Iterative Interim Techniques for Reduced Costs and a Better Mammography Workstation: An Opinion

David V. Beard

The promise of digital mammography has focused considerable resources on the challenges of mammography workstations, but the risk of wasted time and resources in such efforts is very high. Although final testing of the workstation's image quality and ergonomics is common, a number of interim evaluation and refinement techniques can be applied throughout the design and development process. The use of such techniques holds potential not only to save time and money but also to produce a superior workstation.

Copyright © 2000 by W.B. Saunders Company

KEY WORDS: mammography, workstation, diagnostic radiology, observer performance, receiver operating characteristic curve.

MEDICAL IMAGING workstations for soft copy interpretation are complex collections of hardware and software for human operation. No one has yet built one good enough to stand above the crowd. Now corporations, university hospitals, and the National Institutes of Health are focusing considerable resources on the challenges of workstations for mammography because of the recent viability of digital mammography and its need for softcopy display. The very high resolution and other special needs of such displays impose unusual requirements.¹ At the same time, it will be vital to incorporate appropriate human factor designs in the workstation and in its surroundings to facilitate comfortable and accurate diagnostic interpretations for sustained periods of time.

The risk of wasted time and effort in such efforts is very high. The challenges inherent in any workstation and the special difficulties for mammography in the midst of urgency and available funding will cause excessive expenditures without guarantee of success unless testing techniques are applied early and during the process to evaluate the design in an iterative fashion.

Many would think normally of evaluating only a finished workstation. But one does not want to spend hundreds of thousands of dollars developing and building a mammography workstation only to discover the design was fatally flawed. One needs to use simple, fast, and low-cost evaluations throughout the workstation development process to detect flaws early and to insure an acceptable design. Then, a rigorous evaluation of the finished workstation can occur.

This report includes various evaluation techniques that could be used during workstation design and implementation for evaluating workstation ergonomics and for evaluating image quality to create an effective mammography workstation while containing cost.

EVALUATION DURING MAMMOGRAPHY WORKSTATION DEVELOPMENT

The software development lifecycle for a workstation² is a complex process involving 4 steps: (1) *task analysis*, or understanding the user's situation, problems, and goals to be addressed by the workstation; (2) *design*, in which a workstation that will address the radiologist's situation is specified carefully including the hardware components, the interaction design, the internal organization of the software, and the organization and lighting in the reading room³⁻⁵; (3) *implementation*, in which the workstation software (and hardware if needed) is constructed; and (4) *testing*, in which the workstation is evaluated to insure it meets the radiologists' needs. Although the software development lifecycle often is presented as several distinct steps, in actual use, the lifecycle is a highly iterative process with the design phase often discovering flaws in the task analysis and the implementation phase discovering flaws in both earlier segments. The further into this process a flaw is discovered, the more difficult, costly, and time consuming it is to correct the deficiency.

Flaws found only in the final product can be particularly costly and time consuming with a mammography workstation. However, inexpensive *development evaluations* can be applied throughout the software lifecycle. These evaluation techniques should facilitate design improvement or

From the Department of Computer Information Systems, Idaho State University, Pocatello, ID.

Address reprint requests to David V. Beard, Associate Professor of Computer Information Systems, Idaho State University, Pocatello, ID 83209-8020.

*Copyright © 2000 by W.B. Saunders Company
0897-1889/00/1304-0001\$10.00/0
doi:10.1053/jdim.2000.17763*

rejection before the enormous expense of software implementation, prototype construction, and laboratory or field evaluation.

Task Analysis Evaluations

Understanding radiologists' tasks is the essential first step in design. It is important that designers comprehend the differences between screening interpretations, diagnostic workup, and local and remote consultations. Further, any image-viewing task occurs in the context of the clinical operation, so clinic and hospital workflow patterns also must be considered. Observation and possibly videotaping of radiologists in the reading room allows details of the workflow to be understood better by the workstation designers. An eyetracker^{6,7}—a head-mounted device that tracks which images the radiologist is viewing at any instance during the interpretation—can be especially useful for clarifying image-viewing patterns, which are an essential aspect of task analysis. By repeatedly presenting task analysis results to a selected group of radiologists for critique after each revision, the task analysis document can be improved iteratively.

Design Evaluations

The design phase should be driven by a few carefully selected user tasks such as selecting a case, scrolling through the current and comparison examinations, and choosing preset "intensity window" grayscale parameters. Ideally, a single workstation design optimally would facilitate all possible radiologic tasks while still being simple, easy to use, easy to learn, and low cost. But this rarely is the case, and often a choice must be made between extensive features and simplicity. "Feature creep" is an ever-present danger in which more and more functions are added at the expense of increased learning time, interpretation time, interpretation errors, and system costs. It is essential to stay focused on the goal of a workstation that first and foremost works as effectively and simply as film and alternator.

Selection of the workstation's hardware components, such as the monitors, frame buffers, and the underlying computer and network system particularly is important. Calculation or testing by computer scientists can estimate image-display times.⁸ The underlying computer system and related operating system must not only be robust and facilitate rapid display of the mammographic images, but

also must have sufficient ubiquity to facilitate cost-effective maintenance by the radiology department or hospital's computer information systems group.

Monitor grayscale or spatial resolution and the related monitor modulation transfer function (MTF) can be determined from the literature or measured in the laboratory before significant software development.^{9,10} But total workstation image quality also must be established.

IMAGE QUALITY EVALUATION

Assuring acceptable workstation image quality is essential if this technology is to replace film and alternator. Image quality can be established before the interactive portion of the workstation's software is written either from the literature or with receiver operating characteristic (ROC) perception experiments.¹¹⁻¹³ With such an experimental design, radiologists view the displayed images while filling in a detailed findings form. This form subsequently is compared with known truth to determine sensitivity and selectivity for both film and the candidate softcopy display. "Truth" for each case is determined with pathology-proven positives and normals having a multiyear period sans abnormalities.

ROC experiments are costly, so the image quality of the selected individual image-display components (monitor, frame buffer, computer system, and related image transformation, storage, and display hardware or software) should be established from the literature or by experimentation before committing to a final mammography workstation interactive design. Additionally, smaller image quality "pilot studies" can be conducted before committing to the final ROC experiment with its significant costs.

ROC experiments are an excellent method for establishing workstation image quality. However, for a number of reasons, ROC experiments are not appropriate for simultaneously evaluating workstation ergonomics. First, ROC studies are focused on strictly comparing what can be visualized with softcopy as opposed to the established film-displayed images. As such, ROC studies require very specialized observer behavior using a generalized findings form that is designed only for the ROC experiment. In the clinic by comparison, radiologists are focused on addressing a driving issue posed by the referring physician.

Second, clinical interpretations often use a dic-

tated report requiring the radiologist to mentally compose a report while interacting with the mammography workstation. The presence of a clinical question, combined with the use of a dictated report produces very different behavior in the clinic. Third, actual clinical interpretations often involve comparing a current examination with a prior one, generating even more behavior differences. And fourth, although the ROC viewing task generally has no time limits, which is ideal for assessing image quality, the clinical environment often is severely time constrained, again resulting in differing behavior and possibly different interpretation times.

ERGONOMIC TECHNIQUES

Ergonomic design guidelines such as those available from Apple or Microsoft^{14,15} can be helpful during the workstation design phase, especially if applied by experts in ergonomics who are also familiar with radiology interpretation and operations. Such experts can be located through the Human Factors Society (hfes.org), the Association of Computing Machinery's Special Interest Group on Computer Human Interaction (acm.org/sigchi) or the Usability Professional's Association (upasoc.org). Following standard guidelines like these helps insure that the interaction of the mammography workstation is similar to that of word processors and other computer tools familiar to the radiologist. Such similarity can considerably reduce workstation learning time.

The design of the workstation interaction should be an iterative process incorporating designers, radiologists, and hospital administrators. Explaining the workstation design to noncomputer professionals can, however, be difficult because users have trouble visualizing how the tool will be used. A sequence of simple sketches called a *storyboard* can be particularly helpful in showing how a user might perform a particular task (eg, changing intensity window settings) with a particular workstation design. Such comic strip-like storyboards often are used in laying out movies and cartoons. A more sophisticated version of the storyboard involves inexpensive "claymation" videos of computer or even cardboard interface mockups to visualize how a radiologist would use the candidate tool to perform some task.^{16,17}

Time-motion analysis^{18,19} can also be used to compare mammography workstation design alter-

natives. Time-motion analysis involves breaking down a user's intended task, such as selecting the "print" command from a menu, into smaller and smaller subtasks until "atomic" tasks like button presses and mouse movements are encountered; the durations of these atomic tasks are available in the literature—for example, a mouse button press takes 0.1 seconds on average. By summing the durations of all the atomic tasks, the duration of the user's original task can be estimated. Even rough estimates from a time-motion analysis can show the hopelessness of a poor design, and these estimates often are sufficient for choosing between design alternatives.

Even simple laboratory ergonomic evaluations of prototype mammography workstations are costly and time consuming compared with the techniques described above but still considerably less costly than a formal controlled subject experiment with a completed mammography workstation. Such prototype evaluations can range from quick observation sessions to the more formal ergonomic experiment described below.

ERGONOMIC EXPERIMENT

Although the quality of the mammogram displayed on the monitor might wrongly be the only focus of a workstation evaluation, the workstation computer-human interaction, the environment in which the workstation is to be used, and the seating and other ergonomic considerations can be equally important. The menu layout, placement of soft buttons and other features on the monitor, the use of mouse and cursor to select screen items, and most importantly, the functions provided to the radiologists for interpreting the examination all affect a workstations ergonomic qualities. Will room lighting be indirect, dimmable? Is task lighting adequate? Is the viewing angle and distance correct, adjustable? Has external glare been eliminated from the displays? Is the position of one workstation going to create glare for another?

Poor ergonomics not only can affect the examination interpretation time, but also can actually degrade the quality of the radiologist's interpretation or lead to unnecessary fatigue. When dictating a report during a multiple examination case for example, radiologists are forced mentally to keep track of their interpretation plan, clinical findings, and report online, all while trying to interact with a complex software system. Human working mem-

ory can be stressed and even overwhelmed by the combination of a complex case and poor workstation ergonomics, causing radiologists to forget critical findings.^{20,21}

Assuming that the workstation's image quality has been established, the workstation's ergonomics can be evaluated with a relatively inexpensive laboratory study that uses dictated reports. As with many aspects of computer-human interaction, there is a speed-accuracy tradeoff in medical image interpretation. As radiologists increase examination interpretation speed, eventually they will start making more errors. Thus, to accurately compare the average interpretation time for a mammography workstation with film on an alternator, interpretation accuracy must be controlled or held constant while interpretation time is measured.

The least expensive way to do this is to consider the interpretation time for only those examinations that are interpreted correctly. This *error-free interpretation time* can be measured in the following manner. With 1 possible design (within-subject repeated-measures experimental design) a number of mammographers are asked to use both film alternator and the workstation to interpret a set of carefully selected cases, each with a clinical question posed by a referring physician; very complex cases with multiple examinations and many findings seem best able to illuminate ergonomic flaws in the workstation design. For each case, the radiologists are instructed to "dictate a report of acceptable clinical accuracy while working as quickly as possible." The reports later are compared by other expert mammographers to "truth"—typically pathology results or a sufficiently long history of subsequent normal results after the examination. If the interpretation report is considered "clinically acceptable" the associated interpretation time is incorporated into the error-free time calculations.

Ergonomic considerations in the design of the reading room, the height of the table, mouse, and

keyboard, and even the chair on which the mammographer sits, can significantly affect not only interpretation time, but also possibly interpretation accuracy. Glare, visual and auditory distractions, and other interruptions and discomforts should be reduced or eliminated.

MAMMOGRAPHY WORKSTATION COST-EFFECTIVENESS

Although important, workstation interpretation time is only one component of *clinical throughput*, that is, the number of cases that can be interpreted in a given period of time. Clinical throughput also should be measured during a field trial or estimated using a computer simulation model that looks for bottlenecks and excess capacity in the clinic. Such a throughput model uses "workflow" data derived from a film-based clinical operation augmented with workstation-interpretation data from laboratory-timing studies.²²

Further, clinical throughput is one of several inputs into a *lifecycle cost analysis* that radiologists and hospital administrators must consider before making any final-acceptance decision. It is possible that a mammography workstation may require a bit more time to interpret a single case, but may be more cost effective than film because of reduced film costs and better utilization of the radiologist's time when images are shipped from remote locations. Throughput analysis requires industrial engineering skills to construct and verify a model of the clinical operation. Cost analysis requires expertise in managerial accounting.

CONCLUSION

Excellent workstations for viewing digital mammograms are needed and now are being designed. It is hoped that evaluations throughout the design process will be used to detect flaws early, leading, in turn, to more effective workstations at lower cost.

REFERENCES

1. Behlen F, Blume H, Flynn MJ, et al: Report of the Working Group on Digital Mammography: Digital Displays and Workstation Design, Sponsored by USPHS Office on Women's Health and NCI, March 9-10, Washington, DC
2. Mantei MM, Teorey TJ: Cost/benefit analysis for incorporating human factors in the software life cycle. CACM, April, 1988
3. Horii CS: Electronic Imaging Workstations: Ergonomic issues and the User Interface, Radiographics 12:773-787, 1992
4. Horii CS, Horii HN, Mun SK, et al: Environmental designs for reading from imaging workstations: Ergonomic and architectural features. J Digit Imaging 2:156-162, 1989
5. Johnson CW: Impact of working environments upon human-machine dialogues: A formal logic for the integrated

specification of physical and cognitive ergonomic constraints on user interface design. *Ergonomics* 39:512-530, 1996

6. Gale AS, Worthington BS: The Utility of Scanning Strategies in Radiology, Eye Movements and Psychological Functions: International Views. Hillsdale, NJ, Erlbaum, 1983

7. Beard DV, Bream P, Pisano ED, et al: A pilot study of eye movement during mammography interpretation: Eyetracker results and workstation design implications. *J Digit Imaging* Feb, 1996

8. Huang HK, Wong AWK, Zhu X: Performance of asynchronous transfer mode (ATM) local area and wide area networks for medical image transmission in clinical environment. *J Comput Med Imaging Graphics* 21:165-173, 1997

9. Roehrig H, Blume H, Browne M: Performance test and quality control of cathode ray tube displays. *J Digit Imaging* 3:134-150, 1990

10. Krupinski E, Roehrig H: Influence of monitor luminance and tone scale on Observer Detection Performance. *Medical Imaging 98: Image Perception* 3340: 1998

11. Metz CE: Basic Principles of ROC analysis. *Semin Nucl Med* 8:283-298, 1978

12. Metz CE: Evaluation of digital mammography by ROC analysis, in Doi K, Giger ML, Nishikawa RM, (eds): 3rd International Workshop on Digital Mammography, Amsterdam, Elsevier Science (Excerpta Medica International Congress series), 1996, pp 61-68

13. Metz CE: Evaluation of radiologic imaging systems by ROC analysis. *Med Imag Inform Sci* 12:113-121, 1995

14. McKay EN: Developing User Interfaces for Microsoft Windows. Redmond, WA, Microsoft Press, 1999

15. Apple Computer Inc: Macintosh Human Interface Guidelines. Reading, MA, Addison-Wesley, 1992

16. Hackos JT, Redish JC: User and Task Analysis for Interface Design, New York, NY, Wiley, 1998

17. Dix AJ, Finlay JE, Abowd GD, et al: Human-Computer Interaction (ed 2). London, England, Prentice Hall, 1998

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

19. Beard DV, Smith DK, Denelsbeck KM: Quick and Dirty GOMS: A Case Study of Computed Tomography Interpretation. *J Human Computer Interaction* 11:157-180, 1996

20. Beard DV, Molina PL, Muller KE, et al: Interpretation Time of serial chest CT examinations with stacked-metaphor workstation versus film alternator. *J Radiol* 197:753-758, 1995

21. Beard DV, Hemminger BM, Denelsbeck KM, et al: Cognitive load during CT interpretation, SPIE Medical Imaging: Image Perception, 1994

22. Kilpatrick KE, Mackenzie RS, Kisko TM: Expanded function auxiliaries in general dentistry: A computer simulation. *Health Services Research*. vol 7:288-300, 1972