ELSEVIER

Ultrasound in Med. & Biol., Vol. 23, No. 2, pp. 245–260, 1997

0301-5629/97 $17.00 + .00

PII S0301-5629(96)00205-0

●*Original Contribution*

PHANTOMS AND AUTOMATED SYSTEM FOR TESTING THE RESOLUTION OF ULTRASOUND SCANNERS

JASON J. ROWND, ERNEST L. MADSEN, JAMES A. ZAGZEBSKI, GARY R. FRANK and FANG DONG
Department of Medical Physics, University of Wisconsin, Madison, WI, USA

(*Received* 26 *April* 1996; *in final form* 14 *October* 1996)

Abstract—Tissue-mimicking phantoms and an automated system have been developed for testing the resolution performance of ultrasound scanners by determining detectability of low to higher contrast spherical lesions over the entire depth of field. Axial, lateral and elevational resolutions are accounted for simultaneously and equally. Tissue-mimicking spherical simulated lesions are either 3 or 4 mm in diameter and have one of four different intrinsic material contrasts. For each diameter and contrast, there is a set of 109 lesions in a regular array with coplanar centers extending from 0.5–15.5 cm in depth. With the scan slice superimposed on the spheres, the image is frame-grabbed for automated analysis. A diameter-dependent lesion signal-to-noise ratio is computed for each pixel position in the image, excluding a 5-mm boundary. Two universal thresholds, resulting from maximization of agreement between the automated system and human observers, give rise to a depth range, or "resolution zone", over which detection exists for each type lesion. © 1997 World Federation for Ultrasound in Medicine & Biology.

Key Words: **Detectability, Spherical, Focal, Lesion, Contrast, Resolution, Automated, Phantom, Performance, Quality assurance, Ultrasound imaging, Testing.**

INTRODUCTION

The purpose of the work reported here was to develop phantoms and an automated method of image analysis so that low-through-high contrast† focal lesion detection performance of ultrasound imagers could be quickly and quantitatively determined over the entire depth range of the image. The term "focal lesion" implies a clumplike object having a boundary that is confined to the neighborhood of a point in three-dimensional (3D) space. Thus, detectability involves the slice width (or profile) of the ultrasound beam. Although separate measurements of axial and lateral resolutions and slice thickness are often performed on ultrasound scanners (AIUM Standards Committee 1990; Goldstein 1988), results of such measurements do not readily demonstrate whether a focal lesion of a particular contrast and diameter can be visualized.

Automated analysis of images of low-contrast cylinders, oriented perpendicular to the scan plane, has been reported (Lopez et al. 1992). That process involved acquiring a statistically significant number of independent images for each cylinder of given diameter and contrast. The number was usually obtained by maintaining the scan slice perpendicular to the cylinder and translating the scan slice along the axis of the cylinder. The use of the cylindrical target geometry is severely compromised, however, by the fact that the effect of elevational resolution (corresponding to slice thickness) is not accounted for. Thus, the analysis is only valid for assessing resolution regarding objects that possess invariance with respect to translation perpendicular to the scan plane. ("Scan plane" is defined here to mean the plane of symmetry of the scan slice.) Another shortcoming of the method is that obtaining resolution values over a range of depths was not built into the system.

Phantoms for testing the resolution of ultrasound scanners using spherical simulated small focal lesions have been previously reported (Madsen et al. 1991,

Address correspondence to: Dr. Ernest L. Madsen, Department of Medical Physics, University of Wisconsin, 1300 University Avenue, Room 1530, Madison, WI 53706, USA.

† The term contrast in this article refers to the intrinsic material contrast, sometimes called "object contrast," defined in terms of the backscatter coefficients of the spherical lesion material and the surrounding background material, *e.g.*, defining the backscatter coefficient of the material forming the lesions to be B_ℓ and that of the background material to be B_{bg}, the contrast is defined (in dB) as 10 $\log_{10}(B_\ell/B_{bg})$.

Table 1. Weight percents of component materials in the phantoms.

Tissue-mimicking material	Powdered graphite	45 to 53-μm diameter glass beads	Water	Dry agar	*n*-Propanol
Background (reference)	4.0	0.765	84.8	3.7	6.7
N1 spheres	4.6	0.191	84.8	3.7	6.7
N2 spheres	4.7	0.048	84.8	3.7	6.7
N3 spheres	4.8	0	84.8	3.7	6.7
P1 spheres	3.25	1.53	84.8	3.7	6.7

1994). Each section of these phantoms contains many spatially randomly distributed spherical simulated lesions (all with the same diameter and contrast) extending over a depth range from 0.5–15.5 cm. Human observers assess images for proximal and distal limits of detectability for lesions of each contrast and diameter; the depth range between these limits, where lesions are detectable, is called the "resolution zone" for that diameter and contrast lesion. Resolution zones depend on lesion contrast and size, the particular scanning equipment and the scanner console settings. For determining the resolution zone, a spatially random distribution of the spheres is preferred to a known simple distribution, such as one in which spheres have colinear centers and are all at the same distance from the scanning window; in the latter case, the regular spacing might influence detectability decisions of human observers.

For more extensive testing of resolution capabilities, it would be desirable to automate the detection process. This could lead to more objective evaluations of imagers and facilitation of routine quality assurance testing (Lopez et al. 1992). Although images using phantoms with simulated focal lesions randomly distributed in a volume could be digitized and analyzed by computer algorithm, a more efficient approach is to use phantoms in which simulated focal lesions of a given diameter and contrast have coplanar centers so that the plane of symmetry of the scan slice could be positioned to include the centers of all those lesions. A set of phantoms containing eight such layers of equivalent simulated focal lesions was produced. There is one layer for each of eight different combinations of diameter and contrast.

An automated analysis method was developed to compute the resolution zone for focal lesions. It involves rapid determination of detectability of spherical simulated focal lesions over the entire scan depth range. Thus, eventual rapid automated quality assurance is feasible by suitable choice of a limited number of combinations of sphere diameter and contrast.

A pattern recognition algorithm, called the "matched filter method" (Wagner et al. 1983, 1985), has been modified for use in our detectability algorithm. Thus, a lesion signal-to-noise ratio $(SNR)_\ell$, which depends on the expected size of the lesion searched for, is computed for each pair of pixel coordinates in digitized images of the phantoms. Detection exists if $(SNR)_\ell$ surpasses a threshold value. This threshold (one for positive contrast and one for negative contrast) was established by comparison with analyses of images by experienced human observers, since human observers ultimately interpret clinical ultrasound images for diagnostic purposes. Histograms of the number of pixels corresponding to lesion detection vs. depth then allow determination of the resolution zone corresponding to an established cutoff value on the histogram. The cutoff was again established by

Table 2. Ultrasonic properties of tissue-mimicking materials in the phantoms at 22°C.

Tissue-mimicking material	Speed (m/s)	Attenuation (dB cm^{-1} MHz^{-n})		Backscatter (sr^{-1} cm^{-1} MHz^{-m})		Relative backscatter coefficient at 5 MHz (dB)
		α_o	n	$\eta_o \times 10^5$	m	
Background (reference)	1538	0.461	1.07	1.68	3.49	0
N1 spheres	1540	0.445	1.10	0.344	3.69	−5.5
N2 spheres	1539	0.435	1.13	0.0947	3.76	−10.6
N3 spheres	1542	0.418	1.08	0.0121	4.16	−16.8
P1 spheres	1544	0.347	1.21	5.29	3.20	+3.1

The density of all five types of material is 1.04 g cm^{-3}.

The attenuation coefficients were fitted with $\alpha(f) = \alpha_o f^n$, where $\alpha(f)$ is the attenuation coefficient at frequency f and α_o and n are constants. Likewise, the backscatter coefficients were fitted to the relation $\eta(f) = \eta_o f^m$, where $\eta(f)$ is the backscatter coefficient at frequency f and η_o and m are constants. Backscatter coefficients at 5.0 MHz (relative to the background material) are given in the last column in dB.

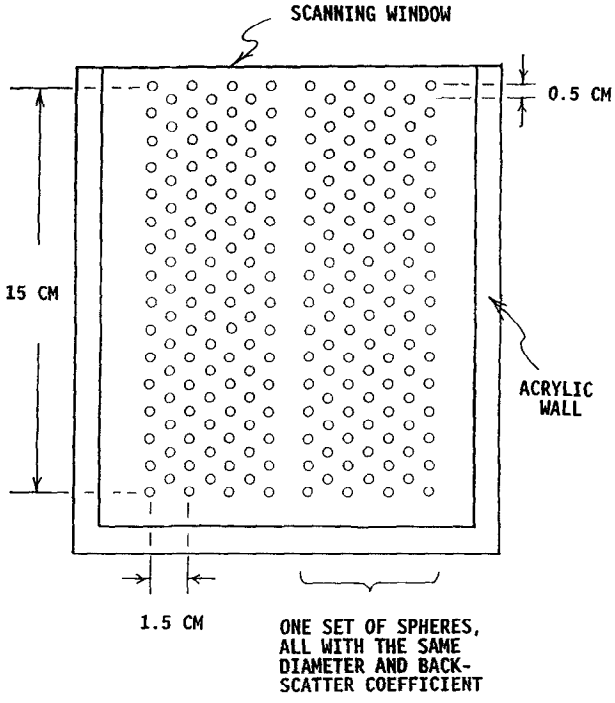


Fig. 1. Diagram of one of the four phantoms showing two sets of spheres with coplanar centers. The scanning window is 50-μm-thick Saran Wrap.

comparison with analyses of images by experienced human observers. We emphasize that the threshold and cutoff referred to above are determined only by the lesion diameter, the number of pixels per mm in the frame-grabbed image and whether the lesions have positive or negative contrast; thus, though human observers were involved in establishing the method for determining thresholds, no further involvement of human observers is necessary. This means that mass production of automated systems is feasible such that, for all subsequent users, results are independent of those users, *i.e.*, objectivity exists.

MATERIALS AND PHANTOMS

Tissue-mimicking materials used in lesion phantoms are composed of distilled water, *n*-propanol, agar (Difco-bacto agar, Difco Laboratories, Detroit, MI, USA), finely powdered graphite (Grade no. 9039, Superior Graphite Co., Chicago, IL, USA) and glass bead scatteres with a sieve-imposed diameter range of 45–53 μm. The specific composition in weight percents of the background material, of which most of each phantom is composed, is shown in the first row of Table 1. There are four values of object contrast in the phantoms. The shape of the objects is spherical. Three object contrasts are negative, designated N1, N2 and N3, and the fourth is positive, designated P1. Contrasts are expressed in decibels with P1 > 0 > N1 > N2 > N3. Specific compositions for the various contrast spheres are shown in Table 1.

Ultrasonic properties were measured at 22°C for each material in the phantoms using test sample cylinders with parallel 25-μm-thick Saran Wrap windows. The method of measurement of attenuation and speed has been described (Madsen et al. 1982), as has the method for measuring backscatter coefficients (Chen et al. 1993). Values are given in Table 2. Ultrasonic speeds were measured at 2.5 MHz and attenuation coefficients at 2.5, 4.5, 6.2 and 8.0 MHz. Attenuation values were fitted to a relation $\alpha(f) = \alpha_0 f^n$ where $\alpha(f)$ is the attenuation coefficient at frequency f, and α_0 and n are constants determined in the curve fitting. In the case of the backscatter coefficient, broadband measurements were made from 3–7 MHz with curve fitting to the form $\eta(f) = \eta_o f^m$, where η_o and m are constants.

The reason for the somewhat different frequency dependencies of backscatter for the different materials (see m values in Table 2) is that a small part of backscatter is due to the graphite particles. The relative backscatter coefficients are therefore slightly dependent on frequency. For example, for the N2 spheres, the backscatter relative to that of the background material in the phantoms is −11.2 dB at 3 MHz and −10.2 dB at 7 MHz.

Four phantoms were produced, each containing

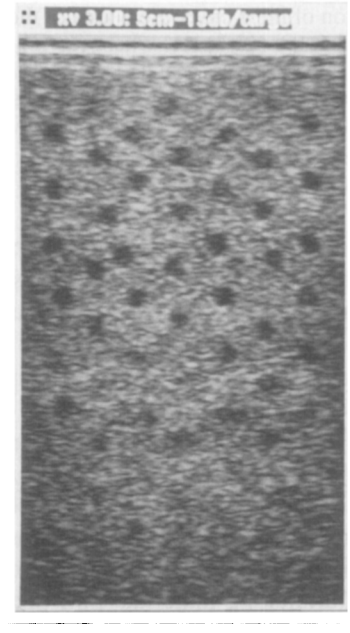


Fig. 2. Image using a 5-MHz linear array with the scan plane superimposed on the plane containing the centers of the 4-mm-diameter, N3 contrast spheres. Spheres close to the transducer and sufficiently distant from it are not detectable.

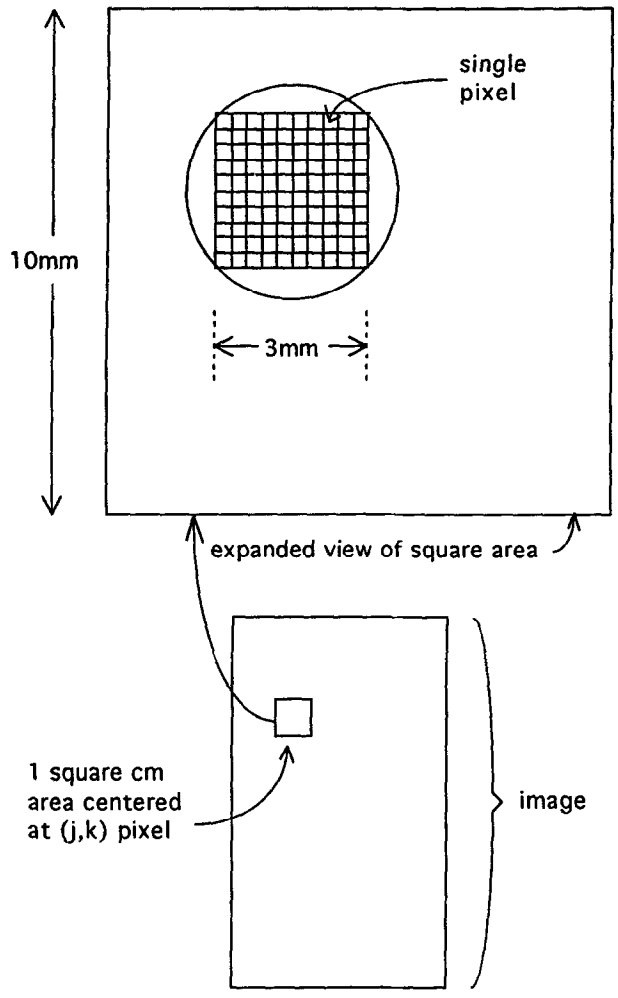


Fig. 3. Depiction of the 1 cm × 1 cm area of the background image centered at the (j,k) pixel where $(SNR)_\ell$ is computed (lower part of figure). An expanded view of the 1 cm × 1 cm area is shown in the upper part, illustrating one of the 3 mm × 3 mm areas for computing the set of S_b's used to obtain σ_b.

two sets of spheres with the same composition and coplanar centers. Figure 1 is a diagram showing the positioning of two sets of spheres in a phantom. Each set includes seven spheres per cm depth. The tissue-mimicking material in each phantom is a rectangular parallelepiped volume with dimensions 14 cm × 17 cm × 6 cm, the 6-cm dimension being perpendicular to the diagram in Fig. 1. The scanning window is 14 cm × 17 cm, and the plane containing sphere centers is 3 cm from each wall. Except for the spheres, the phantom is filled entirely with the background (reference) tissue-mimicking material. Two phantoms contain 4-mm-diameter spheres with contrasts corresponding to N1, N2, N3 and P1 given in Table 1. The other two phantoms are the same, except the sphere diameter is 3 mm.

Manufacture of the phantoms was done in such a way that sphere centers were coplanar. Mass production of spheres of equal diameter was done using two-piece acrylic molds with aligned hemispherical depressions. Next, a 3-cm block of background material was made with one 12 cm × 16 cm side exposed. Two 3-mm-thick acrylic slabs, in contrast with one another over their 12 cm × 16 cm surfaces, were then suspended about 1 cm above the exposed background side. Both acrylic layers had the identical regular array of holes corresponding to the array formed by the spheres, but the two acrylic layers were displaced laterally so that a spherical lesion could be placed into each hole of the upper acrylic layer without falling through. A thin layer of molten background material was then quickly spread over the exposed background slab, and the acrylic slabs were immediately shifted so that the holes lined up and the spheres fell into the thin molten layer. Congealing occurred in a few seconds, before any gravitational sedimentation of glass beads or graphite could occur, and the spheres were stuck into position. Finally, the remainder of the acrylic box, which at the end surrounded the phantom materials, was glued in place over the spheres, leaving a 3-cm space above them. This space was filled with molten background material. Rotation at 2 rpm about a horizontal axis during congealing prevented lack of uniformity of glass bead and graphite distribution. More details regarding temperatures and techniques can be found in an earlier work (Madsen et al. 1991).

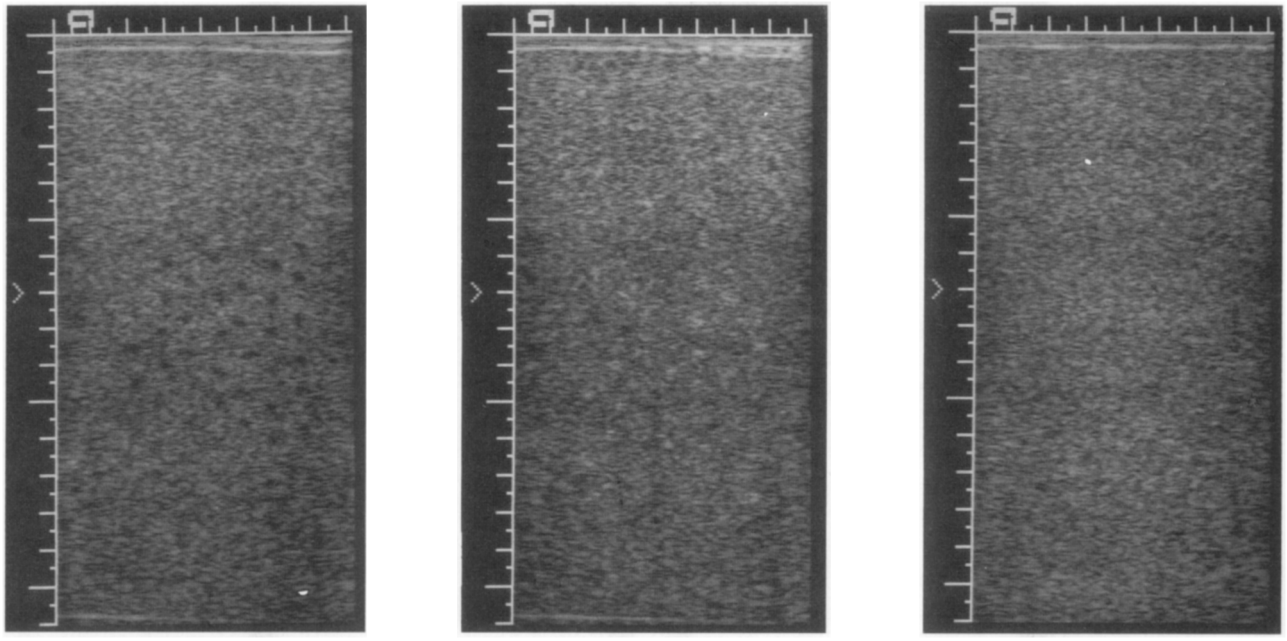
An image using a 5-MHz linear array with the scan plane superimposed on the plane containing the centers of the 4-mm-diameter N3 contrast spheres is shown in Fig. 2. Spheres close to the transducer and sufficiently distant from it are not detectable.

AUTOMATION METHODOLOGY

Image acquisition

Ultrasound B-mode images were made with the scan plane (plane of symmetry of the scan slice) superimposed on the plane containing the spheres. A Macintosh Centris computer equipped with a video frame grabber board (Data Translation DT2255) was then used to digitize the images. The frame grabber was connected directly to the output video signal from the scanner. The public domain software "NIH Image" was used to control data acquisition and storage. The data were archived on a magneto-optical drive.

Image data from the Macintosh were transferred via ethernet to a DECstation 5000/200 workstation in another lab running Ultrix 4.1. The Macintosh TIFF images were converted to portable bit-mapped images using XV version 3.00 on the DECstation. The DECstation image analysis programs were written in C.



a b c

Fig. 4. (a) 3.5-MHz, 7-cm transmit focus, linear array scan (Acuson L382) of the 4-mm-diameter, N3 contrast lesions. The lesions lie in the 2- to 6.5-cm lateral range of the image. (b) 3.5-MHz, 7-cm transmit focus, linear array scan (Acuson L382) of the 4-mm-diameter, P1 contrast lesions. The lesions lie in the 2- to 6.5-cm lateral range of the image. Equipment and console settings are identical to those used in obtaining the image in panel a. (c) 3.5-MHz, 7-cm transmit focus, linear array scan (Acuson L382) of the background region (no lesions). Equipment and console settings are identical to those used in obtaining the images in panels a and b.

Lesion detectability: The matched filter method

Considerable work has been done on detectability of low-contrast cylindrical targets in ultrasound (Hall et al. 1993; Insana and Hall 1994; Lopez et al. 1990, 1992; Smith et al. 1983; Thijssen et al. 1988; Wagner and Brown 1985; Wagner et al. 1983). The concept of the "matched filter" has been developed for ultrasound (Smith et al. 1983; Wagner et al. 1983) and has been used with phantoms containing low-contrast cylindrical targets to do objective contrast-detail studies (Hall et al. 1993; Lopez et al. 1992).

The matched filter technique (Wagner et al. 1983) is conceptually simple when the target is a cylinder of radius a, with axial symmetry of mean object contrast, oriented perpendicular to the ultrasound scan slice. A lesion signal-to-noise-ratio, which embodies the matched filter method, is defined as:

$$(\mathrm{SNR})_{M\ell} \equiv (S_{M\ell} - S_{Mb})/[(\frac{1}{2})(\sigma_b^2 + \sigma_\ell^2)]^{1/2}.$$

The meaning of the terms on the right of this relation is somewhat complicated. Using a part of the image involving echoes from background materials only, define $S_b \equiv (\frac{1}{n}) \sum_{i=1}^{n} P_i^m$ to be the mean of the mth power of the image pixel values, P_i, over an area in the image defined by the projected area of the lesion. (Thus, if the lesion is a right circular cylinder of radius a perpendicular to the scan slice, the projected area is a circle of radius a.) S_{Mb} is the ensemble average or mean (M) of a large number of independent realizations of S_b. If m is 1, then S_b is called the pixel amplitude mean and, if $m = 2$, S_b is called the pixel intensity mean. $S_\ell \equiv (\frac{1}{n}) \sum_{i=1}^{n} w_\ell(r_i)\, P_i^m$ is the weighted mean of the mth power of the pixel value over an equal area in the image when the lesion actually occupies the area of concern. $S_{M\ell}$ is the ensemble average of S_ℓ for M independent realizations. σ_b is the standard deviation of the ensemble of S_b values, and σ_ℓ is the standard deviation of the ensemble of S_ℓ values.

The factor $w_\ell(r_i)$ is a weighting function of distance r_i from the center of the circular area and depends on the expectation value of P_i relative to that of the background material. If $w(r_i)$ is independent of r_i, corresponding to an infinitely long cylinder with uniform ultrasonic properties (including backscatter coefficient), then $w_\ell(r_i) = 1$. If the simulated lesion does not possess invariance of backscatter expectation value with respect to translation perpendicular to the scan slice, then $w_\ell(r_i)$ will generally depend not only on

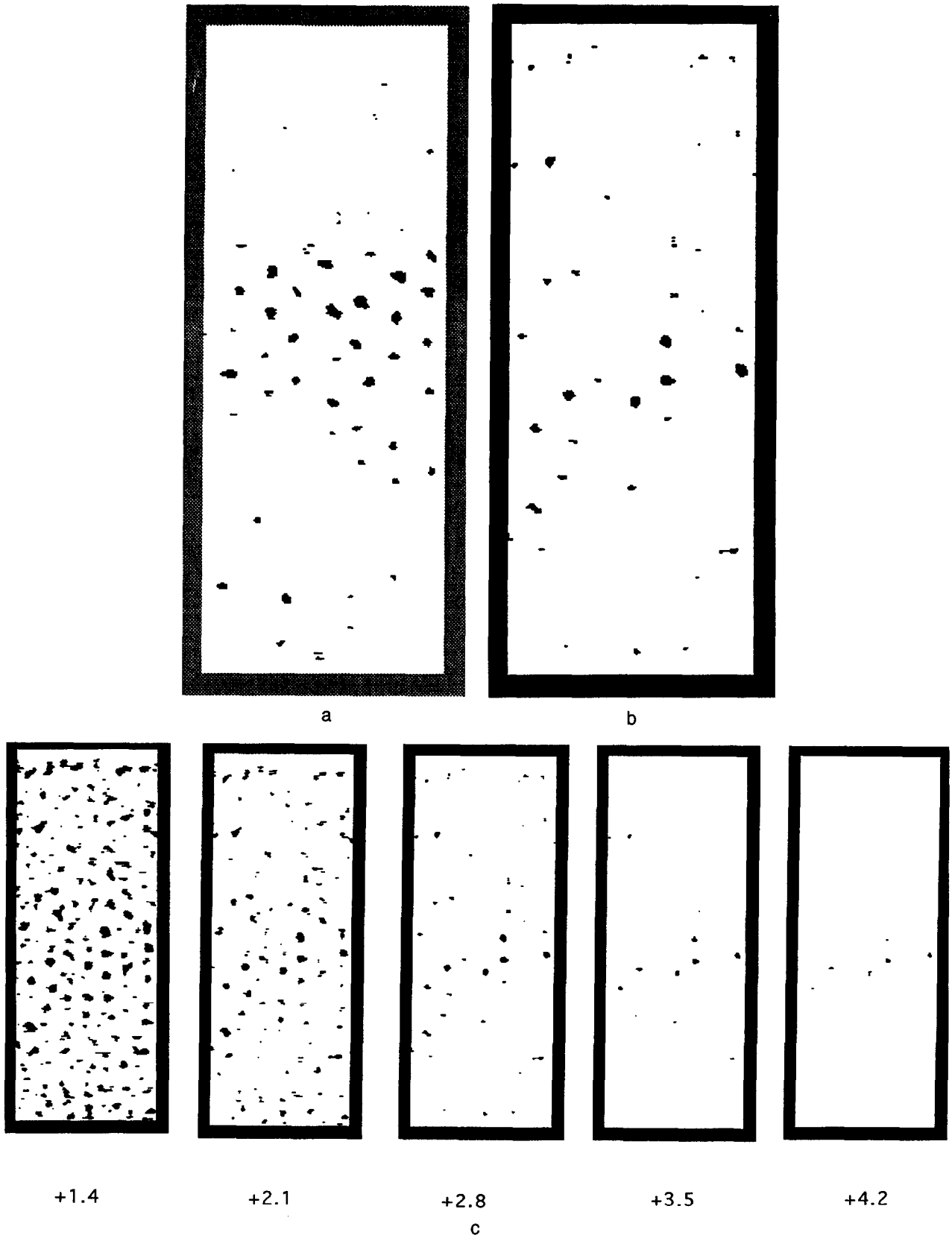
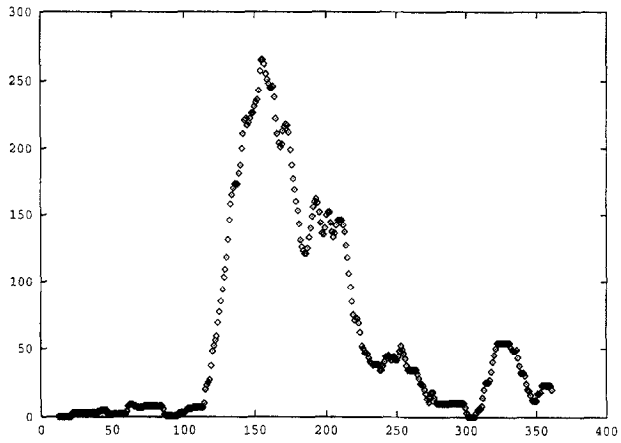


Fig. 5. (a) Map of pixels with $(SNR)_\ell \leq -2.8$ corresponding to Fig. 4a. The 5-mm shaded border indicates image points not included in map. The left edge of the 5-mm strip on the left of the "image" corresponds to 1.0 cm on Fig. 4a, and the right edge of the 5-mm strip on the right corresponds to 7.5 cm on Fig. 4a. (b) Map of pixels with $(SNR)_\ell \geq +2.8$ corresponding to Fig. 4b. The 5-mm shaded border indicates image points not included in map. The left edge of the 5-mm strip on the left of the "image" corresponds to 1.0 cm on Fig. 4b, and the right edge of the 5-mm strip on the right corresponds to 7.5 cm on Fig. 4b. (c) Maps of pixels for five different thresholds on $(SNR)_\ell$, corresponding to Fig. 4b, show the effects of thresholds other than +2.8. The five thresholds are 1.4, 2.1, 2.8, 3.5 and 4.2.



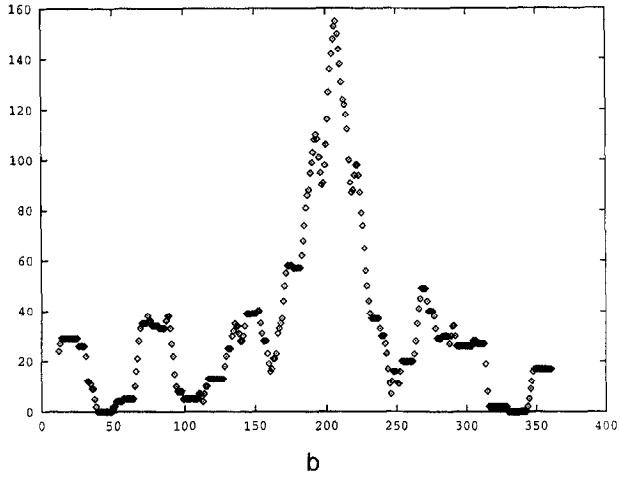


Fig. 6. (a) Histogram, corresponding to Figs. 4a and 5a, of the number of pixels with $(SNR)_\ell \leq -2.8$ included in a 1-cm depth range centered at the pixel depth coordinate indicated on the abscissa. Abscissa value 0 corresponds to 0-cm depth, and 400 corresponds to 16-cm depth. (b) Histogram, corresponding to Figs. 4b and 5b, of the number of pixels with $(SNR)_\ell \geq +2.8$ included in a 1-cm depth range centered at the pixel depth coordinate indicated on the abscissa. Abscissa value 0 corresponds to 0-cm depth, and 400 corresponds to 16-cm depth.

the nature of the variation but also on the local slice profile of the ultrasound beam. $w_\ell(r_i)$ is defined so that those regions of the lesion contributing more to backscatter than others are given higher weight. If the object is a sphere, the weighting function would be chosen to be a monotonically decreasing function of r_i being a maximum at $r_i = 0$ (the center of the projected area of the sphere) and zero for $r_i \geq a \equiv$ radius of the sphere. Suppose, for example, that the simulated lesion is in the center of a beam with a large slice thickness so that the acoustic pressure amplitude is constant over the volume of the sphere; then a reasonable choice for $w_\ell(r_i)$ might be to make it proportional to the thickness of the sphere along the elevational axis at the distance r_i from the center, *i.e.*, $w_\ell(r_i) = C\,(a^2 - r_i^2)^{1/2}$, where C is a constant. $w_\ell(r_i)$ should always be normalized, however, so that $\int_0^a w_\ell(r_i)\, 2\pi r_i\, dr_i = \pi a^2$, where a is the radius of the sphere (and of the projected area πa^2). The normalization condition in this example yields $w_\ell(r_i) = (3/2)[1 - (r_i/a)^2]^{1/2}$. If the slice profile is not sufficiently large that the acoustic pressure can be considered constant over the sphere volume, then its variation should be accounted for in the weighting function; knowledge of slice profile, however, cannot be assumed.

Application of the matched filter method for spherical simulated focal lesions and justification for assuming $\sigma_\ell \approx \sigma_b$

If the slice profile is known, then the discussion in the above paragraph can be applied with no compromise to spherical targets centered in the scan plane (plane of symmetry of the scan slice). However, if the slice profile is not known (which must be assumed in clinical situations) then $w_\ell(r_i)$ must be approximated in some way. Let the radius of a spherical target lesion be R. An example of a reasonable approximation (which we have used) follows. The sphere can be approximated to have the shape of a right circular cylinder with radius $a \approx (\frac{3}{4})\, R$ and length $2R$, where the

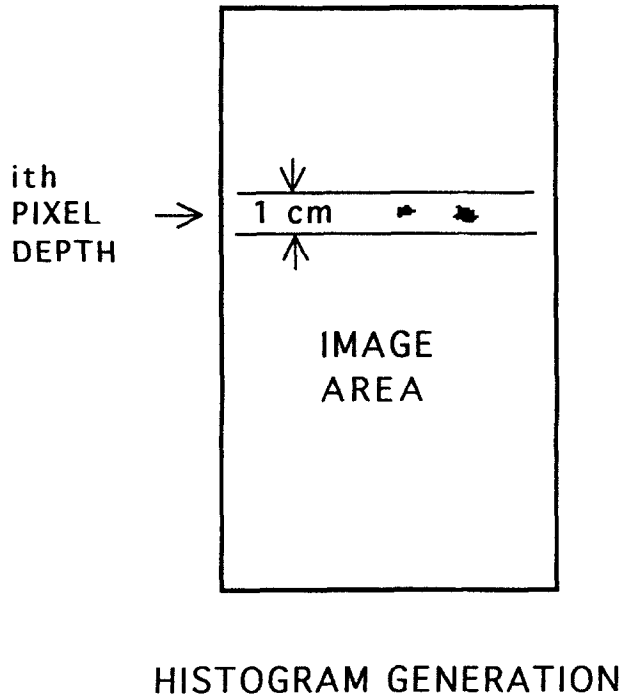


Fig. 7. Example of 1-cm depth area over which superthreshold pixel values are summed in generation of histograms such as those shown in Figs. 6a and b.

Table 3. p and d values (in cm) for the various lesion types using the Acuson 3.5-MHz L382 scan head with transmit focus at 7 cm.

	4 mm, N3		4 mm, N2		4 mm, P1		3 mm, N3		3 mm, P1	
Observer ident.	p	d	p	d	p	d	p	d	p	d
A	5.5	11.0	6.0	11.5	7.0	10.0	6.5	7.5	6.5	7.5
B	5.0	11.0	5.5	12.0	7.5	9.5	6.0	8.0	—	—
C	5.7	11.0	6.0	11.0	7.0	11.0	6.0	8.0	—	—
Automated system	4.8	10.3	5.2	10.7	6.1	11.1	6.4	7.2	5.7	6.6

For 4 mm, N1, 3 mm, N2 and 3 mm, N1, no lesions were detected by observers or the automated system.

axis of the cylinder is perpendicular to the scan plane. (Note that the cylinder volume approximates the actual sphere volume.) For programming convenience, we have gone one step further and approximated the sphere and the cylinder described with a square-ended rectangular parallelepiped, where length perpendicular to the scan plane is $b \approx 2R$ and the side of the square $c = (4/3)R$. We take the projected area of the sphere to be the square area c^2. Thus, n is the number of pixels in the square area, and since the slice profile should not vary greatly over this projected area, $w_\ell(r_i) = 1$. An approximate matched filter detection process can now be defined for spherical lesions. Before completing the model, however, consider a convenient approximation for σ_ℓ.

In an application of the matched filter concept via $(SNR)_{M\ell}$, computation of σ_ℓ for an actual distribution of independent realizations of S_ℓ could be prohibitive in terms of time required. Thus, arguments are presented here for approximating σ_ℓ with σ_b for any object contrast. If the object contrast is small, then $\sigma_\ell \approx \sigma_b$ is obviously reasonable. It can be argued, however, that setting $\sigma_\ell = \sigma_b$ is a reasonable approximation even when object contrast is large since detectability thresholds are of interest. For barely perceptible objects, a "partial volume" effect will generally exist for spherical targets. (Some background material will almost always exist in the volume, which gives rise to the signal since this is a cylindrical volume perpendicular to the scan slice.) Thus, for spherical lesions in the regions of the imaged area where lesions are barely detectable, the approximation $(SNR)_{M\ell} = (S_{M\ell} - S_{Mb})/\sigma_b$ is assumed to be justified, even when object contrast is large.

Modification of the matched filter method to allow $(SNR)_\ell$ "scans" of images of the phantoms

The straightforward way to apply the matched filter concept through $(SNR)_{M\ell} = (S_{M\ell} - S_{Mb})/\sigma_b$ would be to compute $S_{M\ell}$, S_{Mb} and σ_b at some depth in the image plane. Thus, many independent realizations of S_ℓ and S_b at the depth involved would be required, necessitating a scheme to identify the precise location of all lesions. Because the software is ultimately intended for use by clinical personnel and because the lateral and axial positioning of the spheres is not exact (although, very importantly, the sphere centers are definitely coplanar), we have devised a method for avoiding the need to know exactly where the lesions are in the scan plane. This is done by adaptation of $(SNR)_{M\ell}$ to a computer scan technique. We define:

$$(SNR)_{\ell[j,k]} \equiv \frac{[S_\ell(j,k) - S_{Mb}(j,k)]}{\sigma_b(j,k)}$$

where j,k is a set of pixel coordinates at the center of the square areas in the image over which $S_\ell(j,k)$, $S_{Mb}(j,k)$ and $\sigma_b(j,k)$ are computed. Notice that $S_\ell(j,k)$

Table 4. p and d (in cm) for an Acuson 5-MHz L558 linear array with transmit focus at 5 cm.

	4 mm, N3		4 mm, N2		4 mm, N1		4 mm, P1		3 mm, N3		3 mm, N2		3 mm, P1	
Observer ident.	p	d	p	d	p	d	p	d	p	d	p	d	p	d
A	1.1	7.3	0.6	7.4	1.8	6.0	0.5	6.8	1.7	6.2	2.4	5.6	1.7	5.2
B	1.2	9.0	0.7	8.4	1.2	6.5	2.4	6.3	1.7	5.3	3.1	5.7	3.5	4.9
C	1.2	8.6	0.7	7.7	3.6	5.6	2.6	6.5	1.9	5.5	3.1	5.5	4.4	4.4
Automated system	0.5	7.0	0.5	7.3	0.9	5.5	0.7	6.1	0.8	4.4	1.5	3.5	1.1	5.0

No 3 mm, N1 lesions were detected by observers or the automated system.

is a single realization, not the ensemble average $S_{M\ell}$. The value of $S_\ell(j,k)$ is computed over a square area of side $\frac{2}{3}$ times the sphere diameter and centered at the j,kth pixel. The S_bs to determine σ_b are computed over a 1-cm^2 area centered at the j,kth pixel using an image of background only material. The situation is depicted in Fig. 3. σ_b is obtained by computing many values of S_b corresponding to 3 mm × 3 mm areas lying in the 1-cm^2 area for the 4-mm-diameter lesion. (In current practice, to get better estimates of σ_b, four independent background images have been used, so S_bs were computed over four realizations of each 1-cm^2 area.)

Note that it is not necessary that $S_{Mb}(j,k)$ be strictly the ensemble average over areas S_b corresponding to a lesion since, assuming some degree of uniformity of the slice profile, $S_{Mb} \approx$ mean pixel value over a sufficiently large image area. Thus, we take S_{Mb} to be the pixel value average over the entire 1-cm^2 area, which is equivalent to at least 100 mm^2/9 mm^2 ≈ 11 independent ensemble elements for a 3 mm × 3 mm approximated lesion area. Note also that, because there must be a 1 cm × 1 cm area of the image centered at each pixel where a value of $(SNR)_\ell$ is computed, a 5-mm wide border of the image contains pixels for which no $(SNR)_\ell$ is computed.

*Thresholding to determine the proximal (*p*) and distal (*d*) limits of resolution zones*

The 3.5-MHz linear array scans (Acuson L382) of two lesion-containing sections (4-mm diameter, N3 contrast and 4-mm diameter, P1 contrast) and of a background region (no lesions) are shown in Figs. 4a, b and c, respectively. The transmit focus was at 7 cm. The lesions lay in the 2- to 6.5-cm lateral range of the images in Figs. 4a and b. (See the horizontal cm scale at the top of the images.) The two-dimensional mappings of pixels (in black) having $(SNR)_\ell \leq -2.8$ (threshold) for the 4-mm, N3 lesions and $(SNR)_\ell \geq +2.8$ for the 4 mm, P1 lesions derived from the frame-grabbed images are shown in Figs. 5a and b.

The thresholds [$(SNR)_\ell \leq -2.8$ or $(SNR)_\ell \geq +2.8$] were chosen as indicators of detectability after making many comparisons between such depictions and three human observer impressions of images for all eight combinations of lesion diameter and contrast in present phantoms. For example, in Fig. 5c are shown black-on-white "images" corresponding to Fig. 4b; results for thresholds equal to 1.4, 2.1, 2.8, 3.5 and 4.2[‡] are shown. At least 20 depictions such as those in Fig. 5c, involving different scanner configurations and a variety of lesion types, were assessed subjectively. There was agreement among all three observers that the ± 2.8 threshold black-on-white "images," of the kind shown in Figs. 5a, b and c, best corresponded with resolution zones estimated in original B-mode images, such as those shown in Figs. 4a and b.

In Figs. 6a and b are shown histograms, corresponding to Figs. 5a and b, respectively, of the number of pixels beyond threshold in a 1-cm depth range of the image vs. axial pixel coordinate, each 1-cm depth range being centered at the axial pixel coordinate on the abscissa. Figure 7 depicts one of these 1-cm depth intervals. A second threshold (referred to as a cutoff in the *Introduction* section) was established regarding use of the histograms to determine p and d, the proximal and distal limits of depth ranges where lesions are considered to be detected (resolution zones). The threshold was chosen to optimize agreement with averaged p and d values determined by human observers. The threshold value for the pixel number per 1-cm depth interval has been taken to be 7% of the number of pixels contained in the projected area of all lesions in a 1-cm depth range. Note that, since the (lateral) width of a scan depends on the particular scan head, the threshold number of pixels will vary. For 4-mm-diameter lesions addressed in Figs. 5a and b, the width inside the 5-mm border is 5.5 cm, which means that one row of four and one row of three lesions exist in every 1-cm depth range (Fig. 1); thus, the threshold number of pixels is $0.07 \times \pi(2\text{ mm})^2 \times (2.5\text{ pixels/mm})^2 \times 7 \cong 40$ for the images shown in Figs. 4a and b (see also Fig. 6). Using the fact that there are 25 pixels per cm depth in Figs. 7a and b, p = 4.8 cm and d = 10.3 cm for the 4-mm, N3 contrast case and p = 6.1 cm and d = 11.1 cm for the 4-mm, P1 contrast case. Table 3 shows estimated p and d values in cm for three human observers and corresponding values for the automated analysis just described. Dashes indicate no detection at any depth.

The values of p and d for a 5-MHz Acuson L558 linear array focused at 5 cm are shown in Table 4. An image of the 4-mm, N3 lesions made with this scan head is shown in Fig. 2.

Agreement between automated system analysis and human observers seems reasonably good in Tables 3 and 4, although automated system values of both p and d tend to be smaller than those of human observers. Comparable results were found for many other imager configurations analyzed thus far. They are presented in the *Results* section along with a statistical analysis regarding agreement between human observers and the automated method.

[‡] The step size between trial thresholds is about 0.7 instead of, say, 0.5, which may seem odd. The reason this occurred is that, when the threshold comparisons were done, there was a factor of $(2)^{-1/2}$ mistakenly included in the $(SNR)_\ell$ expressions, and we tried threshold values of ±1.0, ±1.5, ±2.0, ±2.5 and ±3.0, which correspond to ±1.4, ±2.1, ±2.8, ±3.5 and ±4.2, respectively, when the corrected $(SNR)_\ell$ is used.

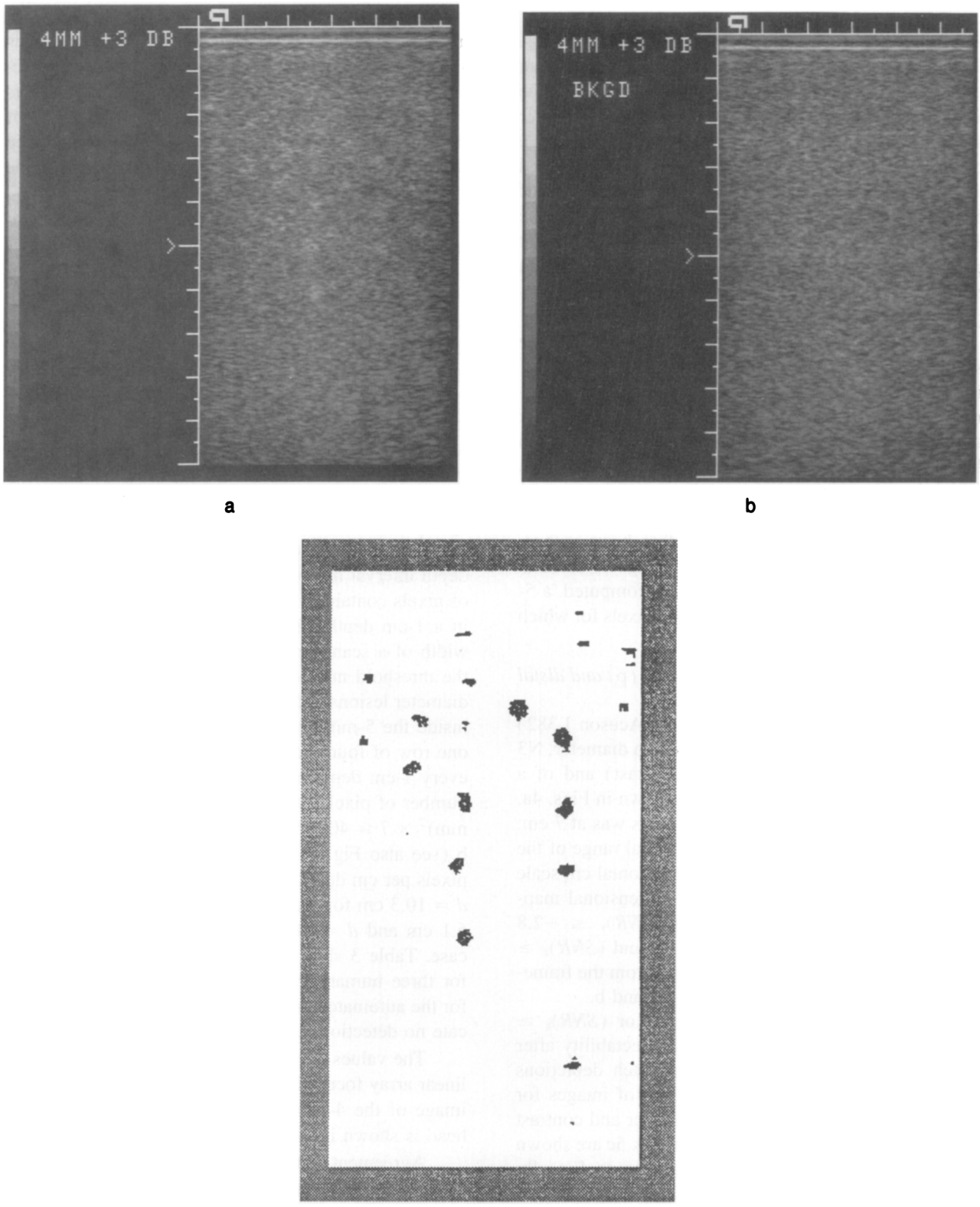


Fig. 8. (a) Image of the 4-mm, P1 spheres made with an Acuson 128 system with a 5-MHz L558 linear array scan head. The transmit focus is at 5 cm. (b) Image of background only using the same equipment and identical scanner settings as used to obtain the lesion images shown in panel a. (c) Mapping of the pixels (in black) for which $(SNR)_\ell \geq +2.8$ when the automated analysis was applied to the image shown in panel a. (d) Histogram corresponding to panels a through c, with the cut-off at pixel count of 100. The proximal cut-off is at 66 pixels depth corresponding to 1.6 cm, and the distal cut-off is at 215 pixels depth corresponding to 5.4 cm. Thus, the resolution zone extends from a depth of 1.6–5.4 cm.

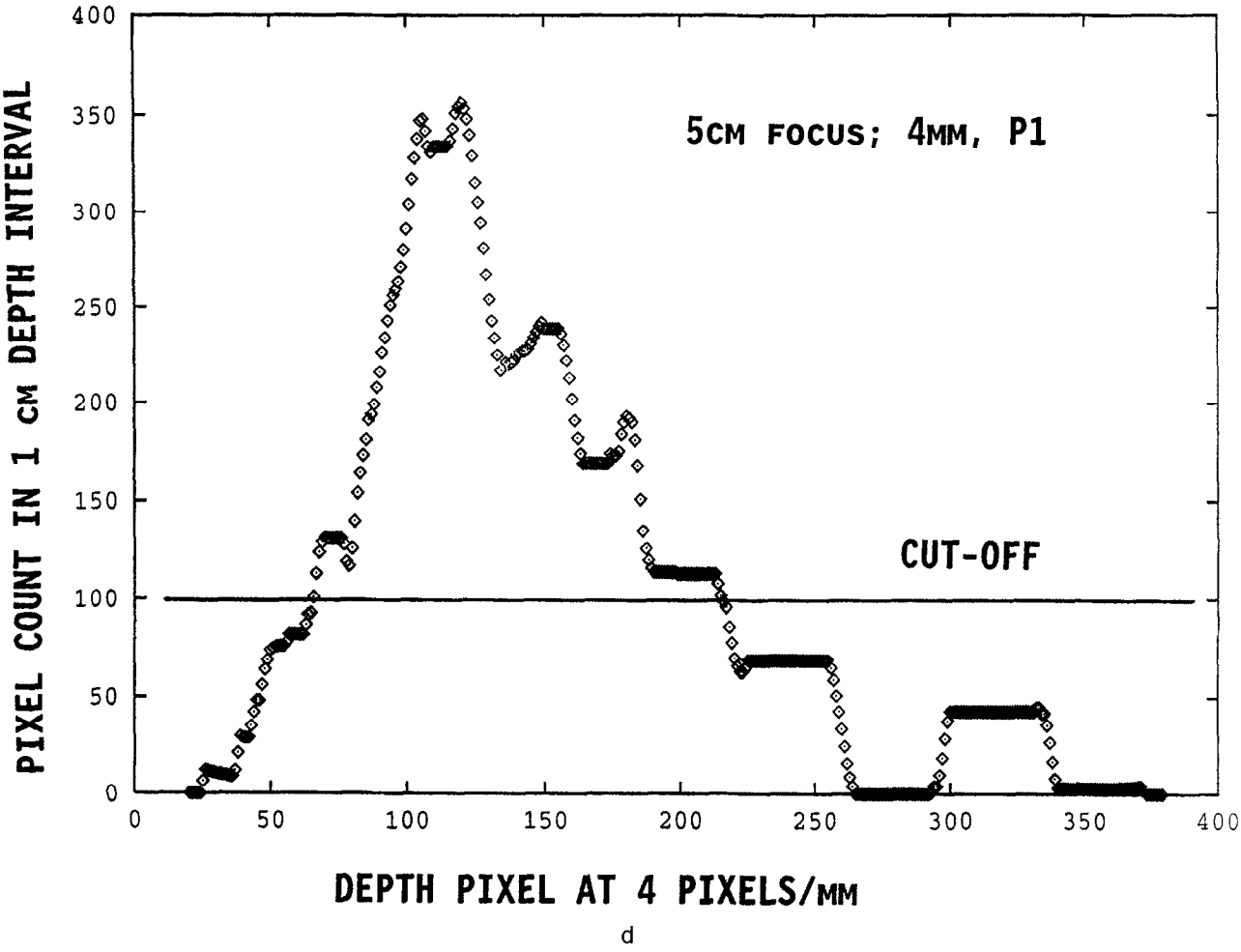


Fig. 8. *Continued.*

RESULTS

Performance assessments of numerous scanner configurations via the automated system and human observers

Perhaps the best indicator of adequacy of the automated method for mimicking human observers is the level of agreement between human observer estimations of resolution zone limits and estimations with the automated method for a large number of distinct scanner configurations. Such a study is presented in Figs. 8 through 15.

Figure 8a is an image of the 4-mm, P1 spheres made with an Acuson 128 system with a 5-MHz L558 linear array scan head. The transmit focus is at 5 cm. (Note: the serial numbers of the scanner and scan head were different from those corresponding to the results shown in Table 4.) For comparison, an image of background material only using identical scanner settings is shown in Fig. 8b. Figure 8c shows the mapping of the pixels (in black) for which $(SNR)_\ell \geq +2.8$, and Fig. 8d is the corresponding histogram with the histogram cutoff at a pixel count of 100, the number of pixels in half the area of a 4-mm-diameter circle. The proximal cutoff is at 66 pixels depth corresponding to 1.6 cm, and the distal cutoff is at 215 pixels depth corresponding to 5.4 cm. Thus, the resolution zone extends from a depth of 1.6–5.4 cm (see also the upper right part of Fig. 11.)

Figures 9 through 16 show the automated system and human observer imaging performance evaluations for a broad range of scanner configurations. All images were made with linear arrays, since the automated system software has not yet been adapted for other type scan heads. The results are shown as bar graphs with configuration parameters shown at the top along with the lesion diameter (4 or 3 mm). Depth into the phantom extends from 0 cm through the depth of field, and there is one column for each of the four contrast values, N3, N2, N1 and P1, from left to right. For each contrast, the bar on the left (cross-hatching from upper

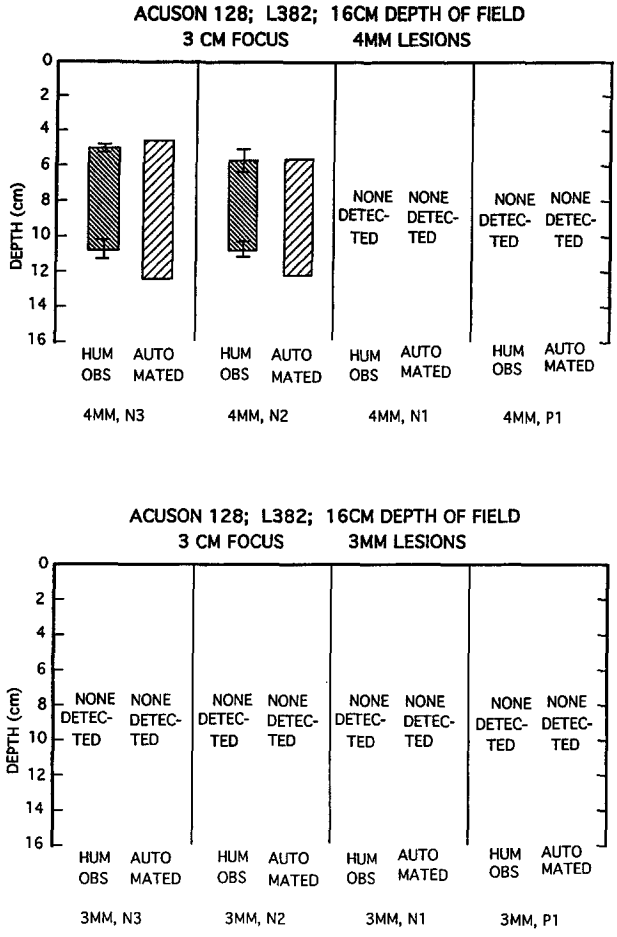


Fig. 9. Resolution zones determined by human observers and by the automated system for an Acuson 128 scanner with 3-MHz L382 scan head, 3-cm transmit focus and 16-cm depth of field. Top: 4-mm lesion results; bottom: 3-mm lesion results. Note that in this case, both the automated system and human observers detected 4-mm, N3 and 4-mm, N2 lesions only.

left to lower right) shows the resolution zone determined by the three human observers; the proximal and distal limits correspond to the means for the three observers, and the "error bars" correspond to standard deviations. The bar on the right, for each contrast, shows the result of the automated system analysis. Identical images were used by human observers and the automated system for each scanner configuration and lesion type. Therefore, these results not only show comparative performance evaluations for different imager configurations, but also provide an extensive comparison of automated with human observer results.

The data shown in Figs. 9 through 15 were analyzed statistically in terms of level of agreement of three parameters: (1) the proximal values of the resolution zones for the automated system and mean of human observers; (2) the corresponding distal values of the resolution zone; and (3) the resolution zone lengths. These results are shown in Table 5. For the six lesion types shown in Fig. 9, where human observers and automated method detected no lesions, perfect agreement was assumed. Also, the following four of the 56 sets of data (resolution zones) were not included in the analysis for the reasons given. For the 3-mm, N2 and 3-mm, N1 cases shown in Fig. 10 and the 3-mm, N1 case shown in Fig. 11, data were not included in the statistical study because only two of three human observers reported detecting any lesions. Data in Fig. 11 for the 3-mm, P1 lesions were not included because it was not clear how to interpret numerically the fact that lesions were detected by human observers, but not by the automated method.

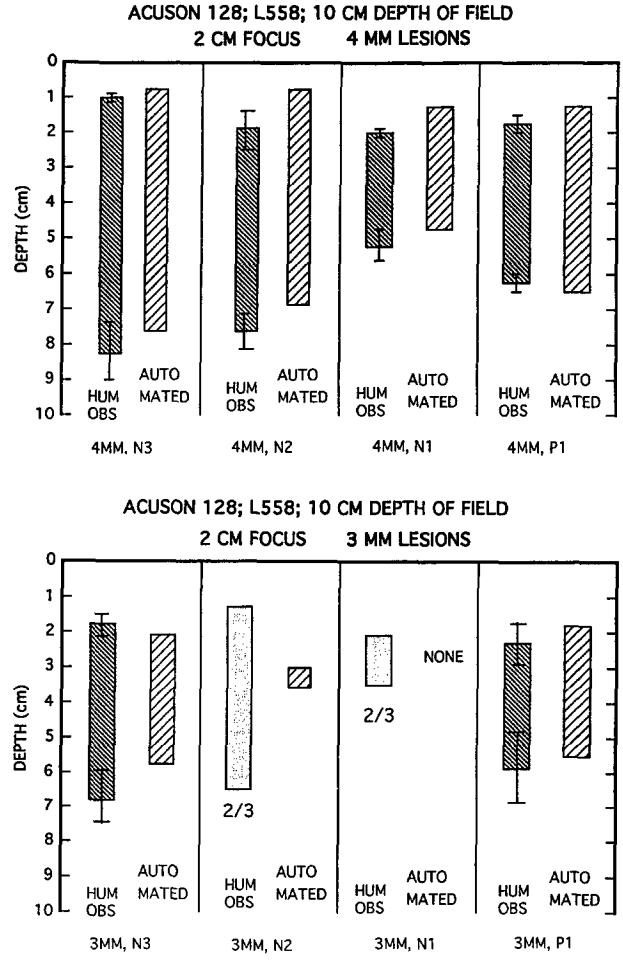


Fig. 10. Resolution zones determined by human observers and by the automated system for an Acuson 128 with 5-MHz L558 scan head, 2-cm transmit focus and 10-cm, depth of field. Note that in the case of the 3-mm, N2 lesions and 3-mm, N1 lesions, two of the three observers detected lesions, and one did not (hence, the $\frac{2}{3}$ note below the related bars); no error bars were placed because of the small number in the means.

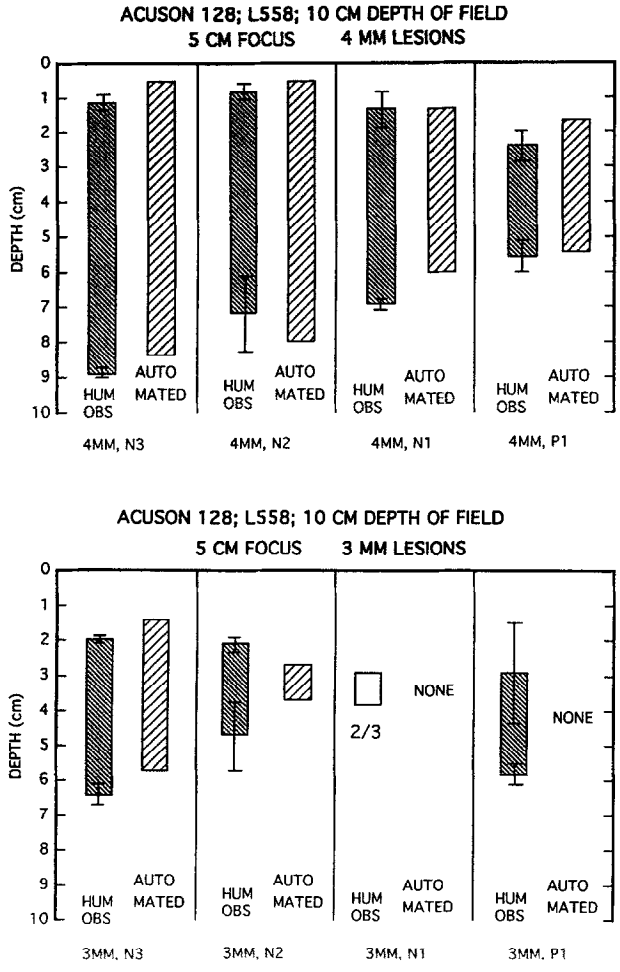


Fig. 11. Resolution zones determined by human observers and by the automated system for an Acuson 128 with 5-MHz L558 scan head, 5-cm transmit focus and 10-cm depth of field. In the case of the 3-mm, N1 lesions, only two of the three human observers detected lesions. Note also that the automated results for 4-mm, P1 lesions correspond to Figs. 8a through d.

DISCUSSION

The resolution zones determined by the three human observers and by the automated system shown in Figs. 9 through 16 demonstrate reasonably good agreement between human and automated results as quantified in Table 5. Statistical analysis of 52 different resolution zones shows that, on average, the proximal and distal depth values for resolution zones given by the automated method are both about 0.3 cm less than those found by the human observers. The resolution zone lengths for the automated method vs. human observers are (on average) in excellent agreement. Standard deviations of the differences, shown in the third column of Table 5, are <1 cm for the resolution zone length and its distal limit, d; the standard deviation for the differences in proximal limit of the resolution zone is <0.4 cm.

The most severe discrepancies in Figs. 9 through 15 occur in cases where all lesions are barely detectable. For example, the human observer proximal and distal values for the resolution zone for the 3-mm-diameter, N2 lesions in Fig. 10 are 1.4 and 6.5 cm, respectively, whereas the automated system detected lesions over a range of <1 cm; however, one of the three human observers detected no lesions at any depth, indicating that lesions were borderline in detectability.

There are at least two aspects of the automated method that are being addressed to improve agreement with human observers. One is that phantoms are being constructed allowing access to 30 lesions per cm depth for a given diameter and contrast. In the present phantoms, there are only seven per cm depth (Fig. 1), which may not be statistically adequate. The new phantoms will allow an estimate of the least number per

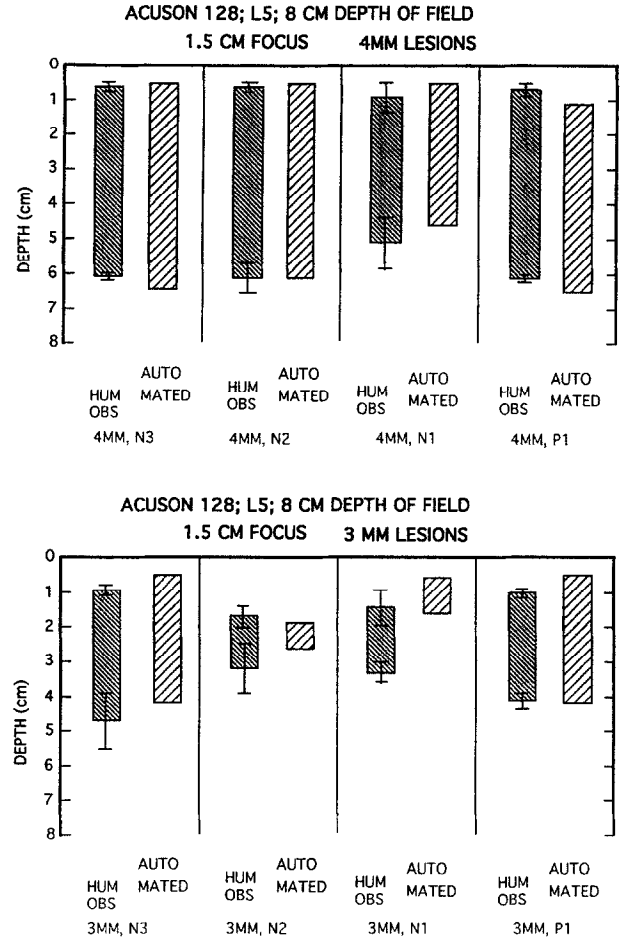


Fig. 12. Resolution zones determined by human observers and by the automated system for an Acuson 128 with ART 5-MHz L5 scan head, 1.5-cm transmit focus and 8-cm depth of field.

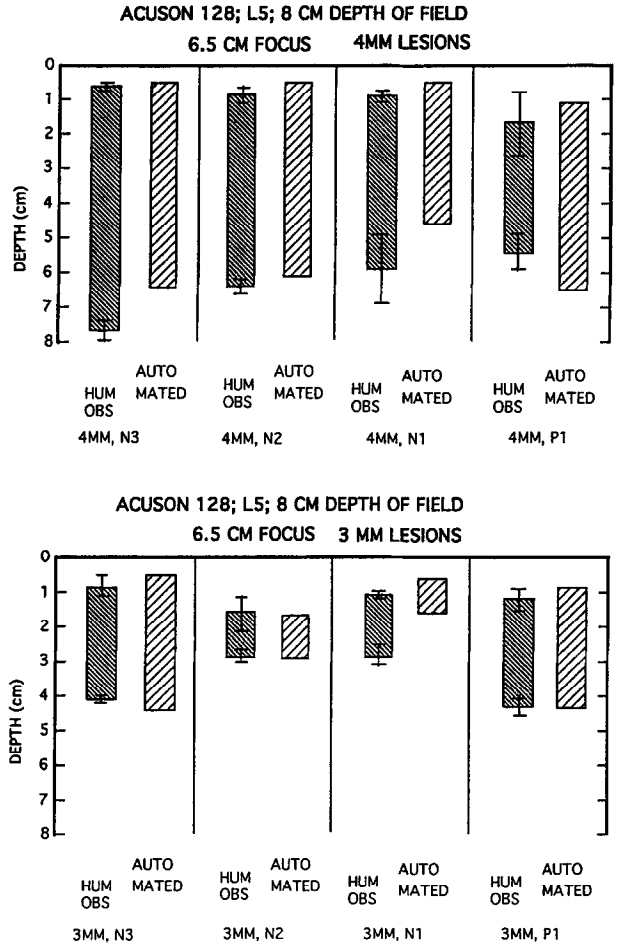


Fig. 13. Resolution zones determined by human observers and by the automated system for an Acuson 128 with ART 5-MHz L5 scan head, 6.5-cm transmit focus and 8-cm depth of field.

cm required for acceptable precision. Another aspect for improvement involves the fact that, over the 1 cm × 1 cm area used in determining σ_b, the local mean pixel value can vary enough that computed σ_bs, which assume a single mean pixel value, can be erroneously high. Significant variations in mean pixel value over 1-cm distances both laterally and axially are apparent near the edges of the image shown in Fig. 2. The method for accounting for this smooth two-dimensional (2D) variation in mean pixel value will be to replace the estimate of σ_b using the standard deviation with the "standard error" according to one definition (Croxton 1953). The idea is to do a 2D bilinear curve fitting for mean pixel value over each 1 cm × 1 cm area and then use local values of S_{Mb} (from the 2D curve fitting) in the computation of σ_b over the area.

It is also possible that the human observer results are wrongly influenced by the regular arrangement of lesions across the image. (Detection of a lesion should be independent of positioning of other lesions.) To account for this possibility, phantoms will be produced in which lesions of a given diameter and contrast are still coplanar, but the arrangement otherwise is pseudo-random, as determined with molding techniques involved in phantom production. It is possible that the thresholds for $(SNR)_\ell$ and related histograms will change somewhat as a result.

The comparative results for three different scanner configurations shown in Fig. 16 indicate an interesting distinction between the Acuson L558 and L5 scan heads. The L558, which is the older model of the two, is superior regarding its greater distal limits of the resolution zones for larger (4 mm) more negative contrast (N3 and N2) lesions; however, for the smaller (3 mm) lesser contrast (N1, P1) lesions, the L5 is superior. A simple explanation might be that the fixed elevational focal depth is greater for the L558 than for

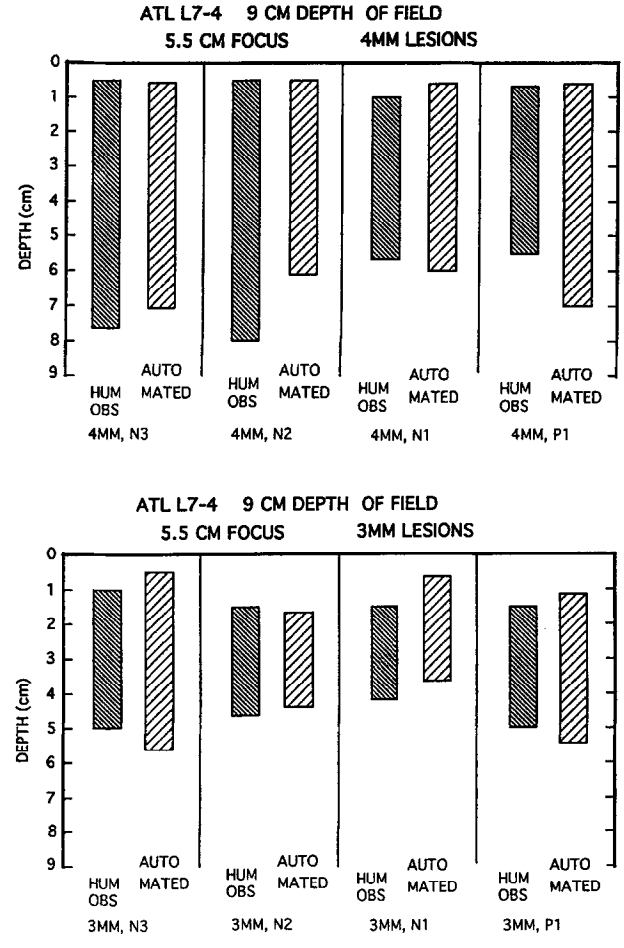


Fig. 14. Resolution zones determined by human observers and by the automated system for an ATL Ultramark 9 scanner with L4-7 linear array scan head, 5.5-cm transmit focus and 9-cm depth of field. There was only one human observer in this case, resulting in no error bars.

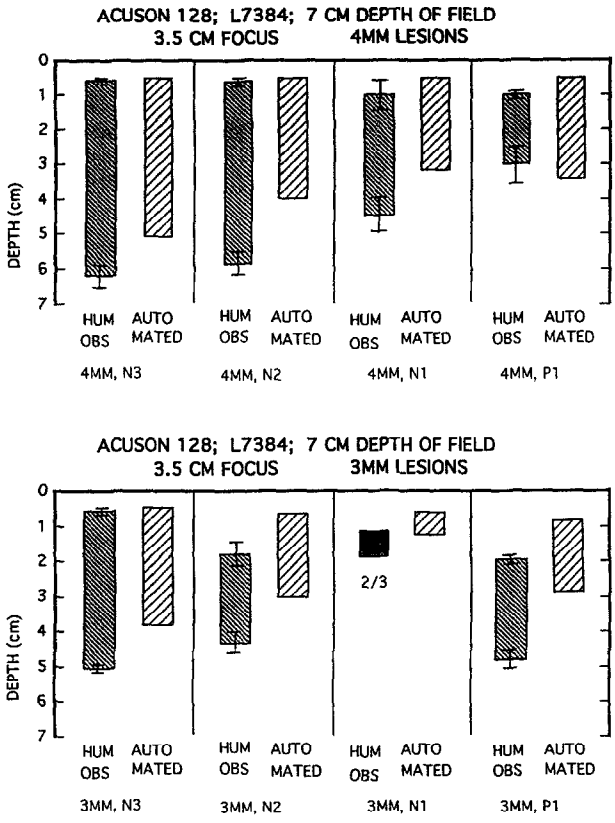


Fig. 15. Resolution zones determined by human observers and by the automated system for an Acuson 128 with 7-MHz L7384 scan head, 3.5-cm transmit focus and 7-cm depth of field.

the L5. Thus, for comparable axial and lateral resolutions, the L5 does not perform as well at greater depths where its elevational beam width is greater, but it performs better more proximally where its elevational focus is, the L558 performing more poorly more proximally because it is not focused there elevationally.

It should be noted that application of the automated method to scanners other than linear arrays is not trivial, which is the reason we have so far limited our work to linear arrays. Linear arrays are suited to square arrays of pixels in frame-grabbed images and to Cartesian coordinates, each depth corresponding to a fixed row of pixels. For scan heads other than linear arrays, the number of pixels in a 1-cm^2 sample area of the image will change greatly as the depth changes. Adaptation to phased or curved arrays will involve introduction of plane polar coordinates and location of the origin of the coordinate system. Frame-grabbed images are inherently Cartesian, and coordinate conversions will be a major programming concern. Another challenge of other-than-linear arrays is that, although the number of lesions per cm depth will be statistically adequate at greater depths, this will not be the case near the scan head for a single image of target lesions; thus, some decision process will be required regarding the geometry of the image and the number of independent images of target lesions.

Automated imaging performance analysis in hospitals, using phantoms and automation such as those described in this work, is feasible due to the wide availability of personal computers that can be fitted with digitizing cards and the increasing availability of workstations interfaced to ultrasound systems. Thus, a package could be developed that could be produced commercially and easily installed on IBM or Macintosh type personal computers or on workstations. Thus, performance comparisons could be made between dif-

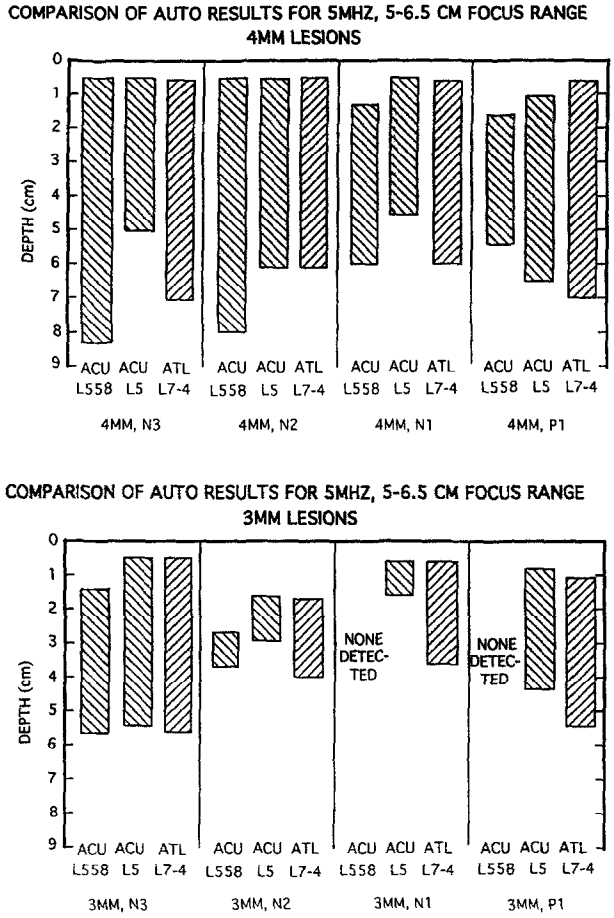


Fig. 16. Comparison of imaging performance for comparable scan heads and scanner settings. For each lesion diameter and contrast, three automated system results (resolution zones) are shown, as compiled from Figs. 11, 13 and 14. All three have about the same depth of field and transmit focus depth. The older 5-MHz Acuson system (L558) shows deeper resolution zones for the 4-mm, N3 and 4-mm, N2 lesions, but is inferior to the other two configurations for the smaller lower contrast cases of 3-mm, N1 and 3-mm, P1. Notice that the ATL L4-7 seems to be the best of the three in the 3-mm, N2, 3-mm, N1 and 3-mm, P1 cases.

Table 5. Means and standard deviations for 52 distinct combinations of scanner configuration and spherical lesion type for the three parameters*.

Parameter identity	Mean	Standard deviation
$p_{auto} - \bar{p}_{h.o.}$	−0.27 cm	0.37 cm
$d_{auto} - \bar{d}_{h.o.}$	−0.33 cm	0.85 cm
$(d_{auto} - p_{auto}) - (\bar{d}_{h.o.} - \bar{p}_{h.o.})$	−0.01 cm	0.89 cm

* Parameters: (1) $p_{auto} - \bar{p}_{h.o.}$ = the proximal value of the resolution zone given by the automated method less that given by the mean for the human observers, *e.g.*, in Fig. 9 for the 4 mm, N3 lesions, p_{auto} = 4.5 cm and $\bar{p}_{h.o.}$ = 4.9 cm, resulting in $\bar{p}_{auto} - \bar{p}_{h.o.}$ = − 0.4 cm. (2) $\bar{d}_{auto} - \bar{d}_{h.o.}$ = the distal value of the resolution zone given by the automated method less that given by the mean for the human observers, *e.g.*, in Fig. 9 for the 4 mm, N3 lesions, d_{auto} = 12.4 cm and $\bar{p}_{h.o.}$ = 10.8 cm, resulting in $d_{auto} - \bar{d}_{h.o.}$ = +1.6 cm. (3) $(d_{auto} - p_{auto}) - (\bar{d}_{h.o.} - \bar{p}_{h.o.})$ = resolution zone (length) from the automated method less that using the human observer means.

ferent makes and models of equipment for purchasing decisions, rapid quality assurance testing could be done, *etc.*

CONCLUSIONS

The level of agreement between human observers and the automated method of performance evaluation is encouraging. With the addition of the refinements described in the *Discussion* section and inclusion of more combinations of lesion diameter (perhaps 1, 2 and 5 mm) and a few more contrast values (perhaps −8 and +6 dB), the phantoms and automated system should be a powerful tool for rigorous 3D (including axial, lateral and elevational axes) contrast-detail determinations. Especially relevant is their use in the development of the new generations of 1.5 and 2D scan heads that allow for elevational focusing.

Phantoms with a limited number of combinations of lesion diameters and contrasts (2, 3 or 4?) and automated analysis should allow rapid clinical periodic performance quality assurance tests. Data acquisitions should take only a few minutes for each scanner configuration and, of course, no time is required for data analysis.

Acknowledgements—This work was supported in part by the University of Wisconsin-Madison Graduate School, The Medical Physics Ultrasound Research Fund and NIH grants R41GM54377 and R01CA39224.

REFERENCES

AIUM Standards Committee. Methods for measuring performance of ultrasound scanners: American Institute of Ultrasound in Medicine Standard. Laurel, MD: AIUM, 1990.

Chen J, Zagzebski JA, Madsen EL. Tests of backscatter coefficient measurement using broadband pulses. IEEE Trans Ultrason Ferroelectr Freq Contr 1993;40:603–607.

Croxton FE. Elementary statistics with applications in medicine and the biological sciences. New York: Dover, 1953.

Goldstein A. Slice thickness measurements. J Ultrasound Med 1988;7:487–498.

Hall TJ, Insana MF, Soller NM, Harrison LA. Ultrasound contrast-detail analysis: A preliminary study in human observer performance. Med Phys 1993;20:117–128.

Insana MF, Hall TJ. Visual detection efficiency in ultrasonic imaging: A framework for objective assessment of image quality. J Acoust Soc Am 1994;95:2081–2090.

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

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

Madsen EL, Zagzebski JA, Frank GR. Oil-in-gelatin dispersions for use as ultrasonically tissue-mimicking materials. Ultrasound Med Biol 1982;8:277–287.

Madsen EL, Zagzebski JA, Macdonald MC, Frank GR. Ultrasound focal lesion detectability phantoms. Med Phys 1991;18:1171–1180.

Madsen EL, Zagzebski JA, Medina IR, Frank GR. Performance testing of transrectal US scanners. Radiology 1994;190:77–80.

Smith SW, Wagner RF, Sandrik JM, Lopez H. Low contrast detectability and contrast/detail analysis in medical ultrasound. IEEE Trans Sonics Ultrason 1983;SU-30:164–173.

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

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

Wagner RF, Smith SW, Sandrik JM, Lopez H. Statistics of speckle in ultrasound B-scans. IEEE Trans Sonics Ultrason 1983;SU-30:156–163.