Ultrasound in Med. & Biol., Vol. 31, No. 3, pp. 351–359, 2005

0301-5629/05/$–see front matter

ELSEVIER

doi:10.1016/j.ultrasmedbio.2004.12.003

● *Original Contribution*

ASSOCIATION OF AUTOMATED AND HUMAN OBSERVER LESION DETECTING ABILITY USING PHANTOMS

JAMES M. KOFLER, JR.,* MARY J. LINDSTROM,† FREDERICK KELCZ,‡ ERNEST L. MADSEN§
*Department of Radiology, Mayo Clinic, Rochester, MN, USA and Departments of †Biostatistics and Medical Informatics, ‡Radiology and §Medical Physics, University of Wisconsin, Madison, WI, USA

(*Received* 23 *February* 2004; *revised* 24 *November* 2004; *in final form* 2 *December* 2004)

Abstract—A set of tissue-mimicking phantoms containing spherical negative contrast simulated lesions was employed to associate an automated method for determining detectability with human observers. Six alternative methods for computing the lesion signal-to-noise ratio (LSNR) were employed for quantifying automated detecting ability. The six methods differ regarding effective lesion area and whether or not gradients in local mean background echo levels were accounted for. The two-alternative-forced-choice (TAFC) technique was used to associate detecting ability of human observers with LSNR values. Although the six methods gave similar results, one method exhibited the least dependency on lesion diameter and is recommended; that method accounts for gradients in local mean background echo levels and employs an effective sphere area of $2/\pi$ times the projected sphere area. A reasonable LSNR detection threshold value of −2.0 was found to apply for nominal transducer frequencies from 4 through 6 MHz and for lesion diameters from 2 through 5 mm. This result allows rapid human-observer-calibrated automated determination of the depth range of detectability as a function of sphere diameter and contrast. (E-mail: elmadsen@wisc.edu) © **2005 World Federation for Ultrasound in Medicine & Biology.**

Key Words: **Automate, Detectability, Phantoms, Human observers.**

INTRODUCTION

Studies have been done to assess detectability by human observers of simulated lesions in ultrasound phantoms as a function of object contrast[1] and lesion size (Smith and Lopez 1982; Lopez et al. 1990; Hall et al. 1993). All of these studies consider the detectability of cylindrical targets in which the axes of the cylinders (or slightly tapered cones) are perpendicular to the scan plane (plane of symmetry of the scan slice). Two of these studies involve a single realization, or a small number of realizations, of each target of a given diameter and object contrast (Smith and Lopez 1982; Lopez et al. 1990). The observer knows that there is a target present and states whether it is detectable or not; thus, there is no possibility that there is no target present. The same situation exists in the work described by Hall et al. (1993) except that many independent realizations can be produced by the observer for each target of a given diameter and object contrast; thus, the effect on detectability of speckle variations in background and cylinder are suppressed.

Our work differs considerably from the earlier work described above in at least three ways:

1. Targets are spheres instead of cylinders perpendicular to the scan plane; thus, 3-D focal lesions are simulated with full account made for the effect of scan slice width.
2. A two-alternative-forced-choice technique is employed; therefore, the observer only knows that there is a lesion present in one of two images.
3. The aim is to associate values of lesion signal-to-noise ratio (LSNR, defined below) with detectability and to deduce a human observer detection threshold value for the lesion signal-to-noise ratio.

Address correspondence to: Ernest L. Madsen, PhD, Department of Medical Physics, University of Wisconsin-Madison, 1530 Medical Sciences Center, 1300 University Avenue, Madison, WI 53706 USA. E-mail: elmadsen@wisc.edu

[1] Object contrast in decibels is defined as 10 $\log_{10}[BSC_{obj}/BSC_{ref}]$, where BSC_{obj} is the backscatter coefficient of the type of material composing the object and BSC_{ref} is the backscatter coefficient of the type of material composing the background, or surroundings, of the object. In these phantoms the object is a sphere.

In addition, large numbers of image pairs are involved in the two-alternative-forced-choice process.

Ultrasonically tissue-mimicking (TM) phantoms with regular arrays of spherical simulated lesions have been reported, along with robust software for computing detectability (Kofler and Madsen 2001). Each sphere in an array has the same diameter and object contrast. The lesion centers are coplanar and a transducer-positioning device allows accurate superimposing of the scan plane on the sphere plane (plane containing the sphere centers). A digitized image is obtained with this transducer position, along with an image of the background material only; the latter image is used as a detectability reference. Using the known regular positions of the spheres, the software automatically locates the centers of all lesions in the lesion image, whether the lesion is detectable or not. Then the software computes the lesion signal-to-noise ratio (LSNR) at each lesion position.

The LSNR is defined (Wagner and Brown. 1985; Lopez et al. 1992; Rownd et al. 1997) as

$$\mathrm{LSNR} \equiv \frac{(S_L - S_B)}{\left[(1/2)(\sigma_B^2 + \sigma_L^2)\right]^{1/2}} \quad (1)$$

where S_L is the average pixel value over the area corresponding to the lesion, S_B is the average pixel value over an equal area of "background" surrounding the lesion, σ_B is the standard deviation of a sufficiently large number of independent realizations of S_B and σ_L is the standard deviation of a sufficiently large number of independent realizations of the target. Note that a better representation of S_B in the numerator is to average pixels over a larger area than that of S_L.

A reasonable approximation of the LSNR value when the lesions are borderline detectable is obtained by assuming $\sigma_L \approx \sigma_B$ (Lopez et al. 1992; Rownd et al. 1997). The approximation is

$$\mathrm{LSNR} \approx \frac{(S_L - S_B)}{\sigma_B}. \quad (2)$$

This approximation for LSNR is employed in this work.

The problem remains to establish realistic threshold values of LSNR for automatic determination of the detectability of a lesion and to determine any threshold dependencies on lesion size, object contrast or instrumental factors (such as frequency). We have chosen to address these problems by letting human-observer-detecting ability be the reference standard using the two-alternative-forced-choice methodology. Only negative contrast lesions were investigated.

Calibration of LSNR values to human observers

Two-alternative forced-choice (TAFC) analysis is a means by which a human sensory process, such as visual detectability, can be quantified (Green and Swets 1996). For visual detectability experiments, a TAFC study consists of a large number of paired images, each pair consisting of a "signal image," which contains the "signal" (such as a lesion representation) plus noise (surrounding background), and a noise-only image. The paired images are presented side-by-side or successively, with the location of the signal within the signal image having been randomized. The ordering of the signal image and noise-only image must be randomly presented for each image pair and visual cues that could bias the observers must not be present. The images are reviewed by requiring ("forcing") observers to specify the signal image for each image pair. The observers may be given some information regarding the nature of the object, such as its shape or size.

The results of a TAFC study yield the fraction of correct observer choices as a function of a measurable parameter, such as lesion signal-to-noise ratio. Thereby, an association is made between a quantifiable parameter (LSNR) and human detecting ability.

The purpose of this study is to associate LSNR values with the fraction of correct choices of human observers for spherical lesion ultrasound images. Thus, association of a human observer "fraction correct" with threshold detectability also establishes the LSNR threshold. Results obtained assuming circular and square signal (target) shapes of various sizes in computing LSNR values are compared. To account for effects of gradients in local mean pixel values of the background on detectability, a gradient-corrected standard deviation noise component for the LSNR computation is also investigated. Additionally, detectability thresholds are investigated with regard to sphere size and transducer nominal frequency.

PHANTOMS

The phantoms employed have been described previously (Kofler and Madsen 2001). The geometry of the phantoms is shown in Fig 1. There are two parallel equivalent scanning surfaces on opposite ends of the phantom that are covered with a 25-μm-thick scanning window. The vertical surfaces of the phantom material are bounded with 1-cm-thick acrylic walls. All spheres in any one phantom have the same diameter and object contrast. The available sphere diameters are 2, 3, 4 and 5 mm with object contrasts of -3, -6, -9 and -14 dB.

IMPLEMENTATION OF LSNR CALCULATION

Determination of the sphere matrix location is done using an automated technique (Kofler and Madsen 2001), assuring that the position of the center of each simulated lesion is known. Six methods for computing the LSNR

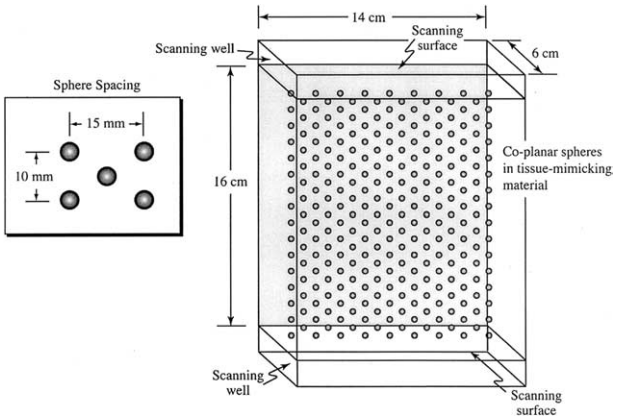


Fig. 1. A schematic of a spherical lesion phantom. Note that each phantom contains simulated lesions that are identical in size and object contrast.

were considered (Fig. 2). The six methods differ regarding the effective lesion signal area and/or the method of calculating the noise (denominator in eqn (2)).

Effective lesion signal area

For cross-sectional images of spherical target objects, the signal component S_L in the matched-filter method (Wagner and Brown 1985) considers a circular region of diameter equal to the projected target diameter. A weighting factor is applied as a function of radial distance from the center of the projected sphere target. For ultrasound images, derivation of an accurate weighting factor is impractical, because it requires a prior knowledge of the ultrasonic intensity distribution in the slice thickness (elevational) direction. However, because of the partial volume effect, a weighting factor might be important. One way to apply a weighting factor is to reduce the area corresponding to the target, referred to as the effective area. Thus, a circular target area that is smaller than the projected target area with a weighting factor of one inside and zero outside might be employed (methods 3 and 6 in Fig. 2). A small square area, which simplifies computational effort, was also used to approximate a weighting function (methods 2 and 5). Methods 1 and 4 employ a circular region with diameter equal to that of the projected target, thus ignoring the partial volume effect.

Area used in computing background signal S_B

S_B in the numerator of eqn (2) equals the mean pixel value over a 1.5 cm square area centered at the target of concern with the effective target area eliminated (see Fig. 2). Note that the image employed is that in which the scan plane is superimposed on the target centers.

Noise calculation using matched filter methods

The matched filter method for computing σ_B in eqn (2) has been implemented by computing the standard deviation of N samples $S_{B,n}$ ($n = 1,2,\ldots,N$) over a 1.5-cm × 1.5-cm area, the latter centered at the coordinates of the target of concern. The 1.5-cm × 1.5-cm area is in a "background only" image. $S_{B,n}$ is the nth sample and is equal to the average pixel value over the effective target area, A_{eff}. These effective target areas are centered at points in a square lattice, with 1-mm nearest neighbor distance, covering the1.5-cm × 1.5-cm square. Since overlap of the $S_{B,n}$ areas occurs, the number of *independent* $S_{B,n}$ values is less than N being, however, at least equal to 1.5-cm × 1.5-cm/A_{eff}. Thus,

$$\sigma_B = \left[[1/(N-1)] \sum_1^N (S_{B,n} - \overline{S_{B,n}})^2 \right]^{1/2}$$

$$\text{where } \overline{S_{B,n}} = (1/N) \sum_1^N S_{B,n}.$$

To increase the accuracy of σ_B (*i.e.*, to reduce the difference between σ_B and the standard deviation of the parent distribution), σ_B is computed for four independent background images and the root-mean-square value taken to be the employed value of σ_B. The final value of σ_B is referred to in this work as the ordinary standard deviation (OSD).

Noise calculation modification to account for background gradients

The matched-filter method assumes signal detection in the presence of isotropic background noise and consequently does not address gradients in *local* mean pixel values in the background (Wagner and Brown 1985).

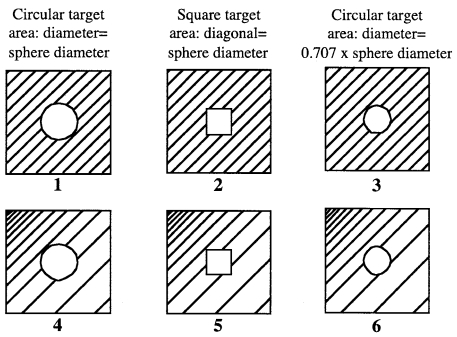


Fig. 2. Illustration of the variations used to calculate LSNR values. The circular image target area in the left column (numbers 1 and 4) equals the projected area of the spherical lesion. Uniform cross-hatching (1–3) corresponds to the area used to compute the noise of the denominator of eqn (2) using the ordinary standard deviation. The graded cross-hatching (4–6) corresponds to the area used to compute the gradient-corrected standard deviation in the denominator of eqn (6).

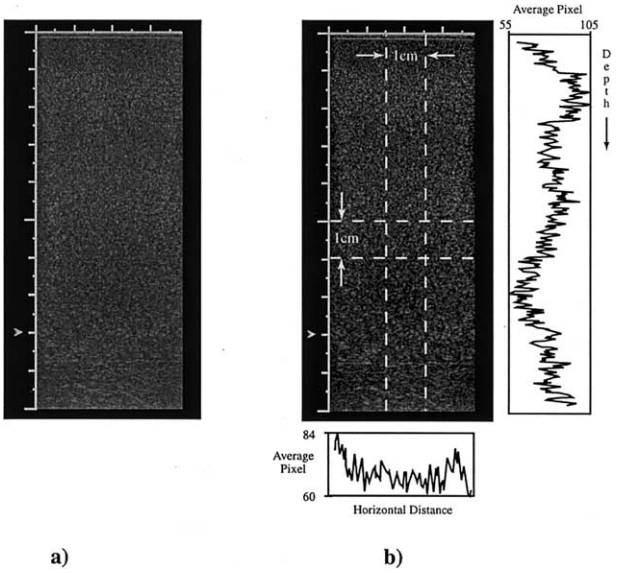


Fig. 3. Depiction of subtle gradients in local mean pixel values. a) Background-material-only image. b) Same image as in a), showing the pixel value profiles in the vertical and horizontal directions; each point in the graph to the right of the image equals the average value of the (horizontal) row of pixels at that depth over the 1-cm length between the vertical dashed lines; the graph below the image is made in the same fashion, relative to the horizontal dashed lines.

Such local mean pixel gradients (Fig. 3) are rather common in ultrasound images, especially sector formats, and it is reasonable to assume that human observers account for this mean gray level gradient in the process of detecting spheres, whereas the OSD noise would be biased too large. Additionally, subtle gradients may not be visually obvious, yet do influence the calculated noise value.

A potential improvement for implementing the matched-filter method accounts for gradients in local mean pixel value (mean grey level) over the 1.5-cm × 1.5-cm area used in computation. Thus, a gradient-compensating method for computing the noise denominator of LSNRs has been implemented (Croxton 1953) in methods 4 to 6 (see Fig. 2). This "noise" is referred to as the gradient-compensated standard deviation, $\sigma_{B,GC}$, and is defined as

$$\sigma_{B,GC} = \left[\frac{\sum (S_{B,n} - S_{B,nc})^2}{N}\right]^{1/2}, \quad (3)$$

where N and $S_{B,n}$ are defined above. The variable $S_{B,nc}$ is given by

$$S_{B,nc} = a + bX_n + cY_n, \quad (4)$$

where X_n and Y_n refer to the coordinates of noise element $S_{B,n}$ and the constant coefficients are determined *via* a 2-D least-squares fit.

$$a = \frac{A[G^2 - F^2] - B[DG - FE] + B[DF - GE]}{W},$$

$$b = \frac{N[DG - FE] - AB[G - F] + B^2[E - D]}{W},$$

$$c = \frac{N[GE - DF] - B^2[E - D] + AB[F - G]}{W},$$

where

$$W = N\left[G^2 - F^2\right] + 2B^2[F - G],$$

and

$$A = \sum_{n=1}^{N} S_{B,n}, \quad B = \sum_{n=1}^{N} X_n, \quad D = \sum_{n=1}^{N} S_{B,n}X_n,$$
$$E = \sum_{n=1}^{N} S_{B,n}Y_n, \quad F = \sum_{n=1}^{N} X_nY_n, \quad G = \sum_{n=1}^{N} X_n^2. \quad (5)$$

The gradient-compensated LSNR (GCLSNR) value is

$$GCLSNR \equiv \frac{S_L - S_B}{\sigma_{B,GC}}. \quad (6)$$

To increase the accuracy of $\sigma_{B,GC}$, a value of $\sigma_{B,GC}$ is computed over the 1.5-cm × 1.5-cm area centered at the target position in four separate background (no targets) images and the $\sigma_{B,GC}$ value employed is the rms value of the four.

TWO-ALTERNATIVE-FORCED-CHOICE STUDY

As shown in Table 1, seven phantoms, each unique regarding lesion diameter and/or object contrast, were imaged using three different transducers, a 6 MHz linear probe, a 5 MHz curvi-linear probe and a 4 MHz sector probe (Acuson models 6L3, 5C2 and 4V1, respectively), using an Acuson Sequoia scanner. For each probe, one set of images was obtained using a focal depth of 2 cm and another set, using a focal depth of 8 cm, yielding a range of sphere image qualities. For computing the "noise," σ_B in eqn (2), or $\sigma_{B,GC}$ in eqn (6), four corresponding images of the phantom background material with no sphere in the scan slice were also acquired for each transducer and focal depth combination. Time-gain compensation controls were adjusted to provide a visually uniform image background. The control settings were identical for each set of images (target plus background-only images). The phantom images were acquired digitally directly from the scanner and transferred to a personal computer for analysis *via* magneto optical

Table 1. The number of image pairs for each sphere type and transducer frequency

Sphere size	Sphere contrast	Transducer frequency	Number of image pairs	Sphere size	Sphere contrast	Transducer frequency	Number of image pairs
2 mm	−14 dB	6 MHz	168	4 mm	−14 dB	6 MHz	14
					−9 dB	6 MHz	21
3 mm	−14 dB	6 MHz	54		−14 dB	5 MHz	941
	−6 dB	6 MHz	60		−9 dB	5 MHz	743
					−14 dB	4 MHz	555
	−14 dB	5 MHz	735		−9 dB	4 MHz	536
	−9 dB	5 MHz	279			Total 4 mm pairs:	2810
	−6 dB	5 MHz	279				
	−14 dB	4 MHz	276	5 mm	−9 dB	5 MHz	304
	−9 dB	4 MHz	276		−9 dB	4 MHz	405
		Total 3 mm pairs:	1959			Total 5 mm pairs:	709
Total number of image pairs (all sphere sizes): 5646							

media. Then, the LSNR value was computed at each sphere location.

TAFC image pairs, such as those shown in Fig. 4, were generated from the digital images using sphere images near the threshold of detectability. For each sphere a 1-cm × 1-cm square image area was selected such that the sphere image was randomly located within its boundaries. The corresponding 1-cm × 1-cm noise-only image was obtained from an area on a background image of the same size and pixel coordinates as that used for the sphere image.

The signal and noise square-shaped images were separated by approximately 10 cm on the displayed pair-composite image. After randomizing whether the sphere image was on the left or right side, each image pair was annotated using dark gray text indicating the sphere size and an assigned random identification number. The criterion for inclusion of a sphere image in the study was that the LSNR value for method 4 (refer to Fig. 2) be between −1.0 and −6.0.

The explicit procedure by which human observers participated in the study was as follows. Image pairs were presented to the reviewer in random order on a computer display. With the exception of the pair-composite image and a small control panel window, the entire computer screen was black and room lighting was minimal. Observers were not restricted regarding eye-to-monitor distance, so that each observer could choose the distance to correspond to best perception. The observer clicked the mouse with the cursor on the choice for the signal image and the screen was made blank before the next image pair composite was displayed. The selected image was recorded *via* software. The control panel window allowed the observer to view previous pairs and change his or her response if desired. Reviewers used this option rarely and then only for a current image pair; *i.e.*, any choice change was made before the next image pair was addressed. A total of 5646 pair-composite images were reviewed in sets containing a maximum of 500 images to avoid reviewer fatigue. A total of three reviewers, each with a minimum of 6 y of ultrasound quality assurance experience, participated independently.

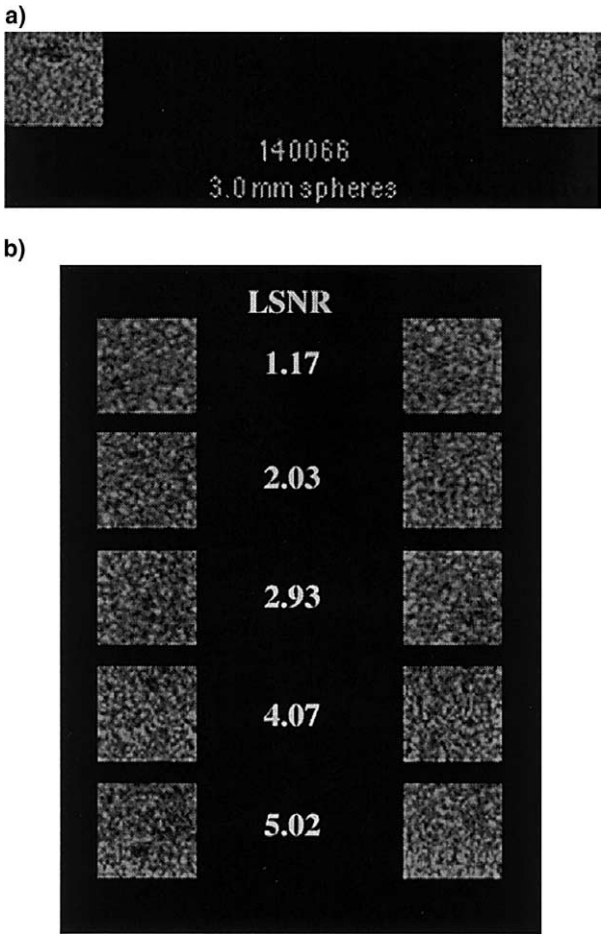


Fig. 4. a) Sample TAFC pair-composite. b) Sample images showing a range of LSNR values. The left-hand column of images contained spheres, whereas the right-hand column was background material only.

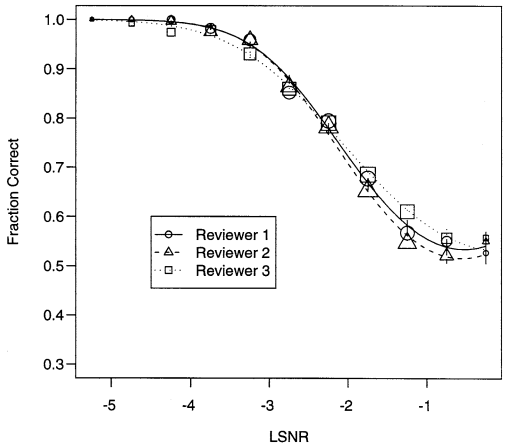


Fig. 5. The TAFC fraction of correct choices *vs.* LSNR ranges for each reviewer. All 5646 image pairs are represented for each reviewer. Note that the size of the squares, circles and triangles increases with the number of TAFC image pairs contributing. The curves result from logistic regression and the vertical bars are standard errors associated with the fitted curves.

Monitor considerations

Thijssen et al. (1988) concluded that detectability of low contrast targets is insensitive to gray level transform. Thus, monitor variability was not considered an issue in this study, particularly since both monitors were routinely used by persons experienced in viewing ultrasound images. Two monitors were employed. That used by reviewers A and B had contrast and brightness settings optimized using a SMPTE (Society of Motion Pictures and Television Engineers) pattern. Corroboration that monitor calibration is not a significant factor was provided by the fact that the TAFC results of observers A and C were more closely matched than those of A and B.

Monitor resolution was set at 2.83 pixels per mm, a value which adequately reproduced image texture in the focal region of the nominal 6 MHz transducer.

Statistical analysis

The effect of reviewer, sphere size, method and frequency on the fraction correct was assessed by using standard logistic regression (Nelder and McCullagh 1989). The logit is defined as $\log_e[p/(1-p)]$, where p is the fraction correct. The logit is modeled as a function of one or more of the following parameters: reviewer, sphere diameter, method of computing LSNR and frequency.

RESULTS AND DISCUSSION

Figure 5 shows the averaged TAFC study results for all 5646 image-pairs and for each of the three reviewers. The LSNR values were computed using method 5 (refer to Fig. 2) and the fitted curves result from logistic regression. The performances of all three reviewers were very similar, with reviewer 2 scoring lower for lower LSNR values. Results obtained for the three reviewers using the other five methods of computing LSNR were correspondingly very similar. Analysis that considers sphere size shows that differences in fraction correct among reviewers are statistically significant for all but the 2 mm sphere size but the differences in fraction correct are small. For example, for LSNR values between −1.5 and −2.5, the maximum difference in fraction correct is for the 5 mm spheres, and that is 0.046. Also, the ordering of reviewers by performance is not consistent among the sphere sizes, leading us to conclude that there is no systematic effect of reviewer. Subsequent analyses will pool data from all three reviewers.

The influence of lesion size on human detecting ability is addressed in Fig. 6. For each LSNR value, the TAFC fraction correct equals the mean for the three human observers. For each of the methods there is a significant effect of size ($p < 0.0001$). These differences are rather small, however. For LSNR values between −1.5 and −2.5, the maximum differences in fraction

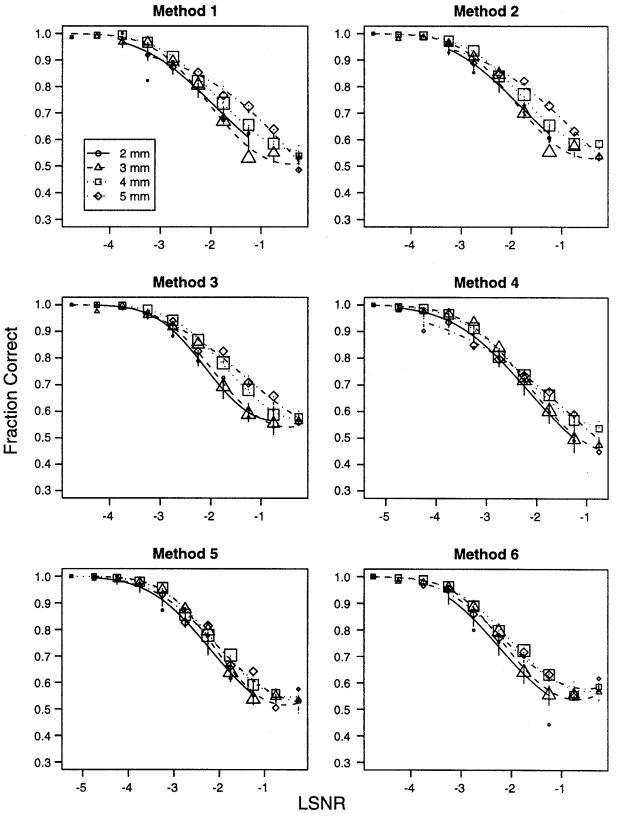
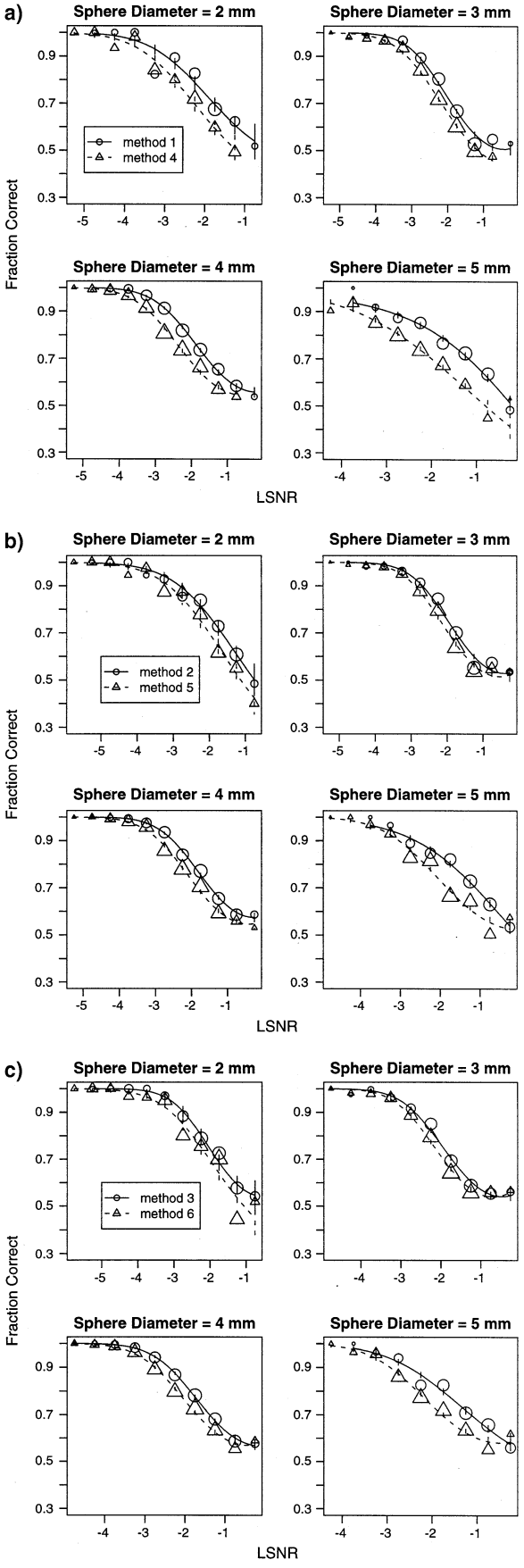


Fig. 6. TAFC results for the four sphere diameters and each of the six methods for computing the LSNR values. Logistic regression curves are shown for each sphere diameter.



correct are 0.138, 0.117 and 0.126 for methods 1, 2 and 3, respectively.

Methods 4, 5 and 6, where the noise $\sigma_{B,GC}$ was computed using the gradient-corrected standard deviation (GCSD), consistently show less dependence on sphere size compared with methods that use the ordinary standard deviation (OSD) (methods 1 to 3). For LSNR values between -1.5 and -2.5, the maximum differences in fraction correct are 0.076, 0.053 and 0.074 for methods 4, 5 and 6, respectively; these are about half the differences for methods 1, 2 and 3. This reduced dependency on sphere size supports the superiority of methods 4, 5 and 6 to methods 1, 2 and 3. Note that, of methods 4, 5 and 6, method 5 exhibits the least dependency on sphere size.

Note that, generally, $\sigma_{B,GC} < \sigma_B$ because of two factors. One is that there is a gradient in local mean pixel value in the image (and correction for such gradients is the purpose of the GCSD method). The other factor is a computational bias and results from the fact that the GCSD technique determines the fitted 2-D plane by, in effect, minimizing $\sigma_{B,GC}$, and the smaller the number of independent samples, $S_{B,n}$, the greater is the minimizing bias. The minimizing bias would result in greater differences between the OSD and GCSD methodologies as the sphere diameter increases. For all sphere sizes, the GCSD curves in Figs. 7 a), b) and c) are below the OSD curves, and the shift decreases from 5-mm to 4-mm spheres and from 4-mm to 3-mm spheres. However, the downward shift for the 2-mm sphere case is comparable with that for the 4-mm case. Thus, the minimizing bias does not appear to be of concern for diameters 2, 3 and 4 mm and is small enough to be ignored for the 5 mm diameter case.

Possible frequency dependence of detectability *vs.* LSNR values is addressed in Figs. 8 and 9. In Figure 8 are shown TAFC fractions correct *vs.* LSNR values for 3 mm diameter lesions. The size of the markers plotted (triangles, circles or squares) increases with number of TAFC pairs represented. The curves resulted from logistic regression. A reasonable detectability threshold for fraction correct is 0.70, corresponding to an LSNR value of about -2.0. For the logistic regression curves at LSNR $= -2.0$, the fractions correct range from 0.68

Fig. 7. The effect on TAFC results for each sphere diameter regarding whether the LSNR noise was taken to be σ_B (OSD) or $\sigma_{B,GC}$ (GCSD) for three different effective target areas used in computing LSNR values a) Target area equals the projected area of the sphere. b) Target area is a square with diagonal equal to the sphere diameter. c) Target area is a circle with diameter equal to $2^{-1/2}$ (0.707) times the sphere diameter.

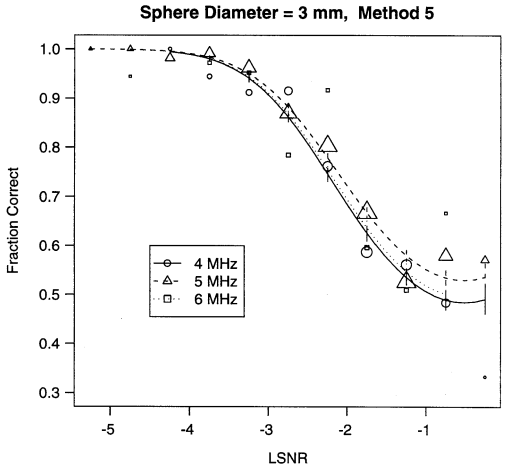


Fig. 8. TAFC curves for 3-mm diameter spheres corresponding to three different nominal frequencies; LSNR values were computed with method 5. The 6 MHz values exhibit considerable statistical variation, due to the limited numbers of TAFC image pairs available. Logistic regression curves are also shown.

through 0.73, with the curve for 4 MHz lying below the 5 and 6 MHz curves.

In Figure 9 are shown curves corresponding to those in Fig. 8 except TAFC pairs for all sphere sizes are included. At LSNR = −2.0, the range of fractions correct is nearly the same as that in Fig. 8, *viz.*, 0.70 to 0.74. In this case, however, the logistic regression curve for 6 MHz lies below the others.

The small and nearly identical ranges of the fraction correct in the reasonable threshold region in Figs. 8 and 9, plus the fact that there is no common ordering of the 4, 5 and 6 MHz logistic regression curves, suggests that there is negligible frequency dependence of detectability based on LSNR values.

Detectability threshold

Choosing a detectability threshold for the LSNR values is somewhat subjective. A reasonable threshold would correspond to a fraction correct in the TAFC curves of 0.7. Referring to Fig. 6 for methods 4 and 5, this fraction correct corresponds to an LSNR value of about −2.0. The corresponding threshold for method 6 has a slightly lower value, approximately −1.8. Notice that a fraction correct of 0.7 also corresponds to a threshold LSNR value of about −2.0 in Figs. 8 and 9.

SUMMARY AND DISCUSSION

An automated method for locating the positions of spherical lesions in the scan plane was employed and images were digitized with targets and without targets (background images). Sphere diameters from 2-mm through 5-mm were represented, along with a range of negative contrasts. A variety of scanner configurations was employed in the study and six methods for computing LSNR values were investigated.

Thousands of image pairs were assessed *via* TAFC by three human observers. Resulting graphs of fraction correct *vs.* LSNR values showed that, when the ordinary standard deviation (OSD) was used to compute the noise denominator in the LSNR, a dependence on sphere diameter was demonstrated. When the gradient-corrected standard deviation (GCSD) was used to compute that noise, negligible dependence on sphere diameter was observed for the three methods of weighting the target area. Since no sphere size (target size) dependence is expected in the LSNR methodology, the GCSD method of computing noise is considered to be superior to the OSD. Also, of methods 4, 5 and 6 (GCSD and different target area weighting), method 5 is most appealing for two reasons: first, the target area weighting is moderate, method 4 involving no weighting and method 6, perhaps too much weighting; second, the level of agreement between logistic regression curves in Fig. 6 is best for method 5.

Further analysis of the TAFC results provided evidence that there is negligible dependence of detectability on the nominal frequency of the transducer.

Regarding the relation of lesion detectability performance in phantoms to quality of clinical images, a manuscript addressing that topic is in preparation.

Practical uncertainty in determining LSNR values

When a spherical lesion phantom is employed in performance testing, it is important to know the practical uncertainty in determined LSNR values near the threshold of detectability. The detectability of a target

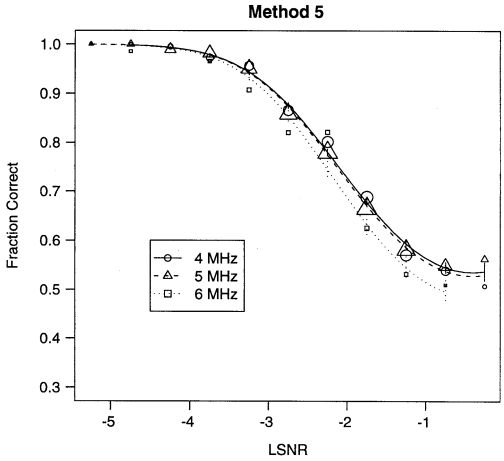


Fig. 9. TAFC results for all contributing spheres differentiated by nominal frequency. LSNR values were computed with method 5. Logistic regression curves are also shown.

is subject to the effect of random fluctuations in echo signals in materials with randomly-positioned scatterers. The fluctuations are due to the fact that the echo signal involving a given A-line and depth depends on interference between scattered echo signals from a finite number of randomly-positioned scatterers. Thus, detectability of a single target sphere can be enhanced or lessened, depending on the particular random configuration of scatterers in and around the target. To overcome the randomness aspect, an average of LSNR values for different spheres of the same diameter and object contrast and at the same depth should be taken as the LSNR value. A practical upper limit on the number of spheres involved in the averaging is about eight (Kofler and Madsen 2001). Assuming that the probability of detection is 0.7 (threshold) and using the binomial distribution function, the uncertainty in the LSNR value (averaged over eight eight realizations) will lie between 0.64 and 0.76. Referring to TAFC curves, such as in Fig. 8, the uncertainty in LSNR value is $\pm$ 0.26 (a spread of 0.52). Referring again to Fig. 8 as typical, the spread in LSNR values for logistic regression curves at a fraction correct of 0.70 is 0.14, which is about 1/4 of the practically determinable LSNR spread. Thus, any statistical significance of frequency dependence corresponding to the curves in Fig. 8 is made negligible by the practical limitation of LSNR determination. The same argument applies to the practical significance of sphere-size dependence of LSNR values (see Fig. 6, method 5); *i.e.*, in practice, dependence of LSNR values on sphere size in the threshold range is also negligible.

Recommendation

Although distinctions between the six LSNR computation methods are not great, method 5 has the edge as the best method and is recommended for computing LSNR values with an LSNR detectability threshold of -2.0, the latter corresponding to a human observer TAFC fraction correct of 0.7.

Acknowledgement—The research described in this manuscript was supported in part by NIH grant R42GM54377.

REFERENCES

Croxton FE. Elementary statistics with application in medicine and the biological sciences. New York: Dover, 1953:163–169.

Green DM, Swets JA. Signal detection theory and psychophysics. New York: John Wiley, 1966.

Hall TJ, Insana MF, Soller NM, Harrison LA. Ultrasound contrast-detail analysis: A preliminary study in human observer performance. Med Phys 1993;20(1):117–127.

Kofler JM, Madsen EL. Improved method for determining resolution zones in ultrasound phantoms with spherical simulated lesions. Ultrasound Med Biol 2001;27(12):1667–1676.

Lopez H, Loew MH, Butler PF, Hill MC, Allman RM. A clinical evaluation of contrast-detail analysis for ultrasound images. Med Phys 1990;17(1):48–57.

Lopez H, Loew MH, Goodenough DJ. Objective analysis of ultrasound images by use of a computational observer. IEEE Trans Med Imaging 1992;2(4):496–506.

Nelder JA, McCullagh P. Generalized linear models, Second Edition. Boca Raton: CRC press, 1989.

Rownd JJ, Madsen EL, Zagzebski JA, Frank GR, Dong F. Phantoms and automated system for testing resolution of ultrasound scanners. Ultrasound Med Biol 1997;23(2):245–260.

Smith SW, Lopez H. A contrast-detail analysis of diagnostic ultrasound imaging. Med Phys 1982;9(1):4–12.

Thijssen JM, Oosterveld BJ, Wagner RF. Gray level transforms and lesion detectability in echographic images. Ultrasonic Imaging 1988;10:171–195.

Wagner RF, Brown DG. Unified SNR analysis of medical imaging systems. Phys Med Biol 1985;30(6):489–518.