Evaluating user interfaces for stack mode viewing

M. Stella Atkins, Arthur E. Kirkpatrick, Adelle Knight, School of Computing Science, Simon Fraser University, B.C. Canada

Bruce Forster, Department of Radiology, University of British Columbia, B.C., Canada

Abstract

The goal of this research was to evaluate two different stack mode layouts for 3D medical images—a regular stack mode layout where just the topmost image was visible, and a new stack mode layout, which included the images just before and after the main image. We developed stripped down user interfaces to test the techniques, and designed a look-alike radiology task using 3D artificial target stimuli implanted in the slices of medical image volumes. The task required searching for targets and identifying the range of slices containing the targets.

Eight naïve students participated, using a within-subjects design. We measured the response time and accuracy of subjects using the two layouts and tracked the eyegaze of several subjects while they performed the task. Eyegaze data was divided into fixations and saccades

Subjects were 19% slower with the new stack layout than the standard stack layout, but 5 of the 8 subjects preferred the new layout. Analysis of the eyegaze data showed that in the new technique, the context images on both sides were fixated once the target was found in the topmost image. We believe that the extra time was caused by the difficulty in controlling the rate of scrolling, causing overshooting.

We surmise that providing some contextual detail such as adjacent slices in the new stack mode layout is helpful to reduce cognitive load for this radiology look-alike task.

Keywords: stack mode image layout, observer performance evaluation.

1 Introduction

Radiologists often need to navigate through long sequences of 2D image slices generated from MR or CT volume data. With current computer systems, it is possible to store and display large sets of images quickly in many different formats. For example, there are up to 1000 images for abdominal CT exams [Reiner et al. 2003]. This evolution in technology must be accompanied by a necessary evolution in radiology workstation software, as the 3D medical data must somehow be viewed on 2D computer monitors.

In order for image viewing software to be valuable, it must display the images in a manner that is useful for the tasks that must be accomplished by a radiologist. Although the technical hurdles of digital radiology have been met, there are human-computer interaction (HCI) problems that remain to be solved [van der Heyden et al. 2001], [Moise and Atkins 2004]. In this paper, we discuss the design of stack mode image layouts and the results of a small evaluation on the software. Stack mode, or cine-mode, is an image layout format for display monitors where images in a volume set are stacked on top of one another. At any given time, only the top image is visible on the screen but images

Medical Imaging 2007: Image Perception, Observer Performance, and Technology Assessment,
edited by Yulei Jiang, Berkman Sahiner, Proc. of SPIE Vol. 6515
65150A, (2007) · 1605-7422/07/$18 · doi: 10.1117/12.707533

can be dynamically traversed using a mouse or suitable pointer, at high refresh rates. Scrolling through images in stack mode presents the images as an animated sequence. Stack mode is thus convenient for viewing multiple 2D slices of 3D objects on the limited area of display screens [Mathie and Strickland 1997, Ellis et al. 2006]

In radiology, stack mode viewing of cross sectional radiological images has become very popular. Because stack mode presents image sequences over time, tumours are more readily appreciated by the human visual system than when the same sequence is presented over space, as the eye is very sensitive to motion. Hence if the viewer maintains their gaze on a specific spatial region as the images change, the 3D relationship of various structures is better seen, and a tumour can be differentiated from a normal blood vessel. The value of image motion in detecting abnormalities in 3D radiological data sets has recently been studied and validated by Krupinski et al. in a different context, rotating 3D images around a z-axis for detection of stenoses in MRA images [Krupinski et al. 2006a]. A few studies have shown that stack mode viewing is much faster then tile-mode, and just as accurate [Ellis et al. 2006], [Kim et al. 2005], [Mathie and Strickland 1997]. However, although there are several stack-based viewers in use today, there has been very little research on the design of interfaces for using stack mode viewers.

We observed that during stack mode viewing there was a lot of forward and backward scrolling to locate the boundaries of abnormalities. To avoid this extra scrolling, we designed an enhanced stack layout (called 3-up) where the user may see the next and previous images in the set while still viewing the current image (see Fig. 1). The size of the previous and next stacked images was made a little smaller than the main image to fit into a 17" monitor, but it has been shown that radiological performance can be improved with large displays [Krupinski et al. 2006b], so we could display them all at the same size in future large screen displays. Our hypothesis was that the new layout would be more efficient for detecting the boundaries of anomalies, because adjacent images were available for immediate comparison to the main image.

Both layouts allow the user to navigate through images using a mouse, in a "click and hold" manner. If the user makes a single right click, the next image is displayed in the centre; conversely, for a left click, the previous image is displayed in the centre. If the user clicks and holds the button down, the images are scrolled through at a fixed rate, about 4 frames per second. The rate can be adjusted up or down using the speed control under the left of the central image (see Fig 1).

Our objective was to determine which layout was more effective for locating anomalies in sets of images. We focussed on the speed and accuracy with which anomalies could be located with each layout.

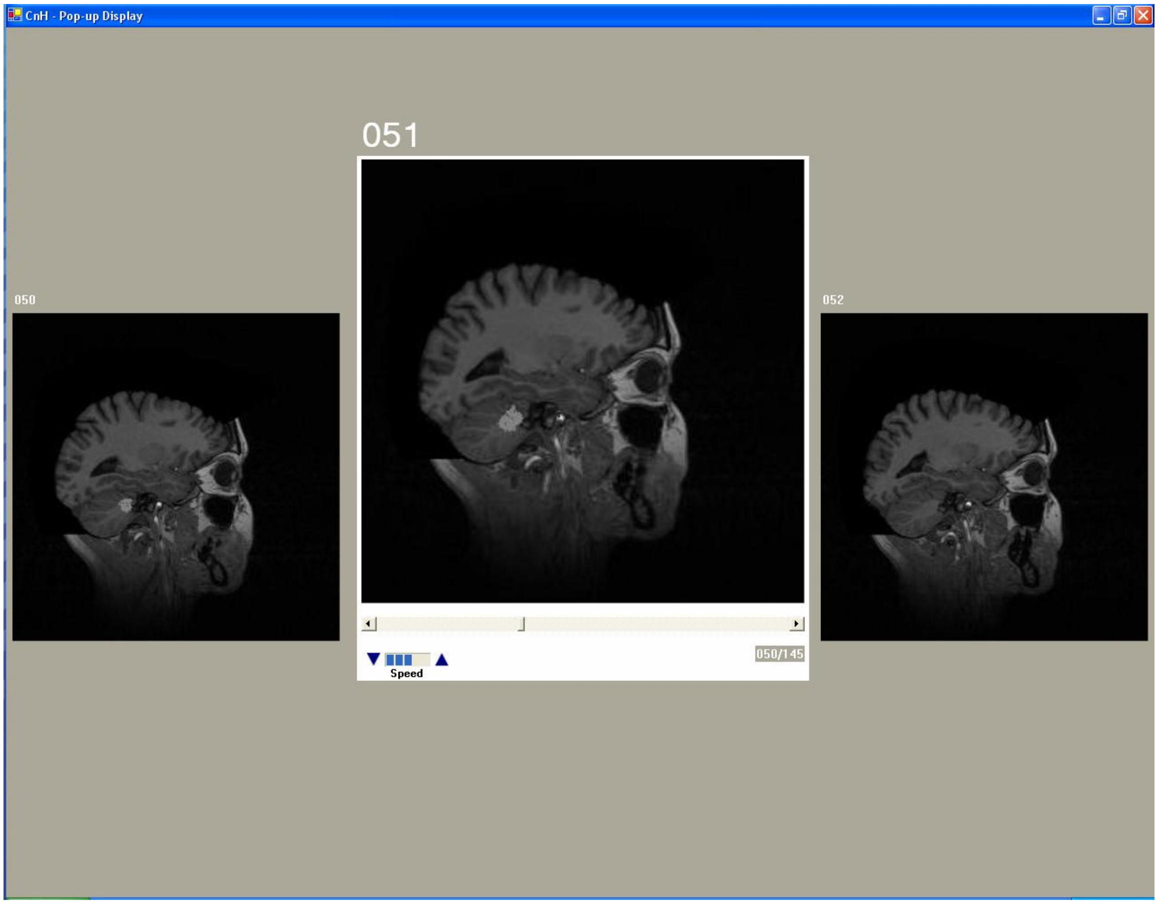

Figure 1: The new stack viewer (3-up layout).

Because radiologists already use "linked" images, they are used to scrolling through one image set and having another automatically scroll at the same rate e.g. for PD and T2 weighted MRI images. This suggests they will tolerate the apparent motion of another image, usually above or below (but here, to the side of) the image they are focussing on. They certainly do not look at more than one image set while they are scrolling, even in a linked set (self report from a dozen radiologists).

2 Materials and Methods

We designed and implemented prototypes of the two layouts, using the "click and hold" mouse interaction technique.

We performed a controlled user study to determine the speed and accuracy of using both kinds of stack viewers. We used students as subjects, with a look-alike radiology task to detect artificial hyper-intense regions representing roughly spherical lesions placed on MRI and CT data slices. We have used students in the past to compare different HCI methods for radiology workstation designs, and have found that in a look-alike radiology task on a stripped-down workstation, the students' and radiologists' performances were indistinguishable [Moise et al., 2005].

2.1 Stimuli

The stimuli were created by overlaying artificial 3D target anomalies on six real MR image volumes of different body parts (four head, one knee, one thigh). The anomalies were light colored (mostly white or gray) and were meant to stand out once seen. All the anomalies had a roughly spherical shape and therefore spanned several 2D image slices. Each image volume had versions with one and two anomalies, for a total of 12 stimuli.

Consecutive image slices containing a typical anomaly are shown in Fig. 2.

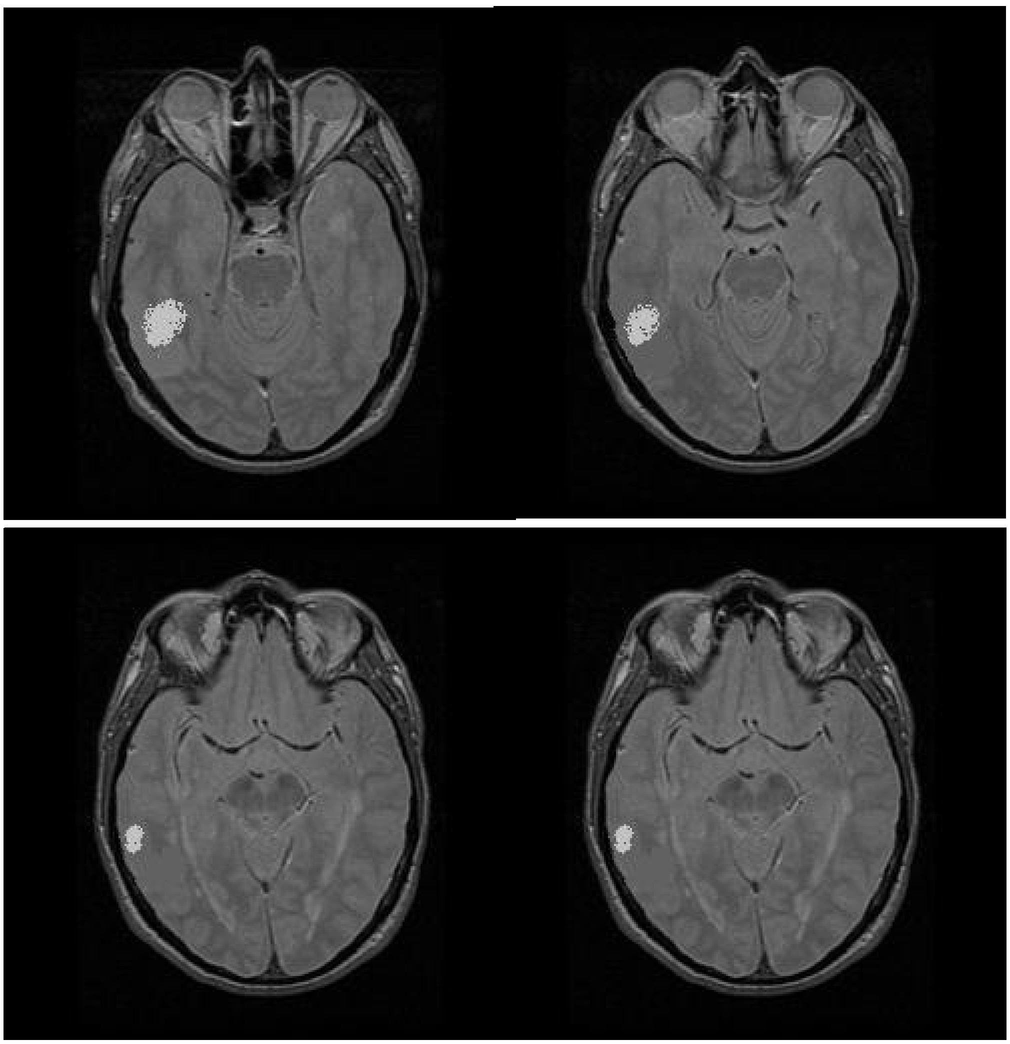

Figure 2. Consecutive image slices containing an artificial anomaly .

2.2 Experiment Design

Design. The design was a within-subjects comparison of two conditions, the 1-up and 3-up layouts.

Task. Subjects were asked to find all anomalies in a volume and specify the first and last slices of each anomaly. Subjects operated the mouse with their preferred hand and spoke the anomaly boundaries aloud, which were recorded by the experimenter. Subjects did not use the keyboard during the task. Subjects were told that they could examine the volumes in either direction, and could report the first and last anomaly slices in either order.

Dependent measures. We recorded the total time for each trial, as well as the range(s) the subject gave for anomalies. A Tobii eye-tracker recorded subjects' saccades and fixations.

Participants. There were 8 graduate student participants, volunteers from neighbouring computer science labs. They were not rewarded for participation.

Protocol. Each subject performed the same anomaly detection task using both layouts (1-up and 3-up). Each subject performed two blocks of 4 trials: one block with 1-up layout, and one with 3-up layout. Each block contained a total of 6 anomalies, embedded in 4 different volumes. Order of layouts was counterbalanced. Each block began with a practice trial, using a stimulus different from the ones used in the main trials. The practice stimulus, which was used in both blocks, had six anomalies.

3 Results and Discussion

3.1 Accuracy and Time

Subjects were equally (100%) accurate with both layouts.

Every subject was slower using the 3-up layout. These results are shown in Fig. 3. The mean time for 1-up was 33.7 secs, and the time for 3-up was 40.2 secs.

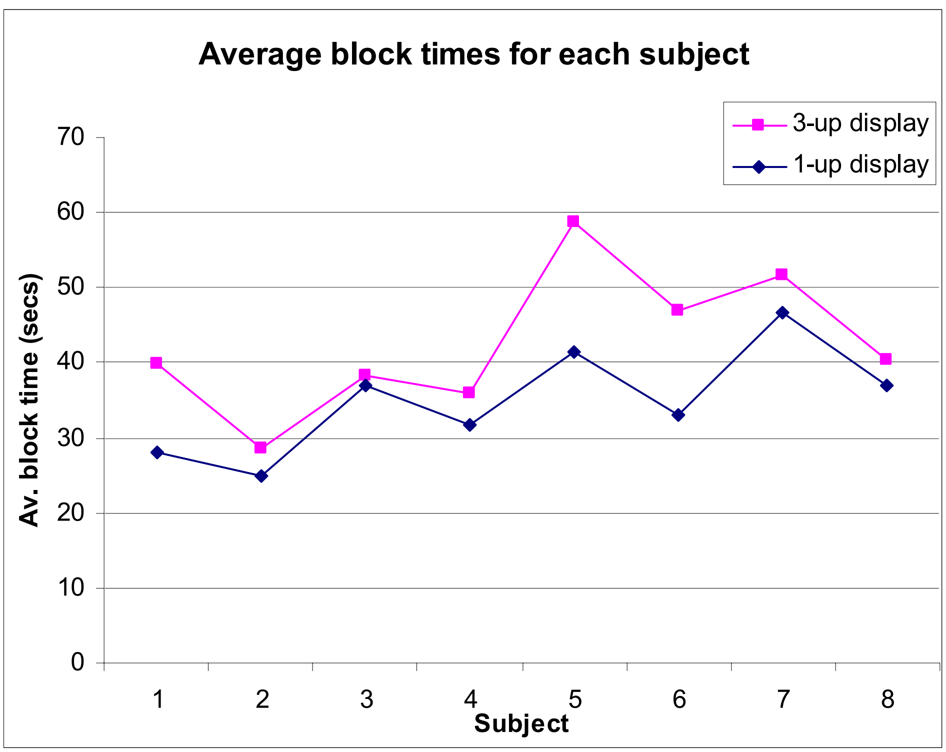

Figure 3. Plot of layout against mean trial time (secs) over all 8 subjects.

In the post-session questionnaire, 5 out of the 8 subjects preferred the 3-up layout over the 1-up layout (despite being slower in every case). Typical comments included:

"3-up is easier to detect the start and end images of the anomaly" (S2 who preferred the 3-up layout).

"I found the peripheral images a distraction since I could not keep them in my visual field" (S1 who preferred the 1-up layout).

The timing results were counter to expectation—i.e. 3-up is slower than 1-up. In our quest to discover why, we studied the eyegaze data.

3.2 Eyegaze Data

Analysis of the eyegaze data included watching animated replays of subjects' gazes and studying static "gaze plots" (screen shots overlaid by the sequence of fixations and saccades). We used the "hot spots" display to define areas of interest (AOI)s, and performed statistics on the AOIs.

The gaze replays show that the user overshot when an anomaly appeared whilst scrolling. The overshoot depends on the scrolling speed, fixed for this task at about 4 frames/second. The overshoot is about 700 msecs, corresponding to the reaction time to stop scrolling once an anomaly has appeared in the peripheral vision. Also the 3-up layout involves more scrolling back and forth, between previous, central and next images, to determine the anomaly boundaries.

Figure 4 shows a typical gaze plot on the 3-up layout, with the areas of interest highlighted. It is seen that the context images on both sides were fixated once the target was found in the central image. Also, multiple saccades between adjacent slices were used during the decision making process. Note that most saccades are between the centre images and those on either side; there are hardly any long saccades between the previous and next images.

We noted that 7/8 subjects used the side images in the 3-up layout, but one subject deliberately chose to focus on just the central image.

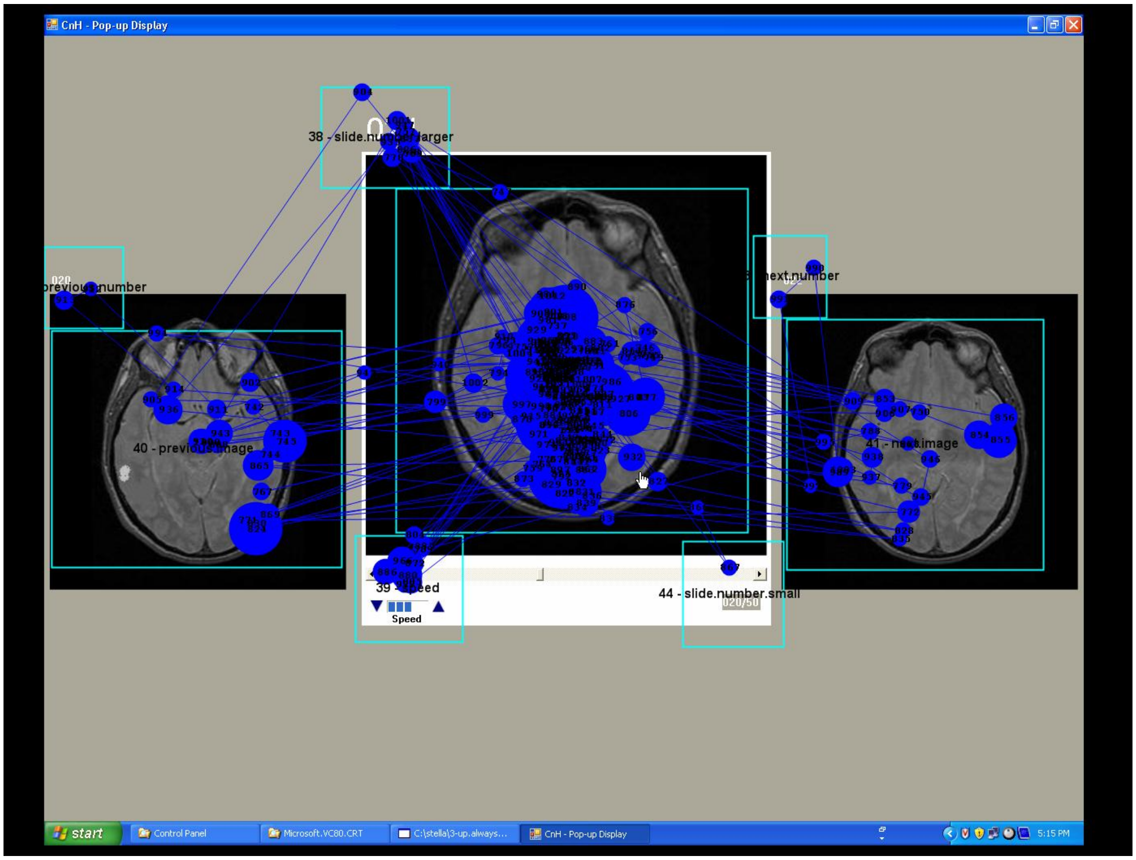

Figure 4. Gaze plot of fixations on the 3-up layout. The circles represent fixations, and the lines show saccades. Areas of interest have been outlined.

Figure 5 shows the hot spot view of fixations on the 3-up layout corresponding to the gaze plot in Figure 4.

These views were used to define the areas of interest shown outlined in Figs. 4 and 5.

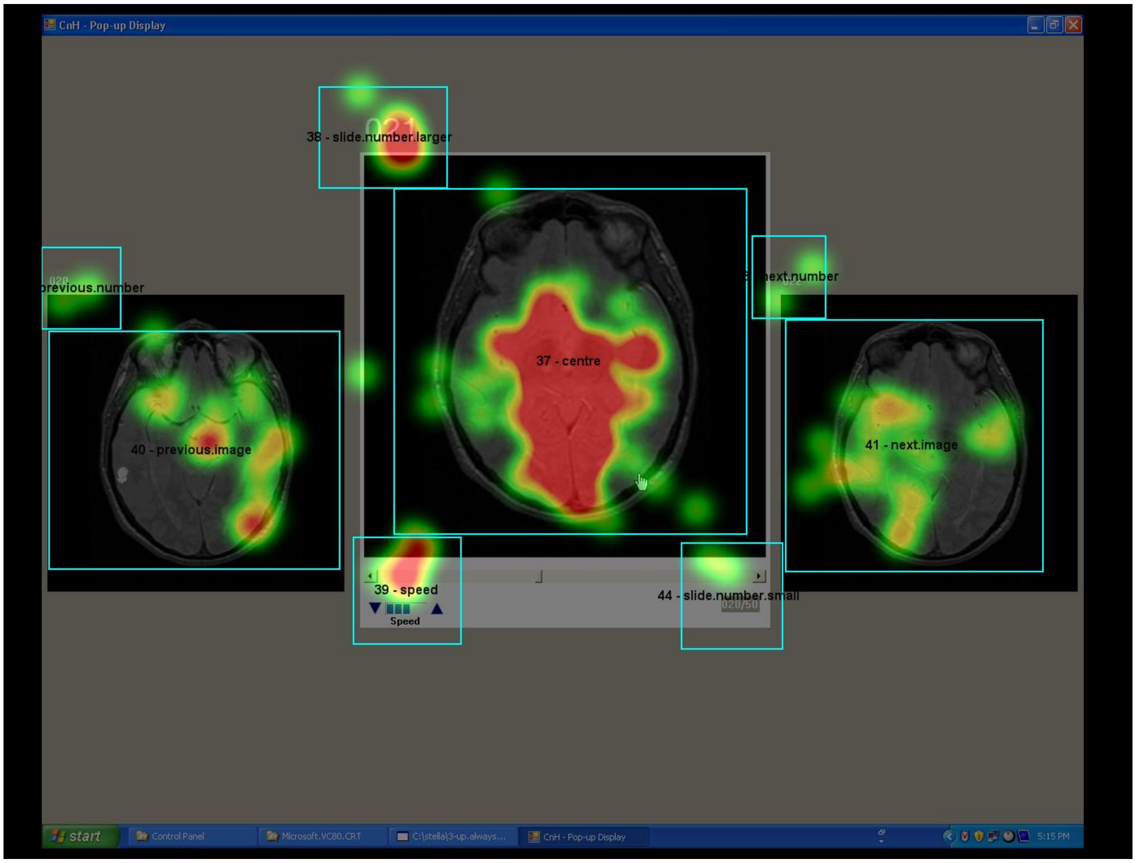

Figure 5. Hot spot view of fixations on the 3-up layout corresponding to the gaze plot in Figure 4.

3.3 Areas of Interest

The gaze time of the fixations on several areas of interest were analysed for each layout is summarized in Table 1.

Table 1. Gaze time and % of total trial times for both layouts for areas of interest.

Layout	Av. Gaze time on centre image		Av. Gaze time on prev. image		Av. Gaze time on next image		Av. Gaze time on slice number	
	secs	%	secs	%	secs	%	secs	%
1-up layout	27.6	82					1.7	5.1
3-up layout	29.9	74	1.3	3	1.8	4.5	1.8	4.5

Average gaze time over all the participants for all the trials on the centre image was 27.6 secs for the 1-up layout, representing 82% of the trial duration. For the 3-up layout, the central image received almost 30 secs, and the side images 3.1 secs, for a total gaze time on the images of 33 secs. Hence the images of the 3-up layout received 5.4 secs more attention than the 1-up layout, over all the trials. This accounts for all the difference in trial times between the two layouts.

4 Summary and future work

We presented 3-up, a new stack mode layout. Our initial hypothesis was that the 3-up stack viewer would allow users to locate anomalies more quickly, but we were uncertain if the accuracy of localization would differ between the two systems. As a secondary issue, we were interested in examining the search strategies employed by users of each viewer.

In the new 3-up layout, the peripheral (previous and next) images are viewed when available (by 7/8 subjects). Furthermore, 5/8 participants preferred the 3-up layout. The preference for the 3-up layout likely reflects the lesser cognitive effort involved in searching for the boundaries of the anomaly. When 3 adjacent slices are displayed, the eye can register where the anomaly is not present; the subject's short-term memory is not overladen with remembering a previously-viewed image.

We believe that the extra time was caused by the difficulty in controlling the rate of scrolling, causing overshooting. There is more back and forth scrolling used for the 3-up, and the eye hand coordination on the mouse, combined with system lag, has a total lag of > 500 msecs.

Future work will validate these results by measuring radiologists performing the same task. We also wish to measure an effective size for the centre image display, so that anomalies are easily spotted in the peripheral vision. We would like to estimate performance tradeoffs in using 3-up or more, such as 5-up layouts.

We would also like to use the task to evaluate different interaction devices, and design and evaluate alternative visualizations as aids to detection. For example, use difference images to highlight differences between consecutive thin slices.

Our hypothesis was contradicted, that our modified stack layout would allow users to locate anomalies more quickly than the regular stack layout. The eyegaze data allowed us to analyse the search strategies employed by users of each viewer and explain these results.

Acknowledgements

Many thanks to Modupe Omueti for designing the stimuli, Jen Fernquist for making all the code improvements, and Susan Villecroze for help with the initial pilot experiments. Thanks to the Canadian Natural Science and Engineering Council for funding this research.

References

1. Ellis, SM, Hu, X, Dempere-Marco, L., Yang, GZ, Wells, AU, Hansell, DM. 2006. Thin-section CT of the lungs: eye-tracking analysis of the visual approach to

reading tiled and stacked display formats. *European Journal of Radiology* 59: 257-264.
2. Kim, Y.J., Han, J.K., Kim, S.H., Jeong, J.Y., An, S.K., Han, Son, Lee, Lee, Choi, Small Bowel Obstruction in a phantom Model of ex Vivo Porcine Intestine: Comparison of PACS Stack and Tile Modes for CT Interpretation. *Radiology* 236: 867-871, 2005.
3. Krupinski, E.A., Roehrig, H., Berger, W., Dalal, S. & Stanton, D.2006a. Potential use of a large screen display for interpreting radiographic images. *Proc SPIE Medical Imaging* 6146, 614607-1:7, 2006
4. Krupinski, E.A., Mincilla, R. Sewell, P. Steiner, E., Widlus, D. 2006b. Value of Image Motion in Detecting Stenoses. *Proceedings SCAR*: pp. 10-11, 2006.
5. Mathie, A.G. and Strickland, N.H. 1997. "Interpretation of CT Scans with PACS Image Display in Stack Mode," *Radiology* 203: 207-209, 1997.
6. Moise, A., Atkins, M.S. and Rohling, R. 2005. Evaluating different radiology workstation interaction techniques with radiologists and laypersons. *Journal of Digital Imaging* 18(2), 116-130. June 2005.
7. Moise, A. and Atkins, M.S. 2004. Interaction techniques for radiology workstations: impact on users' productivity. *Proc SPIE Medical Imaging* 5371: 16-22.
8. Reiner B.I., Siegel, E. L., Siddiqui, K. 2003. Evolution of the Digital Revolution: a Radiologist Perspective. *Journal of Digital Imaging*, 16(4):324-330, Dec. 2003.
9. van der Heyden, J., Inkpen, K., Atkins, M.S. and Carpendale, M.S.T. 2001. Presentation of Tomographic Medical Images. *Artificial Intelligence in Medicine*, 22: 89-109, Dec. 2001