Clinically Relevant Phantom/automated System for Testing Ultrasound Imaging Performance

by EL Madsen, JM Kofler,
MJ Lindstrom and F Kelcz

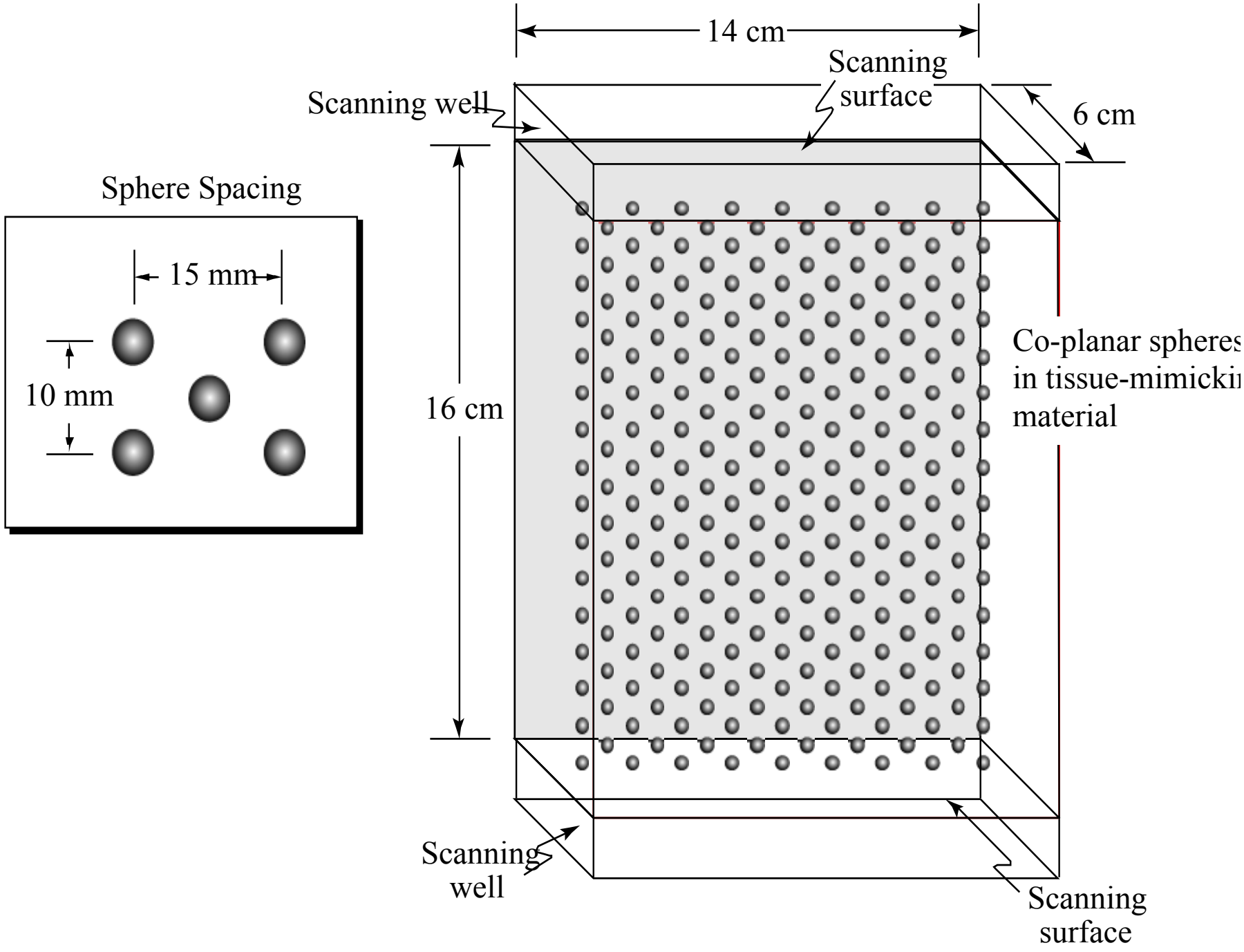
SUBJECTS:

-- spherical lesion phantoms

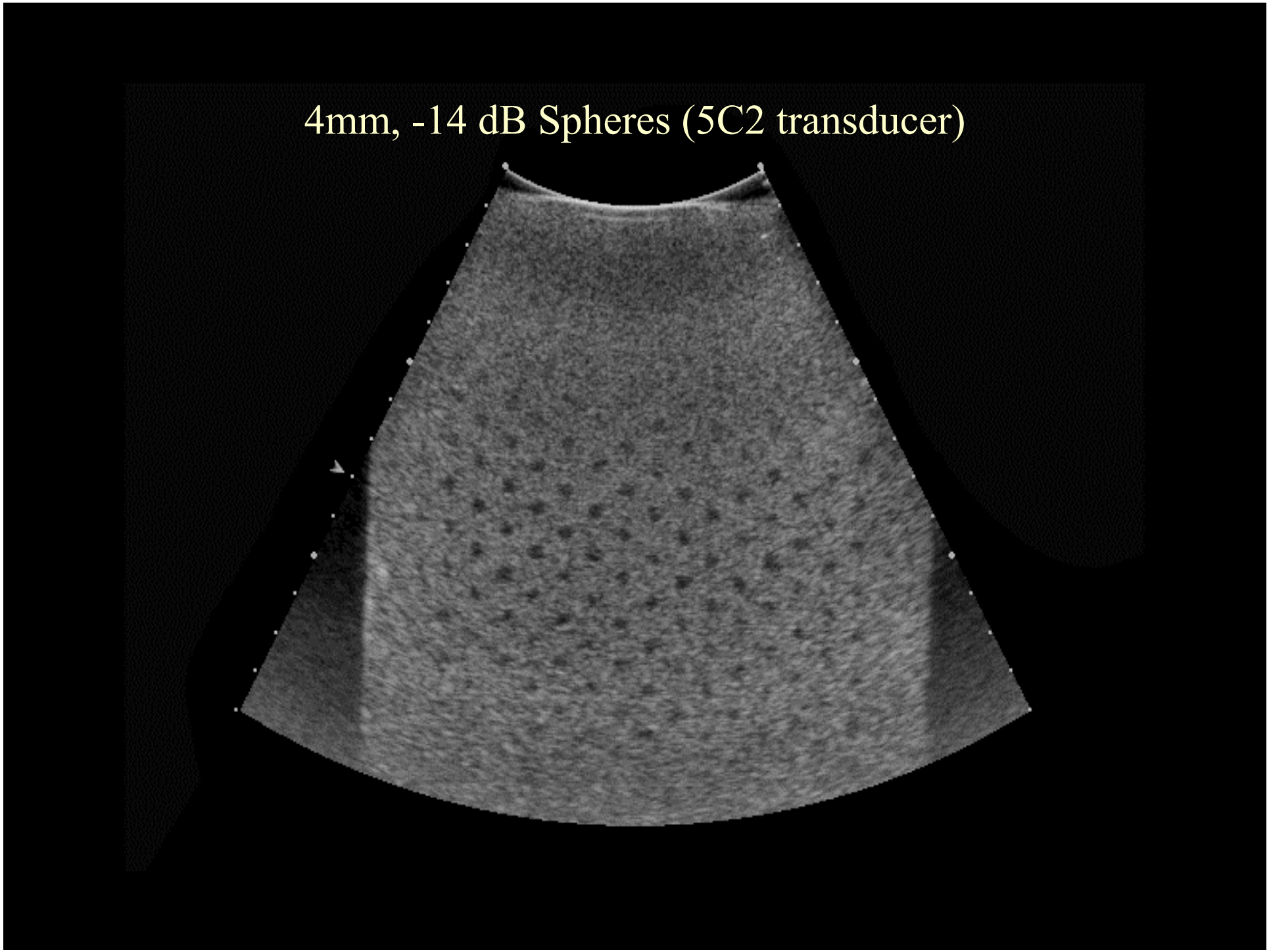
-- automated method for detecting lesions

-- detection threshold *via* human observer correspondence

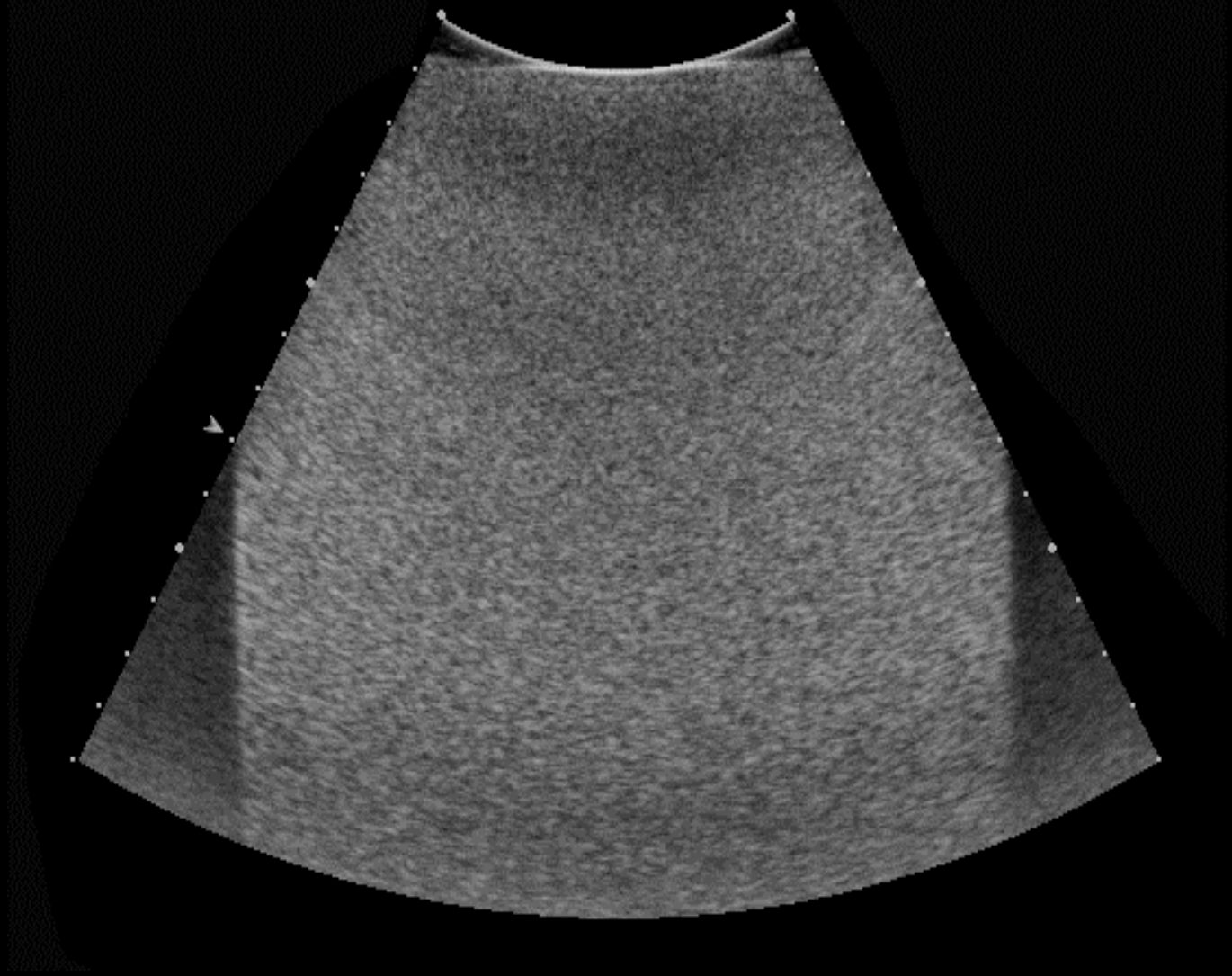
-- clinical relevance of phantom/ automation performance method *via* comparison with clinical performance on thyroid nodules


Sphere Spacing
15 mm
10 mm
14 cm
Scanning well
Scanning surface
6 cm
16 cm
Co-planar spheres in tissue-mimicki material
Scanning well
Scanning surface

Alignment of scan plane on spheres employed a micrometer-driven rotation/translation device attached to the phantom top and holding the scan head.


4mm, -14 dB Spheres (5C2 transducer)

Sample Background Image (5C2 transducer)

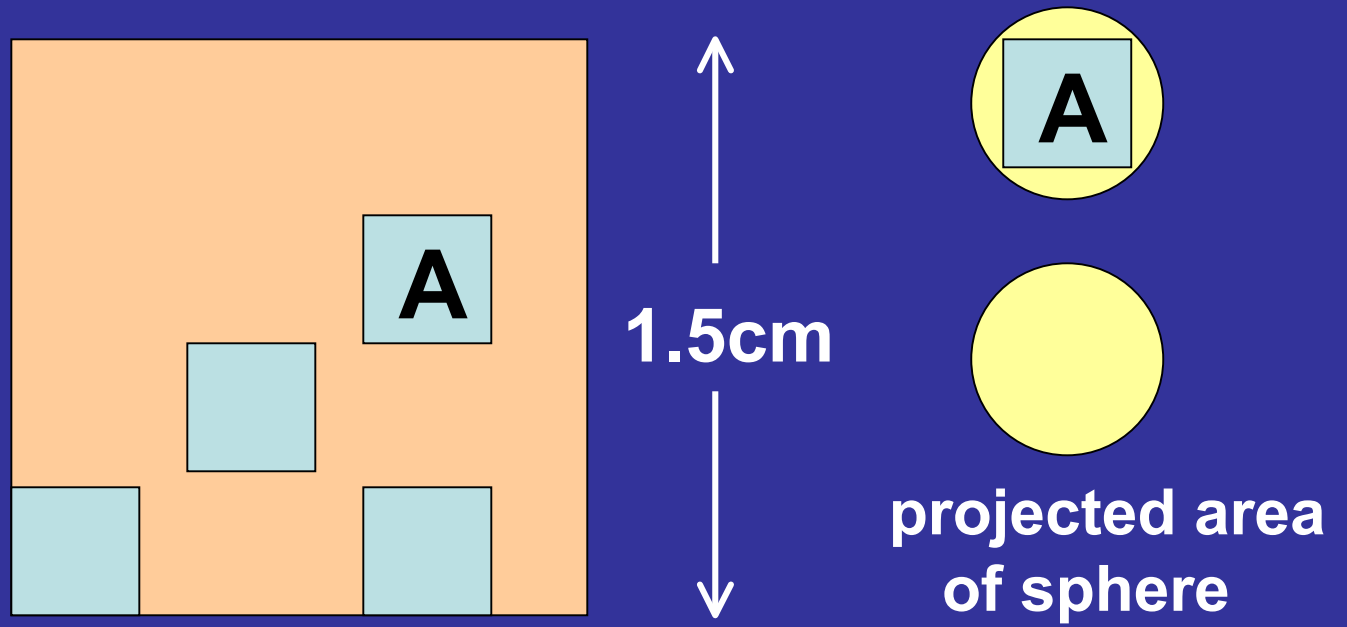


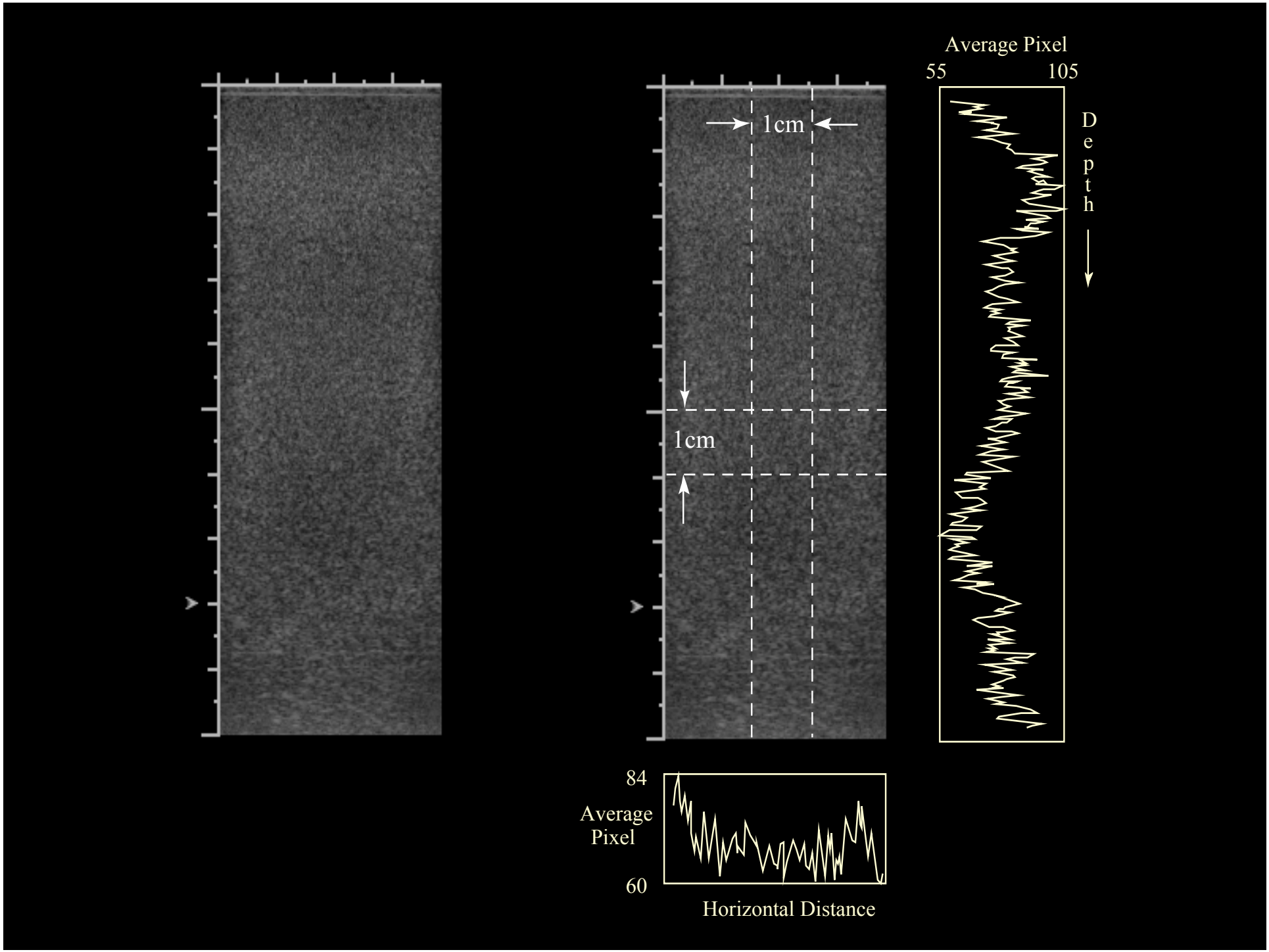
$$\overline{LSNR_{SD}} \approx \frac{\overline{S_L - S_B}}{\sigma_B}$$

$S_L \equiv$ mean pixel value over area A which is centered on the lesion and approximates the projected area of the target lesion

$S_B \equiv$ mean pixel value over a 1.5cmx1.5cm area centered at the lesion and excluding the area A

$\sigma_B \equiv$ standard deviation of mean pixel values over areas A distributed over a 1.5cmx1.5cm area of background image centered at the lesion position

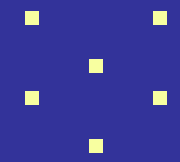



Average Pixel
55
105
1cm
1cm
Depth
84
Average
Pixel
60
Horizontal Distance

$$\overline{LSNR_{GCSD}} \approx \overline{\frac{S_L - S_B}{\sigma_{B,GC}}}$$

$\sigma_{B,GC} \equiv$ **gradient-corrected standard deviation relative to a plane fitted to local mean pixel values**

DETERMINATION OF TARGET POSITIONS

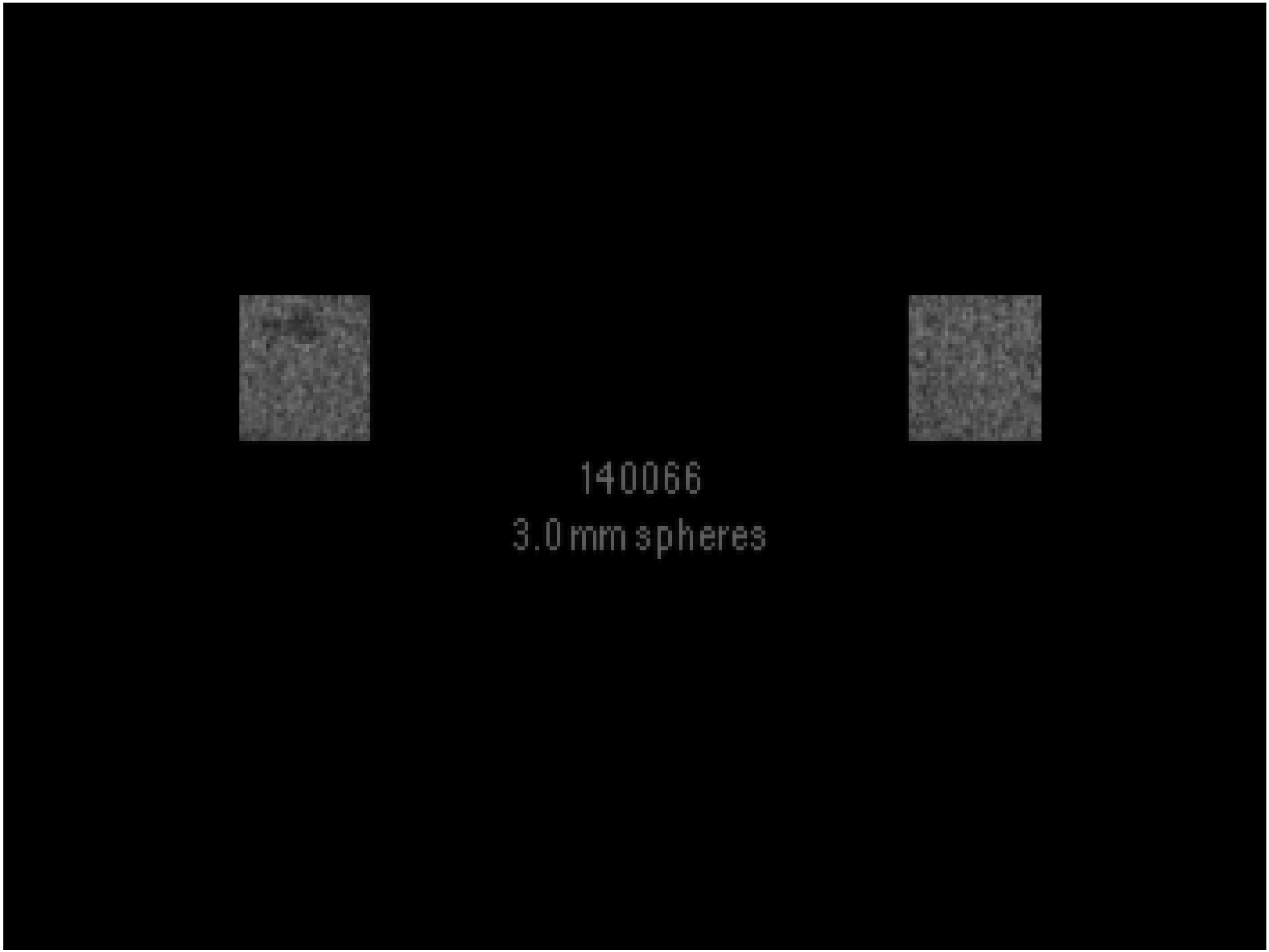
1. using the target image, *ad hoc* LSNR values are computed at relative array positions  and values summed
2. step 1 is repeated for a set of small translations and rotations
3. the most negative sum corresponds to alignment

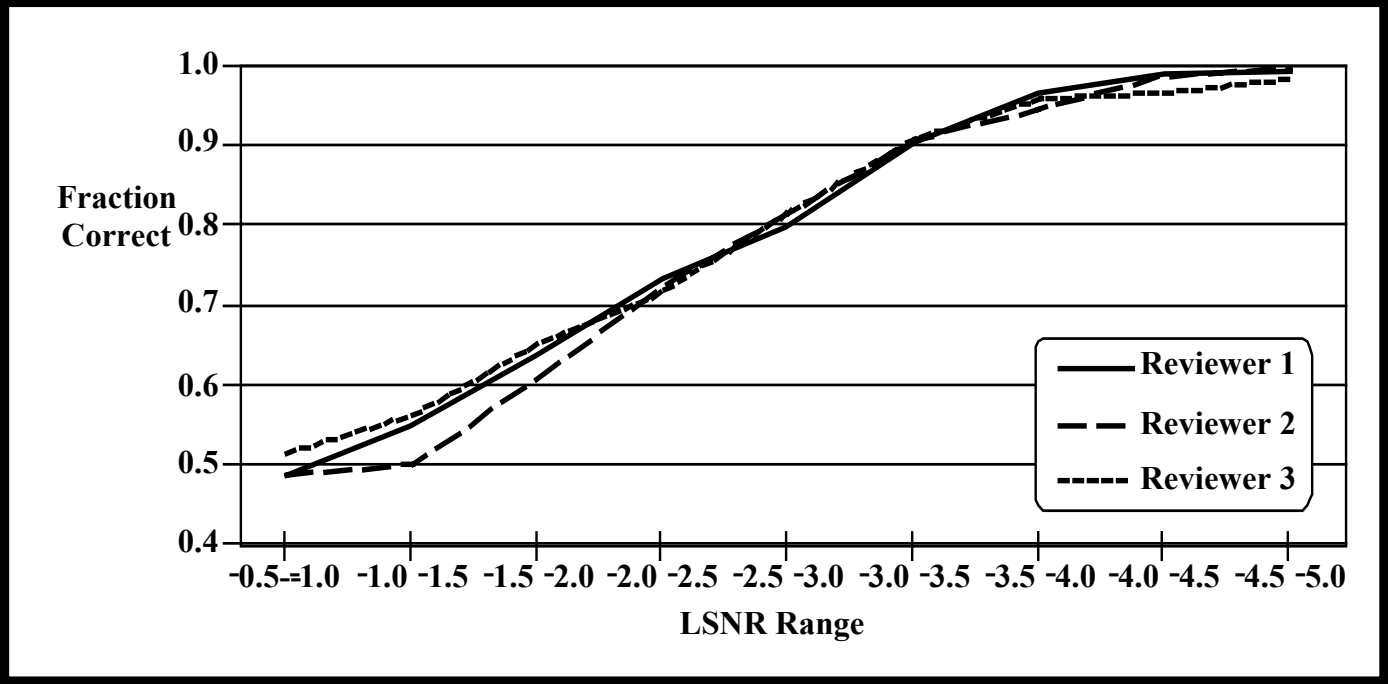
RELATION OF AUTOMATED TO HUMAN OBSERVER DETECTABILITY

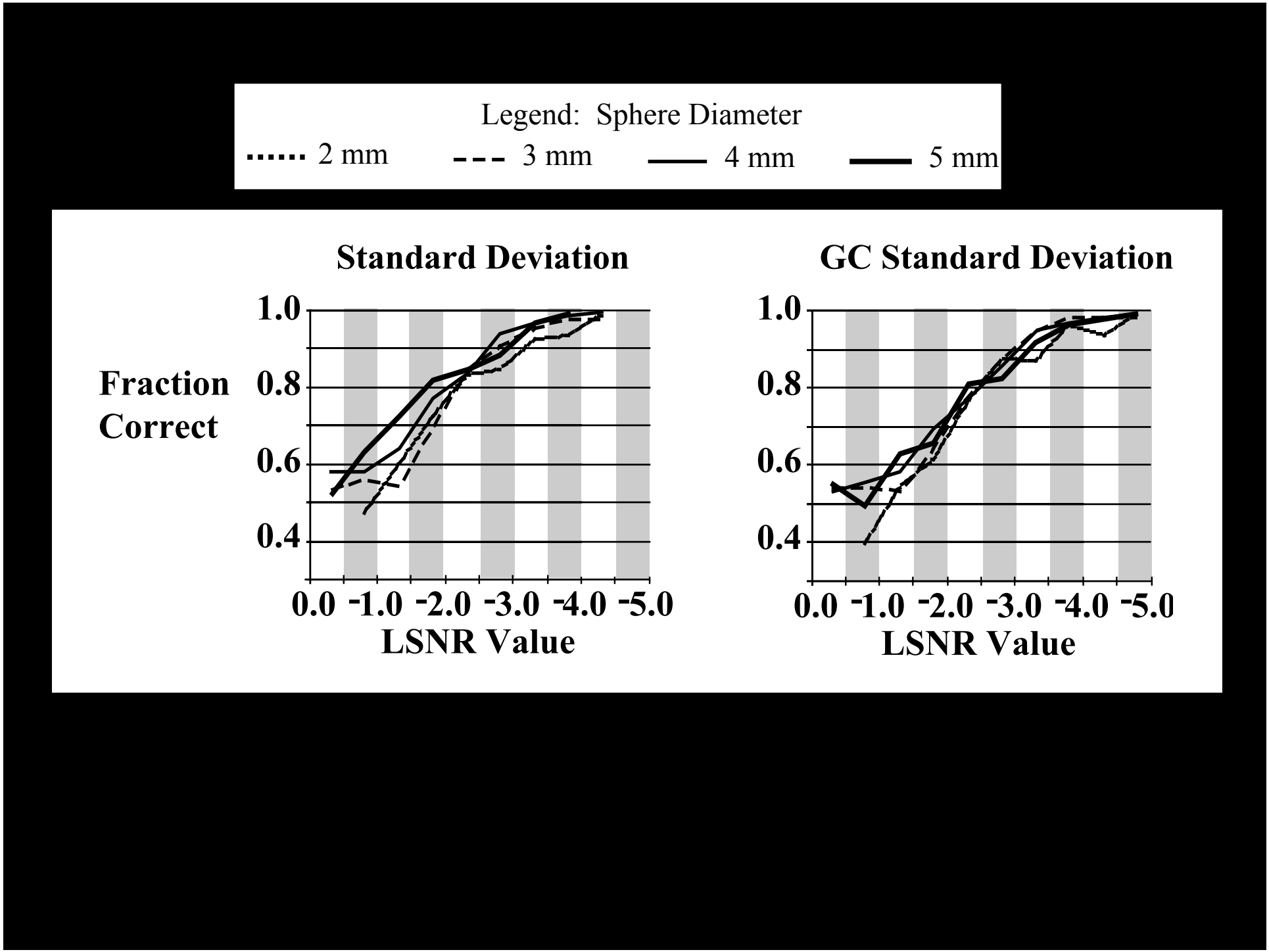
Two-Alternative Forced Choice (TAFC)

- 2, 3, 4 and 5 mm diameter spheres
- Contrasts -3, -6, -9 and -14 dB
- Acuson 128 scanner with:
  - 4 MHz sector (4V1);
  - 5 MHz curved array (5C2);
  - 6 MHz linear array (6L3)

- 5,646 pairs of 1cmx1cm image areas, one containing a lesion with LSNR between 1 and 6 and the other from a background image
- Lesion image on left or right (random)
- 3 human observers *required* to choose image which contains the lesion ≡ Two-Alternative Forced Choice (TAFC)

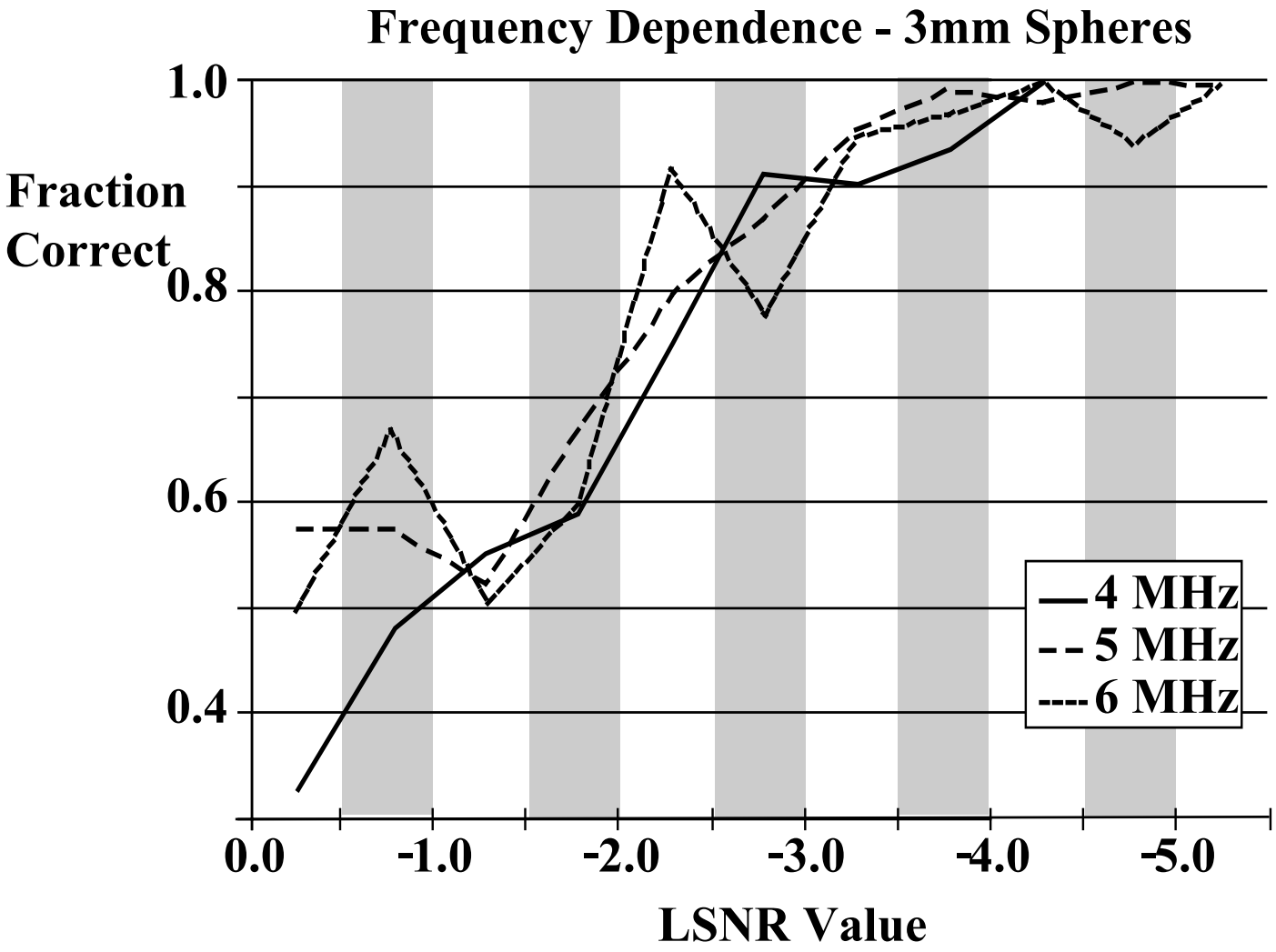

140066
3.0 mm spheres

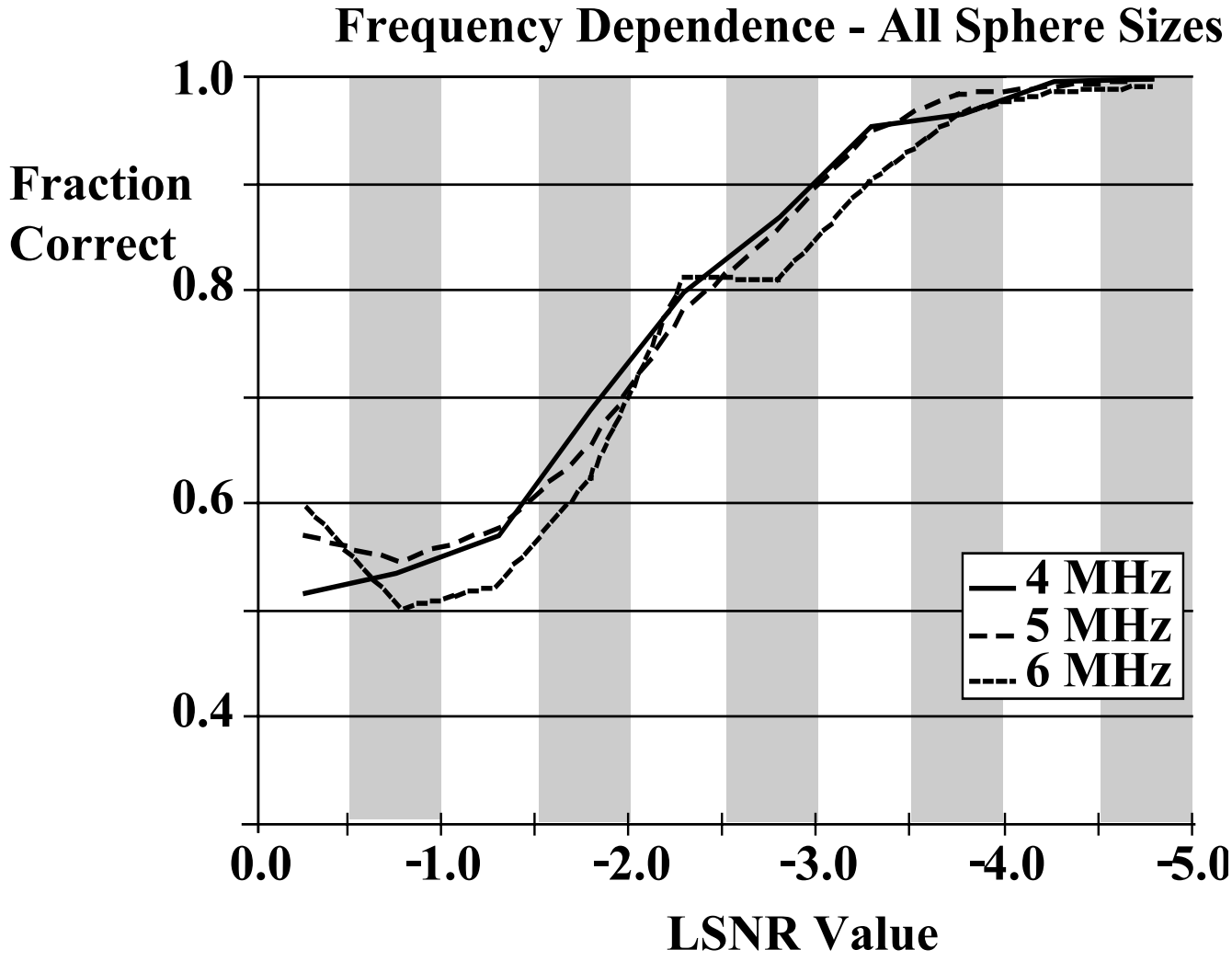

1.0
0.9
0.8
0.7
0.6
0.5
0.4
Fraction Correct
Reviewer 1
Reviewer 2
Reviewer 3
-0.5--1.0
-1.0 -1.5
-1.5 -2.0
-2.0 -2.5
-2.5 -3.0
-3.0 -3.5
-3.5 -4.0
-4.0 -4.5
-4.5 -5.0
LSNR Range


Legend: Sphere Diameter
2 mm
3 mm
4 mm
5 mm
Standard Deviation
GC Standard Deviation
Fraction Correct
1.0
0.8
0.6
0.4
0.0 -1.0 -2.0 -3.0 -4.0 -5.0
LSNR Value

Gradient-corrected standard deviation for LSNR is more independent of lesion size

Thus, choose GCSD over ordinary standard deviation method


Frequency Dependence - 3mm Spheres
Fraction Correct
1.0
0.8
0.6
0.4
0.0
-1.0
-2.0
-3.0
-4.0
-5.0
LSNR Value
4 MHz
5 MHz
6 MHz


Frequency Dependence - All Sphere Sizes
Fraction Correct
1.0
0.8
0.6
0.4
0.0
-1.0
-2.0
-3.0
-4.0
-5.0
LSNR Value
4 MHz
5 MHz
6 MHz

Selection of threshold somewhat subjective

-- fraction correct of 0.7 on TAFC curve
is a reasonable detectability threshold

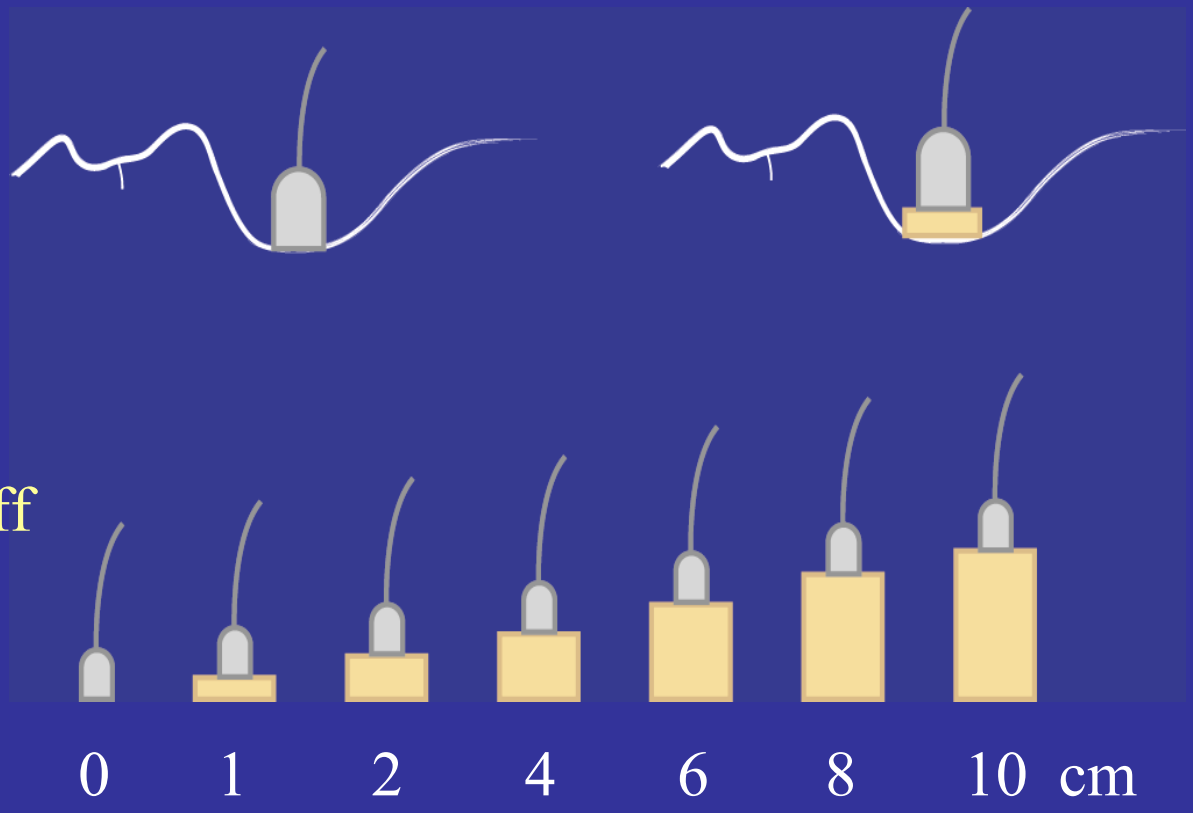
-- corresponds to an LSNR value of -2

STATISTICAL CONSIDERATIONS

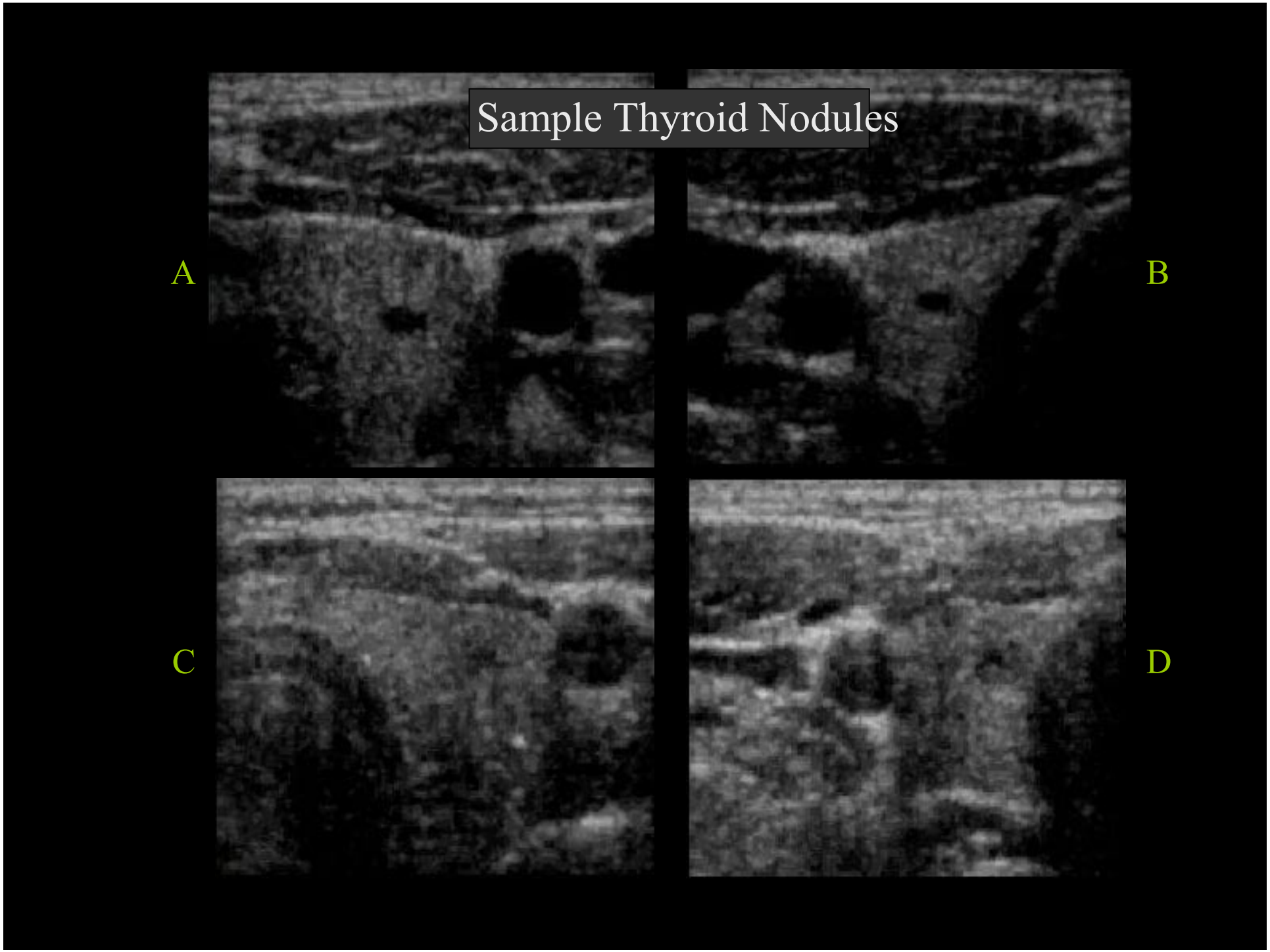
- More than one independent realization of LSNR value needed at a given depth

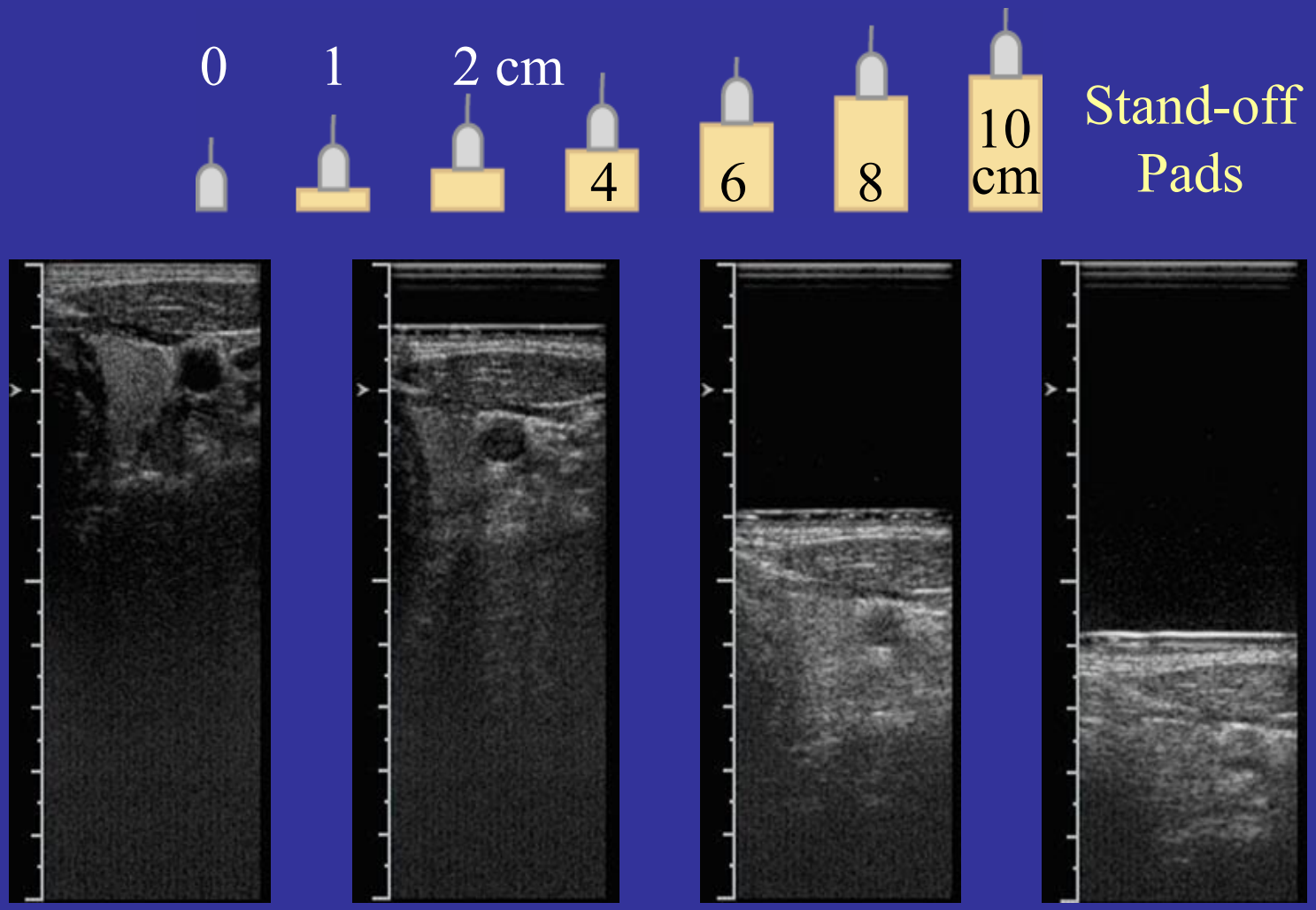
- Average of 8 realizations $\Rightarrow$ standard error in LSNR of $\pm$ 0.3 $\Leftrightarrow$ standard error $\approx \pm$ 0.05 in TAFC fraction correct

Clinical Study-Thyroid Nodules: Methods



Sample Thyroid Nodules




0
1
2 cm
4
6
8
10 cm
Stand-off Pads

1. Acquire patient scans using a range of scan parameters
    - Scored by radiologists
2. Acquire phantom scans using identical scan parameters
    - “Scored” by software
3. Compare scores of radiologists and phantoms

Scan Parameters

3 Transducers*

- 6L3 (6 MHz), 5C2 (5 MHz), 4V1 (4 MHz)
- 2 different focal depths each (2cm, 8cm)
- Other parameters constant

"Scan Configuration": Transducer/Focal Zone Combination

* Acuson (Mountain View, CA)

Patient Scans

IMAGING

-- 4 thyroid nodules

-- Various depths

-- Recorded 6 second cine loops

- Masked and coded
- Single movie (each scan configuration)
- Reference sample (best clip of all)

REVIEW

-- 9 Radiologists

-- Score 1 (not detectable) to 5 (best)

Phantom Scans

IMAGING -- 7 unique spherical lesion phantoms

2mm @ -14 dB; 3mm @ -6, -9, -14 dB;

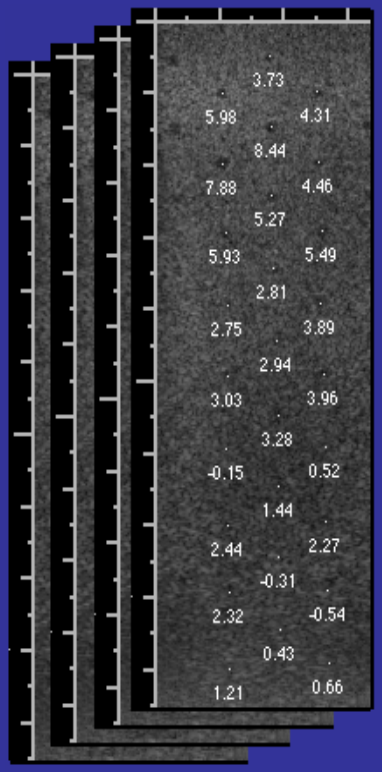
4mm @ -9, -14 dB; 5mm @ -9 dB

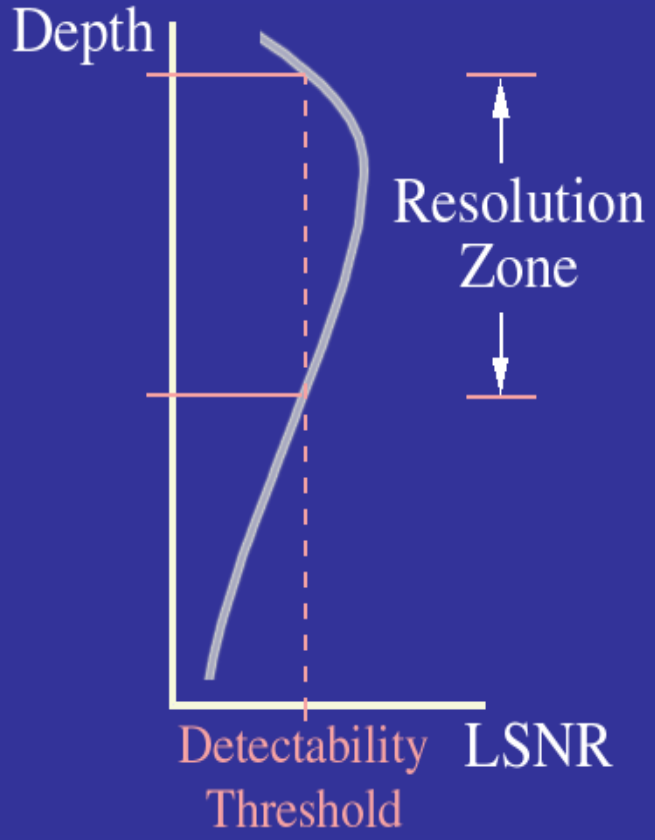
-- at least 8 independent realizations for each sphere size, contrast and depth interval

REVIEW -- Automated lesion signal-to-noise ratio (LSNR)

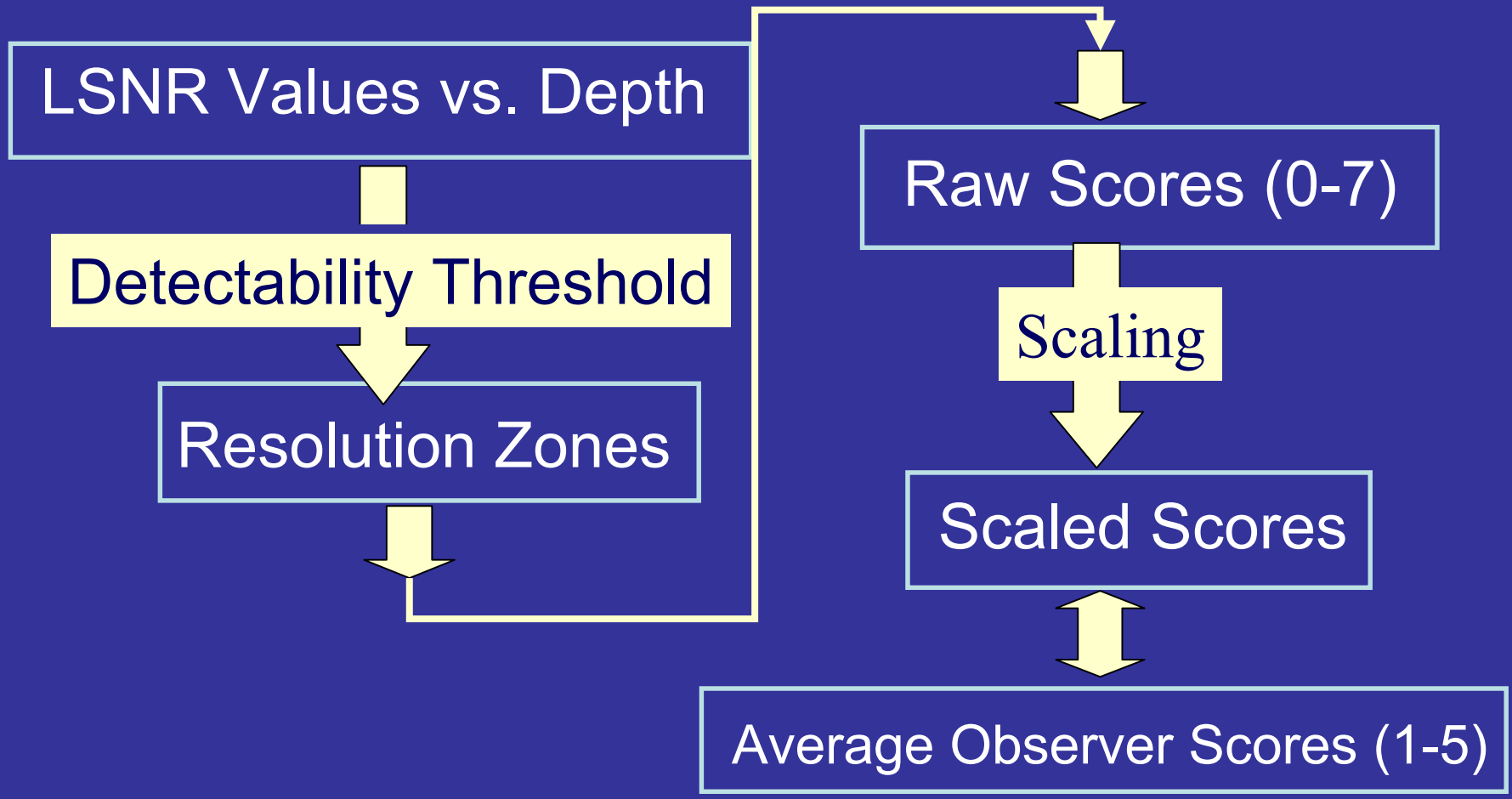
-- Score from “resolution zone”

Resolution Zone

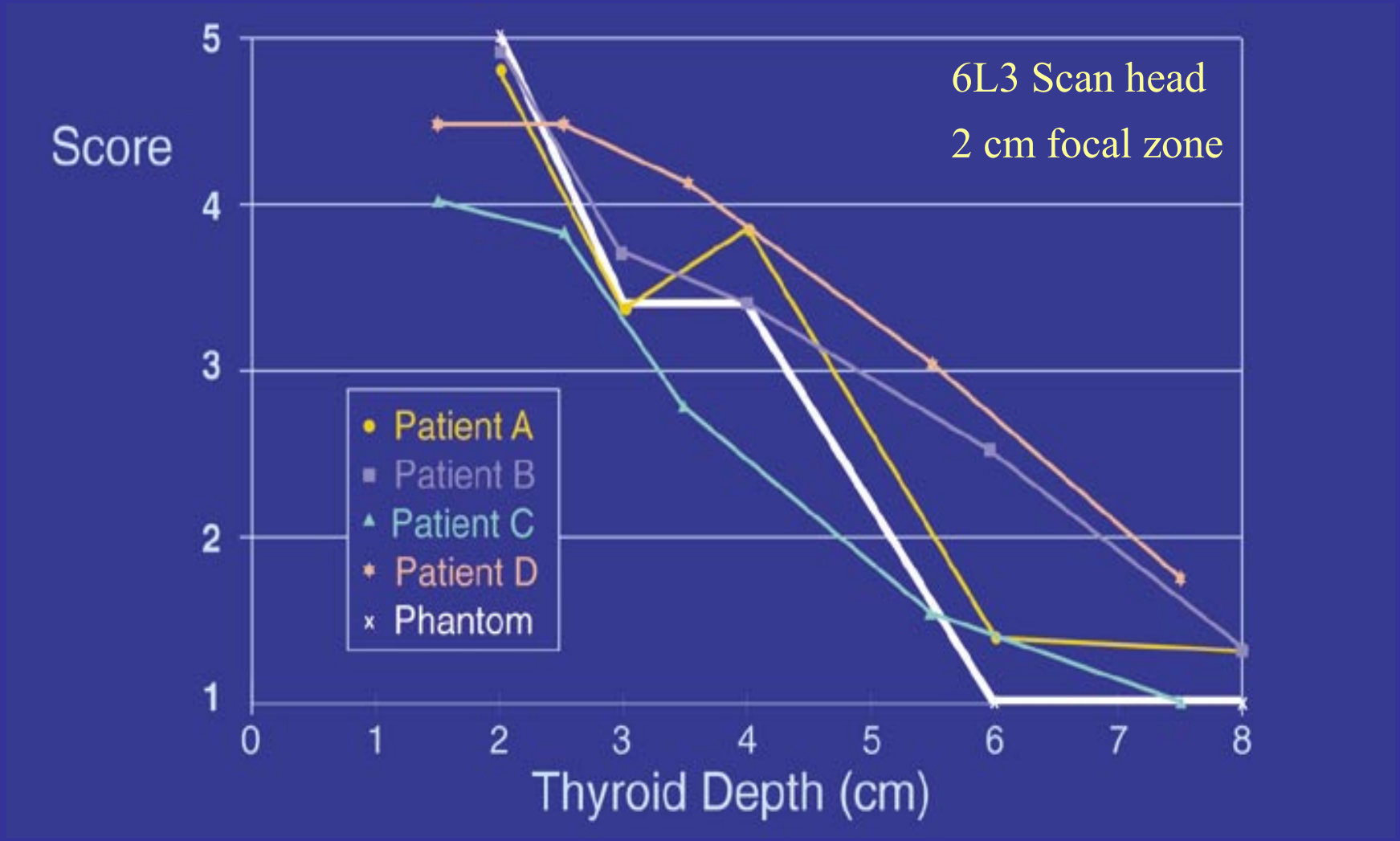




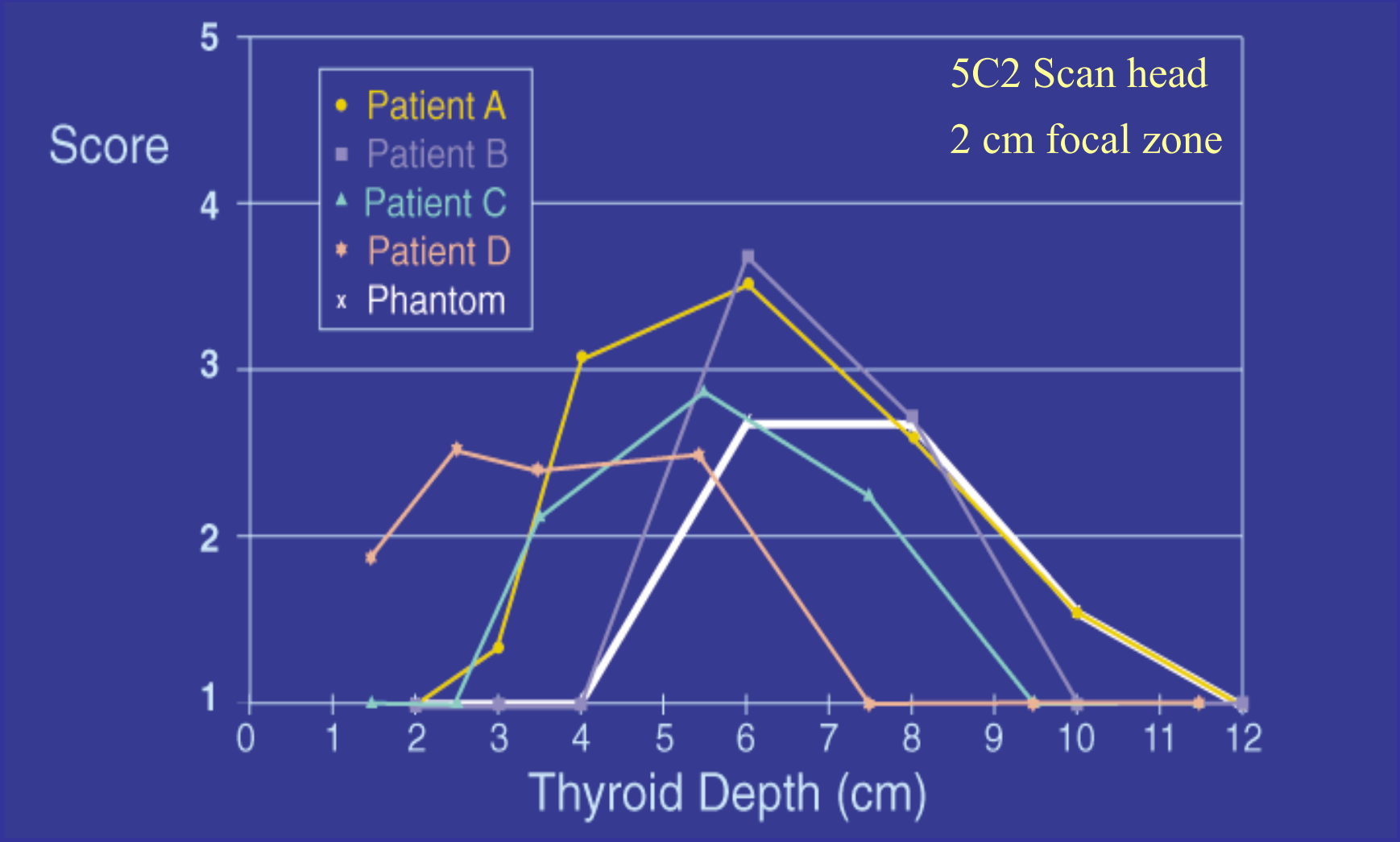
Results: Analysis Review



Thyroid Study: Some Results



Thyroid Study: Some Results



Kappa Coefficient

A statistical measure of agreement between two sets of data which suppresses biasing due to chance agreement.

Kappa Value	Interpretation*
Below 0.00	Poor agreement
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

Landis, J. & Koch, G. G. (1977). The measurement of observer agreement for categorical data

Biometrics, (33), 159-174.

$$\text{KAPPA COEFFICIENT} \equiv K \equiv \frac{p_o - p_e}{1 - p_e}$$

$$p_o \equiv \sum_{i,j}^{n} p_{ij} w_{ij} \qquad p_o \equiv \sum_{i,j}^{n} p_{i\bullet} p_{\bullet j}$$

$$p_{i\bullet} p_{\bullet j} \equiv \sum_{k=1}^{n} p_{ik} p_{kj}$$

$n \equiv$ no. of unique thyroid nodule depictions

$p_{ij} \equiv$ fraction of n scored i by humans and j by phantom/automation

$w_{ij} \equiv$ weighting factor (1 if i=j; 0.5 if $|i-j| = 1$; 0 if $|i-j| > 1$)

Scoring Phantom Images

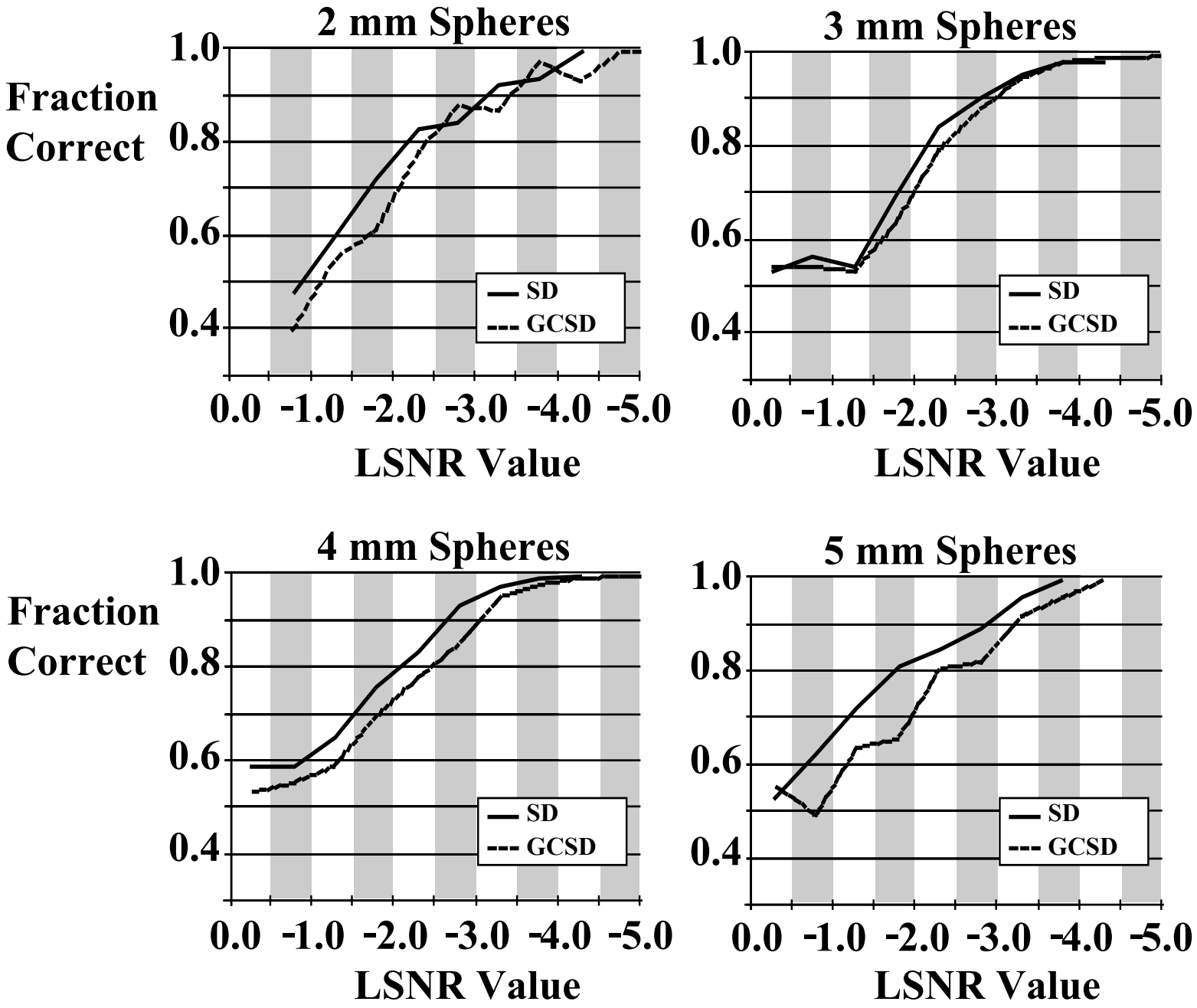
- Function of Depth
- Raw Score
  - Resolution Zone includes depth of interest: 1 point
  - 7 total phantoms, 7 max. points
  - Need to map raw scores to clinical scores (1-5)

Results: Kappa Coefficients

LSNR threshold	TAFC fraction correct	Kappa coefficient	agreement level
-2	0.7	0.33	fair
-3	0.88	0.58	moderate

Conclusions

- Clinical and Phantom/automation agreement is at least “fair” for an LSNR threshold of -2
- Agreement is “moderate” for an LSNR threshold of -3
- Perhaps the LSNR threshold of -3 is the better choice
- More patient studies required


2 mm Spheres
3 mm Spheres
4 mm Spheres
5 mm Spheres
Fraction Correct
1.0
0.8
0.6
0.4
SD
GCSD
0.0 -1.0 -2.0 -3.0 -4.0 -5.0
LSNR Value