Visual detection efficiency in ultrasonic imaging: A framework for objective assessment of image quality

Michael F. Insana and Timothy J. Hall
Department of Diagnostic Radiology, University of Kansas Medical Center, 3901 Rainbow Boulevard, Kansas City, Kansas 66160-7234

(Received 13 August 1993; accepted for publication 29 December 1993)

Experimental methods for estimating detection efficiencies of human and computational observers viewing low-contrast circular targets in acoustic noise are described. Sonographic images were simulated with signal and noise properties specified exactly. These images were presented to observers in two-alternative forced-choice (2AFC) experiments. Relative to the ideal observer of these images, i.e., the prewhitening matched filter, human observers were 60% efficient for detecting targets over a broad range of target energies and for both target polarities. Studies were limited to situations where target diameters were much larger than the correlation length of the noise. In that case, observers unable to decorrelate the noise showed no reduction in detectability as predicted by theory. For example, the efficiency of one computational observer, a nonprewhitening matched filter, was nearly ideal. Its response was proportional to that of the average human observer, which suggests a role for computational observers in image evaluation.

PACS numbers: 43.60.Gk, 43.60.Cg

INTRODUCTION

Many imaging tasks require observers to visually discriminate among regions of similar reflectivity. An example in diagnostic ultrasound is lesion detection. Consequently, one essential feature of high-quality ultrasonography is superior low-contrast detectability. Today's sonographers have the opportunity to reconfigure the instrumentation to match specific diagnostic tasks, e.g., maximize lesion detectability, by selecting the appropriate transducer technology and pre- and postprocessing schemes. Unfortunately, there are no standard criteria for optimization. Evaluations are subjective, for the most part, because objective criteria have not yet been developed for many important clinical imaging tasks. In this paper we discuss objective criteria for evaluating the quality of a sonographic image for the clinically important task of detecting low-contrast targets.

Objective assessment begins by first specifying a relevant task and then determining quantitatively how well the task is performed.[1] Our approach to image assessment is based on classical signal detection theory[2,3] as applied to ultrasonic imaging by Wagner *et al.*[4] and Smith *et al.*[5] The task is visual detection of low-contrast targets in speckle (correlated noise). The procedure begins with the derivation of a test statistic, called a *decision function*, for the optimal detector or *ideal observer* of the image. The ideal observer is a computational observer or mathematical function of the image data that makes optimal use of all available information for the purpose of accomplishing a task. Ideal observers provide the upper limit of detectability by which the efficiency of human observers can be determined. The task required of the ideal observer is defined by the task it is given. The decision function—test statistic—is determined mathematically by the probability density functions (pdf) of the image data. The test statistic is applied to images, a decision is made, and the ability to perform the task is evaluated. Observer performance is specified by the signal-to-noise ratio (SNR). Finally, human visual detection efficiency is computed from SNR values measured for ideal and human observers by the ratio $\mathrm{SNR}_H^2/\mathrm{SNR}_I^2$.[6]

The efficiency for performing important clinical examinations is the ultimate assessment of image quality, and therefore the appropriate criteria for task-specific system optimization. Unfortunately, statistically robust estimates of human observer SNR are labor intensive because of three intrinsic sources of response uncertainty: Between-observer variance, within-observer variance, and between-image variance.[7] Therefore we have also examined the use of computational observers, since they generate no between- or within-observer variances. Suboptimal computational observers are also of great practical interest if they can be computed quickly—ideally at the frame rate for video—and if the corresponding SNR is highly correlated with that of the average human observer. We show that the nonprewhitening matched filter for intensity is one such computational observer that could quickly provide precise estimates of the human observer's ability to visualize low-contrast targets. Computational observers may offer the objective assessment of ultrasonic images necessary to optimize systems for specific diagnostic tasks.

I. METHODS

Observer experiments were performed using simulated ultrasonic B-mode images. The images were generated by applying knowledge of the statistical properties of radio-frequency echo signals, X.[8] Two independent, identically distributed Gaussian white-noise patterns were formed us-

 0001-4966/94/95(4)/2081/10/$6.00

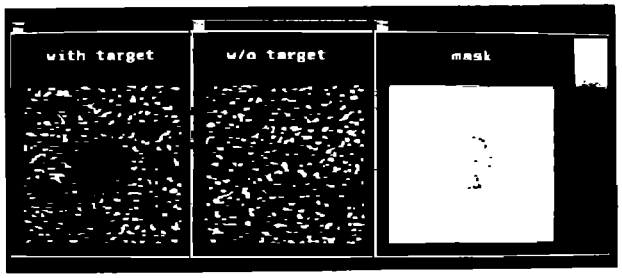

FIG. 1. An example trial of 2AFC detection where the signal (disk) is known exactly (SKE). A speckle-free mask image of the target is displayed along with two speckle fields, where only one contains the target.

ing an algorithm that generates normal random deviates. One pattern represented the real (in-phase) part of the analytic signal of X and the other represented the imaginary (quadrature) part. Circular targets with either positive or negative contrast were introduced by *multiplying* a circular target region in the center of the real and imaginary patterns by the object contrast factor (OCF). For example, the OCF was 1.2 for a 20% increase in echo amplitude in the target relative to the background (1.6 dB). Both patterns were smoothed using a two-dimensional (2-D) Gaussian-shaped smoothing kernel, where the horizontal dimension was 50% larger than the vertical dimension. Finally, to simulate the envelope-detection process of image formation, the real and imaginary components were squared and summed. The square root of the result yielded simulated B-mode image data, y. These images simulate the essential statistical properties of real B-mode images. The 3-D nature of realistic transducer pressure fields and objects was not included in this study.

The probability density function (pdf) of the B-mode image data, where the in-phase and quadrature components of the echo signal are bivariate normal with zero mean and equal variance, ψ, is Rayleigh

$$p(y)=\frac{y}{\psi}\exp\left(\frac{-y^2}{2\psi}\right), \tag{1}$$

where 2ψ is the mean-square scattering amplitude, i.e., the average backscattered intensity I.[4] We verified the existence of the fully developed speckle condition that gives rise to Rayleigh statistics by measuring the first eight even moments for y, e.g., $\langle y^2\rangle$, $\langle y^4\rangle$, etc., using 200 target-free images. Values agreed with those derived by Middleton[9] within the experimental uncertainty.

The task required of the observers is to detect disk-shaped targets of known properties, i.e., the so-called signal-known-exactly (SKE) condition, where target visibility is limited by multiplicative, Rayleigh, colored noise. Two-alternative forced-choice (2AFC) experiments were designed to measure detectability. For each response, observers viewed three images as in Fig. 1. Two were speckle fields, one with and one without a target. The selection of the image—image 1 or image 2—for placement of the target was random, although the location of the target in the image was not. The third is a speckle-free image of the target as it appears in one of the first two images. Observers were asked to decide which of the two speckle images contained the target. This 2AFC design follows the basic paradigm for detection analysis that allows straightforward comparisons between human and ideal observers.[2]

A. Ideal observer

In a 2AFC experiment under SKE conditions, the ideal observer examines two images, y_1 and y_2 in the region where the target is expected, and decides between two hypotheses: The target is present in image 1, H_1, or the target is present in image 2, H_2. The decision is based on the value of the test statistic D, where as derived in Appendix A,

$$D=\sum_{i=1}^{M} y_{1i}^2-\sum_{j=1}^{M} y_{2j}^2, \tag{2}$$

and M is the number of independent samples per target area (see Appendix B).

Equation (2) states that the optimal decision function is a matched filter for intensity. The detection strategy of the ideal observer is to measure the total intensity within the target region for both images. For positive contrast, the ideal observer decides that the target is in the first image when $D\geqslant 0$ and in the second image when $D<0$. The decisions are reversed for negative-contrast targets. The ideal observer operates on independent samples of the noise; that is, it finds a representation in which the noise is uncorrelated and in that representation the noise is spectrally white.[22]

The ability of the ideal observer to succeed at this task is quantified by SNR_I^2, which is computed from the means and variances of the decision function under the two hypotheses:

$$\mathrm{SNR}_I^2=\frac{(\langle D|H_2\rangle-\langle D|H_1\rangle)^2}{\frac{1}{2}(\sigma_{D|H_1}^2+\sigma_{D|H_2}^2)}. \tag{3}$$

The means and variances follow immediately from the moments of the Rayleigh pdf:[4]

$$\begin{aligned}&\langle D|H_2\rangle=-\langle D|H_1\rangle=2M(\psi_s-\psi_n),\\ &\sigma_{D|H_1}^2=\sigma_{D|H_2}^2=4M(\psi_s^2+\psi_n^2),\end{aligned} \tag{4}$$

where $I_n=2\psi_n$ and $I_s=2\psi_s$ denote the intensities in the noise only and in the signal target area, and $\langle\cdots\rangle$ denotes ensemble average. Combining Eqs. (3) and (4) we find that

$$\mathrm{SNR}_I^2=4M\frac{(\psi_s-\psi_n)^2}{(\psi_s^2+\psi_n^2)}, \tag{5}$$

$$=(2M)\left(\frac{\bar{\psi}}{[\frac{1}{2}(\psi_s^2+\psi_n^2)]^{1/2}}\right)^2\left(\frac{\psi_s-\psi_n}{\bar{\psi}}\right)^2,$$

$$=\left(\frac{\pi a^2}{2S_c}\right)\mathrm{SNR}_0^2\,C^2, \tag{6}$$

where

$$\mathrm{SNR}_0^2=\frac{\frac{1}{4}(I_s+I_n)^2}{\frac{1}{2}(\sigma_{I_s}^2+\sigma_{I_n}^2)}=\frac{2\bar{\psi}^2}{(\psi_s^2+\psi_n^2)};\quad \bar{\psi}\equiv\tfrac{1}{2}(\psi_s+\psi_n),$$

$$C^2=\frac{(I_s-I_n)^2}{\frac{1}{4}(I_s+I_n)^2}=\frac{(\psi_s-\psi_n)^2}{\bar{\psi}^2};\quad M\simeq\frac{S_t}{S_c}=\frac{\pi a^2}{4S_c}.$$

The first factor on the right-hand side of Eq. (6) describes how *target diameter, a,* and *spatial resolution, S_c,* affect detectability. The second factor, SNR_0^2, is the square of the point signal-to-noise ratio, i.e., SNR_0 is the mean-to-standard deviation ratio for intensity, and a summary of the *noise properties* of squared B-scan images. (Note the distinction between SNR_I and SNR_0.) The third factor is the square of *intensity contrast* C^2.

SNR_I^2 provides the quantitative index needed to measure the highest possible detection performance, and as such defines the difficulty of the observation task. The detectability index for the ideal observer increases as the target area S_t and contrast C increase. Also, as the resolution cell area S_c decreases, the spatial resolving power of the imaging system increases, providing more independent samples per target area and increased target visibility. Equation (6) describes how compromises in the basic engineering properties that determine image quality (for the most part, beam properties) affect low-contrast detectability of the ideal observer. Because it defines the best-possible performance, Eq. (6) is the standard for measuring observer detection efficiency. SNR_I has already been used to predict how changes in system design affect target visibility.[5,11,12] Any change in the image data or experimental conditions alters the task, and hence the decision function, which can change the strategy of the ideal observer.[13]

Note that to compute SNR_I^2 using the analytic expression, we need to count the number of independent samples per target area available to the observer. Methods for estimating M are discussed in Appendix B.

B. Human observers

Three human observers participated in the 2AFC experiments. Observers were asked to examine the mask image and then choose which of the two speckle images contained the target. Viewing time per image was not restricted, and the observers were not told if the choices were correct, i.e., no feedback.

Observers sat in a darkened room, 45 cm from a 19-in. color workstation monitor. The resolution of the monitor was 1024×864 and had 3.049 pixels/mm. The simulated image format was 128×128 pixels and each was magnified a factor of 2, so that the viewing size was 84×84 mm. One magnified pixel subtended an angle of 0.73 mrad (2.5 min of arc), which is approximately the effective blurring aperture of the eye. Measured full-scale brightness of the monitor was 7.7 cd/m^2 and the average image brightness was 1.0 ± 0.1 cd/m^2.

The fraction of correct responses was our estimate of the probability of a correct response $P(C)$. The figure-of-merit for evaluating human observer performance is SNR_H^2, and is often called d_a^2 in the psychophysics literature. As shown in Appendix C, $P(C)$ and d_a^2 are related by the expression

$$d_a^2=4\{\Phi^{-1}[P(C)]\}^2, \tag{7}$$

where $\Phi(z)$ is the integral of the standard normal distribution and $z=\Phi^{-1}[\Phi(z)]$.

Visual detection efficiency of human observers relative to the ideal observer, η_{HI}, was calculated from signal-to-noise ratio estimates as follows

$$\eta_{HI}=\left(\frac{d_{a,H}}{\mathrm{SNR}_I}\right)^2. \tag{8}$$

Efficiency measures the fraction of the total information an observer extracts from the image in performing the task.[14]

C. Computational observer

Our objective in applying a computational observer to ultrasound images is image evaluation. Such evaluation should be accomplished more quickly and with lower uncertainty than human experts. Also, the performance of the computational observer should be highly correlated with that of the human observer, and directly interpretable in terms of fundamental properties of image quality.

An obvious candidate is the nonprewhitening matched filter (NPWMF), which is realized by implementing the strategy of the ideal observer in Eq. (2), but replacing M with N, the number of pixels per target area. Myers *et al.*,[15] have shown that, like the NPWMF, human observers are unable to prewhiten the noise and may be handicapped by large negative noise correlations when performing detection tasks. In the absence of *negative* correlations, such as those that result from images reconstructed from projections, the reduction in detectability due to correlated noise is considered minimal as long as there are many (>10) independent samples per target area.[14] Expressions analogous to Eqs. (7) and (8) were used to compute $d_{a,C}^2$ and η_{CI} for the NPWMF computational observer from the fraction of correct responses to 2AFC experiments.

II. RESULTS

Measurements of d_a^2 for human and computational observers viewing positive-contrast and negative-contrast lesions are listed in Table I and plotted in Fig. 2 as a function of the task difficulty, SNR_I^2. Larger values of SNR_I^2 correspond to more visible targets. The quantity d_a^2 may be interpreted as the squared signal-to-noise ratio required for the ideal observer to match the detectability index of the test observer, e.g., when $d_a^2=\mathrm{SNR}_I^2$ (dashed lines in Fig. 2) the performance of the test observer is ideal. Individual human observer results are plotted in the figures, whereas average values are listed in Table I. Error bars were computed using the results of Appendix C.

We found it convenient to use another form of Eq. (6) to calculate SNR_I^2 as a function of the object contrast factor (OCF) and M:

TABLE I. Summary of signal-to-noise measurements for the ideal observer (SNR_I^2), the average of three human observers ($d_{a,H}^2$), and the NPWMF computational observer ($d_{a,C}^2$). Observers viewed 200 image pairs at each SNR_I^2 value. OCF is the object contrast factor, a is the target diameter in millimeters, and M is the number of independent samples per target area, Eq. (B6).

Positive contrast				Negative contrast					
SNR_I^2	OCF	$d_{a,H}^2$	$d_{a,C}^2$	SNR_I^2	OCF	$d_{a,H}^2$	$d_{a,C}^2$	a	M
1.172	1.029	0.20±0.12	0.77±0.36	1.246	0.971	0.33±0.15	1.58±0.41	29.5	178
2.133	1.059	0.94±0.24	2.00±0.70	2.401	0.941	1.69±0.43	2.28±0.56	19.7	81
3.352	1.075	1.58±0.41	4.39±1.11	3.889	0.925	1.65±0.43	4.67±1.21	19.7	81
4.689	1.059	2.70±0.69	4.67±1.71	5.277	0.941	3.57±0.87	5.76±1.62	29.5	178
6.770	1.090	3.75±0.97	6.29±2.63	8.093	0.910	4.39±1.11	6.02 ±1.74	23.6	145
8.539	1.102	4.48±1.14	9.41±3.92	10.44	0.898	6.29±1.86	9.94 ±4.41	23.6	115
10.46	1.090	5.42±1.56	12.3 ±10.4	12.51	0.910	6.15±1.79	14.2 ±11.1	29.5	178
13.20	1.122	5.96±1.80	13.6 ±14.0	16.77	0.878	10.2 ±4.70	18.8 ±29.3	24.9	127
14.84	1.130	9.17±3.70	14.2 ±11.1	19.13	0.870	9.17±3.70	18.8 ±29.3	24.9	127
17.00	1.140	8.94±2.49	18.8 ±29.3	22.32	0.860	10.8 ±5.38	23.7 ±77.9	24.9	127
18.42	1.122	14.6 ±12.8	26.5 ±196.	23.39	0.878	10.6 ±5.14	26.5 ±139.	29.5	178

$$SNR_I^2=\frac{4M(1-OCF^2)^2}{1+OCF^4}. \tag{9}$$

Because a simple 2-D Gaussian point-spread function was used in the simulations, it was possible to compute M from the exact expression, Eq. (B6), thus eliminating the bias associated with the usual approximation of M [Appendix B, Eq. (B12)].

The data reduction process is illustrated in Fig. 3 for the computational observer examining 1000 image sets for each hypothesis. In this example, OCF = −0.925 (−0.68 dB), $M=81$ (20-mm target), and $SNR_I^2=3.889$, which may be considered as a moderately challenging detection task. The figure shows that the distributions $p(D|H_1)$ and $p(D|H_2)$ overlap, thereby producing detection errors and a score of 81.4% correct. The magnitude of the means and the variances for the distributions are equal, as they must be for a 2AFC experiment, and given by 10.2 and 134.6, respectively. From Eq. (3), we find that $SNR_C^2=3.09$. As expected, this value is approximately equal to the result of Eq. (7), where for $P(C)=0.814$ we find $d_{a,C}^2=3.19\pm0.35$. In this example and in Figs. 2(b) and (d), the detectability index of the NPWMF computational observer is nearly ideal: $\eta_{CI}=0.82$.

The solid lines in Fig. 2 are least-squares fits to a straight line as weighted by the experimental uncertainties.[16] Slopes, intercepts, and χ^2 values for the fits are listed in Table II. The χ^2 values suggest that the human visual response is a linear function of the ideal observer response. The intercepts are small, so that the slope is a good estimate of detection efficiency. The efficiency for detecting negative-contrast targets ($\eta_{HI}=59\%$) was equivalent to that for positive-contrast targets ($\eta_{HI}=58\%$), and the three human observers responded equivalently, although one (LH) was an experienced sonographer and two were not.

Following the suggestion made by Wagner and Brown,[14] we expanded the definition of human observer efficiency to includes $d_{a,C}^2$, and rewrote Eq. (8) as the product of two factors:

$$\eta_{HI}=\left(\frac{d_{a,H}}{d_{a,C}}\right)^2\left(\frac{d_{a,C}}{SNR_I}\right)^2,$$
$$=\eta_{HC}\eta_{CI}. \tag{10}$$

The first factor, η_{HC}, is the efficiency of human observers relative to the NPWMF. Its value is approximately 0.5 for all targets (Table II).

The second factor, η_{CI}, is the efficiency of the NPWMF relative to the ideal observer. It has been called the observer reconstruction efficiency[14,15] for images reconstructed from projections, because it describes the reduction in detectability caused by correlations in the noise resulting from image formation. Since M is large, we expected the effects of noise correlations to be minimal for pulse-echo ultrasound. Figure 2(b) and (d) and Table II

TABLE II. Summary of linear regression analysis applied to the data in Fig. 2. Slopes are an estimate of visual detection efficiencies, η. Q is a goodness-of-fit measure that depends on χ^2 and the number of degrees of freedom; it is the probability that deviations from the model are due to chance.

	Contrast	Slope	Intercept	χ^2	Q
$d_{a,H}^2$ vs SNR_I^2	Positive	$\eta_{HI}=0.58\pm0.07$	-0.45 ± 0.16	5.44	0.99
(human versus ideal)	Negative	$\eta_{HI}=0.59\pm0.05$	-0.34 ± 0.12	11.7	0.86
$d_{a,C}^2$ vs SNR_I^2	Positive	$\eta_{CI}=1.18\pm0.24$	-0.54 ± 0.51	1.11	1.00
(computational versus ideal)	Negative	$\eta_{CI}=0.84\pm0.19$	$+0.51\pm0.52$	1.68	1.00
$d_{a,H}^2$ vs $d_{a,C}^2$	Positive	$\eta_{HC}=0.49\pm0.05$	-0.16 ± 0.13	3.28	0.95
(human versus computational)	Negative	$\eta_{HC}=0.54\pm0.07$	-0.42 ± 0.21	9.93	0.37

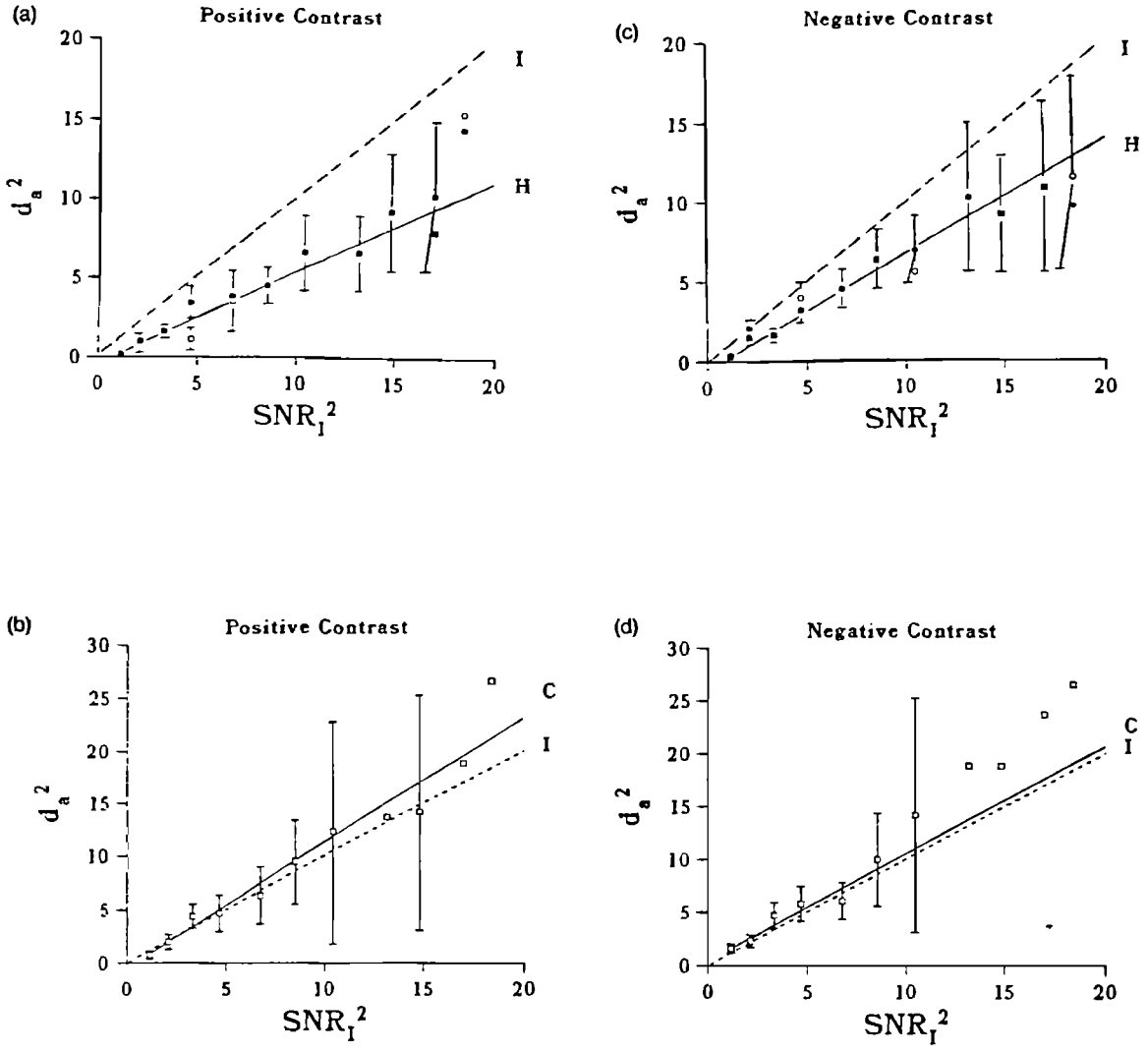

FIG. 2. Plots of the detectability index d_a^2 as a function of the squared signal-to-noise ratio for (a)(c) three human observers [● (MI), O (LC), ■ (LH)] and (b)(d) the NPWMF computational observer (□) of positive-contrast and negative-contrast targets. Dashed lines indicate the detectability index of the ideal observer, whereas solid lines are least-squares fits to a straight line as weighted by the experimental uncertainties.[16] Detection efficiencies are estimated from the slopes, which are summarized in Table II. Error bars denote one standard error according to Appendix C. (Not all points are visible because they overlap.)

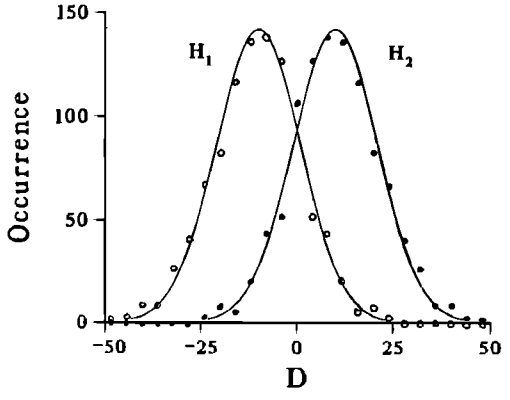

FIG. 3. Histograms of the decision functions for the computational observer under the two hypotheses, H_1 and H_2, and for $SNR_T^2 = 3.889$. Solid lines are Gaussian functions having the same mean and variance as the data.

show that η_{CI} is nearly 1, which indicates that speckle generates no more detection errors than white noise.

III. DISCUSSION

Human observers are approximately 60% efficient at detecting low-contrast targets in ultrasound images when the target is specified exactly. This value is in the range of efficiencies measured by others for many visual detection and discrimination tasks.[17,18] Our observation that target polarity is not a factor determining observer efficiency is consistent with observations made for other digital imaging modalities.[18]

We observed no loss in low-contrast detectability due to the inability of observers to decorrelate the noise. Wagner and Brown[14] show that the SNR for the prewhitening matched filter (the ideal observer for our task) and that of the nonprewhitening matched filter (our computational observer) are approximately equal for colored noise as long as the power spectral density of the speckle is a slowly

varying function at all spatial frequencies in the target power spectrum. This condition is met throughout this study, since there are no negative noise correlations and $M > 10$. We expect noise correlations to become a factor when $S_t \simeq S_c$, or when unusual filter functions are applied to projection data to form ultrasound CT images.[15]

The linear relationship between responses of human observers and the NPWMF suggests that the NPWMF is a reasonable choice for a computational observer. One application of a computational observer in ultrasonic image quality assessment is to compare how various transducer designs or signal processing schemes influence the visibility of low-contrast lesions in the body. Receiver operating characteristic (ROC) analysis is the accepted comparison standard. However, ROC experiments are very costly and the uncertainties are large even for expert observers, so they are rarely used by manufacturers to design instrumentation. In place of expert humans, the NPWMF could be applied to image data for either the 2AFC or yes–no experimental designs. Tissuelike phantoms with well-defined properties are available to provide as many as 100 statistically independent views of a target under SKE conditions.[19] The detectability index for human observers, $d^2_{a,H}$, is then approximately 0.5 $d^2_{a,C}$. In this manner $d^2_{a,H}$ may be estimated very quickly and without the uncertainties associated with the within-observer and between-observer variances.

There are at least two previous examples where computational observers were used to evaluate imaging strategies. Hanson[20] used a computational observer to evaluate image-recovery algorithms applied to images reconstructed from projections. Later, Lopez *et al.*[21] explored applications in sonographic image evaluation. Although the decision function used by Lopez was the total *amplitude* in the target area instead of the *intensity*, he was limited to one noise realization per target, and he used a different experimental design, he too found the responses of human and computational observers to be highly correlated.

These efficiency estimates apply to the SKE condition. If targets are variable—specifically, if parameters such as target amplitude and position are defined statistically—then the detection strategy of the ideal observer must be derived from Eq. (A1) using all the details of the task to formulate the appropriate likelihood functions. Barrett has shown that the optimal *linear* discriminant, i.e., the best performing decision function that is linear in the intensity data, is the Hotelling test discriminant.[1] The Hotelling test statistic may be a better choice of computational observer under signal-known-statistically (SKS) conditions. Linear discriminants are of considerable interest because of their widespread use in many signal processing applications and because of their success at predicting the performance of human observers.

Emphasizing the task specific nature of image evaluation, we conclude that insofar as low-contrast detectability is an essential feature of diagnostic imaging, observer efficiency estimated using the NPWMF provides a precise, easily computed, and objective assessment of image quality.

ACKNOWLEDGMENTS

The authors wish to thank Larry Cook and Linda Harvey for their participation as observers. The important comments and criticisms of Bob Wagner and David Brown are also greatly appreciated. This work was supported by the Whitaker Foundation, NIH DK43007, Siemens Medical Systems, and the Clinical Radiology Foundation at KUMC.

APPENDIX A

The optimal detector of low-contrast targets in ultrasound images is determined using the statistical tools of hypothesis testing.[3] In our 2AFC experiment where the signal is known exactly, e.g., Fig. 1, an observer must decide between two hypotheses:

H_1: the target is present in image 1 and not image 2

or

H_2: the target is present in image 2 and not image 1.

Assume that the joint conditional probability density functions (pdfs) of the data of image 1, y_1, and the data of image 2, y_2, given the two hypotheses, i.e., $p(y_1,y_2|H_1)$ and $p(y_1,y_2|H_2)$, are known. These pdfs are often called likelihood functions. A decision function for the ideal observer D' is the ratio of likelihood functions [Ref. 2, Eq. (3.5)]

$$D' = \frac{p(y_1,y_2|H_1)}{p(y_1,y_2|H_2)}. \tag{A1}$$

This decision function is optimal in the sense that it satisfies the Neyman–Pearson criterion by providing the maximum detection rate for a fixed false positive rate.[3]

The radio-frequency echo signal is modeled as a complex, multivariate, Gaussian random process, so that the B-mode image data are characterized by a Rayleigh pdf.[4,8] If the Rayleigh noise in one image is independent of that in the other, and if we sample the data such that image samples are also independent, then the conditional probabilities in the known target regions of the image are given by

$$p(y_1,y_2|H_1) = \prod_{i=1}^{M} \frac{y_{1i}}{\psi_s} \exp\left(\frac{-y_{1i}^2}{2\psi_s}\right) \prod_{j=1}^{M} \frac{y_{2j}}{\psi_n} \exp\left(\frac{-y_{2j}^2}{2\psi_n}\right) \tag{A2}$$

and

$$p(y_1,y_2|H_2) = \prod_{i=1}^{M} \frac{y_{1i}}{\psi_n} \exp\left(\frac{-y_{1i}^2}{2\psi_n}\right) \prod_{j=1}^{M} \frac{y_{2j}}{\psi_s} \exp\left(\frac{-y_{2j}^2}{2\psi_s}\right), \tag{A3}$$

where the parameters $2\psi_n$ and $2\psi_s$ are the mean-square scattering amplitudes, i.e., the backscatter intensities for the noise-only image, I_n, and for the signal-in-noise image, I_s, in the region in the medium corresponding to that in the image. M is the number of *independent data samples in the target area* available to the decision maker (see Appendix B).

Substituting Eqs. (A2) and (A3) into Eq. (A1), we find

$$D' = \prod_{i=1}^{M} \exp\left[\frac{y_{1i}^2}{2}\left(\frac{1}{\psi_n}-\frac{1}{\psi_s}\right)\right] \prod_{j=1}^{M} \exp\left[\frac{-y_{2j}^2}{2}\left(\frac{1}{\psi_n}-\frac{1}{\psi_s}\right)\right]. \tag{A4}$$

Note that regions outside the target area do not affect the decision function, because outside the target the pdfs are equal and the likelihood ratio is 1. Applying the monotonic transformation

$$D = \left(\frac{2\psi_s\psi_n}{\psi_s-\psi_n}\right)\ln(D') \tag{A5}$$

yields a more economical, yet equivalently performing, form of the decision function (Ref. 2, Section 1.6):

$$D = \sum_{i=1}^{M} y_{1i}^2 - \sum_{j=1}^{M} y_{2j}^2. \tag{A6}$$

Equation (A6) is a matched filter for intensity where we have assumed disk-shaped targets. This expression may be generalized to include nonrandom targets of arbitrary shape by changing the transformation in Eq. (A5) to $D=\ln(D')$, to find

$$\begin{aligned} D &= \sum_{i=1}^{M'} \Psi_i y_{1i}^2 - \sum_{j=1}^{M'} \Psi_j y_{2j}^2 \\ &= \sum_{i=1}^{M'} \Psi_i (y_{1i}^2 - y_{2i}^2) \\ &= \boldsymbol{\Psi}^t \mathbf{z}, \end{aligned} \tag{A7}$$

where $\Psi_i=(\psi_{si}-\psi_{ni})/2\psi_{si}\psi_{ni}$ and M' is the number of independent samples in the entire image. The boldface quantities $\boldsymbol{\Psi}$ and $\mathbf{z}$ are $M'\times 1$ column vectors, $\boldsymbol{\Psi}^t$ is the transpose of $\boldsymbol{\Psi}$(a $1\times M'$ row vector), and $\mathbf{z} = \mathbf{y}_1^2 - \mathbf{y}_2^2$ is the difference image vector. It is clear from Eq. (A7) that the ideal observer's strategy for detection, e.g., the task of determining which of two images contains a target, and for discrimination, e.g., which of two targets is brighter, are equivalent.

Equation (A7) is a valid approximation for correlated image samples if the correlation length of the target function is much greater than the correlation length of the noise, i.e., if M' is large.[22] For small M', the optimal strategy for detecting nonrandom targets is to use the prewhitening matched filter (PWMF) for intensity:[1]

$$D = \boldsymbol{\Psi}^t K^{-1} \mathbf{z}, \tag{A8}$$

where the covariance matrix $K=\langle \mathbf{z}\mathbf{z}^t\rangle$. Equation (A8) is a general expression for the PWMF; for example, in the special case of uncorrelated noise,

$$K^{-1} = \frac{1}{\psi}\begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix},$$

and Eq. (A8) reduces to Eq. (A7) through a monotonic transformation.

APPENDIX B

To compute the signal-to-noise ratio for the ideal observer, SNR_I, we must first estimate the number of independent data samples per target area M. To find M, we focus our attention on the statistics of the image intensity, $I\equiv y^2$. Let I_0 be the average intensity over the target area:

$$I_0 = \frac{1}{S}\sum_{i=1}^{M} s_i I_i, \tag{B1}$$

where s is a weighting function determined by the shape of the target, and $S=\Sigma_{i=1}^{M} s_i$. For unit amplitude, disk-shaped targets, S is the target area and the mean value of I_0 is given by

$$\begin{aligned} \langle I_0\rangle &= \frac{1}{S}\sum_{i=1}^{M} s_i\langle I_i\rangle, \\ &= \frac{1}{M}\sum_{i=1}^{M}\langle I_i\rangle = \langle I\rangle, \end{aligned} \tag{B2}$$

if $\langle I\rangle$ is independent of position in the image. $\langle\cdots\rangle$ is the ensemble average operator.

The variance of I_0 is given by

$$\begin{aligned} \sigma_{I_0}^2 &= \langle (I_0-\langle I_0\rangle)^2\rangle = \left\langle \left(\frac{1}{M}\sum_{i=1}^{M} I_i - \langle I_0\rangle\right)^2\right\rangle, \\ &= \frac{1}{M^2}\left\langle \left(\sum_{i=1}^{M}(I_i-\langle I_0\rangle)\right)^2\right\rangle. \end{aligned} \tag{B3}$$

Since each element I_i in the average I_0 is an independent sample, the cross product terms in the last expression have expectation zero, and combining Eqs. (B2) and (B3) we find the familiar expression for variance of the mean:

$$\sigma_{I_0}^2 = \frac{1}{M^2}\left\langle \sum_{i=1}^{M}(I_i-\langle I\rangle)^2\right\rangle = \frac{\sigma_I^2}{M}. \tag{B4}$$

Consequently,

$$\frac{\langle I_0\rangle^2}{\sigma_{I_0}^2} = \frac{M\langle I\rangle^2}{\sigma_I^2} = M, \tag{B5}$$

since for circular Gaussian statistics $\langle I\rangle/\sigma_I=1$.[4] Therefore the number of independent samples per target are found from the first-order statistics of I_0.

Goodman[23] derived the squared mean and variance of I_0 to find that

$$\begin{aligned} M &= \frac{\langle I_0\rangle^2}{\sigma_{I_0}^2} \\ &= \left(\frac{1}{S^2}\int_{-\infty}^{\infty} R_s(\Delta x,\Delta z)\,|\,\rho_X(\Delta x,\Delta z)\,|^2 d\Delta x\, d\Delta z\right)^{-1}, \end{aligned} \tag{B6}$$

where $R_s(\Delta x,\Delta z)$ is the autocorrelation of the weighting function, s, and a function of the difference coordinates in the range Δz and cross-range Δx directions. The complex coherence factor for echo magnitude, $\rho_X(\Delta x,\Delta z)$,[23] summarizes the correlation properties of rf echo signals, X, and is computed from the autocovariance function normalized by its value at zero, i.e., $\rho_X(\Delta x,\Delta z)=C_X(\Delta x,\Delta z)/C_X(0,0)$, where $C_X(\Delta x,\Delta z)\equiv R_X(\Delta x,\Delta z)-\langle X\rangle^2$.[4] We are

assuming that X is a random process that is stationary to second order.

In general, M is a complicated function of the target shape and resolution properties of the imaging system. For the purposes of studying low-contrast detectability, a very useful approximation is obtained in the limit where the target area is "large" compared to the area of a speckle spot. In that limit, $R(\Delta x,\Delta z)\simeq R(0,0)$ for all non-negligible values of $\rho_X(\Delta x,\Delta z)$, and Eq. (B6) reduces to[23]

$$M\simeq S_t/S_c, \tag{B7}$$

where

$$S_t=\frac{S^2}{R_s(0,0)}=\frac{S^2}{\int_{-\infty}^{\infty}s^2(x,z)dx\,dz}, \tag{B8}$$

$$S_c=\int_{-\infty}^{\infty}|\rho_X(\Delta x,\Delta z)|^2\,d\Delta x\,d\Delta z. \tag{B9}$$

S_t may be interpreted as an *effective* target area. When the target is a disk of amplitude one and diameter a, then $S_t=S=\pi a^2/4$. Goodman found for square targets that the exact expression for M, Eq. (B6), when plotted as a function of the number of speckle spots per target area S_t/S_c, gave approximately the same result for (Gaussian-apodized) circular and (uniformly weighted) square detector apertures. This suggests that some sidelobe energy in the ultrasound beam does not invalidate the use of Eq. (B7). Although Eq. (B7) does breaks down in the near field of the transducer, because adjacent speckle spots are highly correlated, and therefore not independent samples of the image. Equation (B7) is a reasonable approximation in the focal region of the transducer.

In addition, the number of speckle spots underestimates the number of independent samples, so that Eq. (B7) is a biased estimate. For example, S_t/S_c values of 1, 5, 10, 20, 100 produce the following values for M using the exact expression [Ref. 23, Eq. (2.121)]: 2.1, 6.8, 12.4, 23.2, 106.7. To further illustrate the bias, note that a target region much smaller than a speckle spot, i.e., $S_t/S_c\ll 1$, still provides one legitimate sample of the data to the detector. We consider Eq. (B7) a valid approximation when there are 10 or more speckle spots per target, since the bias will be less than 20%.

In our B-mode image simulations, we used a 2-D Gaussian point spread function (psf) given by the equation

$$g(x,z)=\frac{1}{2\pi\sigma_x\sigma_z}\exp\left[-\left(\frac{x^2}{2\sigma_x^2}+\frac{z^2}{2\sigma_z^2}\right)\right]. \tag{B10}$$

It is straightforward to show that for this psf,

$$\rho_X(\Delta x,\Delta z)=\exp\left[-\left(\frac{\Delta x^2}{4\sigma_x^2}+\frac{\Delta z^2}{4\sigma_z^2}\right)\right]$$

and that Eq. (B9) gives

$$S_c=2\pi\sigma_x\sigma_z. \tag{B11}$$

Finally, for disk targets of diameter a, Eq. (B7) yields

$$M\simeq\frac{a^2}{8\sigma_x\sigma_z},\ \text{where } S_t/S_c\geqslant 10. \tag{B12}$$

APPENDIX C

In this section we summarize relationships among the various measures of detectability for two-alternative forced-choice (2AFC) and yes–no (YN) observer experiments.

Both experiments use images with and without targets that are specified exactly, and observers are asked to decide between two hypotheses, H_1 and H_2. In a YN experiment, an observer is presented with one image and asked to decide if the target is absent (H_1) or present (H_2). In a 2AFC experiment, an observer is presented with two images, where one contains a target and one does not, and asked to decide if the target is in image 1 (H_1) or image 2 (H_2). It is important to note that the statistics for the two experiments are different.

First, we examine the YN paradigm. The decision function of the ideal observer D, is found from the likelihood ratio.[2] Presenting the decision maker with many images for both hypotheses and plotting a histogram of the results, we obtain estimates of the probability density functions $p(D|H_1)$ and $p(D|H_2)$, which are often accurately represented by Gaussian functions, as illustrated in Fig. C1. If the task is sufficiently challenging, the pdfs overlap and measureable decision errors result. To assess the observer's performance, we study changes in the true-positive fraction (TPF) of responses as we vary the false-positive fraction (FPF). TPF, FPF pairs are measured according to the following equations once we have selected a decision threshold, D_0:

$$\begin{aligned}\text{TPF}&=\int_{D_0}^{\infty}p(D|H_2)dD=\Phi[(\bar{D}_2-D_0)/\sigma_{D_2}],\\ \text{FPF}&=\int_{D_0}^{\infty}p(D|H_1)dD=\Phi[(\bar{D}_1-D_0)/\sigma_{D_1}],\end{aligned} \tag{C1}$$

where $\Phi(z)$ is the integral of the standard Gaussian random variable: $\Phi(z)\equiv 1/\sqrt{2\pi}\int_{-\infty}^{z}\exp(-t^2/2)\,dt$. Observer performance is entirely specified by the receiver operating characteristic (ROC) curve for the experiment. The ROC curve is a plot of (TPF,FPF) pairs measured for all possible D_0. An important summary measure of performance is the area under the ROC curve:

$$\text{AUC}=\int_0^1\text{TPF}\,d\text{FPF}. \tag{C2}$$

If $p(D|H_1)$ and $p(D|H_2)$ are non-Gaussian, then AUC is our performance metric, and detection efficiency is estimated as follows. First, select a task as defined by SNR_I, and find the average AUC for human observers. Then reduce the visibility of the target, i.e., reduce SNR_I, until the AUC for the ideal observer equals that of the average human. The ratio of the SNR_I^2 values is the detection efficiency.[6]

Fortunately, the Gaussian assumption simplifies performance evaluation significantly. Performance is entirely characterized by the means and variances of the decision function pdfs (see Table CI) using the equation for detectability

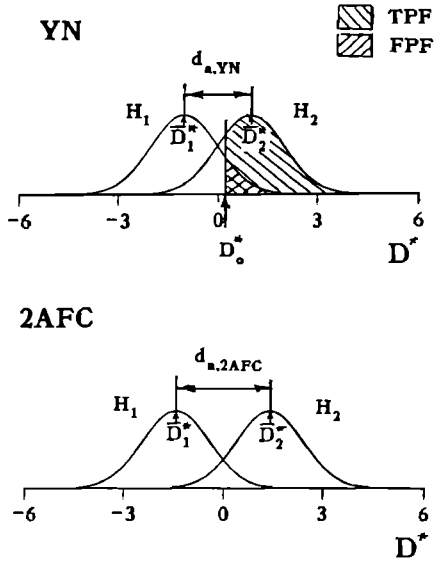

FIG. C1. Conditional probability density functions of equal variance for a binary discrimination task are illustrated for (top) the yes–no (YN) experiment where $d_{a,YN} = 2$ and (bottom) the analogous two-alternative forced-choice (2AFC) experiment where $d_{a,2AFC} = 2\sqrt{2}$. For convenience, the decision axes have been normalized by the standard deviation of the corresponding pdfs, e.g., $D^* \equiv D/\sigma_D$ and therefore $d_a = \bar{D}_1^* - \bar{D}_2^*$.

$$d_a = \frac{\sqrt{2}(\bar{D}_1 - \bar{D}_2)}{(\sigma_{D1}^2 + \sigma_{D2}^2)^{1/2}}. \tag{C3}$$

In addition, if $\sigma_{D1}^2 = \sigma_{D2}^2 \equiv \sigma^2$, then d_a is equal to another popular detectability index, $d' = (\bar{D}_1 - \bar{D}_2)/\sigma$. The relationship between d_a and AUC is given by Swets[10] as

$$\mathrm{AUC}_{YN} = \Phi\left[\frac{d_{a,YN}}{\sqrt{2}}\right]. \tag{C4}$$

The same equation holds for d' in place of d_a.

The analysis of 2AFC experimental data is similar, although each decision is based on the observation of two images. Consequently, the means and variances of the decision function are each twice that of the YN case (Ref. 2 Sec. 3.2.4), and therefore

$$d_{a,2AFC} = \sqrt{2} d_{a,YN}. \tag{C5}$$

Distinctions between the decision functions statistics using the YN and 2AFC experimental methods for equivalent sets of observer data are illustrated graphically in Fig. C1 and quantitatively in Table CI.

Another important relationship between YN and 2AFC methods is a theorem that states: The probability of a correct response in a 2AFC experiment, $P(C)$, equals the area under the ROC curve in a YN experiment [Ref. 2, Eq. (2.11)],

$$P(C)_{2AFC} = \mathrm{AUC}_{YN}, \tag{C6}$$

where we estimate $P(C)_{2AFC}$ by the fraction of correct responses resulting from a 2AFC experiment, $\hat{P}(C)$. Combining Eqs. (C4)–(C6) we find that

$$d_{a,2AFC} = 2\Phi^{-1}[\hat{P}(C)_{2AFC}], \tag{C7}$$

where $z = \Phi^{-1}[\Phi(z)]$.

Following Simpson and Fitter,[24] Hanson[20] shows that the uncertainty in d_a estimates is given by

$$\sigma_{d_a} = \sqrt{4\pi}\ \sigma_{\hat{P}(C)} \exp\left[\left(\frac{d_a}{2}\right)^2\right], \tag{C8}$$

where the nature of the 2AFC experiment dictates that $\hat{P}(C)$ estimates are binomial random variables with standard error

$$\sigma_{\hat{P}(C)} = \{P(C)[1 - P(C)]/N_p\}^{1/2} \tag{C9}$$

and N_p is the number of image pairs per 2AFC experiment.

Finally, we compute the uncertainty in d_a^2 using the approximation $\sigma_{d_a^2} \simeq 2 d_a \sigma_{d_a}$. Here, $\sigma_{d_a^2}$ depends on $P(C)$ and N_p. For any fixed N_p, the smallest uncertainty is obtained near 80% correct ($d_a = 1.7$). Experiments can be designed to minimize uncertainty by adjusting the difficulty of the task so that approximately 80% correct is achieved, and then increasing the number of image pairs to obtain a set uncertainty. If the observer task is too easy

TABLE CI. A comparison of decision function statistics for yes–no (YN) and two-alternative forced-choice (2AFC) observer experiments.

	YN (one image)	2AFC (two images)
Hypotheses	H_1: target absent H_2: target present	H_1: target present in image 1 H_2: target present in image 2
Optimal decision function	$D = \Sigma_{i=1}^{M} I_i$ $(I_i \equiv y_i^2)$	$D = \Sigma_i^M {}_1(I_{1i} - I_{2i})$ $(I_{1i} \equiv y_{1i}^2)$ and $(I_{2i} \equiv y_{2i}^2)$
Means: $\bar{D}_1 \equiv \langle D \mid H_1 \rangle$ $\bar{D}_2 \equiv \langle D \mid H_2 \rangle$	 I_1 I_2	 $I_1 - I_2$ $I_2 - I_1$
Variances: $\sigma_{D1}^2 \equiv \langle (D \mid H_1 - \bar{D}_1)^2 \rangle$ $\sigma_{D2}^2 \equiv \langle (D \mid H_2 - \bar{D}_2)^2 \rangle$	 σ^2 σ^2	 $2\sigma^2$ $2\sigma^2$
Detectability index, d_a	$d_{a,YN} = (I_1 - I_2)/\sigma$	$d_{a,2AFC} = 2(I_1 - I_2)/\sqrt{2}\sigma = \sqrt{2} d_{a,YN}$

$[P(C)=1.0]$ or too hard $[P(C)=0.5]$ then the percent uncertainties become very large for any value of N_p (see Fig. 2).

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

[2] D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics* (Peninsula, Los Altos, CA, 1988).

[3] A. D. Whalen, *Detection of Signals in Noise* (Academic, New York, 1971).

[4] R. F. Wagner, S. W. Smith, J. M. Sandrik, and H. Lopez, "Statistics of speckle in ultrasound B-scans," IEEE Trans. Sonics Ultrason SU-**30**, 156–163 (1983).

[5] S. W. Smith, R. F. Wagner, J. M. Sandrik, and H. Lopez, "Low contrast detectability and contrast/detail analysis in medical ultrasound," IEEE Trans. Son. Ultrason. SU-**30**, 164–173 (1983).

[6] W. P. Tanner and T. G. Birdsall, "Definitions of d' and η as psychophysical measures," in *Signal Detection and Recognition by Human Observers: Contemporary Readings*, edited by J. A. Swets (Wiley, New York, 1964), pp. 147–163.

[7] J. A. Swets and R. M. Pickett, *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory* (Academic, New York, 1982), pp. 80–83.

[8] R. F. Wagner, M. F. Insana, and D. G. Brown, "Statistical properties of radio-frequency and envelope-detected signals with applications to medical ultrasound," J. Opt. Soc. Am. A **4**, 910–922 (1987).

[9] D. Middleton, *An Introduction to Statistical Communication Theory* (McGraw-Hill, New York, 1960), p. 415.

[10] J. A. Swets, "Indices of discrimination or diagnostic accuracy: their ROCs and implied models," Psychol. Bull. **99**, 100–117 (1986).

[11] S. W. Smith, G. E. Trahey, S. M. Hubbard, and R. F. Wagner, "Properties of acoustic speckle in the presence of phase aberration part II: correlation length," Ultrason. Imag. **10**, 29–51 (1988).

[12] G. E. Trahey, J. W. Allison, S. W. Smith, and O. T. vonRamm, "A quantitative approach to speckle reduction via frequency compounding," Ultrason. Imag. **8**, 151–164 (1986).

[13] The SNR_I^2 values computed in this paper are four times that reported by Smith *et al*[5] Eq. (36). One factor of 2 results from the difference between the yes–no experimental method that they used and the 2AFC method that we used (see Appendix C). Another factor of 2 arises from the definition of SNR_I^2. Their definition [Ref. 5, Eqs. (31)–(33)] includes the *total* variance $(\sigma^2_{D|H_1}+\sigma^2_{D|H_2})$, whereas we have adopted the more usual definition by using the *common* variance, $\frac{1}{2}(\sigma^2_{D|H_1}+\sigma^2_{D|H_2})$. Since d_a^2 and SNR_I^2 are each affected by the choice of convention and experimental design, detection efficiency estimates from the two analysis will be equal.

[14] R. F. Wagner and D. G. Brown, "Unified SNR analysis of medical imaging systems," Phys. Med. Biol. **30**, 489–518 (1985).

[15] 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).

[16] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing* (Cambridge U.P., Cambridge, 1986), pp. 504–509.

[17] A. E. Burgess, R. F. Wagner, R. J. Jennings, and H. B. Barlow, "Efficiency of human visual signal discrimination," Science **214**, 93–94 (1981).

[18] P. F. Judy, R. G. Swensson, and M. Kijewski, "Observer efficiency and feature polarity," Proc. SPIE **767**, 310–316 (1987).

[19] T. J. Hall, M. F. Insana, N. M. Soller, and L. A. Harrison, "Ultrasound contrast-detail analysis: a preliminary study in human observer performance," Med. Phys. **20**, 117–128 (1993).

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

[21] H. Lopez, M. H. Loew, and D. J. Goodenough, "Objective analysis of ultrasound images by use of a computational observer," IEEE Trans. Med. Imag. **11**, 496–506 (1992).

[22] There is no explicit whitening step included in Eq. (2). However, when the target diameter is large compared to the speckle correlation area, that is, for $M>10$, the noise power spectrum is approximately white in the range of frequencies that define the detection task [Ref. 14, Eq. (4)]. Therefore, one practical approximation to the prewhitening matched filter of Eq. (A8) is to confine the experiments to the detection of large targets and use the simple matched filter.

[23] J. W. Goodman, "Statistical properties of laser speckle patterns," in *Laser Speckle and Related Phenomena*, Topics in Applied Physics, Vol. 9, edited by J. C. Dainty (Springer-Verlag, Berlin, 1975), pp. 46–51.

[24] A. J. Simpson and M. J. Fitter, "What is the best index of detectability?" Psychol. Bull. **80**, 481–488 (1973).