A Comparison of Four Standard Scintigraphic TV Displays: Concise Communication

Alexander S. Houston

Royal Naval Hospital, Haslar, Gosport, Hants, United Kingdom

Four displays (pseudocolor mappings) available on a standard color TV are compared using two series of images (100 normal brains and 100 normal livers) with 50 computer-simulated lesions superimposed on each set. Four observers viewed the sets of images in such a way that the order of the display methods for both organs formed two orthogonal Latin squares. The observers were asked to locate and rate, on a standard scale, the most apparent area in each image, and ROC analysis was applied to the results.

The "heated-object spectrum" was shown to be a useful display for brain images. It was also shown that the choice of display depends on the organ to be imaged.

J Nucl Med 21: 512–517, 1980

The use of color TV displays* (*1*) for diagnostic purposes in nuclear medicine has frequently been discussed (*2,3*). One particular type of color display, the so-called "heated-object spectrum," has proved useful in ultrasound imaging (*4,5*). A recent study has shown that, for a given observer, this display will produce a greater "perceived dynamic range" than a gray-scale display, i.e., there will be a greater number of "just discernible intensity levels," where this parameter is well defined (*6*). However, a comparison of both displays with 64 intensity or color levels,† using images of a series of liver phantoms with areas of either increased or decreased activity showed that in this case there was little to choose between the methods (AE Todd-Pokropek, unpublished data).

The object of this paper is to compare different displays available on a particular (yet fairly standard) type of TV. Color is produced by three guns (red, green, blue), which may be fired at 16 intensity levels (0–15). Gray is produced by firing the three guns at the same level. Thus only 16 levels of gray are available out of a total of 4096 different color levels.

Received Aug. 7, 1979; revision accepted Nov. 16, 1979.

For reprints contact: Alexander S. Houston, PhD, Dept. of Nuclear Medicine, Royal Naval Hospital, Haslar, Gosport, Hants, U.K.

The questions asked were:

1. Do the greater design variability available for color displays, and the fact that more than 16 color levels may be used in a given display, increase diagnostic capabilities when compared with 16-level (once-cycled) gray-scale displays?
2. Is the "heated-object spectrum" an improvement on existing displays?
3. Do important differences occur when different organs are viewed?

MATERIALS AND METHODS

The method of superimposing computer-simulated lesions onto images of normal organs has been described previously (*7,8*). To investigate various processing and display methods in scintigraphy, we have used a series of 100 brain images (matrix size 64 × 64), 50 with single spherical areas of increased activity superimposed, and another of 100 liver images (64 × 64), 50 with single spherical areas of decreased activity superimposed (*8–10*). One important omission from this survey, however, was the comparison of several types of display on a single color TV.

In the present study a 625-line color picture monitor‡ was used in the color mode. The luminances of the 16

TABLE 1. ESTIMATED LUMINANCE (BRIGHTNESS) OF EACH AVAILABLE GRAY-LEVEL ON THE COLOR TV MONITOR

Gray level	Luminance* (candela/m²)
0	<1
1	<1
2	1.4
3	2.9
4	5.8
5	9.4
6	13
7	19
8	27
9	36
10	45
11	55
12	65
13	76
14	87
15	99

* Luminances <1 candela/m² not measurable on spotmeter.

N.B. Full image area illuminated at 60 cm from spotmeter.

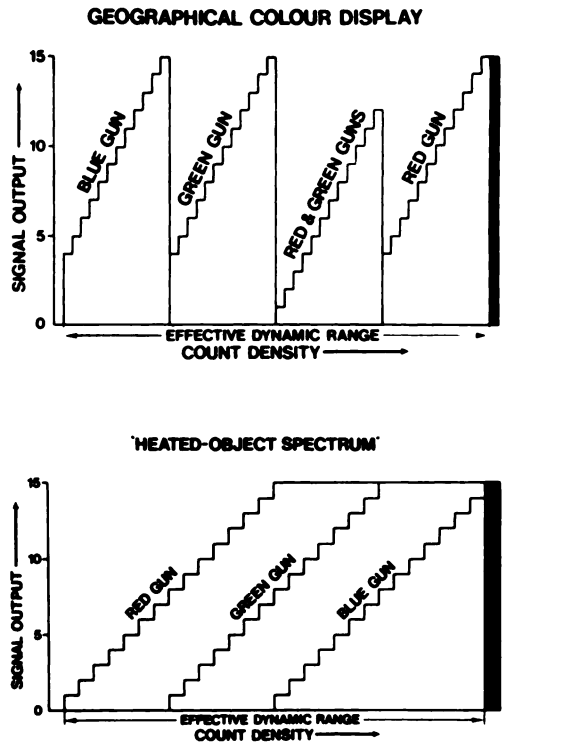

FIG. 1. Signal output, the level at which each gun is fired, is shown as a function of count density for "geographical" and "heated-object spectrum" displays. Blackened regions correspond to all guns firing at level 15, i.e., producing white. "Geographical" display is made up of four distinct phases: (a) blue only; (b) green only; (c) red and green only; and (d) red only; whereas "heated-object spectrum" display is formed by activating and incrementing the level of each gun at appropriate count density until it attains maximum level, at which it remains.

available gray levels were estimated using a spotmeter[11] and are shown in Table 1. Ambient light was well below one candela/m². The luminance contributions of saturated red, green, and blue to white were estimated as 29, 58, and 13%, respectively, these values being in good agreement with figures quoted by Pratt (*1*).

Four TV displays were assessed using both sets of images. These were—

1. *Black-on-white:* 16 levels of gray were used, white corresponding to the lowest count-density range, black to the highest.

2. *White-on-black:* similar to (1) but with the intensity scale reversed.

3. *Geographical:* for this fairly standard display, 50 levels were produced, through red, green, yellow, and blue, by firing the guns as shown in Fig. 1 (upper histogram).

4. *Heated-object spectrum:* 30 levels were produced, through red, orange, yellow, and white, by firing the guns as shown in Fig. 1 (lower histogram).

In all four cases the two end levels could be varied by the operator to produce any desired integral percentages of background subtraction and saturation level—provided, of course, that the latter exceeded the former. All intermediate levels were linearly distributed with respect to count density. The total range provided by these intermediate levels will be called the "effective dynamic range."

Figure 2 shows typical brain and liver scintigrams using the "white-on-black" display where, in each case, the superimposed lesion is indicated by an arrow. The intensity scale is shown as a bar underneath each image.

The views were presented to four observers with varied experience in nuclear medicine. Two of these were clinicians and two were technicians. One from each group had some previous experience of this type of experiment.

The observer viewed 100 images of a given organ and display in two batches of 50 to lessen the effects of fatigue. The four displays of the brain images were presented to the observers in orders forming a Latin square, and subsequently the four displays of the liver images were presented in orders forming a Latin square orthogonal to the first. In this way the effects of observer learning could be studied and effectively eliminated. For each batch, the order of presentation of the images was varied. For each view the observer was allowed to vary the "effective dynamic range" by altering the background subtraction and saturation level until he was satisfied that a decision could be made. The scintigrams were viewed as an image of size 625 cm², the screen height being 27 cm. Although no restrictions on head

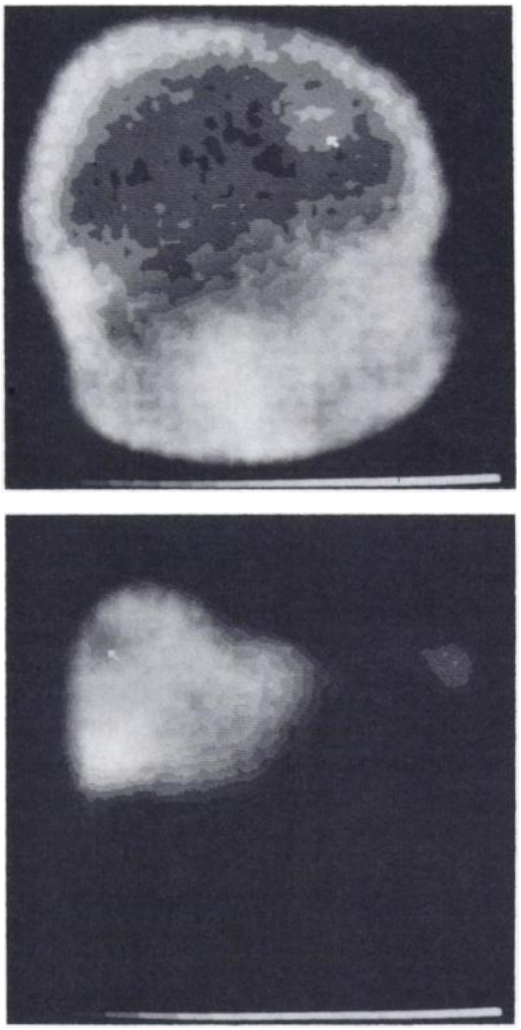

FIG. 2. Typical brain (top) and liver scintigrams are shown using "white-on-black" display. Superimposed lesion is indicated by arrow in each case.

movement were made, the observing distance was about 60 cm in all cases.

The observer was asked to rate the views with the five-category scale used by other authors (*11*), in which a rating of 5 corresponds to "lesion almost definitely present" and a rating of 1 to "lesion almost definitely not present," with 2, 3, and 4 representing intermediate classifications. The observer was also asked to indicate the area of highest apparent abnormal uptake corresponding to this rating. His *a priori* knowledge was that only single increased areas of activity were present in the brain images and single decreased areas of activity in the liver images, each with a probability of 0.5 per image. Positional discrepancies of up to 20 and 27 mm were allowed for brains and livers, respectively.

An LROC curve (*12*) was formed for each run, i.e., one observer viewing one display for one organ, by dividing the results into true positive, correct location (TP(CL)); true positive, incorrect location (TP(IL)); and false positive (FP). The numbers in the first and last

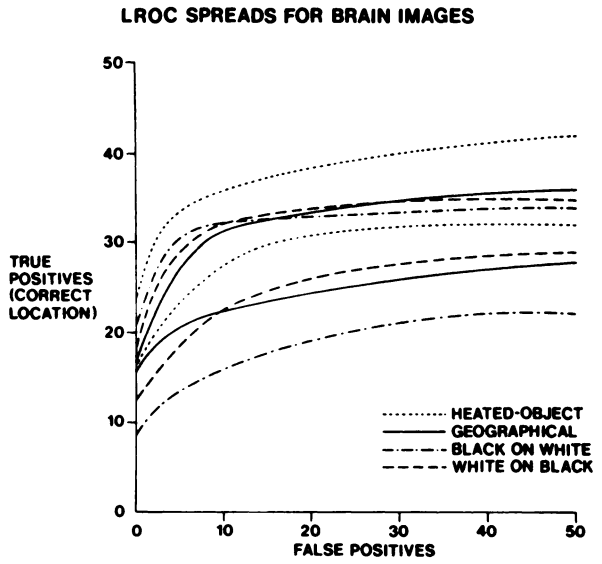

FIG. 3. LROC spreads for the four displays used for brain images, showing limits of LROC curves for four observers.

groups were then plotted against each other for each cumulative rating, i.e., 5, ≥ 4, ≥ 3, ≥ 2, ≥ 1. The curves for the four observers were then combined to form an LROC spread (*11*) in each case.

A points value was also obtained for each run from the formula used in previous studies (*8–10*), i.e.: Points = $3n_{TP(CL)} - n_{TP(IL)} - 2n_{FP} - 50$, where this value is first calculated for each cumulative rating and the maximum value then chosen.

RESULTS

LROC spreads corresponding to the various displays for brain images are shown in Fig. 3. The points totals for brains corresponding to the various runs are pre-

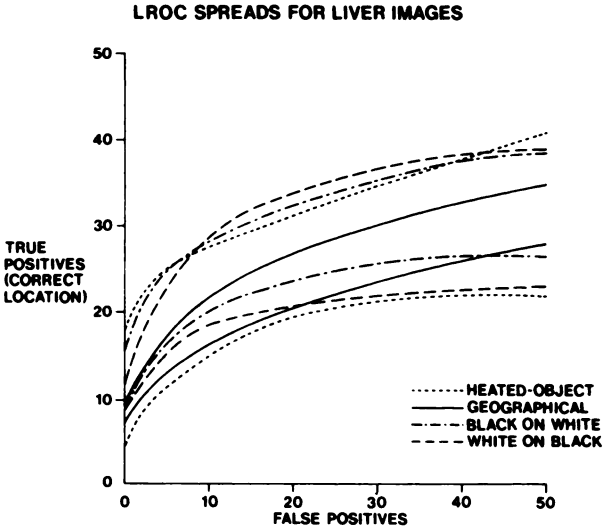

FIG. 4. LROC spreads for four displays used for liver images, showing limits of LROC curves for four observers.

TABLE 2. POINTS TOTALS FOR EACH RUN USING BRAIN IMAGES,* WITH ORDER OF OBSERVATION IN PARENTHESES

Observer	Black-on-white	White-on-black	Geographical	Heated-object
1	−25(1)	−9(2)	2(3)	29(4)
2	28(2)	17(1)	25(4)	31(3)
3	0(3)	6(4)	9(1)	25(2)
4	15(4)	28(3)	7(2)	41(1)
Total	18	42	43	126

* The four displays were viewed by observers in orders forming Latin square as shown, each position in order of observation appearing once and only once in each row or column.

sented in Table 2. The order of observation of each display for a given observer is contained in parentheses. Observers 1 and 3 had no previous experience of this type of experiment and were relatively inexperienced in the field of nuclear medicine. Observer 3 showed some degree of red-green color blindness using the Ishihara test and was graded color-perception 3 using the Holmes-Wright Defence Lantern. The other observers had normal color vision (color-perception 1).

A statistical analysis of the results in Table 2 is shown in Table 3. It appears that the effect of learning has not influenced the results, whereas interobserver and inter-display effects are significant at the 10% level. In the latter case it appears from the data that this is mainly due to one display producing better results than the others.

The root-mean-square error of a single points value was estimated as 11.8 from the Latin square residual value. Using the Scheffé method of comparison, the points total for the "heated-object spectrum" was then shown to be significantly better at the 10% level than the mean points total for the other three displays. This finding is demonstrated in Fig. 3, where both upper and lower limits of the LROC spread are raised for this display. This study does not show any obvious differences among the other displays.

LROC spreads corresponding to the various displays for liver images are shown in Fig. 4. Points totals (order of observation in parentheses) and the subsequent statistical analysis are contained in Tables 4 and 5, respectively. In this case the interobserver effect does not appear to be important, whereas the between-order and between-display effects are significant at the 10% and 15% levels, respectively.

Following an argument similar to that for brain images, the rms error of a single points value was estimated as 9.5 and the points total for the "geographical" display was shown to be worse than the mean points total for the other three displays at the 15% significance level—again using the Scheffé test. From Fig. 4 it is seen that, whereas the lower limit of the LROC spread of the "geographical" display is comparable with the others, the upper limit is somewhat lower than the others. This implies that nobody performed particularly well using this display. Again, no obvious differences were found among the other displays.

DISCUSSION AND CONCLUSION

Any attempt to interpret these results involves a study of the regions of the "effective dynamic range" at which lesions are likely to occur. When questioned, all observers admitted that, in general, a value for background subtraction of about 5% was used throughout, whereas the value for saturation level was about 60% for brains and 100% for livers. Thus lesions tended to be present in the lower-middle of the range for brains and towards the upper end for livers. This would correspond respectively to orange and yellow for the "heated-object spectrum" and to green and red for the "geographical" display.

TABLE 3. ANALYSIS OF VARIANCE USING POINT TOTALS FOR BRAIN IMAGES

Source	Sum of squares	Degrees of freedom	Variance estimate	F ratio
Display	1675.7	3	558.6	3.98
Observer	1745.2	3	581.7	4.14
Order	150.2	3	50.1	0.36
Residual	842.4	6	140.4	
Total	4413.5	15		

TABLE 4. POINTS TOTALS FOR EACH RUN USING LIVER IMAGES,* WITH ORDER OF OBSERVATION IN PARENTHESES

Observer	Black-on-white	White-on-black	Geographical	Heated-object
1	−8(1)	8(2)	−26(3)	4(4)
2	−11(3)	14(4)	−22(1)	5(2)
3	−4(4)	−12(3)	−6(2)	−29(1)
4	14(2)	−2(1)	−12(4)	14(3)
Total	−9	8	−66	−6

* Latin square shown here is orthogonal to that chosen for brains, i.e., if i and j are respective positions in orders of observation for brains and livers, then i and j appear in same place on corresponding Latin squares once and only once.

Pizer and Chan (*6*) showed that observers were more sensitive to small changes in intensity in the orange region of the "heated-object spectrum" than in the yellow region. Moreover, in the version of the display used in the present study, the luminance differences (calculated from Table 1 and the luminance contributions to white of the three guns) were much greater in the orange region. In the case of the "geographical" display, the difference in luminance between adjacent levels of green was twice that for red. For TV gray-scale displays, it is commonly assumed (*14*) that the change in intensity that will be just noticeable is a constant fraction of the background intensity (Weber's law). It would therefore follow from Table 1 that adjacent gray levels should be slightly more readily distinguished at lower intensities.

It might therefore be expected that the "heated-object spectrum" would be much better for brain images than for liver images; the "geographical" display somewhat better; "white-on-black" slightly better; and "black-on-white" slightly worse. Although it is not possible to measure these effects directly using the available data, it can be seen from Tables 2 and 4 that, relative to each other, the results for the various displays are in reasonable accordance with predictions.

The residual errors occurring in this study are greater than those obtained previously (*8*). It is felt that this is due mainly to the observer being allowed to set his own "effective dynamic range," which introduces an additional jitter term.

The total residual error in each case will, of course, include any interactive terms present, e.g., if observers learn at different rates. The individual residuals for the various runs were calculated and the larger ones examined.

Two possible effects became evident. The poor start made by Observer 1, using brain images, could be due to his relative inexperience, while the fact that Observer 3 was partially color-blind might explain why his results for color displays of liver images were inconsistent with those of the other observers. However, the latter effect is not apparent in the results for brain images and is difficult to explain in terms of the colors involved.

The last point raises the question of whether the four observers are a representative sample of personnel training in nuclear medicine. Clearly an increase in the number of observers would be advantageous, and any conclusions should be viewed with this in mind.

In answer to our first two questions, the "heated-object spectrum" appears to be better than the other displays for observing brain images on the aforementioned TV system. Whether this is due to the greater number of levels used, compared with the gray-scale displays, or to the design of this display, is not proved and will be the subject of future research. For both brain and liver images, however, this display gives better results than the "geographical" display, which has even more color levels, indicating the importance of display design. No other conclusions are obvious regarding the use of color displays for liver images, since the "heated-object spectrum" (as designed in this experiment) appears to be

TABLE 5. ANALYSIS OF VARIANCE USING POINT TOTALS FOR LIVER IMAGES

Source	Sum of squares	Degrees of freedom	Variance estimate	F ratio
Display	801.2	3	267.1	2.98
Observer	536.2	3	178.7	1.99
Order	1014.7	3	338.2	3.77
Residual	537.9	6	89.6	
Total	2890.0	15		

roughly on a par with the gray-scale displays. It does appear, however, that the choice of display depends on the organ to be imaged, due to important differences in the required and available information.

FOOTNOTES

* The term "display" is used throughout to infer "pseudocolor mapping" as defined by Pratt (*1*).

† A color level is uniquely defined by its brightness, hue, and saturation.

‡ EMI Prowest PMC 17/9A,

‖ Asahi Pentax III,

ACKNOWLEDGMENTS

The author thanks Surg. Cdr. M. A. Macleod, Surg. Lt. A. J. B. McEwan, MTI R. A. Squires, and MTI C. I. Hamilton, of the Department of Nuclear Medicine, RNH Haslar, for their participation to the experiment. Thanks also to Dr. R. J. Pethybridge of the Institute of Naval Medicine, Alverstoke, for several helpful discussions.

REFERENCES

1. PRATT WK: *Digital Image Processing.* New York, John Wiley, 1978, pp 336-338
2. TODD-POKROPEK AE, PIZER SM: Displays in scintigraphy. In *Medical Radionuclide Imaging.* Vol. 1. Proceedings of a Symposium, Los Angeles, 1976. IAEA, Vienna, 1977, pp 505-537
3. BRILL AB, ERICKSON JJ: Display systems in nuclear medicine. *Semin Nucl Med* 8: 155-161, 1978
4. MILAN J, TAYLOR KJW: The application of the temperature scale to ultrasonic imaging. *J Clin Ultrasound* 3: 171-173, 1975
5. CHAN FH, PIZER SM: An ultrasonogram display system using a natural color scale. *J Clin Ultrasound* 4: 335-338, 1976
6. PIZER SM, CHAN FH: Evaluation of the number of discernible levels produced by a display. Proceedings of the VIth International Conference on Information Processing in Medical Imaging. In Press (abst)
7. HOUSTON AS: Mathematical tumours and their use in assessing data processing techniques in radioiscope scintigraphy. *Phys Med Biol* 19: 631-642, 1974
8. HOUSTON AS, MACLEOD MA: An intercomparison of computer assisted image processing and display methods in liver scintigraphy. *Phys Med Biol* 24: 559-570, 1979
9. HOUSTON AS, MACLEOD MA: An intercomparison of computer assisted data processing and display methods in radioisotope scintigraphy using mathematical tumours. *Phys Med Biol* 22: 1097-1114, 1977
10. HOUSTON AS, SHARP PF, TOFTS PS, et al: A multicentre comparison of computer assisted image processing and display methods in scintigraphy. *Phys Med Biol* 24: 547-558, 1979
11. GOODENOUGH DJ, ROSSMANN K, LUSTED LB: Radiographic applications of receiver operating characteristic (ROC) curves. *Radiology* 110: 89-95, 1974
12. STARR SJ, METZ CE, LUSTED LB, et al: Visual detection and localization of radiographic images. *Radiology* 116: 533-538, 1975
13. PRATT WK: *Digital Image Processing.* New York, John Wiley and Sons, 1978, pp 32-33