Human and Quasi-Bayesian Observers of Images Limited by Quantum Noise, Object Variability, and Artifacts

K.J. Myers[1], R.F. Wagner[1], K.M. Hanson[2], H.H. Barrett[3], J.P. Rolland[4]

[1]CDRH/FDA, 12720 Twinbrook Parkway, Rockville, MD 20857
[2]Los Alamos National Laboratory, Los Alamos, NM 87545
[3]Arizona Health Sciences Center, Univ. of AZ, Tucson, AZ 85724
[4]Dept. of Computer Sci., Univ. of NC, Chapel Hill, NC 27599

ABSTRACT

Many investigators have pointed out the need for performance measures that describe how well the images produced by a medical imaging system aid the end user in performing a particular diagnostic task. To this end we have investigated a variety of imaging tasks to determine the applicability of Bayesian and related strategies for predicting human performance. We have compared Bayesian and human classification performance for tasks involving a number of sources of decision-variable spread, including quantum fluctuations contained in the data set, inherent biological variability within each patient class, and deterministic artifacts due to limited data sets.

1. INTRODUCTION

Although the radiologist may be the expected end user of a clinical image, it is often desirable to have a model observer in order to expedite the system design and optimization phases. One particular model, the ideal Bayesian observer of statistical decision theory,[1,2] is of particular interest because it predicts the best possible performance on a given visual task. It can therefore serve as a benchmark for observer performance - indicating when information needed to perform a given task is readily extracted by the human observer, as well as when either the hardware should be redesigned to better match the human's capabilities, or the human observer should be augmented or even replaced by a machine reader. We have therefore investigated a variety of imaging tasks to determine the applicability of Bayesian and related strategies for predicting human performance.

In all the investigations that we will describe, the task is the classification of an image into one of two possible categories: signal-present (abnormal) or signal-absent (normal). The human observer is modelled as a decision maker that calculates a scalar decision variable and classifies the image based on whether that decision variable is above or below some threshold for calling the images positive or negative. This approach presupposes the existence of two classes of objects, one for the signal-absent and one for the signal-present images, where both generate a spread of decision variable outcomes as shown in Figure 1. Image quality is determined by the degree of separation of these two populations.

0-8194-1461-1/94/$6.00

Several factors contribute to the spread of the decision variables shown in Figure 1 and thus influence diagnostic performance. Even for a single object in each class, the decision variable would be randomly distributed for a set of images because of the random quantum or thermal noise contained in the data set. Inherent biological variability within each patient class serves to spread the distributions further. For images reconstructed from limited data, the decision variables are further spread by deterministic artifacts. And finally, limits to the contrast visible to the human because of internal noise as well as limits to the spatial area of integration of the human observer cause the distributions to be spread still more.[3-6] Each of these factors can be influenced through choices available to the imaging system designer. In the sections below, we shall describe investigations into a number of these sources of decision-variable spread, and offer observations regarding the relationship between human and machine observer performance and the implications for system design.

One source of decision-variability spread that is missing from Figure 1 is the contribution from inter- and intra-observer variability. While a system designer can make choices affecting the contributions of each of the formerly mentioned factors on observer performance, observer variability (from such sources as template jitter or threshold variations) is not generally in the control of medical imaging system designers. Attempts are made to minimize observer variability in clinical practice through adequate observer training. Nevertheless, it is present there and in each of the observer studies we shall describe below. We shall not address inter- and intra-observer variability further here. A number of other investigators have attempted to measure the contribution of observer variability to human performance.[4-6]

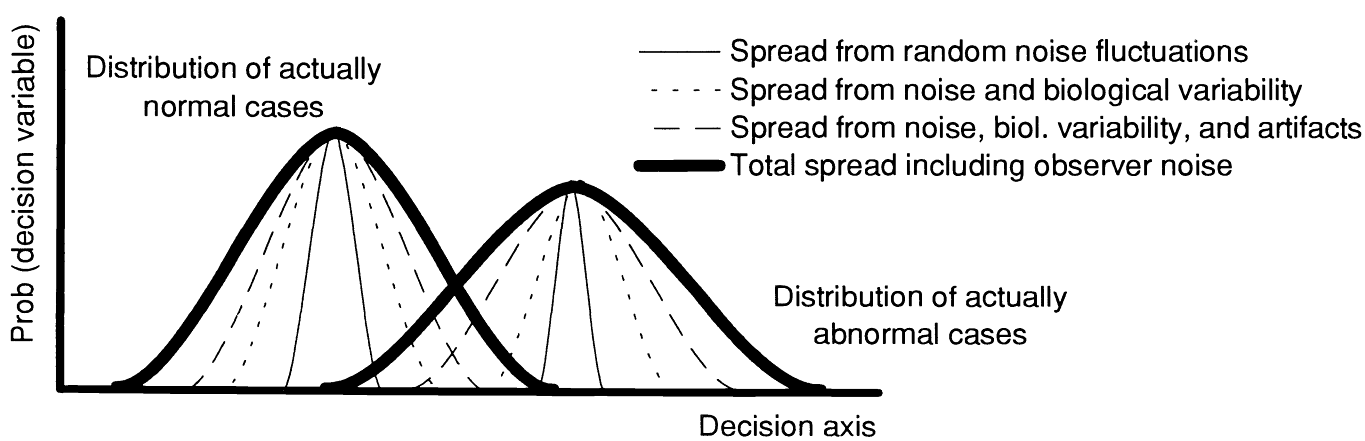

Figure 1. Sources of spread in observers' decision variables.

2. TASKS LIMITED BY PHOTON NOISE

In this section we review some investigations of human performance for tasks represented by the most narrow distributions shown in Figure 1. For tasks in which the signal and background are known and the observer is limited primarily by uncorrelated quantum or thermal fluctuations in the data, the optimal strategy is to perform a linear template-matching operation on the image to calculate the decision variable. For such SKE/BKE (signal-known-exactly/background-known-exactly) tasks in white noise, the ideal observer of Bayesian decision theory has been found to be a good predictor of human performance. Human efficiency relative to the ideal observer has been found to be 30-50% for a range of detection and discrimination tasks.[7-9]

When the image noise is correlated, the ideal strategy is to first perform a prewhitening operation on the data before cross-correlating it with the expected difference signal. A number of investigators have measured human performance in images degraded by colored noise. When the noise is high-pass, such as when it is characterized by a ramp-shaped noise power spectrum as in the case of a computed tomography (CT) imaging system, human efficiency relative to the ideal observer falls to about 20%.[10] High-pass noise characterized by noise power spectra of the form f^n, where n =1,2,3,4, has been shown to cause decreasing human detection performance relative to the ideal observer as n increases.[11] This inability to perform the prewhitening operation has been shown to be consistent with the requirement that human observers process visual information through frequency-selective channels.[12]

The studies we have referred to so far have all considered artifact-free imaging systems with no object variability. The signals and backgrounds were exactly specified. For compact signals (signals with limited spatial extent), human performance can be predicted by an observer that positions a template over the location of the expected signal and performs a linear filtering operation. This model has been shown to be reasonable until the signal extent is so large that some limitation in the integration area of the human is reached.[3] However, signals that are large in spatial extent without needing a large integration area for detection can still be detected quite effectively by the human observer. An example is the detection of a known grid of bright lattice points in a noisy background, which can be effectively detected by the human, possibly by a series of local template matching operations.

3. TASKS LIMITED BY OBJECT VARIABILITY

We now consider tasks represented by the distributions in Figure 1 labelled "quantum noise and object variability." When the task is the detection of 1 of M orthogonal signals of limited spatial extent, human performance has been shown to be again roughly 50% efficient compared to the ideal observer performing the same task.[14-17] That is, the performance penalty experienced by the human observer is no greater than the performance penalty paid by the ideal observer for the same object variability. Nolte and Jaarsma[18] showed that the very complicated (and highly nonlinear) form of the ideal observer's test statistic in this case could be almost matched in performance by an observer that simply performs a sequence of linear filtering operations, one for each of the M signals, and chooses in favor of whichever signal alternative gives the maximum filter output. It is therefore possible that the human performs this non-linear task through the simple mechanics of applying a sequence of templates to the image, one for each signal alternative.

Relatively few studies of more complex forms of object variability have been performed. Revesz et al.[19] investigated lesion detectability on chest radiographs and found good agreement between human performance and a measure of background conspicuity. Swensson and Judy[20] investigated the effect of variable-brightness background levels and areas on the detectibiltiy of disk signals for both white and CT noise. They found little dependence on the type of image noise or the mean brightness of the background.

A more recent study of signal detection with object variability in a Bayesian framework has been performed by Rolland et al.[21] When the object variability is such that the signal is a known Gaussian object on a statistically defined background, Rolland has shown that the detection performance of human observers is modelled quite well by an approximation to the ideal observer called the Hotelling observer. The Hotelling observer does the best job possible for an observer that is constrained to performing only linear operations on the image. For this particular task, which we shall refer to as the "lumpy background" task, the Hotelling template was found to be similar to a difference of Gaussians or the center-surround template that a number of investigators have found to be of use in machine vision. The good agreement between human and Hotelling performance suggests that the human is doing a linear operation on the image with a fairly compact template for this task as well. Yao extended the work of Rolland to show that the Hotelling model could predict human detection performance for images that had non-white noise in addition to background uncertainty, provided the Hotelling observer was constrained to process data through frequency-selective channels.[22,23]

Burgess has also investigated human performance for objects with statistically defined backgrounds of the same nature as those of Rolland and Yao.[24] He has proposed that human performance can be modelled by operating on the image first with a filter that models the frequency response of the eye, followed by a non-prewhitening matched filter for the signal. Further experiments are required to determine which of these observer models is most predictive of human performance. The choice of task is crucial -- when attempting to show the capability of one model for human performance over another, the task should be chosen such that the models give sufficiently different predictions to allow one or the other to be ruled out on the basis of human performance data. We expect the discussion of Hotelling vs. non-prewhitening-matched-filter with eye response to be resolved sometime in the future with experiments chosen carefully to exploit differences in the models. A crucial difference between the two models here is the fact that the Hotelling template is adaptive as exposure level changes, while the Burgess model is not adaptive. As exposure time increases, the quantum fluctuations in the image decrease, increasing the relative weight of the object-variability or conspicuity limitation to the task performance. The Hotelling observer's performance continues to increase with exposure time, indicating the adaptive nature of the Hotelling template. The Burgess candidate model would eventually be expected to saturate. Rolland has shown that human performance can be predicted quite well by the Hotelling model for this case. It will be interesting to see how well the Burgess model will do in an experiment where exposure time is the variable. Experiments that investigate the adaptability of the human observer as a function of other characteristics of the random background, such as correlation length, could also be envisioned.

For non-local signals with uncertainty, human performance drops significantly. We have seen this to be true in a number of investigations: 1) the detection of a "jittery" grid - analogous to the SKE/BKE experiment mentioned in the previous section in which a grid of bright lattice points is to be detected, but where now the locations of the bright points are jittered, or individually randomized, about their means;[13] 2) the detection of random dot patterns as the number of dots grows large;[25-27] 3) the detection of mirror symmetry in random patterns,[28] and 4) the detection of diffuse liver disease.[29] We believe these results are driven by the complexity of the calculation required to derive the ideal decision variable - the human doesn't have the ability to use all the numbers the ideal observer can, perhaps because of memory or time limitations. For tasks in which the human is shown to be markedly inefficient relative to the ideal observer, machine observers should be considered as adjuncts to the human to aid in the extraction of the information needed to perform the task well.

4. TASKS LIMITED BY ARTIFACTS

In the preceding sections we considered tasks performed on images generated by imaging systems that acquired complete (or very close to complete) data sets. Qualitatively, such imaging systems have a fairly localized point spread function (PSF); the image of a point source is point-like although it may be blurred. In addition to containing blur, imaging systems may have PSFs that depart from an ideal delta-function shape either by having long low-amplitude tails (the case for an imaging system with a large scatter component) or by having a non-compact shape that extends anisotropically over a long range (such as a tomographic imaging system with a limited number of views). In the latter case, we say the system has artifacts. As an example, the image in Figure 2 is the reconstruction of a disk object for a noise-free data set from a simulated 2D CT acquiring only 8 views over 180 degrees. The streak artifacts are deterministic, but object dependent.

Few studies have been done to compare human performance to Bayesian performance for even SKE/BKE tasks in images formed by systems with long-tailed PSFs. Rolland et al.[30] have found that human performance is inefficient for detecting disks imaged by an imaging system with a long-tailed PSF. However, human performance can be improved dramatically through post-processing. The long tails of the PSF cause the signal to have a large-area, low-contrast contribution to the image that the human is unable to efficiently detect. When an imaging system has this kind of PSF, post-processing may be necessary to aid the human in extracting all the informaion in the image needed to perform the task.

Just as for long-tailed PSFs, few investigations have been performed comparing human to ideal observer performance on images containing artifacts. Given an image like Figure 2 of a disk object at a known location on a known background, an ideal observer that knew the disk location and background would correct for artifacts as part of a procedure for detecting a disk in that particular location. The performance of the Bayesian observer is calculated using methods applicable to the SKE/BKE tasks described in Section 2. We would expect human performance to degrade for such tasks using images with artifacts because they again cause signal information to become spread out and have low contrast, similar to the effect found by Rolland in her studies with long-tailed PSFs. Further investigation is required to determine if this expectation is true.

Limited data sets become even more problematic when the task is also complicated by object variability. When a disk to be detected has a random location, the artifacts in the image will vary depending on the location of the disk. Even worse, object variability in the form of a random background results in an image that has non-local contributions from all of the non-signal components in the object. One might guess from the non-local signal-uncertain tasks described in Section 3 that the human would be very adversely affected in this situation.
A couple of strategies exist for avoiding artifact limited imaging situations in which the human observer is grossly inefficient at the task required of them. One strategy is hardware design -- seeking an imaging system design that gives a data set that will result in optimal performance of the observer. A second option is to use machines to do post-processing of the images -- machine algorithms can either aid the human observer by generating a post-processed image with an overall PSF that is local, or machines can be trained to perform the task in lieu of the human.

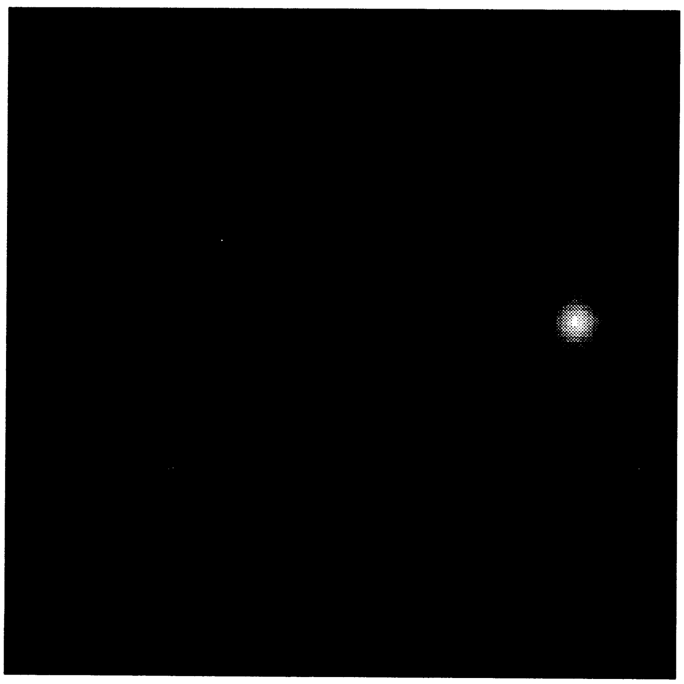

Figure 2. Demonstration of streak artifacts in reconstruction of disk object from noise-free 8-view data set.

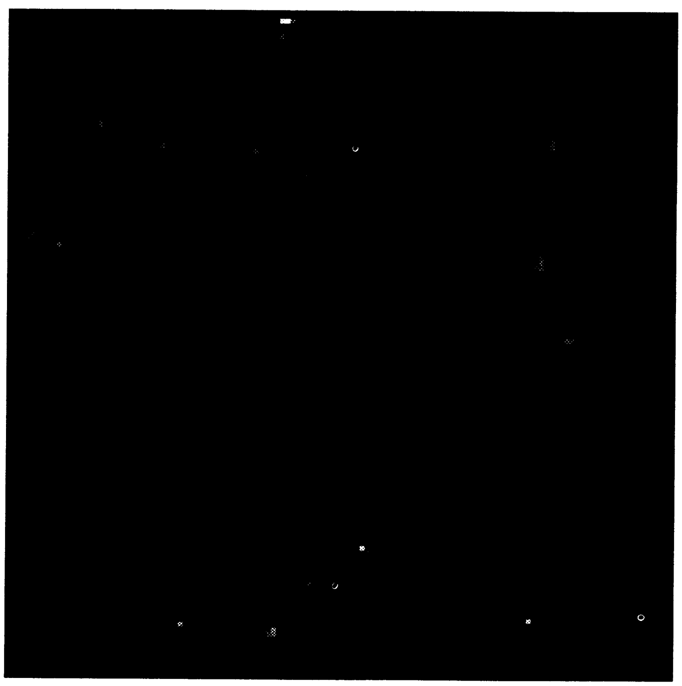

Figure 3. Image reconstructed from data set used to generate Figure2, but with different reconstruction algorithm parameters.

As an example of a hardware design that optimizes observer performance, we return to the study of disk detection in the presence of a lumpy background we first discussed in Section 3. In addition to the finding that human performance tracked the performance of the Hotelling observer, it was also determined that human and Hotelling observer performance could be maximized through a judicious choice of aperture size.[31] It was found that as the "lumpiness" of the background increased, the full-width-at-half-max of the system PSF that gave maximum detectability decreased. Other investigators have found a similar result - while a system with poor resolution might allow decent task performance for a well-specified signal and background, object variability increases the need for system resolution.[32,33]

Now consider the second option of post-processing. We have investigated this option for the task of classifying CT images reconstructed from a limited number of views. Numerous algorithms are available for post-processing such images; each gives different artifact characteristics. For example, Figure 3 is a reconstruction from the same data that produced Figure 2. Both reconstructions were calculated using an algorithm known as MEMSYS3[34,35] but with different settings of certain algorithm parameters. To determine the best settings for an algorithms' variables, and to compare across families of algorithms, it is crucial that the task be specified and the abilities of the intended observer be considered. We have investigated a number of reconstruction algorithms, but will concentrate our attention on the results obtained using the MEMSYS3 algorithm here.[36]

We have investigated human and machine performance for two discrimination tasks on images reconstructed from a limited number of tomographic views. The first task was the detection of low-contrast disks at known locations in scenes also containing many randomly located high-contrast disks.[37] The second task was the discrimination of exactly specified single and binary objects (the "Rayleigh" task) in an image of many such objects placed randomly to generate artifacts in the reconstructions.[38] To calculate the ideal observer's decision variable, the full probability distribution on the data given to the observer under each hypothesis must be known. Thus, for a nonlinear reconstruction algorithm such as MEMSYS3, the Bayesian decision variable can be difficult to calculate and so we have used approximations here. Nevertheless, as Figures 4 and 5 show, the human is remarkably well modelled by a machine observer that performs simple linear operations on the images. These figures represent families of reconstructions derived from MEMSYS3 as a function of the parameter alpha that determines how well the reconstructions are forced to match the acquired data as opposed to maximizing the entropy of the reconstruction.

Given our earlier observations of the poor performance of human observers for non-local tasks with object variability, one might have expected the human to not fare well compared to the ideal observer. One consideration here is that, even though the data sets consist of only 8 views, the MEMSYS3 algorithm tends to result in reconstructions that do not have strong streak-like artifacts regardless of the value of alpha. (Figure 2 contains strong streak artifacts because the algorithm was interrupted before it converged; Figure 3 is a reconstruction obtained by allowing the algorithm to run to completion.) Thus the MEMSYS3 reconstruction of a disk is pretty much a disk even with the kind of object variability we have built into the task -- this makes the use of a simple linear template possibly quite effective.

A more troubling question is whether the true ideal observer's performance curve would be much higher than the machine approximations shown in Figures 4 and 5. The fact that the human does as well as the machine for the detection task at low values of alpha suggests that the machine observer is not truly ideal. After all, even in the simplest tasks described in Section 2 the human typically does no better than 50% efficiency relative to the ideal observer. This outstanding question needs further investigation. Barrett et al. have derived the full probability distribution needed to determine the ideal-observer decision variable for the nonlinear EM algorithm.[39] A similar calculation needs to be done to determine the actual ideal-observer strategy on images generated using MEMSYS3.

When considering artifacts, their impact on task performance, and the concomitant implications for system design, a few more points should be made. First, even though it is often possible to post-process images to reduce artifacts, it should be noted that there is always an attendant noise penalty. Also, while an observer might be able to learn that certain objects generate artifacts of a known general structure, artifacts can lessen clinicians' confidence by putting the burden on them to know what is an artifact and what is biology. And finally, an imaging system is rarely used for a single task. While certain detection or discrimination tasks might be performed reasonably well in the presence of artifacts, those same artifacts can also make performance of estimation tasks on the same image iinaccurate As always, the system design should be for a range of tasks, where all these factors are considered. Barrett et al. have derived figures of merit for both classification and estimation tasks, and have shown how a simple type of object variability - random location - relates the design of a system for a classification task intimately with the design of a system that is optimized for estimation.[40]

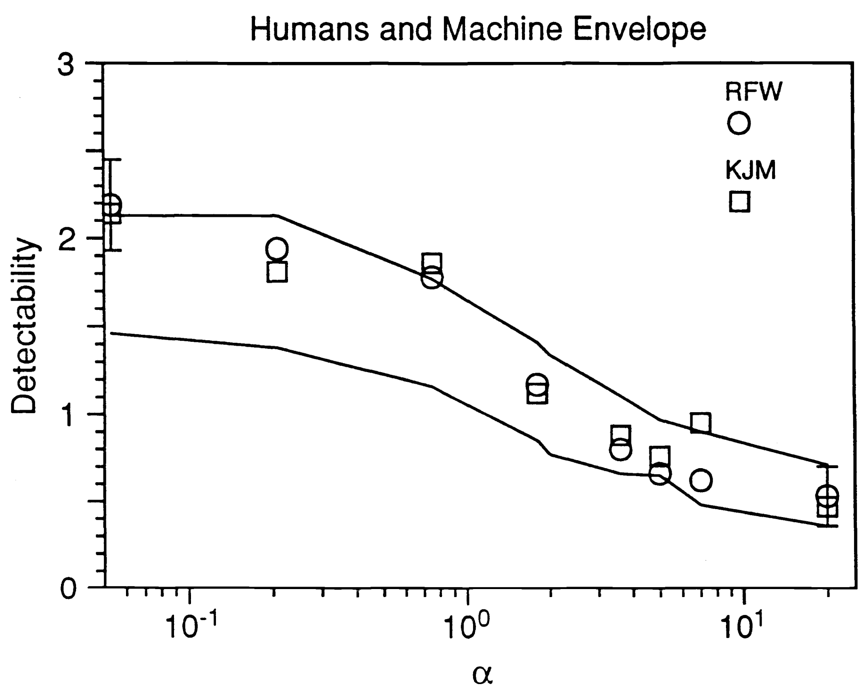

Figure 4. Observer performance for detecting low-contrast disks in tomographic reconstructions from limited, noisy data.

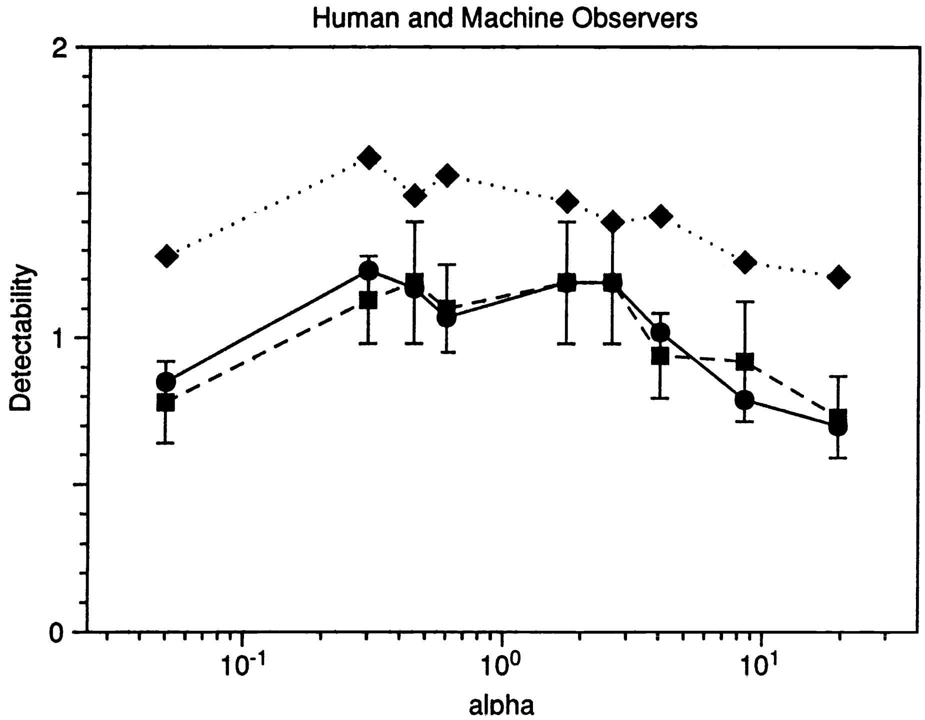

Figure 5. Observer performance for the Rayleigh discrimination task using tomographic reconstructions from limited, noisy data.

5. SUMMARY

Progressively more complicated tasks have been investigated to determine how well machine observers or theoretical models can be used to predict human performance. These investigations have resulted in a better understanding of how to balance the design trade-off between resolution and noise in the presence of a lumpy background, for example, and what choice of post-processing options best serves the observer for the specified task. For tasks in which the optimal strategy can be approximated by one or more linear operations using compact filters, the human has been shown to be fairly efficient. This is true with the caveat that there appear to be limits to the contrast sensitivity and integration area of the human observer. Both these limitations can often be handled by changing the system magnification, in the first case, or by post-processing that alters local or global contrast, in the second case. For more complicated tasks that require either much more number crunching or a large amount of memory, the human often falls off severely in performance compared to better theoretical or machine models. In such cases, the human might be augmented by a machine helper. Or, the system designer might revisit the design to enable the human to more readily extract information required to perform the given task.

6. REFERENCES

1. A.D. Whalen, *Detection of Signals in Noise* (Academic, New York, 1971).
2. H.L. Van Trees, *Detection, Estimation, and Modulation Theory* (Wiley, New York, 1968).

3. A.E. Burgess, K. Humphrey, and R.F. Wagner, "Detection of bars and discs in quantum noise," Proc. SPIE **173**, 34-40 (1979).

4. A.J. Ahumada Jr., "Putting the visual system noise back in the picture," J. Opt. Soc. Am A **4**, 2372-2378 (1987).

5. A.E. Burgess and B. Colborne, "Visual signal detection. IV. Observer inconsistency," J. Opt. Soc. Am. A **5**, 617-627 (1988).

6. D.G. Pelli, "The quantum efficiency of vision," in *Vision: Coding and Efficiency,* C. Blakemore, ed., 3-24 (Cambridge University Press, 1990).

7. A.E. Burgess, R.F. Wagner, R.J. Jennings, and H.B. Barlow, "Efficiency of human visual signal discrimination," Science **214**, 93 (1981).

8. A.E. Burgess and H. Ghanderharian, "Visual signal detection. I. Ability to use phase information," J. Opt. Soc. Am. A 1, **900-905** (1984).

9. L.-N. Loo, K. Doi, and C.E. Metz, "A comparison of image quality indices and observer performance in the radiographic detection of nylon beads," Phys. Med. Biol. **29**, 837-856 (1984).

10. P.F. Judy and R.G. Swensson, "Lesion detection and signal-to-noise ratio in CT images," Med. Phys. **8**, 13-23 (1981).

11. K.J. Myers, H.H. Barrett, M.C. Borgstrom, 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).

12. 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).

13. R.F. Wagner, M.F. Insana, D.G. Brown, B.S. Garra, and R.J. Jennings, "Texture discrimination: radiologist, machine, and man," in *Vision: Coding and Efficiency,* C. Blakemore, ed., 310-318 (Cambridge University Press, 1990).

14. S.J. Starr, C.E. Metz, L.B. Lusted, and D.J. Goodenough, "Visual detection and localization of radiographic images," Radiology **116**, 533-538 (1975).

15. R.G. Swensson and P.F. Judy, "Detection of noisy visual targets: models for the effects of spatial uncertainty and signal-to-noise ratio," Percept. Psychophys. **29**, 521-534 (1981).

16. A.E. Burgess and H. Ghandeharian, "Visual signal detection. II. Signal-location identification," J. Opt. Soc. Am. A **1**, 906-910 (1984).

17. A.E. Burgess, "Visual signal detection. III. On Bayesian use of prior knowledge and cross correlation," J. Opt. Soc. Am. A **2**, 1498-1506 (1985).

18. L.W. Nolte and D. Jaarsma, "More on the detection of one of M orthogonal signals, "J. Acoust. Soc. Am. **41**, 497-505 (1967).

19. G. Revesz, H.L. Kundel, and M.A. Graber, "The influence of structured noise on detection of radiologic abnormalities," Invest. Radiol. **9**, 479-486 (1974).

20. R.G. Swensson and P.F. Judy, "Background area effects on feature detectability in CT and uncorrelated noise," 73rd Annual Meeting of the Radiological Society of North America, Chicago, Ill. (1987).

21. 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).

22. J. Yao and H.H. Barrett, "Predicting human performance by a channelized Hotelling observer model," Proc. SPIE **1768**, 161-168 (1992).

23. H.H. Barrett, J. Yao, J.P. Rolland, and K.J. Myers, "Model observers for assessment of image quality," Proc. Natl. Acad. Sci. USA **90**, 9758-9765 (1993).

24. A.E. Burgess, Statistically defined backgrounds: performance of a modified non-prewhitening observer model," to be published in J. Opt. Soc. Am. A (1994).

25. H.B. Barlow, "The efficiency of detecting changes in density in random dot patterns, "Vision Res. **18**, 637-650 (1978).

26. H.B. Barlow and B.C. Reeves, "The versatility and absolute efficiency of detecting mirror symmetry in random dot displays," Vision Res. **19**, 783-793 (1979).

27. R.K. Maloney, G.J. Mitchison, and H.B. Barlow, "Limit to the detection of Glass patterns in the presence of noise," J. Opt. Soc. Am. A **4**, 2336-2341 (1987).

28. M. Tapiovaara, "Ideal observer and absolute efficiency of detecting mirror symmetry in random images," J. Opt. Soc. Am. A **7**, 2245-2253 (1990).

29. B.S. Garra, M.F. Insana, T.H. Shawker, R.F. Wagner, M. Bradford, and M.A. Russell, "Quantitative ultrasonic detection and classification of liver disease: comparison with human observer performance," Invest. Radiol. **24**, 196-203 (1989).

30. J.P. Rolland, H.H. Barrett, and G.W. Seeley, "Quantitative study of deconvolution and display mappings for long-tailed point-spread functions," Proc. SPIE **1092**, 17-21 (1989).

31. K.J. Myers, J.P. Rolland, H.H. Barrett, and R.F. Wagner, "Aperture optimization for emission imaging: effect of a spatially varying background," J. Opt. Soc. Am. A **7**. 1279-1293 (1990).

32. B.M.W. Tsui, C.E. Metz, F.B. Atkins, S.J. Starr, and R.N. Beck, "A comparison of optimum spatial resolution in nuclear imaging based on statistical theory and on observer performance," Phys. Med. Biol. **23**, 654-676 (1978).

33. M.F. Kijewski, S.P. Mueller, S.C. Moore, F.J. Rybicki, "Generalized ideal observer models for design and evaluation of medical imaging systems," presented at the 77th Assembly of the Radiological Society of North America (Chicago, 1991).

34. J. Skilling, "Classic maximum entropy," in *Maximum Entropy and Bayesian Methods in Science and Engineering, Vol. 1, Foundations*, ed. G.J. Erickson and C.R. Smith (Kluwer, 1988).

35. S.F. Gull and J. Skilling, *Quantified Maximum Entropy - MEMSYS3 Users' Manual,* Maximum Entropy Data Consultants Ltd.. Royston, England (1989).

36. K.M. Hanson, " Method of evaluating image-recovery algorithms based on task performance," J. Opt. Soc. Am. A **7**, 1294-1304 (1990).

37. R.F. Wagner, K.J. Myers, and K.M. Hanson, "Task performance on constrained reconstructions: human observer performance compared with sub-optimal Bayesian performance," Proc. SPIE **1652**, 352-362 (1992).

38. K.J. Myers, R.F. Wagner, and K.M. Hanson, "Rayleigh task performance in tomographic reconstructions: comparison of human and machine performance," Proc. SPIE **1898**, 628-637 (1993).

39. H.H. Barrett, D.W. Wilson, and B.M.W. Tsui, "Noise properties of the EM algorithm: I. theory," to be published in Phys. Med. Biol. (1994).

40. H.H. Barrett, "Objective assessment of image quality: effects of quantum noise and object variability," J. Opt. Soc. Am. A **7**, 1266-1278 (1990).