the linear matched filter is the optimum strategy for the SKE detection task evaluated by North.

Consider again detection of a known (SKE) signal in the waveform, $x(t)$, which is defined over some finite interval of time, $0-T$. The observer must make a binary decision about this waveform: It contains either (1) only noise, or (2) signal plus noise. The decision will accept one of two hypotheses about the state of the world under investigation, denoted by H_i. The observer also needs to know the prior probabilities of occurrence $P(H_i)$ of the two states. Once the waveform has arrived, the observer can use Bayes's theorem to calculate the conditional *a posteriori* probability of the truth of each hypothesis about the state of the world, $P[H_i|x(t)]$, given the new evidence, $x(t)$:

$$P[H_i|x(t)] = \frac{P(H_i)P[x(t)|H_i]}{P[x(t)]} = \frac{P(H_i)P[x(t)|H_i]}{\sum_i P(H_i)P[x(t)|H_i]}. \tag{9}$$

The observer can then adopt several strategies. One is to select the hypothesis that has the maximum *a posteriori* probability (the MAP approach), which can, of course, be extended to deciding between many alternative hypotheses. Sometimes it is preferable to evaluate the evidence independently of the prior probabilities of occurrence. The evidence about the state of the world is conveniently summarized in a positive real-valued scalar quantity, the likelihood ratio, $\lambda[x(t)]$. This is the ratio of the conditional probability densities of the observed data when the signal is present (state, s) and absent (state, n):

$$\lambda[x(t)] = \frac{p[x(t)|s]}{p[x(t)|n]}. \tag{10}$$

The observer computes the likelihood ratio by determining the conditional probability densities under each hypothesis. Once the likelihood ratio is calculated, the observer compares it with a decision criterion, κ, and decides that the signal is present if $\lambda[x(t)]$ is greater than κ. The probabilities of correct detection given signal present, $P(\lambda > \kappa|s)$, and false alarms given no signal, $P(\lambda > \kappa|n)$, are determined by evaluation of the following integrals with the conditional probabilities for $\lambda[x(t)]$ under the two possible states of the world:

$$P(\lambda > \kappa|s) = \int_\kappa^\infty p\{\lambda[x(t)]|s\}\mathrm{d}\lambda[x(t)],$$

$$P(\lambda > \kappa|n) = \int_\kappa^\infty p\{\lambda[x(t)]|n\}\mathrm{d}\lambda[x(t)]. \tag{11}$$

Consider the simplest case of stationary Gaussian white noise. It is assumed that the functions $x(t)$, $s(t)$, and $n(t)$ can be represented by a finite number, M, of discrete samples in either the temporal or the frequency domain. The temporal domain is used here. The collections of samples are denoted by $\{x_m\}$, $\{s_m\}$, and $\{n_m\}$. The noise samples are assumed to be independent and identically distributed with zero mean and variance σ^2. The conditional probability densities of a particular single sample, x_m, of the waveform under the two possible states are described by

$$p(x_m|H_i) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x_m - v_{mi})^2}{2\sigma^2}\right], \tag{12}$$

with means $v_{mi} = s_m$, for signal present, and $v_{mi} = 0$, for noise only.

The conditional probability densities of the collection of samples $\{x_m\}$ for the two states are given by the product of individual sample conditional probability densities:

$$p(\{x_m\}|H_i) = \prod_{m=1}^{M} p(x_m|H_i). \tag{13}$$

The likelihood ratio is evaluated by substitution of the two results for $p(\{x_m\}|H_i)$ in Eq. (10).

$$\lambda[\{x(t)\}] = \frac{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^M \prod_{m=1}^{M} \exp\left[-\frac{(x_m - s_m)^2}{2\sigma^2}\right]}{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^M \prod_{m=1}^{M} \exp\left[-\frac{(x_m)^2}{2\sigma^2}\right]} = \prod_{m=1}^{M} \exp\left[-\frac{(x_m - s_m)^2 + {x_m}^2}{2\sigma^2}\right]. \tag{14}$$

The product of exponentials can be replaced by an exponential of the sum of arguments. Additional simplification can be obtained by evaluation of the natural logarithm of $\lambda[\{x_m\}]$, since any monotonic transformation of the likelihood ratio is also an equivalent decision variable. The final result is

$$\Lambda = \ln(\lambda[\{x_m\}]) = \frac{1}{\sigma^2}\left(\sum_{m=1}^{M} x_m s_m - \sum_{m=1}^{M} {s_m}^2/2\right) = \frac{1}{N_0}\left(\sum_{m=1}^{M} x_m s_m - E/2\right). \tag{15}$$

The signal energy, $E = \Sigma_{m=1}^{M} {s_m}^2$, is constant, and N_0 is the two-sided noise spectral density defined by use of cycles per unit sampling distance. It is immediately recognized that Λ is linearly related to the cross-correlation result, $\Sigma_{m=1}^{M} x_m s_m$, between the received waveform samples and the expected signal samples. So cross correlation with the signal as a template is the optimum signal detection method, given SKE and BKE conditions with white, Gaussian noise. Since matched filtering and evaluation of the signal arrival time is equivalent to cross correlation, we conclude that North was completely correct in his choice of a quantity to maximize.

The observer that uses the optimum Bayesian strategy is known as the ideal observer. SDT analysis of ideal observer procedures and performance can be extended to cases in which the signal parameters are known only statistically through probability distributions. SDT allows one to include a number of sources of uncertainty, including image noise, signal and background parameter variability, prior probabilities of various signals and backgrounds, decision risks and costs, and a variety of decision criteria. A complete discussion can be found in a three-volume set by Van Trees.[31] The pattern recognition literature[33] provides mathematical techniques that can be adapted for digital image applications. Barrett and co-workers[9–11] have presented a number of applications of the SDT approach to image quality assessment.

F. Detectability Index

The above presentations dealt with the ideal observer, its procedures for SKE signal detection, and determination of the probabilities of true-positive responses and false alarms by use of Eqs. (11). One would prefer to obtain a one-dimensional performance scale. One way to do this is to determine true-positive and false-positive rates as the decision criterion is changed with all the other parameters fixed. The plot of the covariation of true- and false-positive probabilities is referred to as the receiver operating characteristic (ROC) curve. This ROC method allows the experimenter to obtain two estimates from the data. One quantity (sometimes called bias) shows where the observer is operating on the ROC curve, given a particular value selected for κ. The other quantity is the area under the ROC curve, A_z, which provides a scalar measure of performance at a given SNR that is independent of the decision criterion. An alternative performance measure, the detectability index d_a, can be obtained by an inverse error function-based transformation[30] of A_z. The ROC method has been the gold standard for evaluation of performance with clinically acquired medical images and comparisons of competing imaging methods. Guidance on dealing with the many subtleties of its use can be found in publications by Metz.[34,35]

An alternative approach is to determine the means, $\langle \Lambda_s \rangle$ and $\langle \Lambda_n \rangle$, and variances, ${\sigma_{\Lambda s}}^2$ and ${\sigma_{\Lambda n}}^2$, of the log-likelihood ratios under the two equally probable states (s, n) of the world and calculate a detectability index, d', using the equation[30]

$$(d')^2 = \frac{[\langle \Lambda_s \rangle - \langle \Lambda_n \rangle]^2}{(1/2)[{\sigma_{\Lambda s}}^2 + {\sigma_{\Lambda n}}^2]}. \tag{16}$$

This detectability index is a normalized Euclidean distance measure that describes the separation of the two population means (signal and no signal present) relative to the average population standard deviation. This equation must be used with great care. If the underlying probability distributions for the log-likelihood ratios are far from Gaussian, the resulting value of d' estimated with Eq. (16) can be seriously wrong. In some cases it may be possible to find an alternative decision variable with Gaussian probability distributions by use of some monotonic transformation of log-likelihood ratio. Then d' can be calculated with the new decision variable means and variances in Eq. (16). For other distributions, the ROC method is a much more reliable approach.[36] Advanced aspects of the use of the ROC approach are discussed by Barrett *et al.*[11]

Very often, the experimenter is interested only in estimating an observer's detectability index directly, with variation in the observer's decision criterion being irrelevant. In this case a more attractive way to evaluate observer performance is to present two or more alternative waveforms (or images) with exactly one containing a signal. The observer is asked to select the most probable alternative. This is known as the M-alternative forced-choice (MAFC) method. For the two-alternative forced-choice method the probability of a correct response, P, is equivalent to the area under the ROC curve.[30] The value of P can then be transformed to an estimate of a detectability index, d', by use of Elliot's tables[29] or their equivalent. The best choice of method, ROC or MAFC, and the best value of M to use depends on details of the experimental conditions.[37]

The relationship between SNR and the detectability index, d', is the source of some confusion. SNR is a measure describing the relative intensities of signal and noise defined, for example, by the SKE signal detection task described above. The detectability index is an objective measure of observer performance, independent of the observer's decision criterion. For the ideal observer doing a SKE and BKE signal detection task, d' is numerically equal to SNR. The equality does not hold for suboptimal observers. For more complex tasks, d' and SNR can be different for all the observers. Engineering texts frequently use d to represent the quadratic quantity presented here as $(d')^2$. The use of d' came from early perception work at the University of Michigan. They chose to use d as in the engineering convention and therefore chose d' as a convenient symbol to represent $\sqrt{d}$.[38]

G. Efficiency

An approach to evaluating nonideal observer performance on an absolute scale by an efficiency measure was suggested by Tanner and Birdsall.[39] They defined observer efficiency, η, using the ratio of input signal energies, E_i and E_t, required by the ideal observer and the observer under test to do the selected task at the same performance level. Occasionally an alternative efficiency definition is used, based on the squared ratio of detectabilities, d_t' and d_i', for the test and ideal observers at a fixed input signal energy. This definition of efficiency can be distinguished from the Tanner–Birdsall definition by the subscript 2:

$$\eta = \left(\frac{E_i}{E_t} \right)_{d'}, \quad \eta_2 = \left(\frac{d_t'}{d_i'} \right)_E^2. \tag{17}$$

If the detectability indices for both test and ideal observers are directly proportional to signal amplitude, then the two definitions are equivalent. However, this condition is not satisfied for many tasks and is rarely found for human observers doing any task, so the two definitions usually give different results. Most researchers use the Tanner–Birdsall definition.

The observer efficiency measure is useful for a number of reasons. (1) It serves as a way of comparing the performance of human observers across a variety of tasks. (2) Aspects of the task that will also reduce ideal observer performance are eliminated from consideration. (3) It provides some guidance as to the type of mechanisms that might be the cause of suboptimal performance. As Rose pointed out,[2] "The gap, if there is one, between the performance expected from fluctuation theory and the actual performance is a measure of the 'logical' space within which one may introduce special mechanisms, other than fluctuations, to determine its performance." (4) One can use SDT analysis of suboptimal observer model performance to search for possible causes of discrepancies between theoretical predictions and human performance. Mechanisms and observer models that predict efficiencies lower than those found by experimental tests of that observer's performance can immediately be rejected. (5) Efficiency measurements give a direct estimate of how much improvement in human performance is possible (by improved display methods or image enhancement, for example).

2. ROSE MODEL

A. Theory

The basic ideas of the Rose model are presented here in a slightly different way from that of his publications, to make a number of important distinctions. Rose was considering the detectability of a flat-topped, sharp-edged signal of area, A, in a uniform background. This choice of signal allowed it to be described by either one scalar (the average incremental number of photons, $\langle \Delta N_s \rangle$, that represent the signal) or two scalars (the signal contrast, C, and area, A) rather than a function, $s(x, y)$. Rose described the photon noise by its statistics in a background region with the same area as the signal: the mean number of photons in the area, $\langle N_b \rangle$, and the variance, ${\sigma_{N_b}}^2$, which equals the mean, since photons have Poisson statistics. It is very important to note, as Rose did, that recorded photons are uncorrelated. For clarity, it is more convenient to use the corresponding mean photon densities, $\langle \Delta n_s \rangle$ extra photons per unit area for the signal, and $\langle n_b \rangle$ expected photons per unit area for the background. Using a signal contrast, $C = \langle \Delta N_s \rangle / \langle N_b \rangle = \langle \Delta n_s \rangle / \langle n_b \rangle$, Rose defined SNR with the equation

$$\mathrm{SNR}_{\mathrm{Rose}} = \frac{\text{mean signal}}{\sigma_{N_b}} = \frac{\langle \Delta N_s \rangle}{\sqrt{\langle N_b \rangle}} = \frac{A \langle \Delta n_s \rangle}{\sqrt{A \langle n_b \rangle}} = C \sqrt{A \langle n_b \rangle}. \tag{18}$$

Now consider the Rose SNR definition in the context of SDT. Two assumptions are necessary to compare the Rose model based on Poisson noise with the previous SDT models based on Gaussian noise. (1) The Rose model neglects the fact that noise in the potential signal location has unequal variances for the signal-present and signal-absent cases. If the signal is present, the variance will equal the mean total, $\langle \Delta N_s + N_b \rangle$. So we need to assume that ΔN_s is very small compared with N_b. Hence the Rose model is an approximation that is valid only in the limit of low-contrast signals. (2) The photon noise has a Poisson distribution, whereas the above SDT approach was based on a Gaussian distribution. We need to assume that photon densities are large enough that Poisson noise can be approximated by Gaussian noise with the same mean and variance.

Other aspects of the Rose model fit conditions of the above SDT model for uncorrelated noise. (1) Rose used completely defined signals at known locations on a uniform background for his experiments and analysis, so the SKE and BKE constraints of the simple SDT analysis were satisfied. (2) He assumed perfect use of prior information about the signal. (3) He used a flat-topped signal at the detector plane (which is only possible when there is no imaging system blur). This assumption meant that integration of photon counts over the known signal area is equivalent to cross-correlation detection. This can be seen by consideration of the following argument. Consider the cross correlation of a signal, $s(x, y)$, and a template, $t(x, y)$, both located at the expected signal location $(x = y = 0)$ in an image with additive, uncorrelated Gaussian noise. The noise functions with and without the signal are $n_s(x, y)$ and $n_n(x, y)$ with zero mean and spectral density equal to $\langle N_0 \rangle$. Let the signal have a flat top and sharp edges (the signal is nonzero only in the range R) with area A and amplitude a above the background. Let the template be an arbitrarily amplified version of the signal, $t(x, y) = \tau s(x, y)$. Using Eqs. (15) and (16), one obtains the SNR defined by SDT:

$$\begin{aligned}
\langle \Lambda_s \rangle - \langle \Lambda_n \rangle &= \left(\frac{1}{N_0}\right) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \langle \{ s(x, y) + n_s(x, y) \\
&\quad - n_n(x, y) \} t(x, y) \rangle \mathrm{d}x \mathrm{d}y \\
&= \left(\frac{a \tau}{N_0}\right) \int \int_R \mathrm{d}x \mathrm{d}y = \frac{a \tau A}{N_0}, \\
{\sigma_{\Lambda_s}}^2 = {\sigma_{\Lambda_n}}^2 &= \left(\frac{1}{N_0}\right)^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \langle n(x, y)^2 t(x, y)^2 \rangle \mathrm{d}x \mathrm{d}y \\
&= \left(\frac{\tau}{N_0}\right)^2 \int \int_R \langle n(x, y)^2 \rangle \mathrm{d}x \mathrm{d}y = \frac{\tau^2 N_0 A}{{N_0}^2}, \\
\mathrm{SNR}^2 = (d')^2 &= \frac{(\langle \Lambda_s \rangle - \langle \Lambda_n \rangle)^2}{(1/2)({\sigma_{\Lambda s}}^2 + {\sigma_{\Lambda n}}^2)} = \frac{\tau^2 a^2 A^2 {N_0}^2}{\tau^2 A {N_0}^3} \\
&= \frac{a^2 A}{N_0}.
\end{aligned} \tag{19}$$

The equality of Rose's SNR approximation and the SDT definition of SNR can be shown by use of Rose's definition of contrast and by the fact that N_0 equals $\langle n_b \rangle$ for Poisson noise:

$$C = \frac{\langle \Delta n_s \rangle}{\langle n_b \rangle} = \frac{a}{N_0},$$

$$\mathrm{SNR}_{\mathrm{Rose}}^2 = C^2 A \langle n_b \rangle = \left(\frac{a}{N_0}\right)^2 A N_0 = \frac{a^2 A}{N_0} = \mathrm{SNR}^2. \tag{20}$$

So the Rose model, when limited to this idealized special case, is a very good approximation to the Bayesian ideal observer model. However, it must be used with care to ensure that all the above conditions and assumptions are satisfied.

Rose used his model to assess the detectability of signals, in that he asked the question, "What SNR was required in order to detect a signal?" His approach was to use a constant, k, defined as the threshold SNR, and to suggest that the value of k must be determined experimentally. The signal was expected to be reliably detectable if its SNR were above this threshold. Once k is selected, the corresponding contrast threshold C_T is given by

$$C_T = \frac{k}{\sqrt{A \langle n_b \rangle}}. \tag{21}$$

Using this definition of threshold SNR has the unfortunate effect of mixing the measure of relative signal strength (SNR) with the observer's decision criterion. However, it followed the convention of the day. This threshold SNR concept is still sometimes used by people unfamiliar with SDT.

B. Determining the Value of *k*

Rose's thoughts on the selection of k are presented in some detail to show that he considered many issues that would subsequently be resolved by the probabilistic approach of SDT. For example, he was searching for ways to include both signal uncertainty and observer inefficiency effects into the value of k. The value of k had been assumed to be unity in earlier publications by Rose[1,40] and de Vries.[14] Rose performed experiments to estimate k by means of highly amplified, static, photon-noise-limited images of disk test object arrays. He concluded that a higher value of k was needed. He stated (Ref. 2, pp. 199–200) that "Some estimates [of k] made recently ... lay in the range of 3 to 7. Additional and more direct evidence is given in the next section that the value of k is not unity but is in the neighborhood of 5." In the other 1948 paper, Rose said[3] that "A recent direct determination by Schade[41] sets the value of k in the range of 3–6, depending on the viewing conditions." The investigation of the value of k raised several issues. Rose observed[3] that "This constant [k] ... is a function of what one calls a threshold. That is, if one asks that each observation have a 90% chance of being correct, the threshold signal to noise ratio will be higher than if one were satisfied with a 50% chance." Rose was also searching for other explanations for the observed value of k, based on human limitations. He noted,[3] "This [value of k] is surprisingly high since mathematical analysis suggests a value close to unity. One way of reconciling the discrepancy is to assume that the eye in looking at the pictures in Fig. 5 does not make full use of the information presented. That is, the eye (and brain) may count only a fraction of the 'quanta' present." This observation illustrated his understanding that human efficiency is limited to both peripheral (eye) and central (brain) considerations. Rose pointed out that the range of validity of a particular value of k is also constrained by other aspects of the human visual system. The value of k increases when signal size is outside the angular range of approximately 1–10 millirad in diameter. For smaller signals, blur due to the point-spread function of the eye mixes signal with nearby noise and thus couples additional image noise into the decision task. For large signals, human spatial integration limits come into play, and signal detection eventually becomes limited by contrast sensitivity. Some experimental results that are pertinent to the value of k are presented in Subsection 2.C.

C. Experimental Methods

Rose used a fixed two-dimensional (2D) array of disk signals with a range of sizes and contrasts in his experiments. The observers were asked to indicate which signals in the array were visible. The signals were always present, so the observer's decision about whether a particular signal was seen was subjective. This method has the potential for significant systematic error due to differences and variations in the observer decision criterion. Objective evaluation of performance in Rose's experiments is not possible within the framework of SDT.

A number of methods were developed in subsequent years to eliminate effects of observer bias and to allow accurate estimates of the underlying intrinsic detectability of signals.[30] One modern experimental approach is to have one known signal location with the signal either present or absent. Observers give a binary response (yes/no) and use a different decision criterion in different blocks of trials to generate a ROC curve. An alternative (and more efficient) ROC-generating approach uses a confidence rating scale for responses. The area under the ROC (A_z) can be used as a performance measure, or it can be transformed into a detectability index (d_a). Another experimental method (forced choice) is done with the signal always present, placed in one of M statistically independent locations. The observer is asked to select the most likely location. The percentage of correct responses or the corresponding detectability index is then used as the performance measure. Design and analysis of these types of experiments are discussed by Green and Swets[30] and by Macmillan and Creelman.[42] Experimental design issues related to images are discussed by Metz[34,35] and Burgess.[37]

The utility of the efficiency measure is illustrated by multiple-alternative forced-choice (MAFC) search experiments[43] that use disk signals with a variety of amplitudes in white noise. In these experiments there was exactly one signal present at one of M statistically independent locations in each image. The observer used a trackball and a cursor to select the most likely location. The SNR necessary to obtain 90% correct responses as a function of M is shown in Fig. 2, for two human observers, the ideal observer, and an observer operating at 50% efficiency. This experiment demonstrates that, for this simple task, human efficiency remained constant at 50% as M increased; the additional complexity of the search task did not cause additional inefficiency. Similar constant efficiency results were obtained for a combination of two-alternative forced-choice detection and MAFC identification experiments with signals chosen from the 2D Hadamard basis function set.[44]

Next, a method of estimating a subjective threshold SNR, k, by use of a signal location identification task is illustrated. A random number (1–16) of disk signals (radius, 4 pixels) were randomly positioned in each image with 392 possible nonoverlapping signal locations. There were 2000 signals present in the set of 256 images. The signal amplitudes were randomly varied to give SNR's in

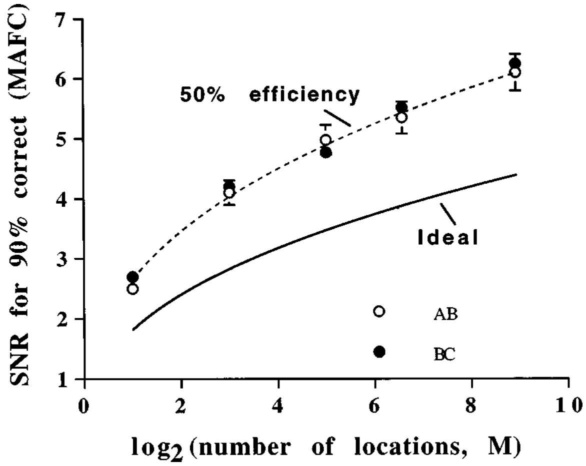

Fig. 2. Results for MAFC disk signal location identification experiments.[43] The figure shows the SNR required for 90% correct detection performance by the ideal observer and by two human observers as a function of the base 2 logarithm of the number of statistically independent signal locations, M. Human efficiency is constant at 50% (SNR = $\sqrt{2}$ higher than ideal). This illustrates one of the benefits of using efficiency as a summary measure of task performance.

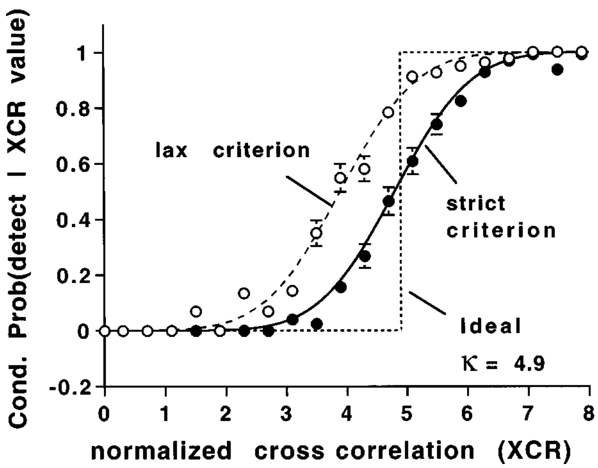

Fig. 3. Results for a free-response experiment for two human observers as a function of measured XCR's between signals and image data. A variable number of disk signals with variable SNR were presented at random positions in white-noise images. The observers located as many signals as possible under two decision strategies: lax (low miss rate) and strict (low false-alarm rate), which produced false-alarm rates of 3×10^{-3} and 2×10^{-4}, respectively. The conditional probability of detection was determined at each quantized XCR value. If one uses a 50% correct detection criterion to define a threshold SNR, then the thresholds are roughly 4 and 5 for the lax and the strict strategies, respectively. With a definition of 90% correct, the thresholds increase to 5 and 6, respectively. These results demonstrate good agreement with Rose's 1948 estimates from subjective signal detection experiments.

the range from 2.7 to 6.3. The image noise level was high enough to cause a large variation in signal appearance for a given signal amplitude (see Fig. 4 below for examples of this effect). Observers used a trackball and a cursor to identify signal locations with two different criterion in separate blocks of trials. Under the strict criterion they were asked to avoid false positives (i.e., avoid identifying locations where there was actually no signal). Under the lax criterion they were asked to avoid missing signal locations (with no concern about false positives). The normalized cross-correlation result (XCR) between the signal and the image data was calculated for each location where a signal was present. Note that the ensemble average, ⟨XCR⟩, for a fixed signal amplitude is the SNR. The XCR values at all the signal locations were sorted into bins, and the conditional probability of signal detection for each quantized XCR value was calculated. The resulting cumulative distributions (averaged for two observers) for the two criteria are shown in Fig. 3. The estimated XCR values for 50% correct were 4.0 and 4.9 for the lax and strict criteria, respectively. The corresponding false-positive rates (based on all possible nonoverlapping signal locations) were 3×10^{-3} and 2×10^{-4}, respectively. The step at XCR equal to 4.9 shows the expected cumulative probability of correct response for a Bayesian observer using a decision criterion value of 4.9. The human results show sigmoidal variation near threshold because of several sources of variability. First, we do not actually measure cross correlations when doing this type of task; we make some sort of noisy and imperfect estimate of signal visibility at each candidate location. Second, we are not able to maintain an absolutely constant decision criterion. The estimated standard deviations of the cumulative normal functions fitted to the human results shown in Fig. 3 are approximately 1.0, which is consistent with previous estimates of total observer internal noise obtained by Burgess and Colborne.[45]

3. IMAGE QUALITY EVALUATION

A. Medical Imaging Perspective

Rose's SNR approach is sometimes used in attempts to evaluate or describe image quality or the performance of imaging system components. There are a number of potential pitfalls along this road. The best strategy is to assume that any measure of image quality must be based on the task or purpose for which the image is used. For example, some images are designed for entertainment (television, cinema, and print), and appearance[46] is usually the most important factor. One seems to achieve maximum success by hiding the inevitable sources of degradation from the user, taking advantage of human visual system limitations. This strategy is employed in a number of lossy image compression methods.[47] Other types of image (reconnaissance and medical, for example) are produced for use by experts, and decision task performance is the important consideration.

Medical images are invariably noisy. Those produced by use of x rays or radioisotopes are noisy because ionizing radiation is harmful to patients, and there is a tendency to minimize the amount of radiation as much as possible. This minimization is limited, of course, by the fact that a misdiagnosis can be the most harmful outcome. Magnetic resonance images are noisy because the signals are very weak, and data acquisition times are limited by a number of considerations. Hence there has been a concerted effort in medical imaging to develop techniques of measuring and describing image noise and its effects. Kundel[48] gives an entertaining historical perspective to definitions of medical image quality. In the early days, individual reports of clinical experience with modified technology were seen as adequate. Then came controlled clinical ROC studies of diagnostic accuracy by means of normal and abnormal images produced by competing technologies. Unfortunately, this method (which is still the gold standard) is very time consuming and expensive because of the difficulty of obtaining images with both subtle abnormalities and known ground truth. It was hoped that a third approach, based on measurements of the physical properties of systems (for example, modulation transfer function and noise power spectrum), would be cheaper and faster and would provide a means of predicting performance before systems were actually built. SDT-based SNR analysis of a variety of medical imaging systems is described by Wagner and Brown.[49] However, it became clear that many other issues, such as mathematical modeling of tasks, patient structure, and the perceptual capabilities of human observers, had to be addressed to provide a link between physical parameters and diagnostic accuracy. A recent report[8] by the International Commission on Radiological Units and Measurements deals in depth with medical image quality assessment. That report breaks down the problem into two parts: (1) the quality of the acquired data that can be assessed and described by physical quantities that characterize the imaging system; (2) the quality of the displayed data, which inevitably involves the perceptual system of the observer. Similarly, assessment is broken down into stages: (1) measuring task performance for an ideal, Bayesian observer when using the acquired data, and (2) measuring human observer performance when using the displayed data.

B. Pixel Signal-to-Noise Ratio

One occasionally encounters a quantity described as pixel SNR, which is denoted here as SNR_p to distinguish it from the SDT-based SNR definition. The usual definition is pixel $\mathrm{SNR}_p = a/\sigma_p$, where a is signal amplitude and σ_p is the noise standard deviation per pixel. There have even been situations in which authors have argued that signals would be undetectable if SNR_p was lower than some threshold value, such as 1, 3, or 5. One possible source of this misunderstanding is that Rose, in his 1973 book,[50] used signals consisting of a single pixel for pedagogical purposes in the introductory theoretical development. Single-pixel signals of interest are rare in images, so unfortunately the pixel SNR concept is overly simplistic and is not very useful. This is illustrated by Fig. 4. Portion A (upper left) shows an array of flat-topped disk signals with diameters of 4, 5, 6, and 16 pixels and amplitudes of 24 gray levels on a constant background of 127 gray levels. Portion B (upper right) was produced by addition of zero-mean white noise with σ_p of 24 gray levels, so $\mathrm{SNR}_p = 1.0$ for all the signals. Given N pixels in a signal, SDT-based SNR $= \sqrt{N}\ \mathrm{SNR}_p$ with values of 3.5, 4.6, 5.7, and 14.4 for the four disk sizes, respectively. Note that the appearance of the smaller disks is quite variable and that there are some spurious noise correlations that could represent small disk signals, except that they are not at the known signal locations. These effects are characteristic of noisy images. The largest disks are easily detected. Now consider portion C

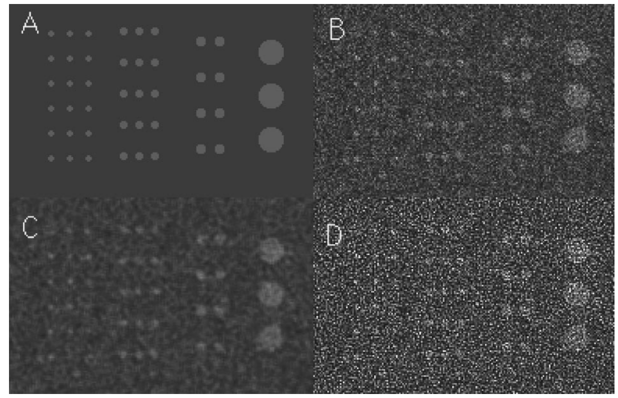

Fig. 4. Demonstration that pixel SNR (SNR_p) is not a good primary measure of image quality. SNR_p is defined as the ratio of peak signal amplitude and noise standard deviation, σ_p, per pixel. This figure also shows that signal appearance is highly variable at low SNR, as is expected, given the statistical nature of the images. A, Signal array of disks (diameters of 4, 5, 6, 16 pixels) with amplitude of 24 gray levels. B, Same signal array with added zero-mean white noise (σ_p of 24 gray levels), so $\mathrm{SNR}_p = 1.0$ for all the signals. The Rose model and the ideal observer SNR's are 3.5, 4.6, 5.7, and 14.4. C, After smoothing of portion B. The low-pass-filtered noise has σ_p of 8.2 gray levels, so $\mathrm{SNR}_p = 2.9$ for all the signals. D, After edge enhancement of portion B, resulting in a σ_p value of 39.6 gray levels and a SNR_p value of 0.6. The filtering would have no effect on ideal observer SNR. The Rose model is not a valid method for calculating SNR with filtered images.

(lower left), obtained by smoothing of portion B. The smoothed image has a reduced σ_p of 8.2 gray levels and increased SNR_p (2.9). The smoothing would have no effect on detectability for the ideal observer and little effect for human observers. Finally, consider portion D (lower right), produced with an edge-enhancement technique. This image has an increased σ_p of 39.6 gray levels, giving a decreased SNR_p (0.6). Clipping of the edge-enhanced image because of the 8-bit display limitation affects approximately 1 pixel in 1000. Again, the ideal observer detectability is unchanged, and the effect on human detectability would be modest. So even though SNR_p changes significantly because of filtering, signal detectability does not. When an image is filtered, the spectral densities of both signal and noise are changed, but the ratio remains constant. Detectability would be unaffected for an observer that can do prewhitening and matched filtering. Human observers can partially compensate for correlations in the noise (which can be described as a partial prewhitening). Hence gentle image filtering has little effect on signal detectability for us. It is possible to perform major filtering that dramatically affects human performance, but this would be unacceptable to expert users of the images. One concludes that SNR_p is not a useful measure for description of detection task difficulty and is probably a misleading metric for image quality evaluation.

A related statement that one occasionally encounters is, "Smoothing makes images less noisy, and edge-enhancement filtering increases noise." This describes appearance. Our subjective estimate of image noisiness uses the entire bandwidth of the visual system. Apparent noise variance is due to noise spectral density integrated over this bandwidth. So the apparent noisiness of most images depends on our viewing distance, among other things. The statement is often of no consequence for signal detection tasks. In most cases the apparent change in noisiness is due to changes in noise power outside the signal bandwidth and has little affect on human performance.

4. HUMAN OBSERVER MODELS

The Rose model was a first attempt at characterizing the ability of human observers to detect signals in noisy images. However, a variety of issues were implicitly included in the empirically determined threshold SNR parameter, k. These included the specification of task accuracy, observer decision criterion, and sources of human inefficiency. The development of SDT and objective experimental techniques for evaluating human visual task performance allowed us to explicitly deal with each of these issues,[30] and the development of digital display technology allowed us to easily compare human performance with the ideal. Human visual efficiency for tasks that use static, noise-limited images turned out to be surprisingly high. In retrospect, high efficiency ought not to have been surprising, given Rose's 1948 estimate of k in the neighborhood of 5 for static, high-contrast, noise-limited images. The subjective detection experimental approach used by Rose cannot be analyzed by SDT, but for comparison consider a binary (yes/no) SKE detection experiment with 50% probability that the signal is present. If the ideal observer maintains a false-alarm rate of 1%, it needs SNR's of 2.3 and 3.6 for 50% and 90% true-positive rates, respectively. Comparison of the Rose result and the binary (yes/no) results suggests that human decision efficiency is between 20% and 50%. This is in reasonable agreement with the value of 50% found for detection of simple aperiodic signals. Of course, one cannot make too much of this, given the ill-defined nature of the subjective task used by Rose. However, if human visual signal detection were as poor as auditory signal detection (1% or less efficiency), then Rose would have had to use a much larger value of k and surely would have taken the discrepancy very seriously.

The Rose model was developed for both static and dynamic noise-limited images. Next, subsequent research on human performance is briefly reviewed, first for static images and then for dynamic image sequences. The investigations described below follow the spirit of Rose's approach in that they compare human and ideal observer task performance and introduce as few mechanisms as possible to account for suboptimal human performance. This presentation, because of space limitations, neglects one important aspect of Rose's work—the topic of photon collection efficiency in the eye. This question has been extensively investigated, and maximum collection efficiency is approximately 10%. This topic has been reviewed by Pelli,[51] and important publications prior to 1980 are in a collection edited by Cohn.[52]

A. Static Images

The first measurements of central (decision-level) visual efficiency were performed by Barlow[53] with computer-generated random dot images. He was attempting to measure spatial tuning curves for dot pattern detection that would be analogous to the contrast sensitivity curves for luminance (gray-scale) patterns. The basic idea was that the small dots were reliably transduced by the retina and that any spatial tuning that was found might be due to receptive fields in the visual cortex. Barlow used Poisson statistics to generate patterns and chose observer efficiency as the measure of pattern detectability. He consistently found efficiencies of approximately 50%, independent of pattern size or shape. Subsequently, Burgess and Barlow[54] used random dot patterns with Gaussian statistics to allow independent adjustment of the mean and the variance. This, in turn, separated effects due to sampling efficiency (which is an estimate of the accuracy with which the appropriate signal detection template is used) from effects due to intrinsic observer variability (internal noise). These measurements suggested that virtually all of the reduction in human dot detection efficiency was due to centrally located internal noise. Sampling efficiency appeared to be approximately 100% for a wide range of random dot patterns. The word sampling was selected[55] to be consistent with the original formulation of decision accuracy by Fisher.[26]

The next investigations of decision efficiency were done with computer-generated gray-scale images and uncorrelated Gaussian noise. Burgess *et al.*[55] demonstrated that humans could perform amplitude discrimination tasks with efficiencies well over 50% with sampling efficiencies as high as 80%. Efficiencies of around 50% were found for detection of aperiodic signals such as disks and compact 2D Gaussian profiles. Efficiency for detecting sinusoidal signals was in the 10–25% range. The decrease in efficiency is probably due to an inability to make precise use of absolute sinusoid phase information[56]: A phase uncertainty of as little as a quarter of a cycle dramatically reduces sine-wave detection accuracy. At the peak of the contrast sensitivity curve (roughly 4 cycles/degree), a quarter of a cycle corresponds to approximately the width of the optical point-spread function of the eye. Pelli argued that a number of other parametric features of sinusoid detection can be explained by uncertainty.[57] A variety of experiments were done to investigate human ability to use prior information about the signal in performing detection tasks, such as use of sine-wave phase knowledge,[56] identification of signal location among a number of possible locations,[43] and signal identification based on spatial profile.[44] These experiments were needed to establish that a cross-correlation-based model is valid for humans. Observer consistency was measured by several techniques,[45] and the results demonstrated that previous estimates of sampling efficiency were systematically low because of an additional, induced form of observer internal noise that is proportional to spectral density of image noise. The efficiency approach has been used to investigate many aspects of noise-limited human vision, including spatial sine-wave detection,[58–60] detection of visual noise,[61] American Sign Language use,[62] recognition of 3D objects in noise,[63] and estimation of fractal dimension.[64] All these experimental results are consistent with the view that humans can be modeled as suboptimal Bayesian observers that use prior information about possible signal shapes and locations in addition to the data displayed in individual image trials.

Recently, the emphasis has been on trying to understand how the shape of the noise power spectrum affects human signal detection performance. The main question is, How effective are humans at compensating for noise correlations? That is, can we prewhiten the noise? Myers *et al.*[65] demonstrated that humans do poorly compared with the ideal observer when the noise has a bandpass spectrum (which produces short-range negative correlations), and they suggested[66] that the reduction in performance might be due to spatial-frequency channels in the visual system. The channels would irreversibly combine contributions from a range of spatial frequencies and would preclude any subsequent attempts to prewhiten the noise. Investigation with low-pass noise did not show such dramatic decreases in human efficiency. It now appears that the effects of low-pass and high-pass noise are very different. For low-pass noise power spectra, prewhitening models are better predictors of human performance than are nonprewhitening models.[67–69] The reason for reduced performance with high-pass noise is yet to be determined.

The ideal-observer-based approach becomes more difficult to use as the complexity of the task is increased. Eventually, ideal observer procedures become nonlinear, and analysis of ideal observer performance becomes mathematically intractable. Hanson[70] suggested that detectability for some tasks can be estimated approximately by inclusion of a power-law frequency weighting to represent task difficulty. Barrett *et al.*[71] developed a convenient quasi-ideal observer approach that is useful for complex tasks. They refer to it as the Hotelling observer, for historical reasons, and it achieves the best possible performance, given the limitation of being forced to use linear procedures. It is identical to the Bayesian ideal observer for tasks in which the ideal observer uses linear procedures. The Hotelling observer approach is related to Fisher linear discriminant analysis used in pattern recognition.[33] The Hotelling approach uses population statistics, while the Fisher approach uses sample statistics.[72] Barrett *et al.*[11] recently proposed a very promising likelihood-generating-function approach for analysis of complex task performance.

B. Image Sequences

Rose[3] explicitly considered temporal effects in his model. This subject was neglected previously in this paper to simplify presentation and is now addressed. In Rose's model, available photon count densities would be proportional to the integration (Rose used the term storage) time of the eye or the image acquisition device. He observed that the human visual impression of the dynamic images in his display was matched by still photographic pictures of the display taken with a 0.25-s-exposure time. Based on this observation and on previous research by Cobb and Moss,[73] Rose adopted an integration time of 0.2 s.

Several studies have investigated signal detection performance in spatiotemporal white noise. For signals with no temporal modulation, detectability for the ideal observer increases indefinitely as the square root of the number of independent frames. Performance improvement will saturate for an observer with a fixed integration time. Some improvement with time is expected for human observers, because of filterlike effects in early stages of the visual system. For noiseless dynamic images the estimated integration time for early stages does not exceed 100–150 ms.[74] The situation for noise-limited images is quite different. Several recent experiments[75–78] have shown that the detection of signals in temporal white noise improves up to approximately 700–1000 ms. Interestingly, these integration times also hold when the signal is moving while the observer's gaze is directed at a fixed point. However, abrupt signal motion and high-velocity motion do disrupt human performance.

These experimental results have been used to extend human observer signal detection models to include the temporal domain. Some include only the early filter stage,[79] while others[76,77,80] include suboptimal integration to 700–1000 ms. There has been limited investigation of whether humans can vary weighting for temporal cross correlation with signals modulated in time. Experimental results by Eckstein *et al.*[81] suggest that humans can adjust to temporally varying signals, although suboptimally.

5. CONCLUDING REMARKS

Rose's 1946 and 1948 papers have been influential in a number of ways. The detective quantum-efficiency metric that he introduced in 1946 has had a long history of application in both human vision and evaluation of imaging system components. The Rose model introduced in 1948 has been used to give several generations of medical imaging scientists an introduction to signal detection in noise. Reading his papers with the benefit of 50 years of slowly developed understanding leaves one with a sense of awe at Rose's insights and innovations. Those of us who work in imaging owe a great deal to these publications, which were the product of a brilliant mind and an unerring intuition. It is hoped that this tutorial paper will give an understanding of how the Rose model can effectively be used if attention is paid to its domain of validity.

ACKNOWLEDGMENTS

I particularly thank Robert Wagner of the Center for Devices and Radiological Health, U.S. Food and Drug Administration, for his assistance and advice at all stages of the development of this manuscript. The content of the paper draws heavily on his many contributions to the medical imaging literature and on his patient explanations, to me personally over the years, of theoretical subtleties. I also thank Craig Abbey, Francois Bouchud, Miguel Eckstein, and Jim Whiting of Cedars-Sinai Medical Center, Philip Bunch of Kodak, Kyle Myers of the U.S. Food and Drug Administration, Robert Webb of Harvard Medical School, and two anonymous reviewers for their constructive criticism of the manuscript. This work was supported by National Institutes of Health grant R01

CA58302, but it does not, to the best of my knowledge, represent an official position of any part of the U.S. Government.

REFERENCES

1. A. Rose, "A unified approach to the performance of photographic film, television pickup tubes and the human eye," J. Soc. Motion Pict. Eng. **47**, 273–294 (1946).
2. A. Rose, "The sensitivity performance of the human eye on an absolute scale," J. Opt. Soc. Am. **38**, 196–208 (1948).
3. A. Rose, "Television pickup tubes and the problem of vision," in *Advances in Electronics and Electron Physics*, L. Marton, ed. (Academic, New York, 1948), Vol. 1, pp. 131–166.
4. A. Rose, "Quantum and noise limitations of the visual process," J. Opt. Soc. Am. **43**, 715–716 (1953).
5. I. Cunningham and R. Shaw, "Signal-to-noise optimization of medical imaging systems," J. Opt. Soc. Am. A **16**, 621–632 (1999).
6. W. W. Peterson, T. G. Birdsall, and W. C. Fox, "The theory of signal detectability," IRE Trans. Inf. Theory **PGIT-4**, 171–212 (1954).
7. D. Van Metter and D. Middleton, "Modern statistical approaches to reception in communication theory," IRE Trans. Inf. Theory **PGIT-4**, 119–141 (1954).
8. P. F. Sharp, C. E. Metz, R. F. Wagner, K. J. Myers, and A. E. Burgess, ICRU Rep. 54, "Medical imaging: the assessment of image quality" (International Commission on Radiological Units and Measurements, Bethesda, Md., 1996).
9. H. H. Barrett, "Objective assessment of image quality: effects of quantum noise and object variability," J. Opt. Soc. Am. A **7**, 1266–1278 (1990).
10. H. H. Barrett, J. L. Denny, R. F. Wagner, and K. J. Myers, "Objective assessment of image quality. II. Fisher information, Fourier cross talk, and figures of merit for task performance," J. Opt. Soc. Am. A **12**, 834–852 (1995).
11. H. H. Barrett, C. K. Abbey, and E. Clarkson, "Objective assessment of image quality. III. ROC metrics, ideal observers, and likelihood-generating functions," J. Opt. Soc. Am. A **15**, 1520–1535 (1998).
12. R. B. Barnes and M. Czerny, "Lasst sich ein Schroteffekt der Photonen mit dem Auge beobachten?" Z. Phys. **79**, 436–449 (1932).
13. S. Hecht, "Quantum relations of vision," J. Opt. Soc. Am. **32**, 42–49 (1942).
14. H. de Vries, "The quantum character of light and its bearing upon threshold of vision, the differential sensitivity and visual acuity of the eye," Physica **10**, 553–564 (1943).
15. O. H. Schade, *Image Quality: a Comparison of Photographic and Television Systems* (RCA Laboratories, Princeton, N.J., 1975).
16. R. F. Wagner, "Decision theory and the detail signal-to-noise ratio of Otto Schade," Photograph. Sci. Eng. **22**, 41–46 (1977).
17. R. F. Wagner, "Toward a unified view of radiological imaging systems. Part II: Noisy images," Med. Phys. **4**, 279–296 (1977).
18. J. W. Coltman, "Fluoroscopic image brightening by electronic means," Radiology **51**, 359–366 (1948).
19. R. E. Sturm and R. H. Morgan, "Screen intensification systems and their limitations," Am. J. Roentgenol. **62**, 617–634 (1949).
20. D. O. North, "Analysis of the factors which determine signal–noise discrimination in pulsed carrier systems," RCA Tech. Rep. PTR6C (1943), reprinted in Proc. IRE **51**, 1016–1028 (1963).
21. N. Wiener, *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series* (Wiley, New York, 1960).
22. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. (McGraw-Hill, New York, 1991).
23. K. R. Castleman, *Digital Image Processing* (Prentice-Hall, Englewood Cliffs, N.J., 1996).
24. J. L. Harris, "Resolving power and decision theory," J. Opt. Soc. Am. **54**, 606–611 (1964).
25. J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures* (Wiley, New York, 1986).
26. R. A. Fisher, *Statistical Methods for Research Workers* (Oliver and Boyd, Edinburgh, 1925).
27. J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," Philos. Trans. R. Soc. London Ser. A **231**, 289 (1933).
28. A. Wald, *Sequential Decision Functions* (Wiley, New York, 1950).
29. J. A. Swets, *Signal Detection and Recognition by Human Observers* (Wiley, New York, 1964).
30. D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).
31. H. L. Van Trees, *Detection, Estimation and Modulation Theory* (Wiley, New York, 1968), Vols. 1–3.
32. C. W. Helstrom, *Elements of Signal Detection and Estimation* (Prentice-Hall PTR, Englewood Cliffs, N.J., 1995).
33. K. Fukunaga, *Introduction to Statistical Pattern Recognition* (Academic, New York, 1972).
34. C. E. Metz, "ROC methodology in radiological imaging," Invest. Radiol. **21**, 720–732 (1986).
35. C. E. Metz, "Some practical issues of experimental design and data analysis in radiological ROC studies," Invest. Radiol. **24**, 234–245 (1989).
36. D. G. Brown, M. F. Insana, and M. Tapiovaars, "Detection performance of the ideal decision function and its MacLaurin expansion," J. Acoust. Soc. Am. **97**, 379–398 (1995).
37. A. E. Burgess, "Comparison of receiver operating characteristic and forced choice observer performance measurement methods," Med. Phys. **22**, 643–655 (1995).
38. J. A. Swets, BBN Technologies, Cambridge, Mass. 02138 and Brigham and Women's Hospital, Boston, Mass. 02115 (personal communication, 1998).
39. W. P. Tanner and T. G. Birdsall, "Definitions of d' and η as psychophysical measures," J. Acoust. Soc. Am. **30**, 922–928 (1958).
40. A. Rose, "The relative sensitivities of television pickup tubes, photographic film and the human eye," Proc. IRE **30**, 293–300 (1942).
41. O. Schade, "Electro-optical characteristics of television systems," RCA Rev. **9**, 5–37 (1948).
42. N. A. Macmillan and C. D. Creelman, *Detection Theory: A User's Guide* (Cambridge U. Press, Cambridge, UK, 1991).
43. A. E. Burgess and H. Ghandeharian, "Visual signal detection. II. Signal location identification," J. Opt. Soc. Am. A **1**, 906–910 (1984).
44. A. E. Burgess, "Visual signal detection. III. On Bayesian use of prior knowledge and cross correlation," J. Opt. Soc. Am. A **2**, 1498–1507 (1985).
45. A. E. Burgess and B. Colborne, "Visual signal detection. IV. Observer inconsistency," J. Opt. Soc. Am. A **5**, 617–627 (1988).
46. S. Daly, "The visual differences predictor: an algorithm for the assessment of image fidelity," in *Digital Images and Human Vision*, A. B. Watson, ed. (MIT, Cambridge, Mass., 1993).
47. A. B. Watson, ed., *Digital Images and Human Vision* (MIT, Cambridge, Mass., 1993).
48. H. L. Kundel, "Images, image quality, and observer performance," Radiology **132**, 265–271 (1979).
49. R. F. Wagner and D. G. Brown, "Unified SNR analysis of medical imaging systems," Phys. Med. Biol. **30**, 489–518 (1985).
50. A. Rose, *Vision—Human and Electronic* (Plenum, New York, 1973).
51. D. G. Pelli, "The quantum efficiency of vision," in *Vision: Coding and Efficiency*, C. Blakemore, ed. (Cambridge U. Press, Cambridge, UK, 1990), pp. 3–24.
52. T. E. Cohn, ed., *Visual Detection*, Vol. 3 of Collected Works in Optics (Optical Society of America, Washington, D.C., 1993).
53. H. B. Barlow, "The efficiency of detecting changes in den-

sity in random dot patterns," Vision Res. **18**, 637–650 (1977).
54. A. E. Burgess and H. B. Barlow, "The efficiency of numerosity discrimination in random dot images," Vision Res. **23**, 811–819 (1983).
55. A. E. Burgess, R. J. Jennings, R. F. Wagner, and H. B. Barlow, "Efficiency of human visual discrimination," Science **214**, 93–94 (1981).
56. A. E. Burgess and H. Ghandeharian, "Visual signal detection. I. Ability to use phase information," J. Opt. Soc. Am. A **1**, 900–905 (1984).
57. D. G. Pelli, "Uncertainty explains many aspects of visual contrast detection and discrimination," J. Opt. Soc. Am. A **2**, 1508–1530 (1985).
58. D. A. Kersten, "Spatial summation in visual noise," Vision Res. **24**, 1977–1990 (1983).
59. R. Nasanen, H. Kukkonen, and J. Ravamo, "Spatial integration of band-pass filtered patterns in noise," Vision Res. **33**, 903–911 (1993).
60. J. Rovamo, J. Mustonen, and R. Nasanen, "Modeling contrast sensitivity as a function of retinal illuminance and grating area," Vision Res. **34**, 1301–1314 (1994).
61. D. Kersten, "Statistical efficiency for the detection of visual noise," Vision Res. **27**, 1029–1040 (1986).
62. M. Pavel, G. Sperling, T. Reidl, and A. Vanderbeek, "Limits of visual communication: the effect of signal-to-noise ratio on the intelligibility of American Sign Language," J. Opt. Soc. Am. A **4**, 2355–2365 (1987).
63. B. S. Tjan, G. E. Legge, W. L. Braje, and D. Kersten, "Human efficiency for recognizing 3-D objects in luminance noise," Vision Res. **35**, 3053–3069 (1995).
64. T. Kumar, P. Zhou, and D. A. Glaser, "Comparison of human performance with algorithms for estimating fractal dimension of fractional Brownian statistics," J. Opt. Soc. Am. A **10**, 1136–1146 (1993).
65. K. J. Myers, M. C. Borgstrom, H. H. Barrett, D. D. Patton, and G. W. Seeley, "Effect of noise correlation on detectability of disk signals in medical imaging," J. Opt. Soc. Am. A **2**, 1752–1759 (1985).
66. K. J. Myers and H. H. Barrett, "Addition of a channel mechanism to the ideal-observer model," J. Opt. Soc. Am. A **4**, 2447–2457 (1987).
67. J. P. Rolland and H. H. Barrett, "Effect of random background inhomogeneity on observer detection performance," J. Opt. Soc. Am. A **9**, 649–658 (1992).
68. A. E. Burgess, X. Li, and C. K. Abbey, "Nodule detection in two component noise: toward patient structure," in *Medical Imaging 1997: Image Perception*, H. L. Kundel, ed., Proc. SPIE **3036**, 2–13 (1997).
69. A. E. Burgess, "Visual signal detection with two-component noise: low-pass spectrum effects," J. Opt. Soc. Am. A **16**, 694–704 (1999).
70. K. H. Hanson, "Variations in task and the ideal observer," in *Application of Optical Instrumentation in Medicine XI: Medical Image Production, Processing, Display and Archiving*, R. H. Schneider and S. J. Dwyer, eds., Proc. SPIE **419**, 60–67 (1983).
71. H. H. Barrett, J. P. Rolland, J. Yao, and K. J. Myers, "Model observers for assessment of image quality," Proc. Natl. Acad. Sci. USA **90**, 9758–9765 (1993).
72. T. W. Anderson, *Introduction to Multivariate Statistical Analysis*, 2nd ed. (Wiley, New York, 1984).
73. J. P. Cobb and R. E. Moss, "The four variables of visual threshold," J. Franklin Inst. **205**, 831 (1928).
74. D. H. Kelly, "Retinal inhomogeneity. I. Spatiotemporal contrast sensitivity," J. Opt. Soc. Am. A **1**, 107–113 (1984).
75. J. S. Whiting, E. Carterette, D. Honig, and N. Eigler, "Observer performance in dynamic displays: effect of frame rate on visual signal detection in noisy images," in *Medical Imaging 1991: Image Perception*, H. L. Kundel, Proc. SPIE **1453**, 165–176 (1991).
76. M. P. Eckstein, J. S. Whiting, and J. P. Thomas, "Detection and discrimination of moving signals in Gaussian uncorrelated noise," in *Medical Imaging 1996: Image Perception*, H. L. Kundel, ed., Proc. SPIE **2712**, 9–25 (1996).
77. P. Xue and D. L. Wilson, "Pulsed fluoroscopy detectability from interspersed adaptive forced-choice measurements," Med. Phys. **23**, 1833–1843 (1996).
78. M. J. Tapiovaara, "Efficiency of low-contrast detail detectability in fluoroscopic imaging," Med. Phys. **24**, 655–664 (1997).
79. R. Aufrichtig, C. W. Thomas, P. Xue, and D. L. Wilson, "Model for perception of pulsed fluoroscopy image sequences," J. Opt. Soc. Am. A **13**, 3167–3176 (1994).
80. D. L. Wilson, P. Xue, K. N. Jabri, and R. Aufrichtig, "Perceived noise versus display noise in temporally filtered image sequences," J. Electron. Imaging **5**, 490–495 (1996).
81. M. P. Eckstein, J. S. Whiting, and J. P. Thomas, "Role of knowledge in human visual temporal integration in spatiotemporal noise," J. Opt. Soc. Am. A **13**, 1960–1968 (1996).