Purdue University

Purdue e-Pubs

Weldon School of Biomedical Engineering Faculty Working Papers

Weldon School of Biomedical Engineering

1995

Clinical ROC Studies of Digital Stereo Mammography

Tarek Z. Elabbady

Charles F. Babbs
Purdue University, babbs@purdue.edu

Valerie P. Jackson

Edward J. Delp

David M. Chelberg

See next page for additional authors

Authors

Tarek Z. Elabbady, Charles F. Babbs, Valerie P. Jackson, Edward J. Delp, David M. Chelberg, Zygmunt Pizlo, and Leslie A. Geddes

Brief Report

Draft 8-17-95

Clinical ROC Studies of Digital Stereo Mammography

TAREK Z. ELABBADY, Ph.D.

CHARLES F. BABBS MD, Ph.D.

VALERIE P. JACKSON, MD

EDWARD J. DELP, Ph.D.

DAVID M. CHELBERG, Ph.D.

ZYGMUNDT PIZLO, Ph.D.

LESLIE A. GEDDES ME, Ph.D.

From the Hillenbrand Biomedical Engineering Center

and the Schools of Science and Electrical Engineering,

Purdue University, West Lafayette, IN 47907, USA

Address for correspondence:

Charles F. Babbs MD, PhD
Biomedical Engineering Center
1293 Potter Building
Purdue University
West Lafayette, IN 47907-1293, USA
Tel: 317-494-2995
Fax: 317-494-0811
Email: babbs@ecn.purdue.edu

Running head: ELABBADY ET AL.: STEREO MAMMOGRAPHY

Key words (not in title): breast, computer simulation, imaging, mammography, phantom

Abstract

The objective of this study was to explore and document the diagnostic utility of digital stereo mammography for the detection of localized breast cancer in women. In it we characterized the ability of experienced mammographers, general radiologists, and non-radiologists to detect three types of tumor masses embedded within a heterogeneous background of normal tissue elements in numerically simulated digital mammograms. The simulated mammograms were displayed to the subjects on a high resolution video display, both in stereo mode and in mono mode. Half of the mammograms contained a single tumor, ranging from 0.3 to 0.8 cm in maximal diameter. Each reader rated 120 images (60 in stereo and 60 in mono) as to the probability of abnormality on scale of 1-5. Observer responses were evaluated using receiver operating characteristic (ROC) analysis to characterize any difference in diagnostic performance between the two viewing modes. The synthesized mammograms and the digital display were highly rated by the participant radiologists as promising tools for future research. The results of ROC analysis, however, indicated no significant difference in tumor detection when the same readers utilized the stereo mode versus the mono mode (A_z mono = 0.833 versus, A_z stereo = 0.826). The results were similar for readers of all 3 experience levels--mammographers, general radiologists, and non-radiologists.

Introduction

The major role for mammography is the early detection of breast cancer in asymptomatic women. Breast cancer is now projected to affect one woman in nine,[1] accounting for 32% of incident cancers.[1] Prior to metastatic spread, breast cancer is a regional disease which can often be cured by surgery or radiation. After metastatic spread, however, it becomes a generalized disease that is resistant to aggressive regimens of chemotherapy. The probability of metastasis is directly related to the size of the primary lesion.[2] Hence, a highly effective means to diminish breast cancer mortality is earlier diagnosis.[3]

Stereo perception, or stereopsis, refers to the impression of visual depth created by binocu-

lar parallax or disparity in the images cast upon the right and left retinas.[4] The two dissimilar images are fused in the visual centers of the brain to give a three dimensional appreciation of depth. The fact that we can see well with one eye indicates that monocular cues such as linear perspective, occultation, shading, shadow, and texture can also provide a sense of depth.[5,6,7] However, in the domain of x-ray imaging, radiologists have few or minimal monocular depth cues. In mammography in particular, the observer may not know the exact shape or form of the tumor being searched for. In this case binocular stereo vision may be of great benefit in depth perception; because it provides the most powerful depth cue for scenes viewed close at hand, which can function independently of the recognition of objects.[8] Stafford Warren in 1930 was the first to report the use of a stereoscopic technique for breast radiography in 119 patients, who then underwent surgery. He stated, "In many of the cases, there was no uniformity of opinion in the preoperative clinical diagnosis.... the opinion from the mammogram, on the other hand, was often very definite and most frequently correct." Technical difficulties and high radiation dose requirements of the day, however, dampened enthusiasm for stereo techniques in mammography. Recent improvements in the technology of radiologic detectors, computers, and video displays[9,10] have made realistic high quality three dimensional x-ray imaging possible. Moreover, in non-medical applications stereoscopic displays have been shown to provide significant improvements, not only in seeing where things are, but also in seeing what things are, especially in an unfamiliar or complex scenes. Makee, Levi and Bowne have suggested the stereopsis is especially useful to break camouflage. Nakayama, Shimojo and Silverman showed that stereoscopic depth plays an important role in delineating and linking parts of an object that is partially occluded by other objects, would be the case for early breast cancers embedded in a matrix of normal parenchyma. Accordingly, we conducted a study to explore the impact of stereo viewing upon diagnostic performance of trained readers, faced with the task of breast cancer detection.

Material and Methods

Overview

To systematically investigate the effectiveness of stereo viewing for the detection of abnormalities, we developed novel software[11] that creates high resolution, virtual digital mammograms from a computational models of the human breast that include branched lobulated ducts and suspensory ligaments, embedded in fatty subcutaneous tissue (Figure 1). They may also include any of three types of computer simulated tumors (fibroadenomas, invasive ductal carcinomas, and intraductal carcinomas (Figure 2). Virtual mammograms are generated by computing x-ray transmission through a mathematically defined, three dimensional tissue space according to Beer's Law, using fast ray-tracing algorithm (Figure 3). The advantages of working with computer generated images, rather than actual clinical radiographs, are that (1) the true normal / abnormal state of the images is known exactly, because the abnormalities are deliberately created and mathematically defined; (2) the size distribution of possible tumors is unlimited, and the nature, background, and context of the abnormalities can be systematically varied to determine under what circumstances perception and diagnostic performance are most and least influenced by particular display techniques; (3) the cost of obtaining images for analysis (at any desired resolution prior to display) is minimal; (4) for many psychophysical experiments, the complete clinical process of image acquisition, display, and interpretation can be simulated at a computer-based work station for efficient, objective data collection and analysis; and (5) the fundamental questions regarding the various digital radiographic techniques as aids to human perception and diagnostic performance can be answered qualitatively without exposing human subjects to additional radiation.

Video Display

To display the synthesized mammograms in either stereo mode or mono mode, we used a SPARCstation 20 with a 20 inch video monitor having 1,280 dots horizontally and 1024 lines vertically (Sun Microsystems, Inc. Mountain View, CA). The monitor is synchronized to time

gated liquid crystal (LCD) eyeglasses (StereoGraphics, San Rafael, CA) that alternately present the right stereo image to the right eye and the left stereo image to the left eye at a rate up to 120 times/sec (Figure 5). The system is equipped with a ZX graphics accelerator, which uses dual video frame buffers to display alternately the right and left images to the screen at a refresh rate of 1/56 second/image. When the LCD lenses are activated via an infrared signal, they become opaque within approximately 5 msec. For one video frame the right eye sees the right-sided image (the left eye vision being blocked by the opaque lens). For the next video frame, the right eye vision is blocked and the left eye sees the left sided image. This display technology permits comfortable viewing of stereo pairs without special training or other apparatus.

Subjects and experimental design

A total of 18 readers participated in the experiment: 5 experienced mammographers, 6 general radiologists, and 7 non-radiologists (the authors, and their graduate students). Each reader participated in two one hour sessions. Each session included 60 images; 30 were displayed in stereo and 30 were displayed in mono. The experiment was designed in UNIX/XWINDOWS environment in a fully automated manner, and the reader's only interaction was through the computer mouse. The reader always wore LCD eyeglasses, and the glasses were switched on during both stereo and mono viewing to maintain constant image brightness across conditions. An introduction to the experiment and instructions for responding examples were presented on the screen, including training images that gave examples of the different types of abnormalities. Every training image was displayed first in the mono mode and then in the stereo mode, and then the tumor was highlighted. These training images served to acquaint the reader with the visual display and the nature of abnormalities to be identified in a given session. The reader could spend as much time as needed to scroll back and forth among the training images. At the end of the training session the reader was automatically instructed to start the experiment.

A total of 120 synthesized images was generated, including 60 mono images, and 60 stereo pairs. Half of the images contained a single tumor, either a fibroadenoma, an invasive ductal carcinoma, or an intraductal carcinoma. Methods of creating tumor models and the details of

tumor composition and geometry have been described previously. The tumors ranged between 0.3 and 1.0 cm in maximal diameter.

The images were presented in two sessions, separated by a rest period. Each session contained 60 images, 15 normal and mono, 15 abnormal and mono, 15 normal and stereo, and 15 abnormal and stereo. The results from the two sessions were combined before ROC analysis. In one session the stereo images were displayed first, followed by the mono images; while in the second session the mono images were displayed first followed by the stereo images. The order of viewing the two sessions was picked randomly for each reader to eliminate any training bias. The time limit was set to one minute per image to simulate the time pressure of clinical practice.

Standard 5-response format for ROC studies[12] was utilized: for each image the reader was instructed to use the mouse to select one of the following ratings.

/ 1--Normal / 2--Probably Normal / 3--Not Sure / 4--Probably Abnormal / 5--Abnormal /

The rating scale was displayed as long as the reader viewed the image. Once the reader selected a rating, an on-screen feedback message was displayed to inform the reader whether the previous image was normal or abnormal. Such feedback is a standard element in psychophysical experiments; without feedback, subjects' prior experiences, which are not under the experimenter's control, may have an effect on the experiment.[13] Readers were informed at the beginning that approximately half of the images in each session were abnormal.

Data Analysis

Receiver operating characteristic (ROC) analysis has been accepted as the most rigorous and objective means of comparing diagnostic imaging modalities in radiology.[14,12,15] In mammography-related research, ROC analysis has been used to characterize the accuracy of mammography,[16] and to compare the performance of mammography and palpation.[17] ROC analysis has also been utilized in a study on the effect of attention-cueing on breast cancer detection performance.[18,19] The ROC curve is a plot of the true positive fraction of readings or "hit rate" as a function of the false positive fraction of readings or "false alarm rate". In the present

study true positive and false positive fractions were derived from 5-response rating data as described by Swets and Pickett[15] and by Metz[12] on the basis of a simple response threshold model. According to this conceptual model, a radiologist or an observer decides to render a positive or negative diagnosis by comparing his or her confidence concerning each image with an internal decision criterion, which can vary among observers. If confidence in a positive diagnosis exceeds this decision criterion, the image is read as positive and vice versa. More stringent decision criteria lead to lower false positive fractions, and less stringent criteria lead to higher false positive fractions. The advantage of ROC analysis is that one can compare diagnostic performance of observers studying two displays, regardless of differences in individual decision criteria by plotting the entire ROC curve over a range of false positive fractions. If for the same observers, the curve for the new technology lies above the curve for the conventional technology, there is objective evidence that the new technology permits a greater fraction of correct diagnosis, regardless of variability in observers' bias for or against making a false positive diagnosis.

To obtain estimates of smooth ROC functions, two adjustable parameters of binormal ROC curves were fit to experimental data using a maximum likelihood parameter estimation scheme.[15] Next, the A_z index, which represents the area under the binormal ROC curve in the unit square, was computed to describe the overall diagnostic performance of the group of observers for stereo vs. mono viewing formats. This index is a commonly used summary of the diagnostic accuracy of an imaging system. Paired t statistics were computed to test the null hypothesis that mean A_z for stereo images is the same as mean A_z for mono images. To minimize occurrences of degenerate data sets, readers were instructed at the beginning of the experiment to use all categories and to distribute their responses uniformly over the rating scale.[20,21]

Results

Fifteen of 18 readers completed both one hour sessions, and 3 of the 18 did one session only. Table I, lists the A_z indices and the corresponding standard errors for both the stereo and mono sessions. The A_z index values for the three observer groups (mammographers, general

radiologists, non-radiologists) were not statistically different. Figures 5 through 8 show the complete ROC curves for groups of experienced mammographers, general radiologists, non-radiologists, and the combined performance for all 18 readers. The shapes of the composite ROC curves are typical for a visual detection experiment in radiology,[20,21] indicating that an appropriate perceptual task was required of the readers, neither too trivial nor too difficult. There was no significant difference in ROC curve parameters between the stereo and mono modes for any of the three subgroups of readers or for all readers combined.

Subjectively, in response to follow-up questionnaires, the radiologists had no preference between stereo and mono displays. The computational models and digital displays were favorably rated by radiologists, who encouraged use of such tools in future research. Four of 18 readers, however, did report eye strain during stereo viewing, most likely as a result of a departure from the habitual accommodation/ convergence relationship.[22] The eyes are focused on, or accommodate for, the plane of the screen, but are converged in accordance with the value and sign of the screen parallax. Only when screen parallax is zero will accommodation and convergence correlate normally. Departures from the normal accommodation/convergence relationships can be minimized by on-screen horizontal translation of the stereo pairs to shift the zero parallax point to the plane of the screen.[22] In these modified displays half the depth of the mammogram appears in front of the screen and half the depth of the mammogram appears behind it. In preliminary tests this adjustment reduced eye strain for the four readers who complained of it.

Discussion

In theory, stereo displays offer potential advantages for target recognition in complex scenes and in breaking camouflage. In the present study we utilized complex, computer simulated images mimicking clinical mammograms to test the diagnostic utility of stereo displays for early breast cancer detection. The study focused on the relative, rather than the absolute detectabilities of abnormalities using the two viewing modes. Objectively, there was no difference in over-all observer performance by ROC analysis. Thus, the enhancement of depth information provided by stereo, however, did not appear to improve the ability of observers to detect

spherical, irregular, or linear tumor-like abnormalities 0.3 to 1.0 cm in maximal dimension,

embedded in a 6 cm thick matrix of glandular elements and fat lobules. We conclude that tumor

features revealed be stereopsis are not especially important for early breast cancer detection.

Table I: ROC curve areas, A_z, with their standard deviations for the three groups of readers (mammographers, general radiologists, and non-radiologists) and the combined data.

	Mammographers	General Radiologists	Non-Radiologists	All Subjects
Mono A_z	0.869	0.837	0.778	0.833
SD	0.022	0.022	0.030	0.014
Stereo A_z	0.878	0.821	0.789	0.826
SD	0.03	0.026	0.028	0.014

References

1.   Boring, CC, Squires, TS, and Tony, T. Cancer statistics, 1992. *CA Cancer J Clin* 1992; 42: 19-86.

2.   Tubiana, M. The screening of breast cancer: the oncologist's point of view. In: JM Bigot, Ed. *Proceedings of the 17th International Congress on Radiology.* Amsterdam; Exerpta Medica, 1990: 355-358.

3.   Fisher, ER and Paulson, JD. Identification of early changes in breast cancer. In: E Grundmann and L Beck, Ed. *Early Diagnosis of Breast Cancer -- Methods and Results.* New York; Fischer, 1982: 65-73.

4.   Dodwell, PC. Binocular vision and pattern coding. In: *Visual Pattern Recognition.* New york; Holt, Rinehart and Winston, 1970: 120-136.

5.   Livingstone, M and Hubel, D. Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science* 1988; 240: 740-749.

6.   Todd, JT and Akerstrom, RA. Perception of three-dimensional form from patterns of optical texture. *Journal of Experimental Psychology* 1987; 13: 242-255.

7.   Troscianko, T, Montagnon, R, LeClerc, J, Malbert, E, and Chanteau, PL. The role of colour as a monocular depth cue. *Vision Res* 1991; 31: 1923-1930.

8.   Julesz, B. *Foundations of Cyclopean Perception.* Chicago; University of Chigago Press, 1971

9.   Lipton, L. *CrystalEyes Handbook,* Stereographics Corporation 1991.

10.   Woods, A, Docherty, T, and Koch, R. The use of flicker-free television products for stereoscopic display applications. *Proceedings SPIE Conference on Stereoscopic Displays and Applications II* 1991; 1457: 322-326.

11.   Babbs, CF, Chelberg, DM, Hsu, J, and Delp, EJ. A computational phantom to create digital images for research and training in diagnostic radiology. *Investigative Radiology* October,

1993; Submitted

12. Metz, CE. ROC methodology in radiologic imaging. *Invest Radiol* 1986; 21: 720-733.

13. Sperling, G, Dosher, BA, and Landy, MS. How to study the kinetic depth effect experimentally. *Journal of Experimental Psychology: Human Perception and Performance* 1990; 16: 445-450.

14. Metz, CD. Some practical issues of experimental design and data analysis in radiological ROC studies. *Investigative Radiology* 1989; 24: 234-245.

15. Swets, JA and Pickett, RM. *Evaluation of Diagnostic Systems -- Methods from Signal Detection Theory.* New York; Academic Press, 1982

16. Goin, JE, Haberman, JD, Linder, MK, and Lambird, PA. Analysis of mammography: a blind interpretation of BCDDP radiographs. *Radiology* 1983; 148: 393-396.

17. Gohagan, JK, Spitznagel, EL, McCrate, M, and Frank, TB. ROC analysis of mammography and palpation for breast screening. *Investigative Radiology* 1984; 19: 587-592.

18. Astley, SM and Taylor, CJ. Combining cues for mammographic abnormalities. *Proceedings First British Machine Vision Conference* September 1990; 253-258.

19. Astley, S, Hutt, I, Adamson, S, Miller, P, Rose, P, Boggis, C, Taylor, C, Valentine, T, Davies, J, and Armstrong, J. Automation in mammography: computer vision and human perception. *Proceedings SPIE Conference on Biomedical Image Processing* 1993; 1905:

20. Straub, WH, Rockette, H, King, JL, Obrchowski, NA, Good, WF, Feist, JH, Good, BC, and Metz, CE. Training observers for receiver operating characteristic (ROC) studies. *SPIE Medical Imaging IV: PACS System Design and Evaluation* 1990; 1234: 126-130.

21. Dawood, RM, Todd-Pokropek, A, Craig, JOMC, Highman, JH, and Porter, AW. The optimal use of rating scales in ROC analysis. *SPIE Medical Imaging IV: PACS System Design and Evaluation* 1990; 1234: 952-956.

22. Lipton, L. *Foundations of the Stereoscopic Cinema -- A Study in Depth.* New York; Van Nostrand Reinhold, 1982

(a) Fat Lobules   (b) Ducts   (c) Suspensory ligaments of Cooper
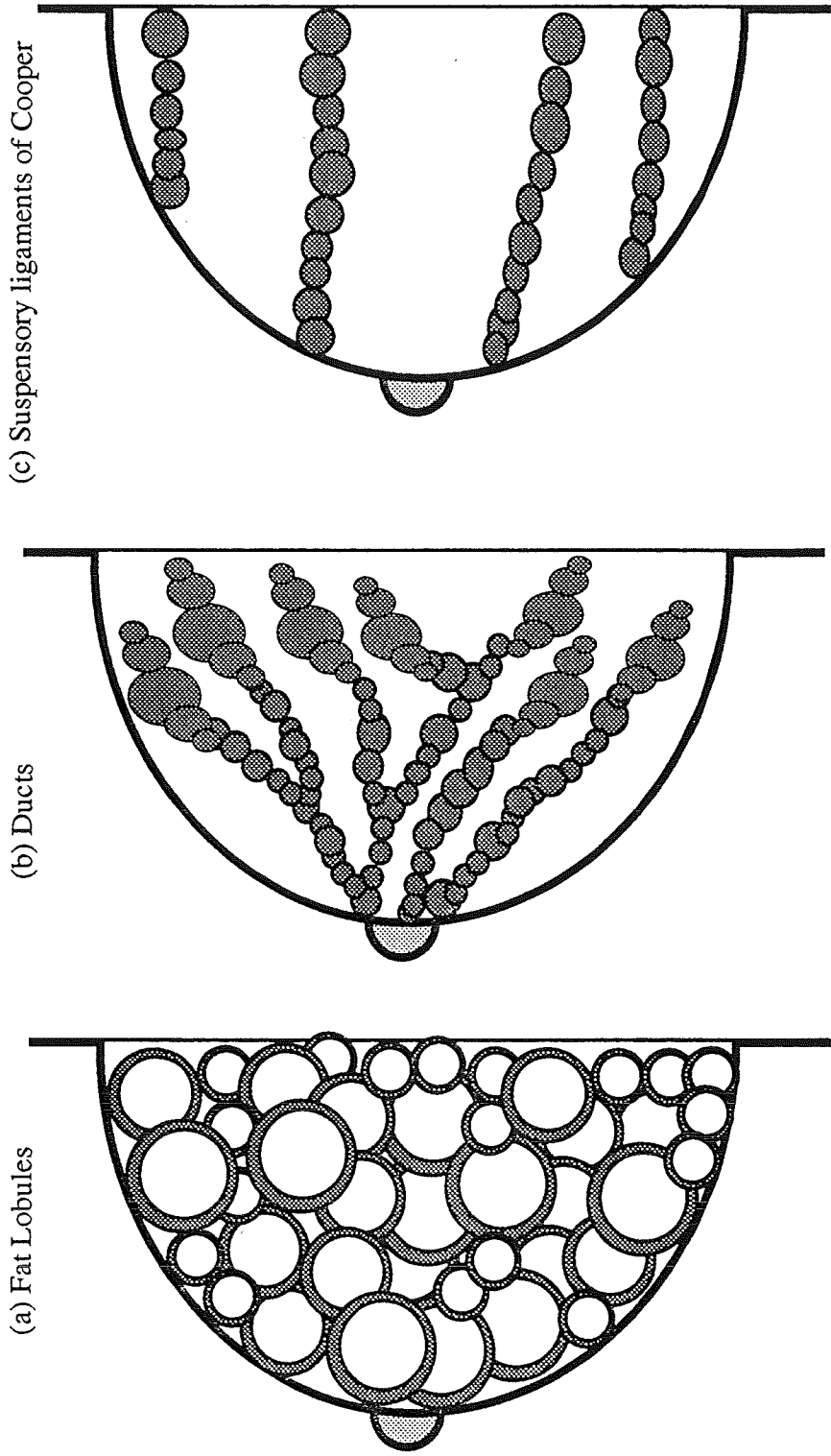
Fig. 2. Scheme for creation of soft tissues of the breast, including fat lobules, branched lobulated ducts, and suspensory ligaments. Fat lobules are created as thin water density shells, surrounding fat density cores. Glands are created as chains of partially overlapping spheres grown from the base of the breast near the chest wall toward the nipple. Fusiform enlargements of chain elements constitute terminal lobules and lactiferous sinuses. Suspensory ligaments are created as telescoped chains of water density spheres.
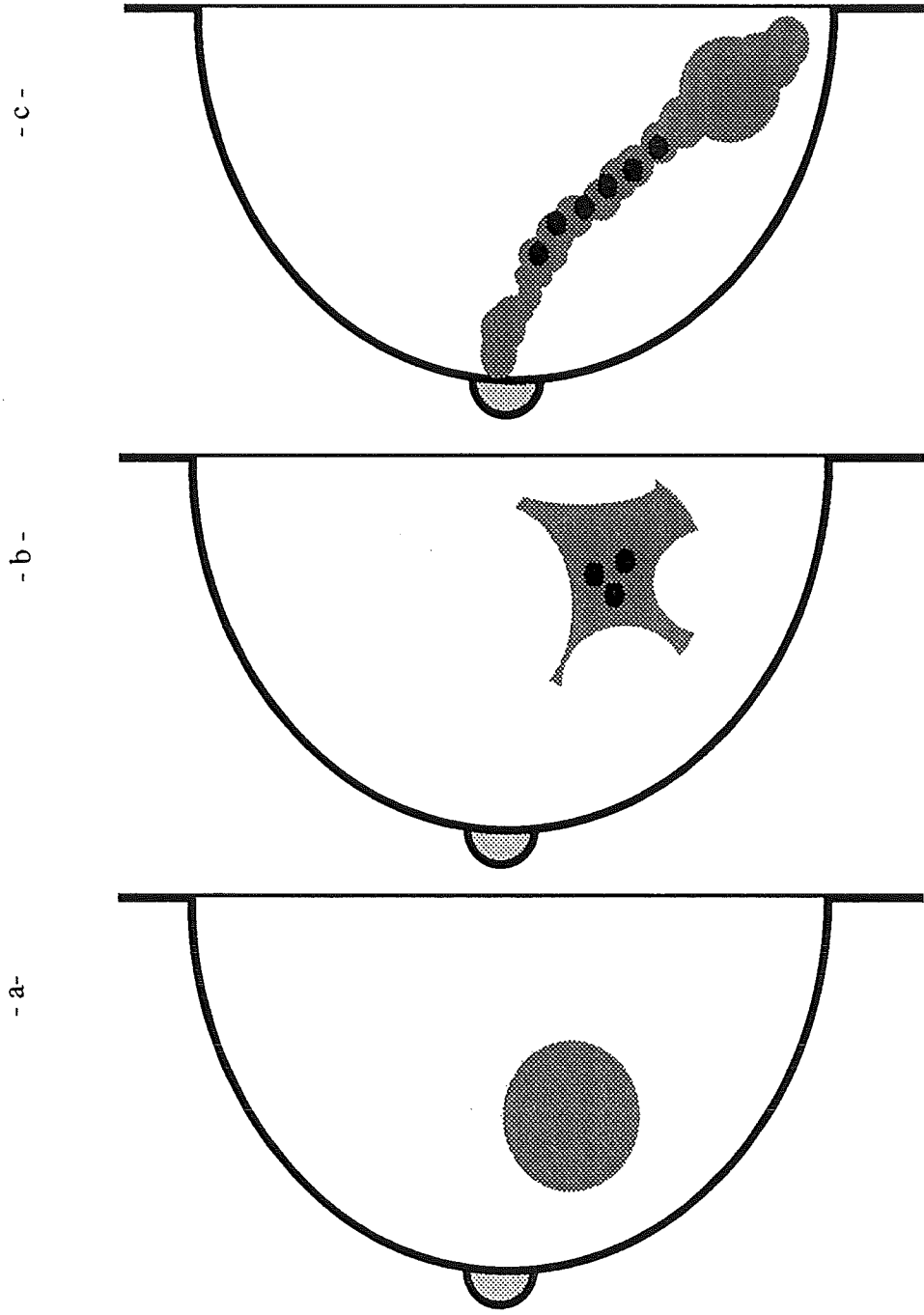
FIG 1:

FIG 2:

Fig. 2. Schemata for modeling 3 types of tumors "enlarged here for clarity". (a) "fibroadenomas" are spherical enlargements of existing duct elements that push away nearby normal elements. (b) "invasive ductal carcinomas" are stellate masses, created by erosion of larger central, water density sphere by overlapping smaller peripheral spheres of fat density. When radii of the peripheral sphere is about 0.5xxx that of the central spheres and when the number is about xxx, then the residual water density object becomes cratered and speculated in appearance, not unlike invasive cancer. 20 (c) "intraductal carcinoma" are indicated by random microcalcifications within normal ducts.
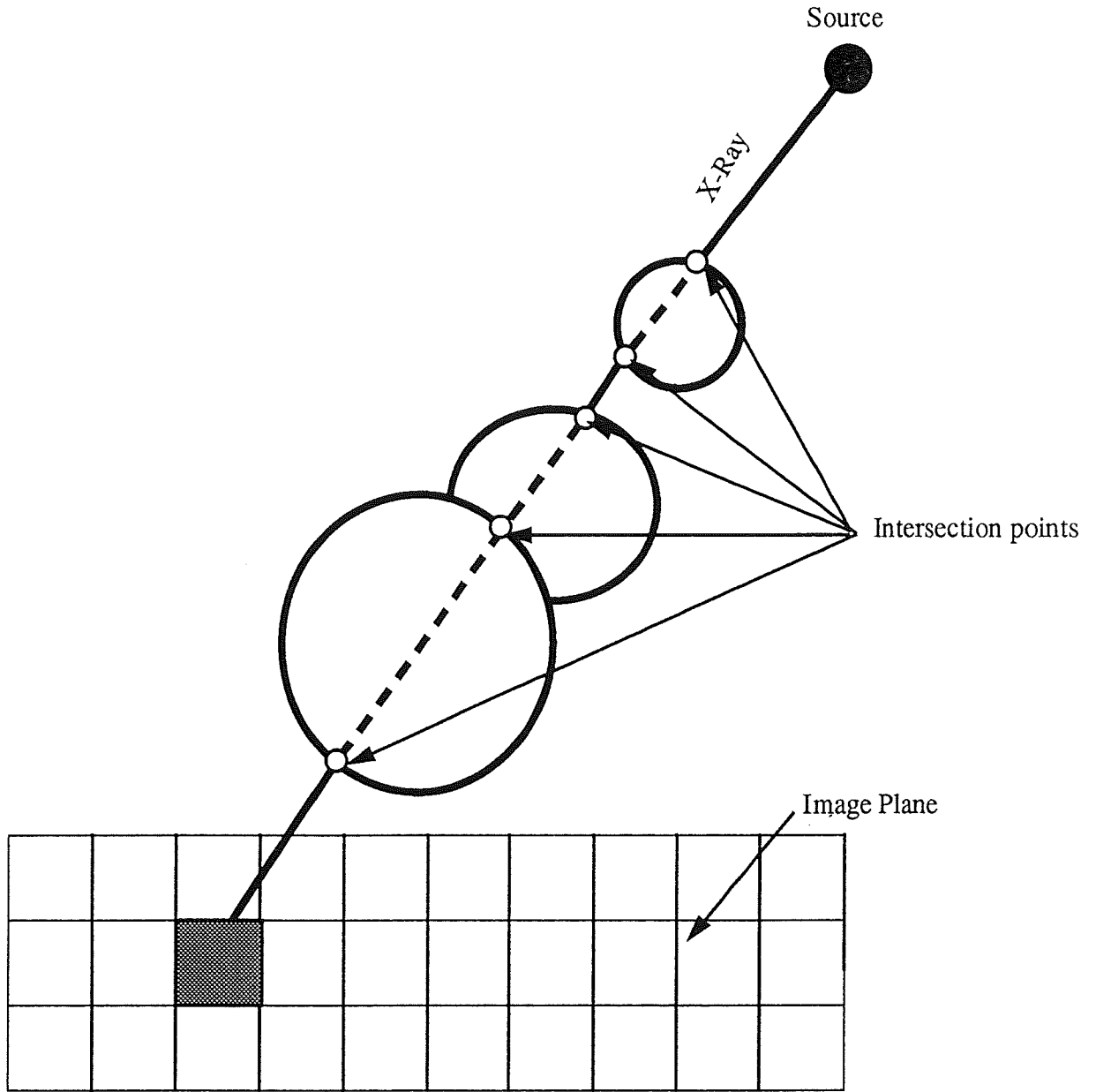
- a -

- b -

- c -

Source

X-Ray

Intersection points

Image Plane

FIG 3   Fig 4. Approach to ray tracing for simulated x-rays. Rays are tested for intersection with a list of spherical primitives. Intersection points are calculated and sorted. Primitives appearing later in the list overwrite primitives appearing earlier in the list. Attenuation of the ray intensity is computed by Beer's lay form the final sorted list of intersection points and the linear attenuation coefficients of corresponding primitives.
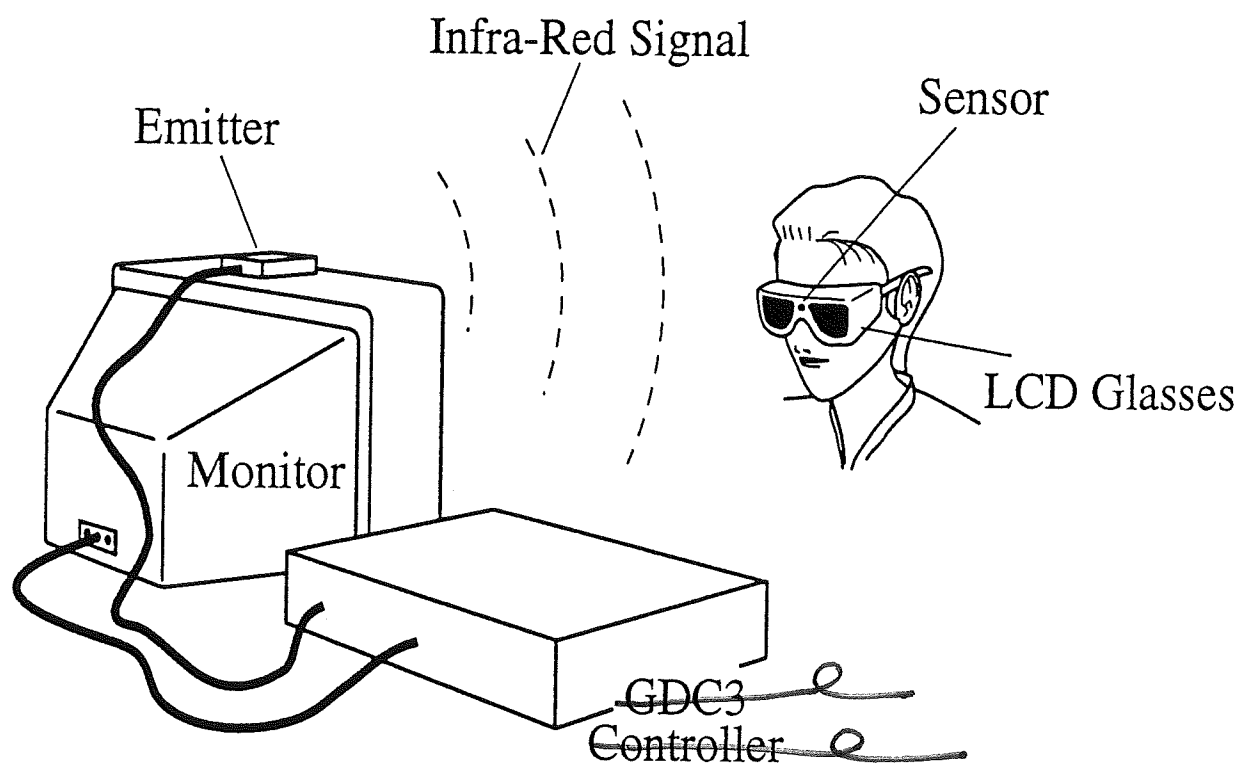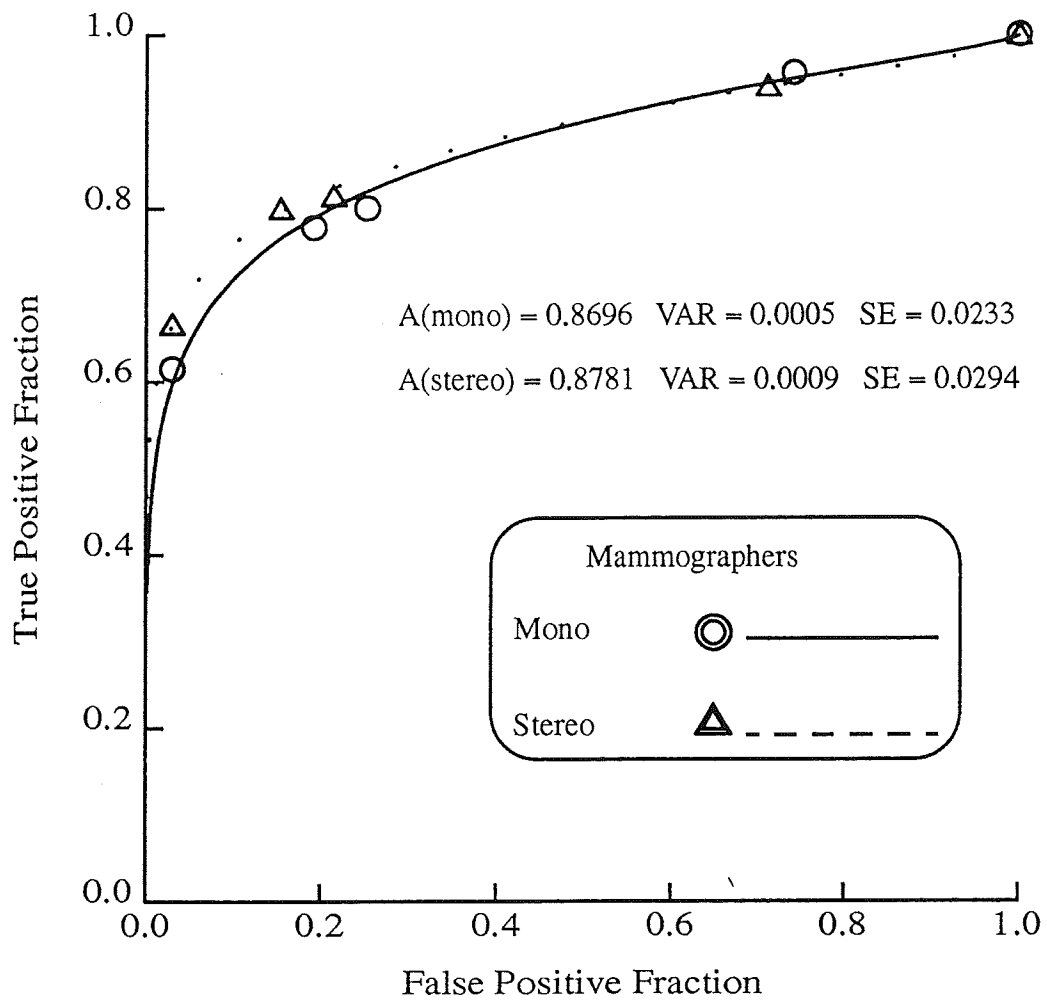
Infra-Red Signal

Emitter

Sensor

Monitor

LCD Glasses

GDC3
Controller

FIG 4:   ~~Figure 5:~~ Experimental Apparatus.

The graph shows a ROC curve plotting True Positive Fraction versus False Positive Fraction.

A(mono) = 0.8696   VAR = 0.0005   SE = 0.0233

A(stereo) = 0.8781   VAR = 0.0009   SE = 0.0294

Mammographers

Mono ⊚ ——————

Stereo △ — — — —

FIG 5!   Fig. 2. ROC curves for subjects with mammography experience only (N=5).

A(mono) = 0.837 VAR = 0.0005 SE = 0.023

A(stereo) = 0.821 VAR = 0.0007 SE = 0.0261

General Radiologists

Mono ◎ ————

Stereo ▲ — — — —
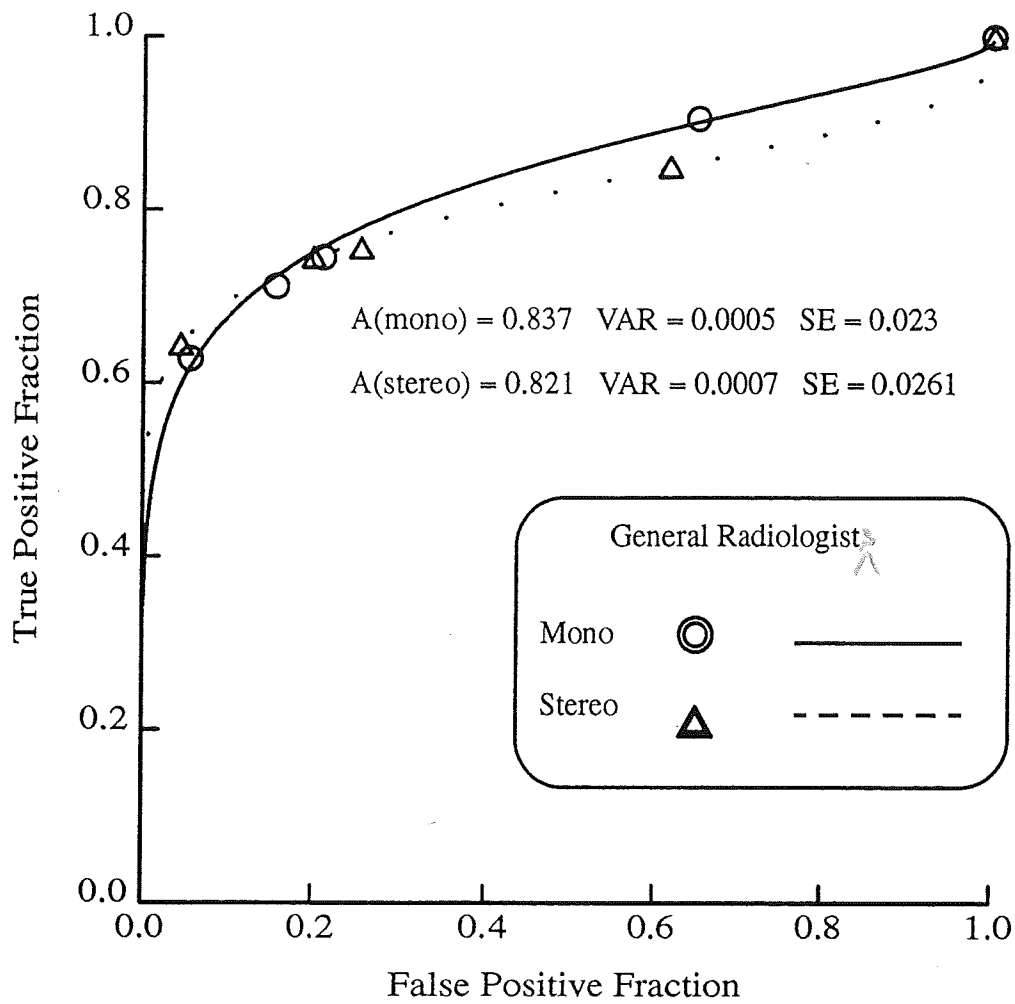
FIG 6: ROC curves for general radiologists only.

A(mono) = 0.7676 VAR = 0.0009 SE = 0.0301

A(stereo) = 0.7887 VAR = 0.0008 SE = 0.0282
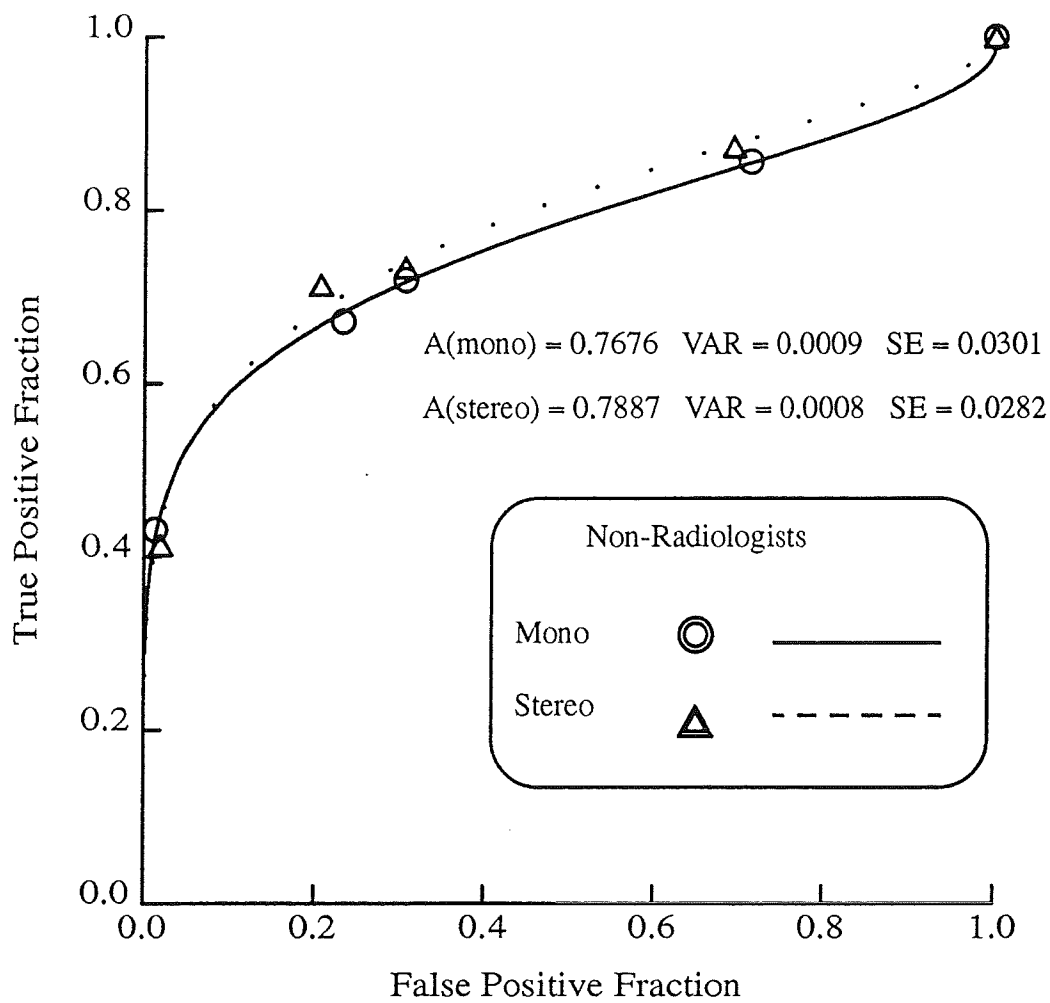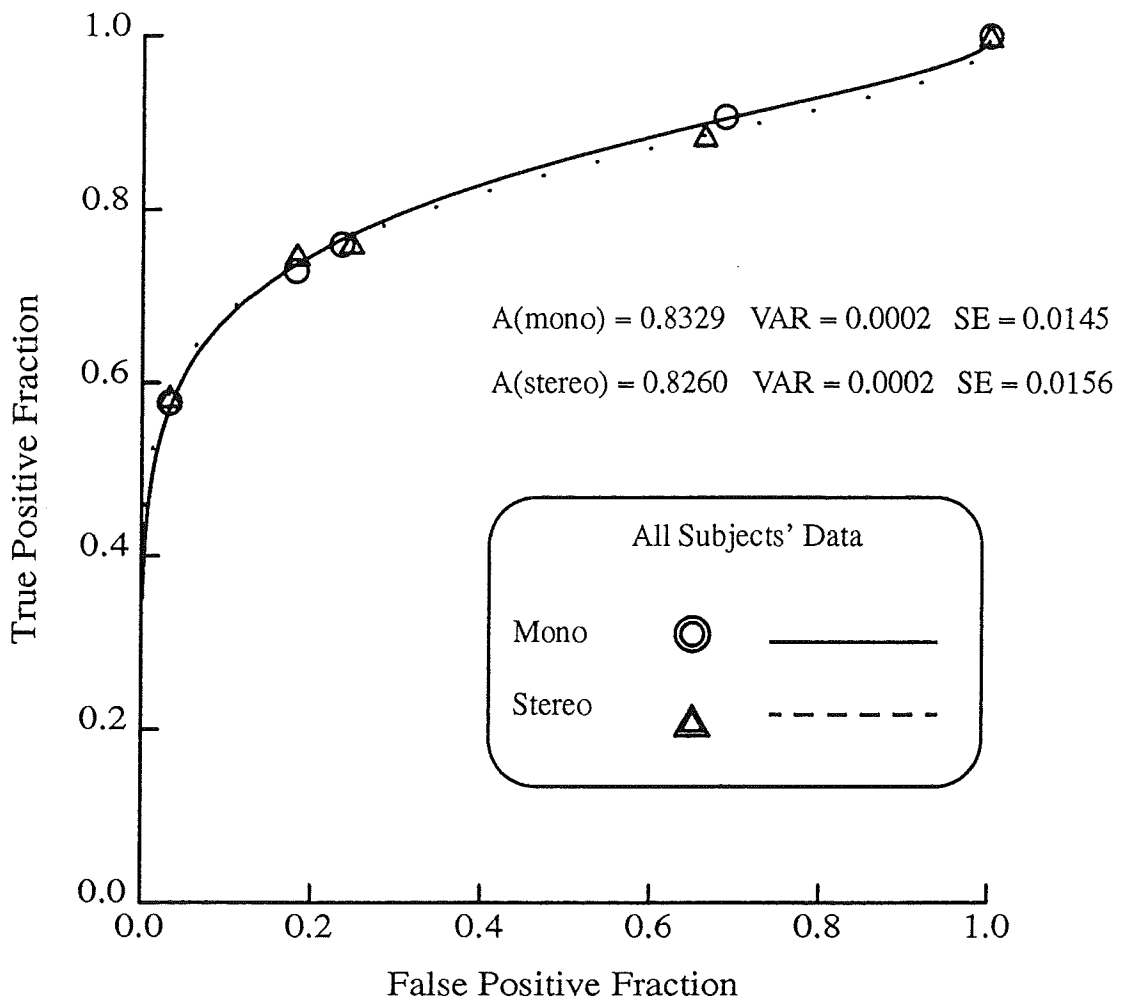
Non-Radiologists

Mono

Stereo

FIG 7: ROC curves for non-radiologists only.

FIG 8: ROC curves for all
subjects combined.