Computed Tomography Interpretations With a Low-Cost Workstation: A Timing Study

David V. Beard, Bradley M. Hemminger, Etta D. Pisano, Kevin M. Denelsbeck, David M. Warshauer, Matthew A. Mauro, Bernadette Keefe, William H. McCartney, and Claire B. Wilcox

An ergonomically simple prototype workstation with two 900 × 1,100-pixel monitors capable of displaying eight full-resolution computed tomography (CT) images in 0.2 seconds, was compared with film for interpretation of computed tomographic images of the chest and abdomen. The hardware platform for this workstation cost less than $11,500 in 1993. A repeated-measures experiment was used to generate average interpretation times of 6.17 minutes for the workstation and 6.03 minutes for the film, including loading and unloading films, with three of the four subjects averaging about a minute longer for each workstation interpretation. All dictated reports were of clinically acceptable accuracy. All radiologists stated that workstations based on this design would be an acceptable clinical tool. However, observation suggested human working-memory strain among infrequent CT readers that could indicate the need for additional training. These data suggest that low-cost workstations can have practical application in interpretation of digital medical images such as CT, with the possibility of small increases in interpretation time.
Copyright © 1994 by W.B. Saunders Company

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

THE POTENTIAL benefits of picture archiving and communication systems (PACS)[1,2] necessitate a workstation which can facilitate fast and accurate computed tomography (CT) interpretations. Electronic image display of sufficient quality is possible with CT and the other digital modalities.[3-12] However, whereas existing commercial and experimental radiology workstations provide sufficient digital modality interpretation accuracy, most do not facilitate interpretations fast enough for clinical use.

To this end, we developed a workstation prototype, called FilmStrip,[13] which uses a single Megascan 2,048- × 2,560-pixel monitor (E-systems, Dallas, TX) to display 12 full-resolution CT images in 0.11 seconds. A good visual mental model[14,15] or metaphor is essential to a well-designed computer human interaction, so FilmStrip uses the mental model of a light box in which the images are organized in a vertical "filmstrip" with "up" and "down" buttons used for scrolling. Twelve CT images are organized in a 3-column × 4-row filmstrip with images displayed at close to the same size as film. Experimental results[13] showed FilmStrip to be as fast as film with identical accuracy for interpretation of single CT chest studies.

However, the hardware alone for a FilmStrip workstation with two 2,560 × 2,048 monitors, each capable of displaying 12 full-resolution images in 0.11 seconds, can cost over $85,000 in 1993, making it difficult to justify the cost of the technology despite the potential for rapid interpretations. Further, previous eyetracker results indicate that it may be sufficient to display only eight full-resolution images for clinically-viable CT interpretation.[16]

To address this issue, we developed FilmStrip-Let, a low-cost prototype CT workstation. FilmStripLet is ergonomically similar to the Film-Strip design described above, but it runs on a hardware platform costing less than $11,500 in 1993. FilmStripLet uses two 900 × 1,100 pixel monitors to display simultaneously 8 full-resolution CT images in 0.2 seconds, with four images on each monitor. To evaluate the clinical utility of FilmStripLet, we conducted an observer study to gather information on the relative interpretation times of film and Film-StripLet for single CT chest cases. This observer study uses the same design and protocol of the experiment previously used to evaluate Film-Strip and Film.[13]

MATERIALS AND METHODS

Equipment

Cases were interpreted with a conventional-film light box and the FilmStripLet prototype CT workstation. We constructed the FilmStripLet prototype to provide the interpre-

tation speed of FilmStrip using low-cost commodity-priced hardware (Fig 1). FilmStripLet, constructed using the X windowing system and C++, is implemented on a Sun Sparc 2 (Sun Microsystems, Mountain-View, CA) with 64 Mbyte of main memory, two 900 × 1,100 8-bit framebuffers, and two 19-in diagonal Sony Trinitron color monitors. These two monitors allow a total of eight full-resolution CT images to be displayed simultaneously, with four images on each monitor. Each of these images measures 15 cm × 15 cm. During the experiment, the two monitors were placed so that 14 cm separated the edges of each monitor's cathode ray tube.

FilmStripLet uses a vertical filmstrip mental model similar to FilmStrip. The workstation keyboard is configured into a three-button "control panel," with buttons for "up" and "down" as well as for toggling through the preset intensity windows. Each operation scrolls the filmstrip up or down by one screen-full of four images so that during an "up" scroll the four images on the right monitor are moved to the left monitor and four new images are displayed on the right monitor. Thus the radiologist can always display an image simultaneously with its pre- and post-neighbors.

To insure that there is no confusion about button location in the dimmed lighting of the interpretation environment, the keyboard's large space bar was used for the "down" button, all the letter keys were used together as a single "up" button, and the horizontal row of function buttons at the top of the keyboard was used as a single button to cycle through a set of intensity window presets. The vertical alignment of the "up" and "down" buttons allowed kinesthetic correspondence of scrolling control and resulting function (Fig 1).

Image display speed is critical. FilmStripLet running on a Sun SPARC 2 is able to scroll the workstation and display the next set of four images in under 0.17 seconds; we would expect somewhat better performance using newer UNIX workstations such as the Sun SPARC 10 at somewhat increased hardware costs. Whereas a Sparc 2 cannot intensity window a 12-bit image and move it into the framebuffer in this amount of time, it can pixel-blit an 8-bit image onto the screen very quickly; a "pixel-blit" is a fast data-movement operation that bypasses most of the operating system overhead. Because the FilmStrip system[13] showed the effectiveness of preset intensity windowing for CT workstations, we precompute lung- and soft-tissue intensity windows for all the slices in a CT study and stored these images in two large 8-bit X window pixmaps, allowing preset intensity windowing as well as 0.2 second movement of the images onto the monitors.

We believe that fixed intensity windows can be practical and even advantageous under clinical conditions. These 8-bit fixed intensity windowed images of all the slices can either be computed before the radiologist initiating the interpretation session, or can be computed when fetched for the interpretation session. With currently available low-cost main memory and fast workstation processors, both schemes are reasonable,[17] particularly if a well-designed prefetching schema[18] is used.

Observers

Four board-certified radiologists participated in the observer study. Observers 1 and 2 were very experienced with CT interpretation, and had participated in the previous FilmStrip experiment,[13] so they were familiar with the three-button keyboard configuration. Observers 3 and 4 only interpreted CT studies infrequently. All four radiologists had experience with word processors on personal computers, and all were under 45 years of age. Two additional observers participated in the auxiliary observer study, detailed in the results section.

Cases

One training case, two normal cases, and eight abnormal CT chest cases with multiple findings were used during the
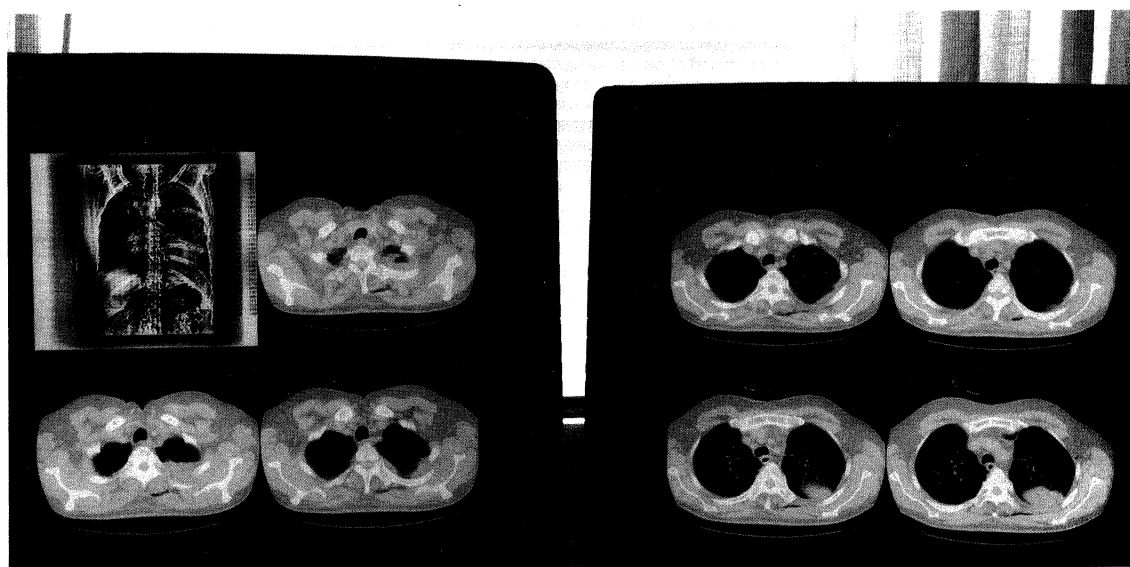

Fig 1.   Low-cost FilmStripLet with two monitors.

study. Most of these chest cases included a portion of the abdomen. All cases were generated using a Technicare 2060 CT scanner, with an average of 37 slices per case. The CT study scout view was included in the upper left-hand corner of the filmstrip. To control for the effect of case on outcome, the 10 cases were paired for difficulty and divided into two equal sets. The original requisition form with the patient history and the clinical question was typed onto a 3 × 5-in card and provided to the radiologist during the sessions. The original films were used for light-box viewing. These same cases were interpreted by observers 1 and 2 during a previous workstation timing study. However, because over one year had elapsed since that experiment, we did not expect any learning effect. Neither of these radiologists remembered the cases. Lung- and soft-tissue intensity window values were selected for each CT study by a radiologist who did not participate as an observer.

Displaying four 512 × 512 pixel full-resolution CT images on a 900 × 1,100-pixel monitor can be a challenge. To this end, we removed about 20 rows of pixels from the tops and bottoms of each image allowing two rows of these images to fit into 900 vertical pixel rows. No clinically significant information was contained in the removed pixels rows. Slightly larger monitors of the 1,280 × 1,024-pixel configuration would have eliminated the need for this step.

Design

A counterbalanced, within-subject experimental design was used, with each observer reading one group of cases using film, and the other group using FilmStripLet in separate sessions. Each case was read exactly once by each observer with the same number of readings for workstation and film. Case presentation, observer, and method order were controlled.

Procedure

Both Film and FilmStripLet sessions were conducted in a controlled laboratory environment[19] under dimmed-lighting conditions. The radiologists were told to work as quickly as possible to produce a report of clinically acceptable quality. For each film case, the radiologist was given a patient film folder containing the single CT chest study and a 3 × 5-in card with information from the attending clinician's requisition form. The radiologists loaded the images onto the light box, interpreted the films, and dictated the report using a dictation machine. With the workstation, the experimenter showed FilmStripLet's three-button interaction to the radiologist taking about 2 minutes, and the radiologist was allowed to try out the interaction for a minute or two. Then the radiologist was asked to interpret the test CT study including dictating an interpretation report. The total time for training was about 7 minutes. For each workstation trial, the appropriate study was displayed on the monitor and the radiologist was given the requisition information and instructed to begin the interpretation. The radiologist then moved through the images, changed preset intensity windows as needed, and dictated a report. The manually gathered timing data had a precision of about one second.

Data Collection and Analysis

We report film interpretation time including the time to load and unload the films from the light box; loading and unloading averaged about one minute. To insure valid interpretation times, we measured interpretation accuracy as follows: A grader developed a findings list for each dictated report and then graded the report as either acceptable or unacceptable with an acceptable report having all findings relevant to the requisition, all "critical" findings, and no incorrect findings. Dictated reports were used to provide a cognitive task similar to an actual clinical interpretation; although more precise and far easier to grade, findings forms typically used with receiver-operating characteristic (ROC) analysis act as an aid or augmentation to human working memory and thus, may distort the results of an otherwise clinically relevant observer study.

RESULTS

Data

Table 1 shows the interpretation times for each observer and interpretation method. Workstation interpretation time averaged 6.17 minutes (6 minutes and 10 seconds), whereas film interpretation time (including loading and unloading the light box) averaged 6.03 minutes. A simple two-sample t-test was used to calculate the 95% confidence interval (CI) for the difference of the mean interpretation times for film and workstation. Observed difference was -0.13 minutes (-7 seconds) with a standard error of 0.96 minutes. The 95% CI for the difference of the means is therefore -0.13 ± 3.05 minutes or simply $(-3.18, 2.92)$ minutes.

Interpretation accuracy with both film and workstation was identical, with no unacceptable reports dictated for either film or workstation, and two omitted nonsignificant findings—one each from both a film and a workstation report. These two missing nonsignificant findings were both generated by the observer 4. The findings had been detected during the interpretation of the case, but were not dictated on the final report, perhaps because of working-memory

Table 1. Interpretation Times (min:sec) for FilmStripLet and Film

	Film	FilmStripLet	Mean
Observer 1	3:45	4:43	4:14
Observer 2	7:15	4:31	5:53
Observer 3	7:17	8:44	8:01
Observer 4	5:55	6:44	6:20
Mean	6:03	6:10	

strain. The observer remembered and reported the findings to the experimenter while interpreting subsequent cases.

Critical Observations

All the observers stated that they expected to be satisfied with the interpretations they would produce using FilmStripLet in the clinic. All the observers said that given the PACS advantage of not losing films, they would be willing to use a workstation similar to FilmStripLet. These are encouraging results, particularly given the low cost of the workstation hardware. However, the radiologists did not feel as comfortable with the FilmStripLet workstation as they did with film. Further, the radiologists who had used our FilmStrip prototype workstation—using a single 2,560- × 2,048-pixel monitor—felt that Film-Strip's larger single monitor provided a more effective interpretation than that facilitated with FilmStripLet's two 900- × 1,100-pixel monitors.

The radiologists were observed (and confirmed by verbal protocol) to use mainly one of FilmStripLet's two 900- × 1,100-pixel monitors during the interpretations. One possible reason for this behavior was the distance between the monitors and the corresponding time to move the eyes between an image on the first monitor and an image on the second. Not only did the general use of only a single monitor increase the number of scrolling operations during an interpretation, but this behavior also caused considerable "thrashing" behavior, or scrolling back and forth multiple times over the same images. This was particularly a problem while viewing the various organs in the abdomen, and several radiologists noted the amount of interaction required to scroll back and forth over the same slices while viewing each of the abdomen's organs in turn. Previous results[16] have suggested that eight CT images are sufficient to avoid this back-and-forth eye movement, so the radiologists' typical use of a single monitor displaying only four images may have been the cause of this behavior. This back-and-forth eye movement behavior not only further increased the duration of the interpretation, but also increased the "cognitive load" under which the radiologist had to operate. The use of preintensity-windowed images seemed effective and the

radiologists liked being able to quickly switch back and forth between the various forms of the same image.

Several other minor observations were noted. First, as with our FilmStrip study, several radiologists wanted to use their right (strong) hand to interact both with the workstation and with the dictation machine requiring the dictation microphone to be transferred to the left hand for many workstation interactions. Careful training in using the left (weak) hand for one of the interactions or a foot control might make a small difference in interpretation time. Second, whereas we used color rather than gray-scale monitors to display the CT images, there were no complaints or comments about this affecting image quality. Note however, that the high-quality monitors and color maps were carefully adjusted before conducting the observer study, and the CT images were very large (15 cm × 15 cm).

Finally, we had not provided a marking or "grease pen" feature allowing the radiologist to denote CT images containing interesting anatomical features for subsequent description in the interpretation report. Whereas the lack of a grease pen did not appear to be a problem for the frequent CT readers, it was mentioned several times by both infrequent readers, and observer 4 appeared distracted by the omission. Both observers 3 and 4 used paper to record interesting findings while interpreting with both film and FilmStripLet. Signs of working-memory strain were noted with both observers 3 and 4 during FilmStripLet interpretation, including the use of written notes and omissions in the dictated reports that were remembered later during the dictation.

Auxiliary Observer Study

Infrequent CT readers appeared to have more difficulty with FilmStripLet interaction than frequent readers. It is possible that radiologists who do not interpret CT frequently do not have well-established "cognitive templates" for CT interpretation with either film or workstation, and thus operate under considerable working-memory strain during interpretation with the unfamiliar workstation environment. Time motion and cognitive analysis using goals, operators, methods, and selection rules (GOMS)

computer-human interaction models[20-24] suggest that FilmStripLet interaction will strain human working memory more than would film interaction, so it is not unreasonable to expect workstation interpretation speed and/or accuracy to be a function of familiarity with CT interpretation in general, as well as familiarity with workstation interaction.

To this end, two other radiologists who were well versed in chest radiographs and nuclear medicine studies, respectively, but who had not interpreted CT images during the last five years, were observed using both film and the FilmStripLet workstation to determine how experienced radiologists who are familiar with dictation, anatomy, and other radiologic tasks, but who lack the cognitive patterns for CT interpretation, would operate under the considerable cognitive working memory load of less-familiar CT interpretation. Neither observer 5 nor observer 6 had much experience interacting with computers, and neither had participated in the Film/FilmStripLet study described above. Their timing and accuracy data were not included in the results for observers 1 through 4. Cases, equipment, environment, and procedure were the same. Both observers 5 and 6 showed clear signs of working-memory strain during workstation interpretation and required from 3 to 10 minutes longer to interpret each CT study with workstation than with film.

DISCUSSION

Whereas average workstation interpretation time was very close to average film interpretation time (0.13 minute difference of means) and in fact were not significantly different, there was insufficient experimental power to conclude that film and workstation interpretation times are equivalent.[25] Further, the individual observers' times (Table 1) suggest that they are not; whereas observer 2 was about 3 minutes faster with FilmStripLet than with film—greatly contributing to the very large confidence interval—observers 1, 3, and 4 were each about 1 minute faster with film than with FilmStripLet; note further that these film times included loading and unloading the films. Observer 2 interpreted CT cases first using film followed by FilmStrip, and observation, verbal protocol, and the length of the dictated reports suggested that Observer

2's slow interpretations during the first session may have been caused by the newness of participating in a controlled subject experiment. Thus, based on examination of our very limited data, it would not be unreasonable to expect FilmStripLet interpretations to be longer than film interpretations despite average interpretations only differing by 0.13 minutes. Identical accuracy was noted in the dictations, with all reports being of acceptable clinical accuracy. However, there was insufficient statistical power from this experiment to conclude that the film and workstation interpretations result in equivalent accuracy, though other studies have shown that to be the case with other CT workstations.[11]

CT, magnetic resonance, and other digital modalities are not read in the per-day volumes of plain chest films, so a small increase in interpretation time over film should not have a significant impact on clinical throughput. Thus we conclude that a clinical workstation based on the FilmStripLet design would provide a viable low-cost platform for clinical CT interpretations. However, the low cost of the hardware would be somewhat offset by the likelihood of an increase in required training as compared with FilmStrip and its 2,560 × 2,048-pixel monitor.

Whereas FilmStripLet provided eight full-resolution images, in general, the radiologists used only four images on one monitor. Radiologists might learn to use both monitors if the distance between the cathode ray tubes were reduced. We suspect that a somewhat larger monitor would be very useful if faster interpretations are desired; larger monitors (1,600 × 1,280 pixels) simultaneously displaying six full-resolution CT images are likely to be available at commodity prices in the next few years and might prove sufficient for a single-screen low-cost CT workstation. Manufacturers could place two 1,024 × 1,024-pixel monitors into a single cabinet allowing the tubes to be closer together. However, we would expect that the additional cost and lower sales volumes of this packaging would offset any savings from the use of low-cost monitors.

There are some differences between our laboratory experiment and a typical clinical interpretation that effect how these results should be interpreted. (1) Absolute (as opposed to rela-

tive) interpretation times for both film and workstation were effected by the complex abnormal cases we selected. (2) Each interpretation session lasted less than one hour. It is possible that after longer interpretation sessions, mental and eye fatigue might effect relative performance or comfort. (3) Experimental sessions were conducted under dim lighting and were controlled for interruptions and other cognitive distractions. We would expect workstation interpretations to be somewhat worse relative to film under clinical conditions because of the limited luminosity of the monitors and the workstation's possible susceptibility to working-memory strain. (4) On the other hand, we conducted these experiments after only a few minutes of training. Because interaction familiarity can greatly reduce cognitive load and computer-human interaction speed, we would expect Film-StripLet to improve considerably relative to film

after several weeks of training and clinical experience. Based on the results of this study, we speculate that infrequent CT readers will require more training to become comfortable with workstations than more frequent CT readers regardless of computer experience. (5) Film-StripLet is a prototype designed to evaluate the effectiveness of a simple interaction, fast response time, and two $900 \times 1,100$-pixel monitor configuration. Actual workstations in the clinic would also require connections to hospital and radiologic information systems, measurement tools, and a small number of other features. These additional features might effect workstation interpretation times. (6) Radiologists may not behave the same under experimental and clinical conditions, though by testing both methods under the same conditions, and counter balancing the design, this effect should have been minimized.

REFERENCES

1. Arenson RL, van der Voorde F, Stevens JF: Improved financial management of the radiology department with a microcosting system. Radiology 166:255-259, 1988

2. Beard DV, Parrish D, Stevenson D: A cost analysis of film image management and four PACS based on different network protocols. J Digit Imaging 3, 108-118, 1990

3. Pizer SM: Psychovisual issues in the display of medical images, in Hoehne KH (ed): Pictorial Information Systems in Medicine. Berlin, Germany, Springer-Verlag, 1985, pp 235-250

4. Pizer SM, Beard DV: Medical image workstations: State of science and technology. J Digit Imaging 2:185-193, 1989

5. Arenson RL, Chakraborty DP, Seshadri SB, et al: The Digital Imaging Workstation. Radiology 176:303-315, 1990

6. 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 174:769-773, 1990

7. Johnston RE, Yankaskas BC, Perry JR, et al: Agreement experiments: A method for quantitatively testing new medical image display approaches. Proceedings of SPIE Medical Imaging IV: PACS System Design and Evaluation 1234:621-630, 1990

8. Beard DV: Designing a radiology workstation: A focus on navigation during the interpretation task. J Digit Imaging 3:152-163, 1990

9. Brown JJ, Malchow SC, Totty WG, et al: MR examination of the knee: Interpretation with multiscreen digital workstation vs hardcopy format. Am J Radiol 157:81-85, 1991

10. Berbaum KS, Franken EA, Honda H, et al: Evaluation of a PACS workstation for assessment of body CT studies. J Comput Assist Tomogr 4:853-858, 1990

11. Straub WH, Gur D, et al: Primary CT diagnosis of abdominal masses in a PACS environment. Radiology 1:739-743, 1991

12. Beard DV, Hemminger BM, Denelsbeck KM, et al: A time-motion comparison of several radiology workstation designs and film. SCAR 92 Symposium for Computer Assisted Radiology, Baltimore, MD, June 1992, pp 588-594

13. Beard DV, Hemminger BM, Perry JR, et al: Single-screen workstation vs. film alternator for fast CT interpretation. Radiology 187:1-6, 1993

14. Rumelhard D, Norman D: Analogical processes in learning, In Anderson JR (ed): Cognitive Skills and Their Acquisition. Hillsdale, NJ, Erbaum, 1981, pp 335-359

15. Young R: The machine inside the machines: User's models of pocket calculators. Int J Man-Machine Stud 15:51-85, 1981

16. Beard DV, Johnston RE, Toki O, et al: A study of radiologists viewing multiple computed tomography studies using an eyetracking device. J Digit Imaging 3:230-237, 1990

17. Hemminger BM: Design of useful and inexpensive radiology workstations. SCAR 92 Symposium for Computer Assisted Radiology, Baltimore, MD, June 1992, pp 615-624

18. Wendler TH, Grewer R, Monnich KJ, et al: Design considerations for multi modality medical image workstations, in Hohne KH (ed): Pictorial Information Systems In Medicine. Berlin, Germany, Springer-Verlag, 1985, pp 401-420

19. Rogers DC, Johnston RE, Hemminger BM, et al: Effect of ambient light on electronically displayed medical images as measured by luminance-discrimination thresholds J Opt Soc Am 4:976-983, 1987

20. Gilbreth FB: Motion Study. New York, NY, Van Nostrand Reinhold, 1911

21. Card SK, Moran TP, Newell A: The Psychology of Human-Computer Interaction. Hillsdale, NJ, Erlbaum, 1983.

22. Fitts PM: The information capacity of the human motor system in controlling the amplitude of movement. J Exp Psychol 47:381-391, 1954

23. 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

24. Beard DV, Hemminger BM, Denelsbeck KM, et al: How many screens does a CT workstation need? J Digit Imaging 7:69-76, 1994

25. Metz CE: Quantification of failure to demonstrate statistical significance: The usefulness of confidence intervals. Invest Radiol 28:59-63, 1993