

Eye Movement During Computed Tomography Interpretation: Eyetracker Results and Image Display-Time Implications

David V. Beard, Etta D. Pisano, Kevin M. Denelsbeck, and R. Eugene Johnston

Stacked displays hold the potential for accurate interpretation of multiple computed tomography (CT) studies on a low-cost workstation. But can such a display scroll as quickly as radiologists can move their eyes to the next image on a film? To address this question, eye-movement duration during CT chest interpretation was recorded using an electronic eye tracker. Adjacent eye movements (± 1 image in sequence) averaged 0.54 seconds. Time motion analysis indicates that a CT workstation using a stacked approach with a 0.2-second image display time and a simple interaction can display the next image in less than 0.4 seconds, so a stacked approach should allow a low-cost workstation to facilitate acceptable interpretation of multiple CT or magnetic resonance studies. However, nonadjacent eye movement is likely to take longer and radiologist behavior may be affected.

Copyright © 1994 by W.B. Saunders Company

KEY WORDS: computed tomography (CT), image display and recording, diagnostic radiology observer performance, images, display.

SEVERAL LABORATORY studies have shown that computed tomography (CT) workstations that array images in a mosaic on a 2,500- × 2,000-pixel monitor can facilitate interpretations that are as fast and as accurate as with film and alternator.^{1,2} However, the monitors used in these studies currently are too expensive. One lower-cost approach using 1,024- × 1,024-pixel monitors, involves showing only a single image per study with previous and subsequent images available at the press of a button. With such a stacked or superimposed display, the user is provided with a mental model or metaphor^{3,4} of a stack of images.

Stacked displays hold the potential for accurate interpretation of up to four of a patient's 512- × 512-pixel CT studies on the same low-cost 1,024- × 1,024-pixel monitor.² But such displays have been rejected in the past because radiologists could move their eyes to the next image on a mosaic display far faster than a workstation could display that image. Because the average CT interpretation could involve 75 scroll operations, a multisecond scroll could increase interpretation times by several minutes.

However, Straub et al² has shown experimentally the potential of the stacked approach when

implemented with a 0.1-second image display time. Thus, a stacked display now has the potential to provide rapid access to multiple CT or magnetic resonance (MR) studies using a low-cost computer workstation. But how fast does the workstation have to display the images for the stacked approach to be effective? Should vendors spend additional money to provide image display times of 0.1 seconds, or would 0.5 seconds or 1.0 seconds be sufficient?

Time-motion analysis⁵⁻⁸ suggests that the time to scroll the stacked display to the next image should be at least as fast as the time for radiologists to move their eyes from one CT image to the next. Our approach was to measure the duration of radiologist image-to-image eye movement during CT interpretation with the use of an optical eyetracker device. Simple time-motion models⁸ were then used to estimate roughly the longest image display time that would allow a scrolling operation to be no longer than an image-to-image eye movement.

MATERIALS AND METHODS

We measured the eye movements of radiologists during interpretation of CT cases by analyzing eyetracker video tapes to determine the time for a radiologist to go from foveally viewing an anatomic object on one CT chest image to foveally viewing the same anatomic object on another.⁹ Additionally we noted how frequently the eye movements were to the next or previous image in the study's sequence. Others have used eyetrackers to determine eye-movement behavior during various radiographic image interpretations.¹⁰⁻¹⁶

Equipment, observers, and cases. Eye movement during CT chest interpretations was recorded using the Eye Mark Recorder Model V (EMR-V) (Inst Technology Systems, Northridge, CA) (see Fig 1). The eyetracker consists of a head-goggle unit and a camera controller unit. The goggle unit, which is mounted on the radiologist's head, contains

From the Departments of Radiology and Computer Science, University of North Carolina School of Medicine, Chapel Hill.

Supported in part by National Institutes of Health Grant No. R01 CA44060.

Address reprint requests to David V. Beard, PhD, Department of Radiology, 509 Old Infirmary Bldg., School of Medicine, University of North Carolina, Chapel Hill, NC 27599-7510.

*Copyright © 1994 by W.B. Saunders Company
0897-1889/94/0704-0003\$3.00/0*

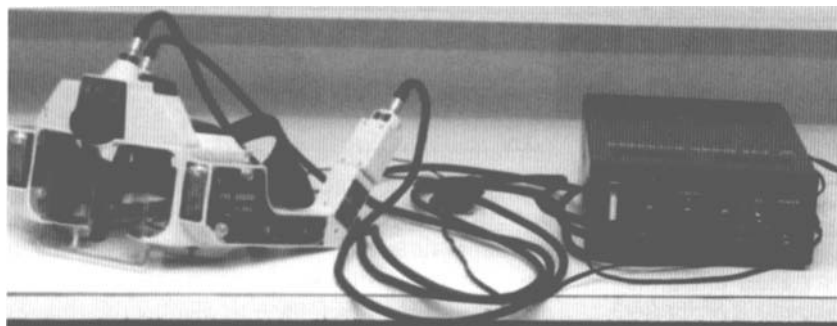


Fig 1. The Eye Mark Recorder Mark V eyetracker used to determine eye movement during this study is shown.

optics and electronics for tracking the location of both eyes. In use, an infrared light-emitting diode light source (950 μm wavelength), which cannot be sensed by the eye, projects a light spot onto the observer's cornea. This light spot is reflected from the cornea, detected by a metal-oxide semiconductor video camera, and sent to the camera controller unit for processing. In addition to the imaging devices for each eye, there is a third field-of-view (FOV) video camera positioned on the center of the observer's forehead when wearing the eyetracker; this FOV camera observes the central portion of the observer's field of view. The camera controller superimposes eye position indicator spots from each eye onto the video image from the FOV camera and sends the resulting National Television Standards Committee video data to a monitor and video recorder. The eyetracker's FOV is 45 degrees vertical and 60 degrees horizontal, with about 1% to 2% error. The instrument tracks with an accuracy of 0.6 degrees. The EMR-V allows the radiologist complete freedom of movement, allowing observation of a more clinically realistic interpretation.

Two board-certified radiologists experienced with CT participated in the study. Neither wore eye glasses. Three CT chest cases were interpreted by each radiologist during the data collection sessions. A fourth calibration and training case was used as well. The cases averaged 45 slices and were printed on 14- \times -17-films with both mediastinum and lung-intensity windows.

Procedure, data collection, and analysis. Each radiologist viewed the one training and calibration case and interpreted the three chest cases during a single observation session in an experimental environment controlled for interruptions, lighting, and sound. Images were displayed on a conventional four-over-four light box. A eyetracker calibration test was conducted before and after each interpretation to insure that the eyetracker data from that interpretation was valid.

The VHS-format eyetracker tape was transferred to U-Matic tape and analyzed frame-by-frame using a Sony BVE-910 editing console driving a Sony VO-9850 Videocassette Recorder (Sony, Park Ridge, NJ). U-Matic tape offers very good single-frame viewing with a temporal resolution of 30 frames per second, allowing timing accuracy to within 1/15th of a second. Eye movements were judged to be instances where successive fixation occurred on the same anatomic object on two CT images, eg, between the left kidney on one image to the left kidney on the another image. Eye-movement duration was determined by counting the number of frames between the last frame of definite

fixation in the first image and the first frame of definite fixation in the second. In our analysis, we included all eye movements in which we could clearly determine which anatomic object was being foveally viewed in a given image.

RESULTS

Frame-by-frame analysis of about 40 minutes of video tape yielded 196 eye-movement data points. 109 of the data points were from radiologist no. 1 with the rest from radiologist no. 2. Sixty-nine percent of the eye movements were to adjacent images, while 31% of the movements were to nonadjacent images. Combining both subjects, adjacent eye movements had a mean of 0.54 seconds with a median of 0.52 seconds and a mode of 0.53 seconds. Nonadjacent eye movements had a mean of 0.54 seconds with a median of 0.5 seconds and a mode of 0.46 seconds. All eye movements together had a mean of 0.57 seconds with a median of 0.53 seconds and a mode of 0.53 seconds.

Figure 2 plots the distribution of adjacent eye movements, nonadjacent eye movements, and all data for both subjects; overlines indicate repeating decimals. Note that adjacent data includes eye movement from the end of one row to the image at the beginning of next and vice versa; most of the adjacent outliers are from this type of eye movement. A considerable number of nonadjacent eye movements were to images directly above or directly below the initial image. These eye movement are likely to be as fast as adjacent eye movements to images on the same row; these vertical eye movements most likely account for the rapid nonadjacent eye movement times.

DISCUSSION

We believe that for a stacked display to be viable, the sum of the hand motions and key-

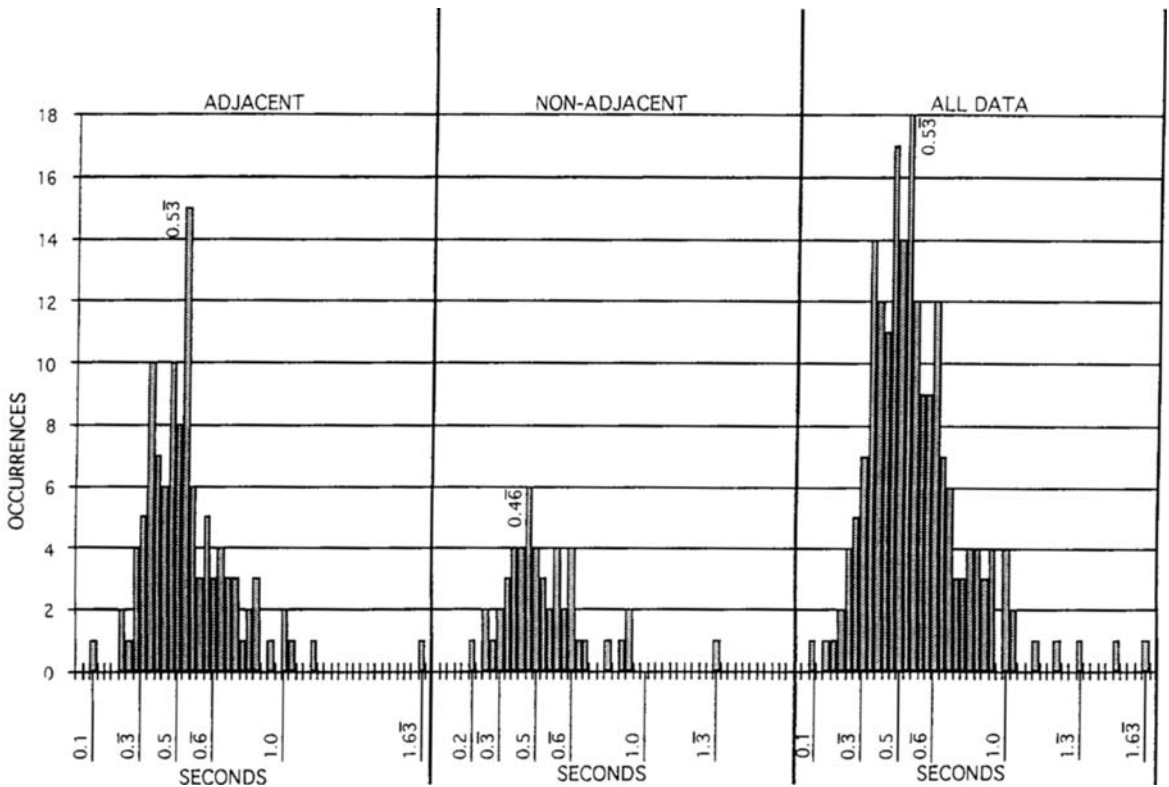


Fig. 2. Plot of the distribution of adjacent eye movements, nonadjacent eye movements, and all data for both subjects.

stroke times for the image display operation combined with the image display time for that operation cannot exceed the average 0.6-second interimage eye-movement duration determined above.

Using the Keystroke time-motion model, developed by Card et al.,⁸ and a very simple workstation interaction, we calculate that the time to scroll the display will be roughly the sum of the time to press a "next image" button and the time for the system to display that image. (Eye movement is not a factor in this analysis of a stacked display because the radiologist can go from foveally viewing an anatomic object on one image to foveally viewing the same object on the next image without significant eye movement.) The button press will take about 0.2 seconds.⁸ Therefore, if we want the total scroll time to be less than or equal to 0.6 seconds, the image display time should be less than or equal to 0.4 seconds. If two button presses were involved, the image display time would have to be less than 0.2 seconds and so forth.

Time-motion models can be accurate to within

0.1 second.⁸ Nevertheless, there is a great deal of potential error in the above calculations. First, whereas the average time to press a button is 0.2 seconds, this could vary from 0.05 seconds to over a minute, depending on the user's hand positions, reaction time, and the size and shape of the button.⁶ Second, users often pause for up to 1.5 seconds to think about a complex computer interaction before undertaking it.⁸ We do not expect mental pauses to be significant with a stacked display because of the simplicity of the interaction, but a mental pause of 0.2 seconds before every scroll operation is quite possible.

Thus, assuming a 0.2-second mental pause, we conclude that for a stacked display to be viable, an extremely simple human-computer interaction is required with a single button press (0.2 seconds) combined with a 0.2-second or less image display time to stay at or less than the 0.6-second goal.

To apply the above calculations to a real case, we measured the image display time of a Sun

SPARC II workstation (Sun Microsystems, Mountain View, CA) and found that a 512- × 512-pixel CT image could be displayed in about 0.2 seconds which should be sufficient for a viable stacked display. We estimate that a SPARC II or similar workstation would cost about \$5,000 if configured with sufficient memory for this application.

Note that the eyetracker tapes also show that over 30% of the time, radiologists often move their eyes between nonadjacent slices while viewing the anatomy. Although there are a number of methods for scrolling through the images in a study, the most likely consists of two buttons, one for scrolling up and the other down through the study. These buttons may have an auto-scroll feature that allows automatic scrolling while the button is held down after an initial 0.2 seconds or so, but for movements of two to four slices, individual button presses will be faster than using the auto-scrolling feature.

Accessing images that are further than one image from the current slice will be increasingly time consuming with a mosaic display because of the increased number of 0.2-second button presses and 0.2-second image display times required.

Stacked display implications. These results suggest stacked display approach should allow low-cost commodity-priced workstations to allow for acceptable display and interpretation of multiple CT or MR studies. Viewing of adjacent images is likely to be more a factor than with film. However, movement to more distant images may be slowed affecting total interpretation time and making it more difficult for the user to develop a three-dimensional (3D) mental model of the viewed anatomy. Additional training may be needed for radiologists using stacked displays, and accuracy studies may be required that focus on 3D spatial comprehension.

REFERENCES

1. Beard DV: Designing a radiology workstation: A focus on navigation during the interpretation task. *J Digit Imaging* 3:152-163, 1990
2. Straub WH, Gur D, Good WF: Primary CT diagnosis of abdominal masses in a PACS environment. *Radiology* 1:739-743, 1991
3. Rumelhard D, Norman D: Analogical processes in learning, in JR Anderson (ed): *Cognitive Skills and Their Acquisition*. Hillsdale, NJ, Erlbaum, 1991, pp 335-359
4. Young R: The machine inside the machines: User's models of pocket calculators. *Int J Man-Machine Stud* 15:51-85, 1981
5. Gilbreth FB: *Motion Study*. New York, NY, Van Nostrand, 1911
6. Fitts PM: The information capacity of the human motor system in controlling the amplitude of movement. *J Exp Psychol* 47:381-391, 1954
7. Meyer DE, Abrams RA, Kornblum S, Wright CE: Optimality in human motor performance: Ideal control of rapid aimed movements. *Psychol Rev* 95:340-370, 1988
8. Card SK, Moran TP, Newell A: *The psychology of human-computer interaction*. Hillsdale, NJ, Erlbaum, 1983
9. Beard DV, Johnston RE, Toki O, Wilcox C: A study of radiologists viewing multiple computed tomography studies using an eyetracking device. *J Digit Imaging* 3:230-237, 1990
10. Tuddenham WJ, Calvert WP: Visual search patterns in roentgen diagnosis. *Radiology* 76:225-256, 1961
11. Tuddenham WJ: Visual search, image organization, and reader error in roentgen diagnosis: Studies of the psychophysiology of roentgen image perception. *Radiology* 78:694-704, 1962
12. Llewellyn-Thomas E: Search Behavior. *Radiol Clin North Am* 1:403-417, 1969
13. Kundel HL, Wright DJ: The influence of prior knowledge on visual search strategies during the viewing of chest radiographs. *Radiology* 93:315-320, 1969
14. Kundel HL, LaFollette PS: Visual search patterns and experience with radiological images. *Radiology* 103:523-528, 1972
15. Kundel HL, Nodine CF, Carmody D: Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. *Invest Radiol* 13:175-181, 1978
16. Carmody DP, Nodine CF, Kundel HL: Global and segmented search for lung nodules of different edge gradients. *Invest Radiol* 15:224-233, 1980