

David V. Beard, PhD • Paul L. Molina, MD • Keith E. Muller, PhD • Kevin M. Denelsbeck, MS
Bradley M. Hemminger, MS • J. Randolph Perry, MD • M. Patricia Braeuning, MD
Deborah H. Glueck, MS • W. Dean Bidgood, Jr, MD • Matthew Mauro, MD
Richard C. Semelka, MD • Ann S. Willms, MD • David Warshauer, MD • Etta D. Pisano, MD

Interpretation Time of Serial Chest CT Examinations with Stacked-Metaphor Workstation versus Film Alternator¹

PURPOSE: Interpretation time of serial staging chest CT cases, which each contained current and previous examinations, with a simple prototype workstation called filmstack was experimentally compared with interpretation time with a film alternator.

MATERIALS AND METHODS: The filmstack displayed a "stack" of sections for each examination; user controls allowed rapid selection of preset attenuation windows and both synchronized and unsynchronized scrolling. Eight radiologists were timed as they used the filmstack and the film alternator to interpret four ergonomically complex serial CT cases.

RESULTS: All reports dictated on the basis of findings with filmstack and film were of acceptable clinical accuracy. The time to examine a case with filmstack was significantly faster than the time with film, including the time to load and unload the alternator (99% confidence [$P = .01$]). There was no statistically significant difference in interpretation time between filmstack and prehung film.

CONCLUSION: Use of a low-cost stacked CT workstation with a single $1,024 \times 1,024$ monitor is an effective means of interpreting cases that require comparison of multiple CT examinations.

Index terms: Computed tomography (CT), image display and recording • Diagnostic radiology, observer performance • Receiver operating characteristic curve (ROC) • Workstation

Radiology 1995; 197:753-758

PICTURE archiving and communication systems (PACS) may eventually reduce costs (1) and improve logistics by eliminating film images, providing simultaneous access by multiple physicians, and allowing use of a variety of three-dimensional and contrast-manipulation display algorithms (2). Computed tomography (CT) is a candidate for early filmless operations because, unlike radiography, CT is inherently digital and its relatively small images completely fit onto available commodity-priced monitors.

CT examinations are often compared to one or more previous CT examinations, and these additional images greatly complicate CT workstation interpretation. Similarly, loading and unloading the film onto the film alternator can be increasingly time-consuming as the number of comparison examinations in a series increases. Many radiology departments have allied personnel preload both the current study and the most appropriate comparison examination to maximize radiologist productivity. However, comparison examinations may be incorrectly selected, hung in the wrong order, or unavailable until just before interpretation, so film hanging still occupies radiologists' time in even the most efficient departments.

Generally, PACS workstations for interpretation of CT examinations have used a mosaic arrangement of the CT sections that is analogous to the 3×4 or 3×5 grid used with film display of a CT examination. Such a mosaic display allows the radiologist's eyes to move quickly back and forth

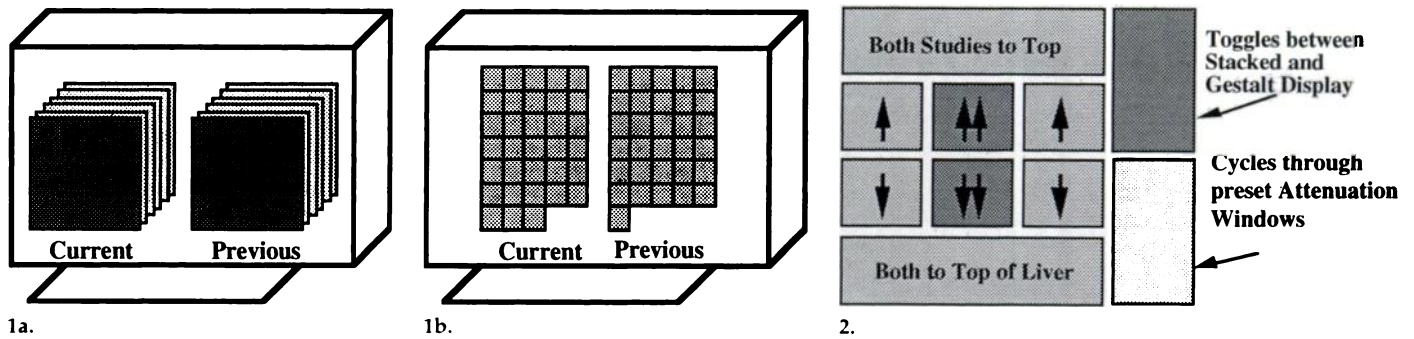
between a "cluster" of two to eight images (3), which facilitates comparison. An alternative to the mosaic arrangement is the metaphor of a stack of CT images sequentially displayed one at a time on a monitor. Use of "forward" and "backward" controls allows the radiologist to visualize the entire examination one image at a time. Because the images are superimposed and "aligned" from one section to the next, radiologists can view longitudinally oriented structures such as blood vessels without having to move their eyes from a given location on the monitor.

Stacked displays have advantages compared with mosaic displays for workstation presentation of CT examinations. Since only one image from each examination is displayed at the same time, a much smaller monitor can be used, which reduces the cost of the workstation. Further, a single low-cost $1,024 \times 1,024$ monitor can hold up to four "stacked" CT examinations simultaneously for comparison. Stacks for each examination are on the same monitor, so comparison images are close together rather than up to several feet away, as can be the case with film or mosaic workstations. There are also disadvantages to stacked displays. Only a single image is displayed at one time, so radiologists may find it difficult to obtain an understanding, or a "gestalt," of a large three-dimensional section of anatomy. Further, coordination of stacks of serial cases can be difficult.

A number of researchers have evaluated the speed and accuracy of use of workstations for interpretation of a single CT examination (4-12). Interpretation with a CT workstation can be as accurate as interpretation with film, as long as the workstation has monitors of good quality, appropriate attenuation window settings, sufficiently rapid image-display times, and acceptable computer-human interactions. Beard et al (11) found that with

¹ From the Departments of Radiology (D.V.B., P.L.M., B.M.H., J.R.P., M.P.B., W.D.B., M.M., R.C.S., A.S.W., D.W., E.D.P.), Computer Science (D.V.B., K.M.D.), and Biostatistics (K.E.M., D.H.G.), University of North Carolina School of Medicine, Chapel Hill. Received March 9, 1995; revision requested, April 26; revision received, June 5; accepted June 12. Supported in part by National Institutes of Health grant no. R01 CA44060. Address reprint requests to D.V.B., Idaho State University, Computer Information Systems, College of Business, Campus Box 8020, Pocatello, ID 83209-8020.

² RSNA, 1995



Figures 1, 2. (1a) Stacked display. Screen layout contains images from the follow-up and initial chest CT examinations. (1b) Gestalt display. Screen layout contains a mosaic of miniature versions of all the images from both the follow-up and initial examinations. (2) Configuration of keyboard for stacked display. Double-arrow buttons scroll both sets of images at the same time.

use of 2,500 × 2,000-pixel monitors and 0.11-second image display times, a single CT examination could be interpreted as quickly with a workstation as with a film alternator. Straub et al (10) conducted a detailed receiver operator characteristic (ROC) study to compare the accuracy of interpretation of an abdominal mass in a single CT examination with a 2,048 × 2,048-pixel mosaic workstation, a stack-metaphor 2,048 × 2,048-pixel workstation, and a film alternator: Results with both the stacked and mosaic methods were obtained with accuracy and speed equal to those achieved with film. Also, Beard et al (13) used an eye-tracker and time-motion analysis (14) to predict that as the image display times are reduced to about 0.3 second, results with the stacked display are achieved about as quickly as they are with the mosaic display. However, interpretation speed was not sufficient with previous workstations that used less expensive 1,024 × 1,024 monitors, regardless of image display speed (12). To our knowledge, the feasibility (or efficacy) of workstation interpretation of serial CT cases, rather than of just a single CT case, has been largely unstudied.

MATERIALS AND METHODS

Experimental Goal

We conducted a timing experiment to evaluate a simple prototype stack-metaphor CT interpretation workstation we called filmstack for use in interpretation of serial CT cases that resulted in dictated reports. Two hypotheses were evaluated. First, error-free interpretation of serial staging chest CT cases with the filmstack workstation is faster than interpretation with a film alternator, including the time to load and unload the film. Second, filmstack interpretation time is faster than interpretation of prehung film with a film alternator, not including the time to load and unload the film. Error-free interpretation time includes interpretation times

only for those interpretations that meet acceptable clinical standards of accuracy. Error-free time is a technique used in psychology studies to hold accuracy constant to obtain a valid comparison of response time (15).

Equipment

Serial cases containing both initial and follow-up CT examinations were interpreted with a conventional horizontal film alternator and with filmstack. This prototype workstation has a SPARC 1+ computer with a single 36 × 28-cm, 1,100 × 900-pixel, 66-Hz, noninterlaced, landscape-oriented monitor (Sun Microsystems, Mountain View, Calif). filmstack was implemented with the C programming language and the X windowing system. This hardware and software configuration displays the next or previous image in a filmstack in 0.2 second, which is approximately the time a human requires to press a button. Figure 1a depicts the filmstack screen layout, with the film stacks of the current examination on the left and of the previous examination on the right.

Pilot studies indicated that a number of radiologists were troubled by their inability to see an overview or gestalt of the images. To address this preference, we added an alternate screen layout that presented gestalt views for both the current and previous examinations (Fig 1b). The gestalt view consists of a five-column mosaic of reduced-size (10 × 10-pixel square) images from both examinations. A button on the control panel toggled the display between the gestalt view and the stacked view of both the current and previous examinations.

Figure 2 shows the configuration of the control panel. By using self-sticking tags on the keys of the conventional Sun workstation keyboard, we were able to rapidly rearrange and reconfigure the keyboard during prototyping and pilot studies. Six functions were included in the final version of the keyboard used in this experiment. First, a vertical button pair was provided to scroll the follow-up examination up and down, as well as a corresponding vertical button pair for the initial examination. Second, a vertical button pair between the follow-up and initial button

pairs was provided to scroll both examinations in a synchronized manner. Third, a button below the scrolling buttons moved both examinations to the topmost image that contained a portion of the liver. This "liver" button proved very useful because it synchronized the images in both examinations regardless of the anatomic position of the examination. Once the examinations were synchronized with the liver button, the radiologist could scroll up and down and the examinations would move in parallel, generally in synchronized fashion. We do not know of any current technology that can automatically recognize which section contains the top of the liver or some other anatomic landmark in a CT examination, so a technologist had to select this section at the same time that the preset attenuation window settings were chosen. Fourth, a long horizontal button above the scrolling vertical button pairs moved both examinations to their topmost image. Fifth, a button toggled between preset soft-tissue, lung, and liver windows. This attenuation window operation took 0.2 second. Findings in previous studies (7,11) supported the effectiveness and speed of providing only preset attenuation windows for general CT interpretation. Sixth, a button toggled between the stacked and the gestalt display.

Observers

Eight board-certified radiologists, five men and three women who read CT scans in our department, participated in the experiment. All observers had some experience with computers such as word processors or the information system in our radiology department. All had used consoles on CT machines and several had participated in previous CT workstation evaluations, but none had experience with filmstack.

Cases

Four pairs of serial "staging" chest CT examinations (lung apices through the liver) with multiple abnormal findings were selected for the experiment. Ergonomically complex cases with potentially long interpretation times were deliberately chosen to maximize any interpretation-

time differences between film and filmstack. Such cases were expected to be particularly difficult for interpretation with filmstack, owing to the large number of multisection visual comparisons required between the serial examinations.

A homogeneous set of cases was chosen to reduce variance. Findings among these abnormal cases were as follows: Case A was a postmastectomy evaluation of a pulmonary nodule suspicious for metastasis, postirradiation changes in the left upper lobe, and diffuse fatty infiltration of the liver. Case B was a posttransplantation evaluation of a lung with interval increased mediastinal adenopathy, fluid collections, bilateral pleural effusions that tracked into the fissures, and patchy right-upper-lobe posterior-segment consolidation. Case C exhibited right-lower-lobe collapse, development of a large right pleural effusion, and persistent right-upper-lobe posterior-segment parenchymal consolidation. Case D was of multiple bilateral pulmonary metastases. All examinations had been performed with a Somatom Plus CT scanner (Siemens; Iselin, NJ) with contiguous 1-cm CT sections and an average of 35 sections per scan.

To help control for the effect of varying difficulty, the four cases were paired and then divided into two groups of equal difficulty. The original report for the initial examination as well as the original CT requisition form for the follow-up examination (which contained the patient history and the referring physician's clinical question) were obtained for each case. The original scans were used for viewing with the film alternator. Preset values for attenuation windows in filmstack viewing were chosen to match as closely as possible the windows used to obtain the original scans.

Experimental Design and Procedure

A counterbalanced, within-subject experimental design was used, with each observer reading two cases with a film alternator and two cases with filmstack. Each case was read exactly once by each observer, and each case was read the same number of times with filmstack and with a film alternator, so no radiologist read the same case with both display methods. Case presentation, observer, and method were all controlled. Four observers started with filmstack, whereas the other four started with the film alternator. Alternator and filmstack interpretations were systematically intermixed to minimize the effect of ordering.

Both filmstack and film alternator reading sessions were conducted under conditions controlled for lighting, sound, distractions, and interruptions (16). Observers were instructed to work as quickly as possible while producing dictated reports of clinically acceptable accuracy. For each film case, the observer was given a folder that contained the scans for the initial and follow-up examinations, as well as

a typed sheet of 8.5" x 11" paper that contained the dictated report for the initial examination and the requisition information for the follow-up examination. Observers loaded the scans onto the alternator, viewed the images, and dictated a report with a familiar dictation machine.

Before interpretation of the first workstation case, the seat adjustment, keyboard, and monitor placement were checked, the keyboard arrangement was explained, and the observer was trained for 2 to 3 minutes by scrolling through a case, adjusting contrast, etc. Then the observer was given two serial staging chest CT training cases to interpret, for a total of approximately 20 minutes of practice. One training case involved changes in size and number of liver metastases, and the other training case involved a pleural effusion.

On the start signal, the observer was given the typed sheet of paper that contained the initial report and the follow-up requisition information, and the images were displayed on the workstation. The observer then scrolled back and forth in the examinations, changed attenuation windows as needed, and dictated a report.

All observers had years of experience interpreting CT cases displayed on film. Given finite available observer time, it was not possible to match this film experience in a laboratory setting. Results of pilot studies indicated that only a small amount of training was sufficient for our observers to accurately interpret cases with filmstack.

Data Collection

Verbal comments—referred to as verbal protocol—made by the observers during the study were recorded by the experimenter and the observers were debriefed in an unstructured interview at the end of each session to gather comments and assess their experience with workstation and film interpretation. The experimenter used a stopwatch to measure the time to interpret each case. Load and unload times as well as interpretation times were measured in the film alternator cases. The stopwatch times have an accuracy of 1 second.

Whether film load and unload times should be included in a comparison of film and filmstack interpretation is open to debate, and the answer depends on the operational characteristics of the reading room to which the results are applied. Thus we report both total film time (including the time to load and unload films) as well as prehung film time (without the time to load and unload films) to allow the reader to make institutionally specific comparisons. Radiologists often perform some image "reading" while they load scans onto the view box, so we might be slightly underestimating the actual time to interpret prehung films.

Data Analysis

It was not the purpose of this experiment to evaluate interpretation accuracy,

but we needed to ensure that only accurate interpretations were included in the mean interpretation times for workstation and film. Accuracy of the 32 reports (the eight observers generated one report for each of the four cases, which each contained scans from an initial and a follow-up CT examination) was determined by a physician grader who was not an observer in the experiment and who was blinded to both interpretation method and the observer. The grader had available the original scans, all the dictated reports generated during the experiment, and the original copies of the initial CT report and follow-up requisition form.

Each experimental report was transcribed, with any information as to observer or method removed, and then evaluated by the grader in the following manner: First, a list of findings was generated for each report. Second, a list of findings was generated for each case on the basis of all available data. Third, the findings for each case were divided into critical findings, which might influence patient-care decisions; relevant findings, which related to the clinical question on the follow-up requisition form; and additional noncritical findings. Fourth, each report was determined to be clinically acceptable if it contained no incorrect findings and contained all critical and relevant findings.

The assumption of the general linear multivariate model is that the errors have a multivariate normal distribution. Box plots, tests of normality, normal plots, and analysis of the means, standard deviations, skewness, and kurtosis revealed that the various time measurements satisfied this assumption fairly well.

Two general linear multivariate models were fitted. The first model was used to test the hypothesis of no difference between filmstack time and total film time. The second model was used to test the hypothesis that there was no difference between filmstack time and prehung film time. The Wilks λ was chosen as the test statistic. In this case, this is equivalent to a paired-data t test. A Bonferroni corrected α level of .025 was used for each test to ensure an experimental type 1 error rate of .05.

RESULTS

Accuracy Data

Interpretation accuracy with a film alternator and filmstack were identical with no unacceptable reports generated with either method. Thus, all the interpretation times can be considered error-free task times. While perfect accuracy is unusual in a reaction-time experiment, it is reasonable and in fact expected with this radiologic image interpretation task. Radiologists, during residency, fellowship, and their work experience, have been trained to a near-perfect accuracy as-

ymptote to always generate a clinically acceptable report. In the past (11), radiologists have been observed to slow down, take notes, use a grease pen, and take other steps to ensure they maintain that standard of clinical acceptability.

Time Data

Case A interpretation time averaged 6.74 minutes, case B averaged 5.61 minutes, Case C averaged 6.10 minutes, and case D averaged 8.38 minutes, with the averages computed with two filmstack times and two prehung-film times for each observer. The Table details the interpretation times by observer and by display method. Load-Unload Time is the sum of the time to load the images onto the alternator and the time to unload the images from the alternator back into the film folder. Total film time of 10.84 minutes was significantly longer than the time to interpret the serial CT cases with filmstack of 6.68 minutes ($P = .0103 < .025$ [ie, the 99% confidence interval on the difference of the means does not cover zero]).

There was no significant difference between prehung film time of 6.75 minutes and filmstack time of 6.68 minutes ($P = .9183$), and four of the eight radiologists interpreted faster with filmstack than with prehung film. A 95% confidence interval on the difference between the mean prehung film time and the mean filmstack time would be (-1.43 minutes, +1.6 minutes). A confidence interval of ± 1.5 minutes may seem quite large, but the natural variation among interpreted cases was over 8 minutes, with interpretation times ranging from less than 4 minutes to almost 12 minutes, so a reduction to ± 1.5 minutes is still fairly useful. Going further and proving that the mean filmstack time and the mean prehung film time are likely to differ only by a clinically unimportant difference (ie, proving the null hypothesis) requires a much more powerful experiment. Retrospective power analysis indicated that even if we had used 12 cases and 24 radiologists, we would have been able to detect a 30-second difference only 36% of the time.

Additional exploratory analyses were conducted to support the above conclusions. A general linear multivariate model was fitted to examine the effects of the order of case presentation and the method of display. There was no significant interaction between the order in which an observer read the cases and the display

Mean Interpretation Times

Observer	Mean Time (min)				
	Load-Unload	Total Film	Prehung Film	Stacked Display	Prehung-Film and Stacked Display
1	2.80	7.54	4.75	4.07	4.41
2	8.10	18.61	10.52	7.62	9.07
3	3.66	9.34	5.68	5.04	5.37
4	3.33	8.23	4.89	5.32	5.11
5	4.37	12.61	8.24	6.71	7.48
6	3.85	14.42	10.57	10.83	10.69
7	4.57	10.07	5.50	6.92	6.21
8	2.05	5.87	3.82	6.90	5.36
Mean	4.09	10.84	6.75	6.68	6.38

method ($P = .8010$). This means that the interpretation time for an observer's first and second cases was about the same, whether the cases were presented on film or on filmstack. There was also no significant effect of order ($P = .7029$). This means that the first film or filmstack case took nearly the same amount of time to complete as the second. There were no missing data.

Observations and Observer Comments

All observers believed that filmstack would provide acceptable interpretations and that they would be willing to use a workstation with a stacked display in clinical practice. However, half of the observers stated that they would still be more comfortable with film, though all liked the fact that an electronic workstation would handle film management and manipulation better.

A number of observers commented that they sometimes had trouble "getting the big picture" (developing a spatial gestalt of the three-dimensional information contained in the CT examinations) with filmstack. One radiologist noted that film was better for determining the change in size of a pleural effusion from the initial to the follow-up examinations because this task necessitated moving the eyes rapidly back and forth over four to six images in the follow-up examination to develop a notion of the craniocaudal extent of the effusion. We believe this finding is likely to apply to spatial comprehension of a variety of longitudinally oriented structures. While several observers made use of the gestalt view (Fig 1b), most did not. Several observers suggested that the 100×100 -pixel images were too small to provide sufficient information for an adequate spatial understanding of

the data. Many observers used the "both to top of liver" key to synchronize the examinations and then scrolled back and forth, keeping the examinations synchronized.

Several experienced radiologists dictated findings "on the fly" or as they discovered them on film but sometimes memorized the findings and dictated them all at one time with filmstack. When asked, these radiologists stated that it took too much time to scroll back and forth through the examination or to write notes and that it was easier to just remember the findings and dictate them at one time. On the other hand, another radiologist reported trouble keeping track of findings with the stacked display, because the spatial location of the images on the alternator normally assisted memory of critical findings. Another radiologist used written notes more with filmstack than with a film alternator. Finally, one radiologist who had been skeptical at first about the stacked display stated that the section-by-section synchronized viewing of the initial and follow-up examinations was ideal for determining change in number and extent of multiple liver lesions.

DISCUSSION

Interpretation Time

After 20 minutes of training, average interpretation time with the stacked display was significantly faster than with total film (including the time to load and unload the film) and was slightly faster than with prehung film. Four of eight radiologists were faster with the stacked display than with prehung scans. Further reductions in interpretation time with a stacked display may occur after several minor changes are made to the user interface and radiologists receive addi-

tional training and become more familiar with both the interaction and how best to integrate a workstation into their clinical operations. Given these results and our power analysis, we would be surprised if, under clinical conditions, replacement of even prehung film with a stacked display would result in any significant reduction in operational throughput.

Our experimental results showed interpretation time with stacked display was faster than with total film, and these results are strengthened further by our decision to limit training and to use complex serial CT cases that compound the difficulties of interpretation with a stacked display. One of our main strategic decisions was that whenever we were forced to choose between two experimental design alternatives, we would be biased in favor of the existing film display and against the new stacked display workstation. This strategy would result in a stronger case for the new technology if our experimental hypothesis was found to be true and the examination would be kept within tractable means. Thus we limited our subject training because two cases were sufficient for rate interpretations and because we could never train radiologists as well with the new workstation as they had been trained in the film approach they had been using in the clinic for years. We also chose serial staging chest CT examinations for our test cases because these are a particularly difficult and common type of multiple-examination CT case (the longer the interpretation, the more differences in display-method interpretation time are accentuated) and because we believed the cases would be fairly difficult to interpret with a stacked display workstation, again with bias in favor of film when we had to make a choice.

Beard et al (11) reported that the time to load and unload the scans in one chest CT case averaged 1.18 minutes versus the mean of 4.09 minutes reported for the similar cases in this experiment (Table). Assuming CT cases of roughly similar complexity in the two experiments, doubling the number of scans to be examined in a case may roughly triple the film-handling time. Four minutes per case across 10 or 20 CT cases per day can add up to a considerable amount of radiologist time. Unless radiologists have considerable free time in their daily routine, it is clearly advantageous either to use an electronic workstation or to have allied personnel accurately preload alternators.

Interpretation Accuracy

The purpose of this experiment was to analyze error-free response time for serial CT cases. As such, the experiment was not designed to have sufficient statistical power to draw conclusions with respect to interpretation accuracy. However, given the lack of unacceptable reports with the stacked display, combined with the excellent accuracy results of Straub et al (10) with another workstation using the stack metaphor, we believe that a stacked display such as filmstack is capable of providing acceptable accuracy for interpretation of both individual and serial CT examinations.

Radiologic Accuracy versus Task-Time Experiment Design

Task-time experiments have a number of major differences from accuracy experiments. First, task-time experiments have a continuous response variable, typically seconds or minutes, whereas accuracy experiments have a binary or discrete response variable, typically an "acceptable or unacceptable" rating. Experiments with continuous response variables necessitate use of considerably fewer cases and/or observers for the same statistical power. Second, variance in observer task time is almost always much greater than variation between case and image-display method (11), so it is more important to use a larger number of observers than a larger number of cases. Third, and perhaps most importantly, accuracy and task-time laboratory results are put to very different purposes in the clinic and so have different levels of acceptable variance. Interpretation accuracy directly affects patient outcome, so radiologists correctly demand very strong experimental results before they accept a new technology. On the other hand, interpretation task time for a new display method is only one of many factors in overall clinical productivity. The "noise" from operational efficiency, case mix, and most importantly, individual differences among radiologists, mean that interpretation task times can have larger variances than those for accuracy results and still be clinically relevant.

Caveats

Observers.—The radiologists had approximately 20 minutes of training with stacked display compared with several years of training with a film alternator. We would expect the

stacked display times to improve with additional training, but we would not anticipate much change with the film times. We would expect radiologists who are less familiar with computers to eventually achieve similar performance after additional training.

A number of radiologists were uncomfortable with the stacked display interpretation method for some cases. We suspect that three-dimensional spatial inferences—such as visualization of longitudinally oriented structures—may be more complex with stacked display than with a mosaic approach. On the other hand, the stacked approach combined with the synchronization function, which allowed both examinations to be scrolled in parallel, appears to be a superior method for section-by-section comparison of CT examinations.

Reading Time.—Since radiologists could view the images while they were loading the films onto the alternator, it is possible that we have slightly underestimated the amount of time a radiologist would require to read prehung films. The prehunging of films was found to be experimentally problematic during pilot studies, owing to variance in preferred film-arrangement configurations.

Fatigue.—Experimental sessions lasted about 1 hour, whereas clinical interpretation sessions can last all morning or all afternoon. Thus, it is possible that in routine practice, mental and eye fatigue with either film alternator or stacked display might affect relative performance.

Implications for CT Workstation Design

The stacked-metaphor approach provides an effective means for interpretation of multiple CT examinations and allows interpretations that are as fast if not faster than with film. However, on the basis of observations and comments by radiologists, we believe that if cost were not a factor, use of a 2,500 × 2,000-pixel, high-brightness monitor with a 0.1-second image display time would be superior to use of a 1,024 × 1,024-pixel monitor for interpretation of CT examinations. CT images might have to be mathematically interpolated to a slightly larger size to compensate for use of smaller pixels, as are often found with high-resolution monitors. This interpolation would still allow 12 or more CT images to be displayed on a single monitor. The 2,500 × 2,000-pixel monitor could be configured for a simultaneous mosaic display of 12 CT

images from the follow-up examination with the option of displaying either the initial or the follow-up examination with the stacked metaphor. A 2,500 × 2,000-pixel monitor could also be used to display chest radiographs and other large images.

Nevertheless, we believe there are a number of circumstances in which use of a 1,024 × 1,024-pixel monitor and the stacked display approach should be considered. First, 2,500 × 2,000-pixel monitors are likely to remain more expensive and physically larger than 1,024 × 1,024-pixel monitors for some time, so stacked display can provide effective CT interpretation at a much lower cost in hardware and physical footprint. There are numerous places at a medical center where this cost trade-off might be advantageous, including in radiologists' offices and in nonradiology clinics. Second, given that the stacked display provides effective CT visualization with considerably less screen display area, it would be advantageous for display of comparison examinations. Third, we believe that use of a stacked display with a synchronization function could provide a superior means of side-by-side examination

of sections from serial examinations. Software would have to be developed to take into consideration varying section thicknesses and intervals. ■

Acknowledgments: We thank Joseph K. T. Lee, MD, and R. Eugene Johnston, PhD, for their help in the preparation of the manuscript.

References

1. Arenson RL, van der Voorde F, Stevens JF. Improved financial management of the radiology department with a microcosting system. *Radiology* 1988; 166:255-259.
2. Pizer SM. Psychovisual issues in the display of medical images. In: Hoehne KH, ed. *Pictorial information systems in medicine*. Berlin, Germany: Springer-Verlag, 1985; 235-250.
3. Beard DV, Johnston RE, Toki O, Wilcox C. A study of radiologists viewing multiple computed tomography studies using an eyetracking device. *J Digital Imaging* 1990; 3:230-237.
4. Arenson RL, Chakraborty DP, Seshadri SB, et al. The digital imaging workstation. *Radiology* 1990; 176:303-315.
5. Foley WD, Jacobson DR, Taylor AJ, et al. Display of CT studies on a two-screen electronic workstation versus a film panel alternator: sensitivity and efficiency among radiologists. *Radiology* 1990; 174:769-773.
6. Johnston RE, Yankaskas BC, Perry JR, et al. Agreement experiments: a method for quantitatively testing new medical image display approaches. *Proc SPIE* 1990; 1234: 621-630.
7. Beard DV. Designing a radiology workstation: a focus on navigation during the interpretation task. *J Digital Imaging* 1990; 3:152-163.
8. Brown JJ, Malchow SC, Totty WG, et al. Examination of the knee: interpretation with multiscreen digital workstation vs hardcopy format. *AJR* 1991; 157:81-85.
9. Berbaum KS, Franken EA, Honda H, et al. Evaluation of a PACS workstation for assessment of body CT studies. *J Comput Assist Tomogr* 1990; 4:853-858.
10. Straub WH, Gur D, Good WF, et al. Primary CT: diagnosis of abdominal masses in a PACS environment. *Radiology* 1991; 1:739-743.
11. Beard DV, Hemminger BM, Perry JR, et al. Single-screen workstation versus film alternator for fast CT interpretation. *Radiology* 1993; 187:565-569.
12. Beard DV, Hemminger BM, Pisano ED, et al. CT interpretations with a low cost workstation: a timing study. *J Digital Imaging* 1994; 7:1-7.
13. Beard DV, Pisano ED, Denelsbeck KM, Johnston RE. Eye-movement during CT interpretation: eyetracker results and image-display-time implications. *J Digital Imaging* 1994; 7:189-192.
14. Card SK, Moran TP, Newell A. *The psychology of human-computer interaction*. Hillsdale, NJ: Erlbaum, 1983.
15. Fitts PM. The information capacity of the human motor system in controlling the amplitude of movement. *J Exp Psychol* 1954; 47:381-391.
16. Rogers DC, Johnston RE, Hemminger BM, Pizer SM. Effect of ambient light on electronically displayed medical images as measured by luminance-discrimination thresholds. *J Opt Soc Am* 1987; 4:976-983.