Evaluating Interaction Techniques for Stack Mode Viewing

M. Stella Atkins,[1] Jennifer Fernquist,[1] Arthur E. Kirkpatrick,[1] and Bruce B. Forster[2]

Three interaction techniques were evaluated for scrolling stack mode displays of volumetric data. Two used a scroll-wheel mouse: one used only the wheel, while another used a "click and drag" technique for fast scrolling, leaving the wheel for fine adjustments. The third technique used a Shuttle Xpress jog wheel. In a within-subjects design, nine radiologists searched stacked images for simulated hyper-intense regions on brain, knee, and thigh MR studies. Dependent measures were speed, accuracy, navigation path, and user preference. The radiologists considered the task realistic. They had high inter-subject variability in completion times, far larger than the differences between techniques. Most radiologists (eight out of nine) preferred familiar mouse-based techniques. Most participants scanned the data in two passes, first locating anomalies, then scanning for omissions. Participants spent a mean 10.4 s/trial exploring anomalies, with only mild variation between participants. Their rates of forward navigation searching for anomalies varied much more. Interaction technique significantly affected forward navigation rate (scroll wheel 5.4 slices/s, click and drag 9.4, and jog wheel 6.9). It is not clear what constrained the slowest navigators. The fastest navigator used a unique strategy of moving quickly just beyond an anomaly, then backing up. Eight naïve students performed a similar protocol. Their times and variability were similar to the radiologists, but more (three out of eight) students preferred the jog wheel. It may be worthwhile to introduce techniques such as the jog wheel to radiologists during training, and several techniques might be provided on workstations, allowing individuals to choose their preferred method.

KEY WORDS: Stack mode image navigation, user interaction devices, mouse scrolling interaction, jog-shuttle wheel

INTRODUCTION

Radiologists often navigate through long sequences of 2D image slices generated from MR or CT volume data. For example, up to 1,000 images are generated by abdominal CT exams, and each image must be viewed[1]. With current computer systems, it is possible to store and display such large image sets quickly, in many different formats. This evolution in technology must be accompanied by a necessary evolution in radiology workstation interaction, as the radiologists must explore this 3D data using 2D displays.

In radiology, stack mode viewing of cross sectional radiological images has become popular. This popularity may be because the technique exploits the visual system's sensitivity to motion. By presenting image sequences over time, stack mode makes abnormalities more apparent than do techniques that present the sequence over space. Hence if the viewer maintains their gaze on a specific spatial region as the images change, the depth relationship of various structures is better seen, and, for example, a tumor can be differentiated from a normal structure. 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 MR angiography images[2]. A few studies have shown that stack mode viewing is much faster then tile-mode and just as accurate[3–5].

In order for image-viewing software and user interaction hardware to be valuable, it must display

the images in a manner useful for radiological tasks[6]. Although many of the storage and display hurdles of digital radiology have been met, many human–computer interaction problems remain[7–9].

Some recent work has addressed these problems by making the presentation of the data more nearly 3D. Teistler et al.[10] developed a 3D visualization paradigm using a 3D mouse that can be lifted to provide rotational degrees of freedom. The 3D mouse is built by using a standard mouse and an electromagnetic motion-tracking sensor. Prototypes of this system have been judged to be useful by radiologists. However, once the appropriate viewing orientation has been chosen from the 3D image, navigation through the set of 2D slices still requires the mouse to be used in regular 2D mode.

Wang et al.[11] compared three display techniques: a stereoscopic display, slice-by-slice, and maximum intensity projections (MIP), for detecting lung nodules from CT lung volumes. They found that the stereoscopic display provided higher detection and classification performance with less interpretation time but the differences were not statistically significant in their study of eight radiologists viewing a total of 91 anomalies. They also observed differences in navigation patterns between the slice-by-slice and the stereoscopic displays. Their focus was the display method; they used the same customized programmable keypad for navigation in every condition.

So, although there are some efforts to develop 3D viewers for medical data, the vast majority of volumetric medical data is viewed as stacks of 2D slices. However, despite the many stack-based viewers in use today, there has been very little research on the design of interaction techniques for stack mode viewers.

One study reports the use of alternative interaction devices for navigating through large CT data sets[12]. Sherbondy et al. compare four devices: a trackball, a tablet with two different software interface designs, a jog-shuttle wheel made by Contour Design, and a mouse. Jog-shuttle wheels have two rings (see Figure 1): the inner ring or 'jog' rotates through 360 degrees and provides precision frame by frame control. The outer ring or 'shuttle' is rubberized and spring-loaded. It facilitates fast forward and rewind. Each subject (four radiologists) looked for artificial targets in five different large CT data sets; each data set was viewed using a different interaction technique. Results showed that the trackball was signif-

Fig 1. The Contour Design Shuttle Xpress jog and shuttle wheel used in our experiments.

icantly slower and least preferred than the other methods, but there was no significant difference among the other methods. The trackball used in the study required subjects to hold down a button while rotating the ball. Its poor performance was hypothesised to be because users have to make large repetitive motions to traverse large numbers of slices; the authors speculate that it is possible that a different interaction technique which did not require the button might improve the ratings of the trackball. Sherbondy et al. used a ShuttlePro jog-shuttle wheel by Contour Design in a hybrid mode, where they mapped the displacement of the outer shuttle wheel (in terms of its rotation from 0°) to control the rate of fast scrolling, and the jog wheel for fine adjustments. They learnt that this was effective, particularly when the outer shuttle wheel with a fast scroll was used with the inner jog wheel performing a finer scroll. Their results also showed that there was little overshoot with the jog-shuttle wheel. We were interested to see how using the inner jog wheel alone would be useful for scrolling in radiology tasks.

Another study by Weiss et al.[13] required six radiologists to evaluate six alternative user interface devices (UIDs), including five-button and eight-button mice, a gyroscopic mouse, a multimedia (a jog and shuttle) controller, a handheld mouse-and-keyboard combination device, and a gaming joy-

stick. Each participant assessed each device during the real-time daily imaging interpretation of magnetic resonance, computed tomography, and general X-ray studies over a two-week period and completed a detailed questionnaire on the ease of use, comparative utility as an alternative device to mouse and QWERTY keyboard, efficiency, workflow, and the ease of customized programming. In this qualitative study, no clear interaction device emerged as the leader; instead, some specific functionalities of the devices were praised, suggesting combining these devices for two-handed operation.

We observed radiologists at work and noted that they typically use a regular scroll-wheel mouse for navigating through stacked images, in conjunction with a special mode of operation called "click and drag" whereby multiple images can be quickly scrolled through, by holding down the right (or left) mouse button, and moving the mouse upwards (away from the user) or downwards (towards the user), to indicate scrolling direction. However, both the scroll wheel and click and drag have disadvantages. The wheel is slow for scrolling through huge numbers of slices, and can cause strain on the scrolling finger. The click and drag method is hard to use for fine tuning to an exact image slice.

We hypothesised that we could combine the best of mouse scroll and mouse drag techniques with a device often used in video editing, the jog-shuttle wheel (see Figure 1 showing the Shuttle Xpress jog wheel we used, by Contour Design). Video editing has characteristics similar to stack mode viewing. The editor scrolls through a long sequence of still images, sometimes quickly scanning, and other times searching for a single specific image. The jog wheel is well-adapted for both operations: by rotating the finger in the inner jog wheel, users can scroll quickly (much faster than they can with a mouse scroll wheel), yet they can also slow down and scroll one frame at a time. Furthermore, the lateral circular motion used by the jog wheel can be sustained indefinitely, whereas there are physical limits to how far the mouse scroll wheel and click and drag can be taken in a single gesture. We wanted to see if these characteristics would benefit radiologists viewing images in stack mode.

To test this hypothesis, we compared the performance of participants (nine radiologists, and also eight naïve students) using three interaction techniques for the common stack mode display: a jog wheel and the two mouse-based methods which the prior work cited above had shown to be most effective: mouse scroll wheel and drag mouse/scroll wheel.

As a secondary issue, we were interested in comparing the interactions and search strategies used by radiologists with those used by naïve students. In an earlier work[14], we showed that radiologists and laypersons had similar interaction performance when expert knowledge was not required, so we hypothesized that we could use naïve laypersons to evaluate different image navigation techniques, with results that would generalize to radiologists performing clinical tasks. For example, we had observed a lot of forward and backward scrolling during a radiology look-alike task with naïve students as subjects[9] and backward and forward scrolling in stack mode by radiologists has been similarly noted by others[11–13].

As a final, tertiary, issue, we were interested whether a group of naïve students would have similar performance and preferences to radiologists for the three techniques.

MATERIALS AND METHODS

We implemented a stripped-down stack mode viewer, and developed a look-alike radiology task of searching for artificial target stimuli anomalies implanted in consecutive slices of medical 3D image volumes. We chose artificial stimuli rather than actual clinical cases for two reasons. First, such stimuli could be smaller than actual clinical cases, reducing load time and diagnostic time, allowing our radiologists to complete the session in under an hour. Second, the artificial data sets allowed us to use naïve students as a comparison group.

The viewer supported three interaction techniques:

1. In the "wheel" technique, the user navigates through the image stack by using the mouse scroll wheel with speed proportional to scrolling. Scrolling downwards displays the next images; scrolling upwards displays the previous images. The mouse used in the study had a single finger scroll range of seven clicks, the number usually found on PACS workstations.

2. In the "drag/wheel" technique, the user clicks and holds the right button down while moving the mouse downwards, to display the next images at a speed proportional to the rate of movement. Conversely, if the user moves the mouse up-

wards, the previous images are displayed. Fine adjustments to displaying the previous or next image slices can be made using the scroll wheel as described above.

3. In the "jog" technique, the user places a finger on the space on the jog wheel, and then rotates the finger left or right to navigate to previous or next images, with speed proportional to the rate of turning.

Using this software, we performed a controlled user study to determine the speed and accuracy of the three interaction techniques.

The software also logged mouse button presses up and down, mouse movement, and mouse wheel scrolling. Rotation of the jog wheel was mapped to appear to the software as mouse scroll-wheel events, albeit arriving at a rate and consistency that would be impossible from an actual mouse wheel.

The experiments were performed in a darkened room simulating a clinical PACS workstation environment, using an IBM Thinkpad laptop 15″ LCD display with resolution of 1,400×1,280 pixels. The laptop was running Windows XP operating system.

Stimuli

The stimuli consisted of MRI data slices modified with artificial hyper-intense spherical regions representing lesions. They 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. There were on average 60 image slices in each volume, with a maximum of 101 slices in one of the test sets, and were a combination of transverse, sagittal, and coronal data sets. 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 Figure 2.

Experiment Design

Design The design was a within-subjects comparison of three interaction techniques.

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 and jog-

shuttle wheel with their preferred hand and spoke the slice numbers of 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 Total time for each trial was recorded by the software. The range(s) the participant reported for anomalies were compared with the actual ranges to yield a measure of accuracy. The software log files were analyzed to derive the path taken by the participant through the stack. Each participant also answered a post-experiment questionnaire in which they stated their preferred technique and gave free-form comments on the techniques and the study as a whole.

Participants There were nine radiologists; three experts with more than 5 years of experience using PACS workstations, and six younger trainee radiology fellows, all volunteers. They were not rewarded for participation.

Protocol Each subject performed the same anomaly detection task using all three interaction techniques. Each subject performed three blocks of five trials. Each block contained a total of eight anomalies for an average of 1.6 anomalies per trial. Most volumes had about 60 slices; two volumes had 95 slices. Order of interaction techniques 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 all three blocks, had six anomalies. Workplace lighting was consistent for all sessions and participants were offered the chance to adjust the chair height and location. An armrest was provided on the chair, although not all participants took advantage of it. The experimental protocol was approved by the Simon Fraser Research Ethics Board.

RESULTS AND DISCUSSION

Means were taken for each participant for each condition. Normal quantile plots showed that the resulting values were approximately normally distributed and had no outliers, so all data points were used without transformation.
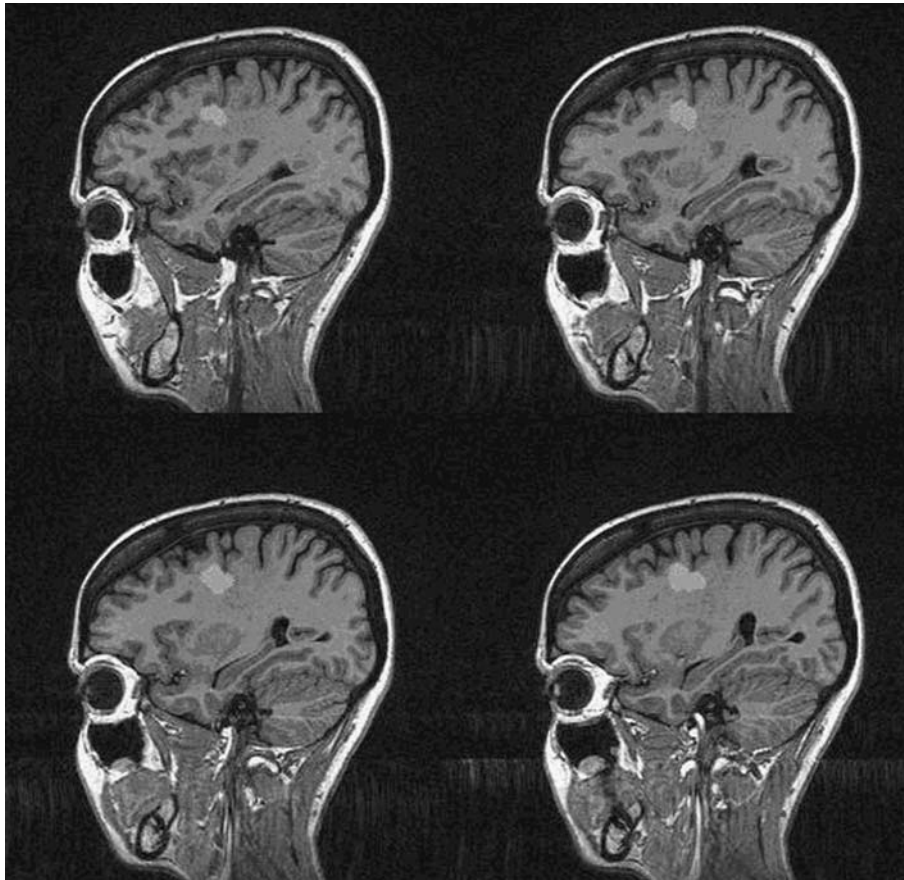
Fig 2. Consecutive image slices containing an artificial anomaly.

Separating Trials into Locate Pass and Review Pass

All but one of our radiologists examined the images in two distinct phases. First, they did a careful scan from the front to the back of the images, calling out any anomalies. We call this scan the *locate pass*. While viewing an anomaly in the locate pass, they would often move back and forth between close slices, confirming its range. The locate pass was mostly a straight run through the other slices.

Upon reaching the last slice, they would do a second scan, from back to front, much more quickly than the first. We call this the *review pass*. The review pass was most often a straight run from back to front, although in a few trials they backed up momentarily, and in a few cases they did not complete the scan all the way to the front slice. The lone exception to this pattern was one radiologist who performed a locate pass but no review pass. The review pass was strictly confirmatory for the eight who performed it; all radiologists called out every anomaly during the locate pass.

Figure 3 illustrates the two passes. The figure plots the displayed slice over time for one trial. We call such a plot a *navigation chart*. Gray bars indicate the slices that contain an anomaly (slices 31–35). In this trial, the participant spent the first 6.0 s scanning for the first anomaly, then about 3.8 s exploring the anomaly, then 1.6 s scanning to the final slice, ending the locate pass. The review pass is much faster, covering the same range of slices in 34% of the time, 3.8 s.

Every trial was divided into a locate pass and a review pass by visually inspecting the navigation charts and log files. The locate pass was defined to start at the beginning of the trial and end when the last slice was reached. The review pass was defined to begin when the last slice was left and ran to the end of the trial.
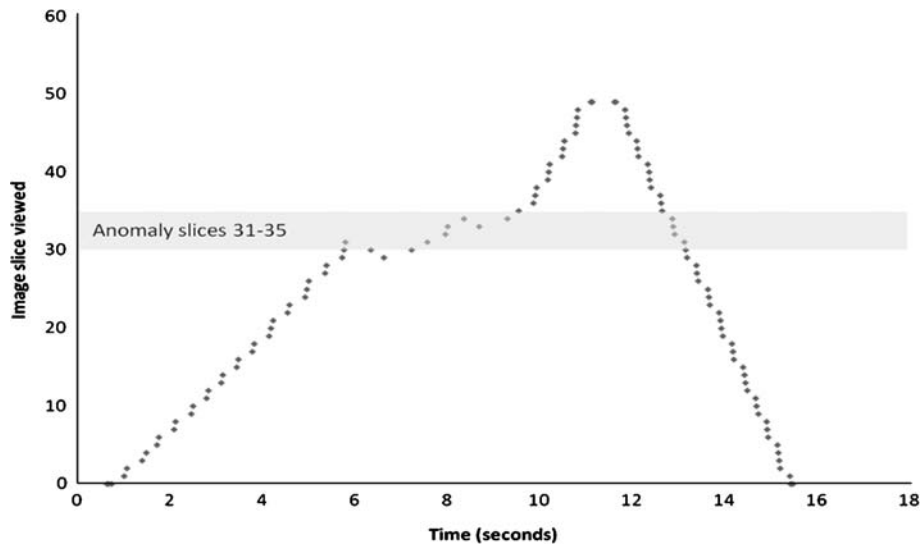
Fig 3. A typical navigation chart (P7, Trial 13, wheel technique). The locate pass takes the first 11.1 s, followed by a 0.5-s pause, and then the review pass takes 3.8 s.

Accuracy and Time

The nine radiologists were 100% accurate with all three techniques. Technique had little direct influence on trial time (over all conditions, $M = 33.2$ s, SD $= 11.5$ s; condition means: drag/wheel 30.5 s, wheel 33.0 s, jog 36.1 s). This was confirmed by analysis of variance, which found nonsignificant effects for technique (within-subjects, $F(2,6) = 0.50$, $p = 0.63$), order of technique (between-subjects, $F(5,3) = 1.59$, $p = 0.37$), and the order×technique interaction (mixed, $F(10,6) = 2.50$, $p = 0.14$). The low F-value and nonsignificance of order indicates that there were little to no asymmetric learning effects across the techniques. Such effects would arise, for example, if performing the task with the wheel first mis-trained participants to perform poorly with drag/wheel in the next block. Such an outcome would confound the results and make it difficult to estimate the actual speed of the techniques.

By itself, the nonsignificance of the technique effect is ambiguous. On the one hand, it could arise because technique made little actual difference in performance. On the other hand, a genuinely large effect could have a nonsignificant test because there was large variation across individual trials, masking the real effect of technique. Proportion of variance measures distinguish the two cases. Technique only accounted for 3% of the variance, while the order× technique interaction accounted for 78% and the variation across individual trials provided the re-

maining 19%. Thus, there was genuinely little difference between techniques.

The overwhelming variance due to order×technique shows that the most important effect of all was the inconsistency of techniques across participants. No technique was consistently the fastest for the participants. In fact, of the six possible rankings of speed of technique, participants were evenly distributed over five—the only ranking that never occurred was jog (fastest), wheel, drag/wheel (slowest).

Times for the locate pass showed similar variability (over all conditions, $M = 23.0$ s, SD $= 7.2$ s), and again technique only made a modest contribution (condition means [SD]: drag/wheel 20.9 s [5.2], wheel 23.8 s [8.8], jog 24.3 s [7.5]).

Nature of Interaction

To discover the source of the very high individual differences in trial times among the participants, we studied details of the navigation interactions during the trials. There were two main activities involved in the locate pass: navigating images till an anomaly was found, and identifying and naming the slices with anomalies (called the *thinking time*). Thinking time was defined as the time when the slice containing the anomaly was first seen, to the last time the anomaly was seen.

The rate of forward navigation was computed as the rate of navigation for the portions of the locate pass where the user was scrolling forwards, ex-

cluding the thinking time. The rate of backward navigation was computed using the entire review pass, as there were no periods of think time in that pass.

Examination Strategies

Most navigation charts reveal an interaction pattern similar to the one displayed in Figure 3, where image slices were examined in sequential order with a few backwards and forwards navigations on images with an anomaly. Figures 3, 4, and 5 show three different subjects performing with the same stimulus, with different interaction techniques. Figure 3 shows a fast subject (P7) using the wheel. P7 had a forward rate of 6 slices/s up to the anomaly at slice 31. Thinking time is 4.0 s. Note that the forward scroll steps were bursty—steps of two or three images—because the participant made small scrolling movements before moving their finger. This bursty navigation is typical of the wheel. In the review pass, the steps were made in groups of three or four images and the wheel was moved faster, achieving 13.0 slices/s.

Figure 4 shows a trial for the same stimulus by another radiologist using the drag/wheel technique. This participant (P5, slower than average) had a forward rate of 4.3 slices/s up to the anomaly at slice 31, followed by a thinking time of 9.0 s, and a backward rate of 8.6 slices/s in the review pass. Note that the drag/wheel proceeds burstily for this participant (also for P3); these participants moved the mouse a short distance, then lifted it up and repositioned it back to center. However, most other participants maintained a smooth navigation using

drag/wheel, moving the mouse in a few long sweeps. For all participants, the backwards rate of navigation in the review pass was much smoother and faster than their forward rate using the drag/wheel technique.

Figure 5 shows a trial for the same stimulus by another radiologist using the jog wheel. This participant (P9, average speed) had a forward rate of 8.0 slices/s up to the anomaly at slice 31, and a backward rate of 11.0 slices/s. Thinking time is 5.2 s. Note the smooth scrolling, both forwards and backwards.

Thinking and Navigation Times

Thinking time was defined as the time from when the first slice containing the anomaly was seen, to the last time the anomaly was seen during the locate pass. The thinking times were derived manually from the raw data and the navigation charts of each participant, so we used just a sample of six typical trials (trials 3, 6, 8, 13, 14, and 15) containing a total of ten anomalies. We assume that the interaction technique had a negligible effect on thinking time, as the rate of slice traversal was very low when anomalies were viewed, so we were not concerned about the thinking time per interaction technique. The mean thinking time per trial was 10.4 s, with a range of 7.8–13.7 s.

The mean navigation time per trial is obtained by subtraction of the mean thinking time from the mean trial locate pass time. The mean navigation time per trial is 12.6 s, with a range of 2.7–22.3 s.

Figure 6 shows a graph of the mean thinking time and mean forward navigation time per trial, for each participant. It is seen that there is much more indi-
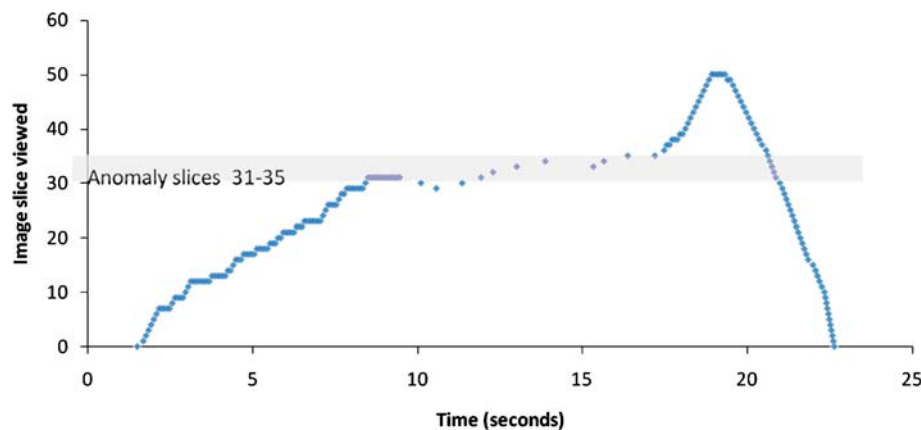

Fig 4. The same stimulus as in Figure 3, with the drag/wheel technique (P5, Trial 13). Forward speed is 4.3 slices/s, backward is 8.6 slices/s. Thinking time to confirm the anomaly is 9.0 s.
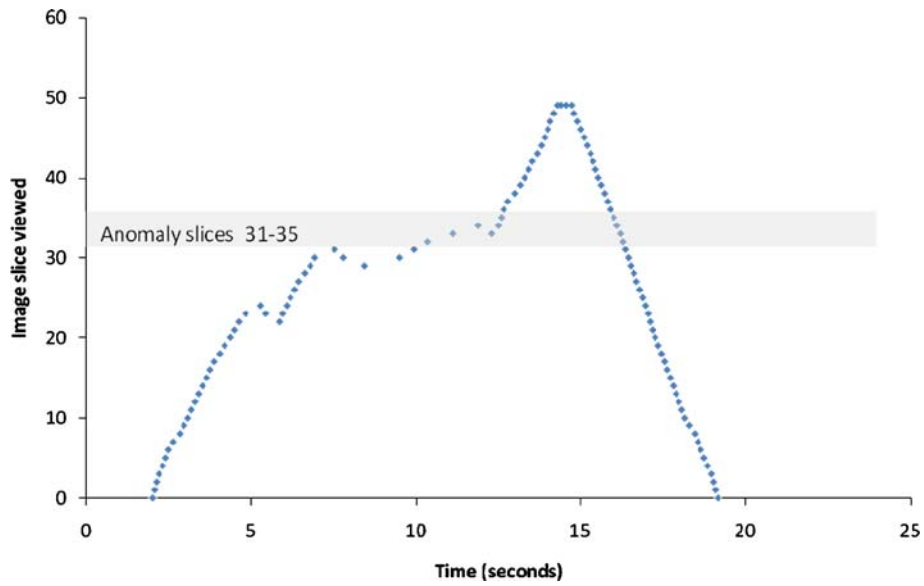
Fig 5. The same stimulus as in Figures 3 and 4, with the jog technique (P9, Trial 13). Forward speed is 8 slices/s, backward speed is 11 slices/s. Thinking time to confirm the anomaly is 5.2 s.
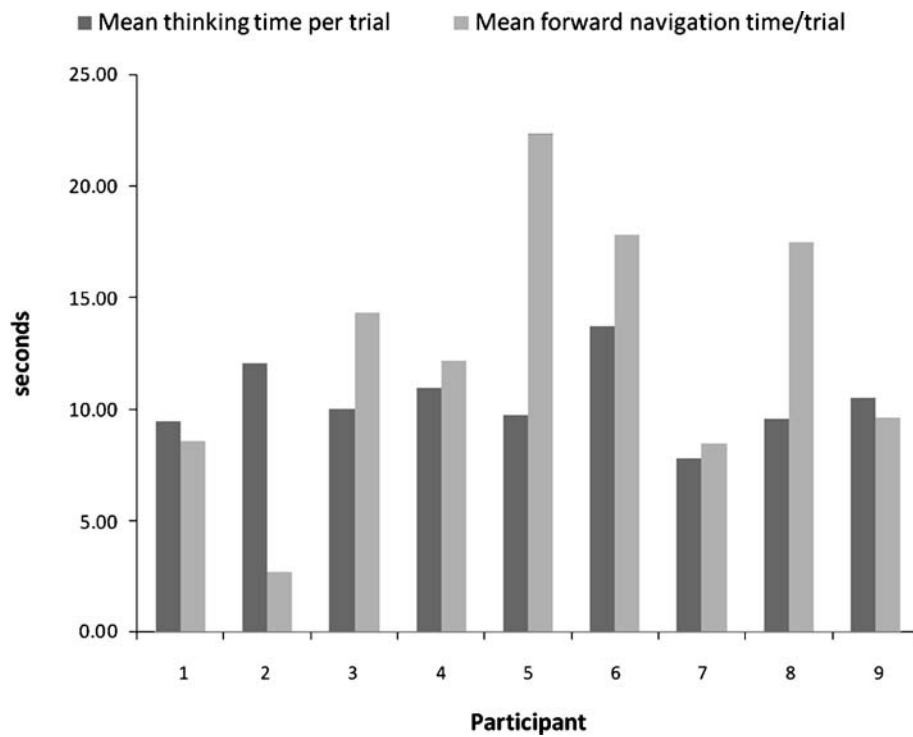

Fig 6. Mean thinking and forward navigation times per trial, for each participant.

vidual variation in the navigation times than in the thinking times; furthermore, most participants spent more time navigating than thinking. We conclude that the very high inter-subject variation in trial times arises mainly because of the large differences in the forward navigation times.

Rates of Forward Navigation

Figure 7 shows the mean rate of forward navigation in slices/second over all the trials for each interaction technique, for each participant. The mean rate of forward navigation is 8.2 slices/s, with a range of 3.2–17.7 slices/s. In Figure 7, it is seen that most radiologists (except P2, the fastest radiologist) navigate fastest with drag/wheel and slowest with the wheel. P2 is a clear outlier; we analyze his performance individually in Section 3.5.2. For the main body of eight radiologists, technique had a significant effect ($F(2,14)=39.9$, $p<0.001$) explaining 85% of the total variance. The techniques had significantly different means (drag/wheel 9.4 slices/s, wheel 5.4 slices/s, jog 6.9 slices/s). Technique had no signi-

ficant effect on backward navigation time ($F(2,16)=$ 1.1, $p=0.368$, $\eta^2=12\%$).

It is interesting that the techniques were not significantly different in trial time, even though their rates of forward navigation were different. It is not simply a matter of a high error variance, as even the ordinal rank of the techniques varied widely across participants. As noted, drag/wheel was the fastest technique for forward navigation for all techniques other than the (outlying) P2. These conflicting results show that participants moved through more slices using drag/wheel than they moved through using the other techniques, nullifying the more rapid navigation of the drag/wheel technique. The navigation charts show that they "backed up" more often during the locate pass when navigating fast.

Three Slow Navigators

Why do P5, P8, and P6 have the slowest forward navigation rates, and why is P2 so much faster than the others? We use the navigation charts to answer these questions.
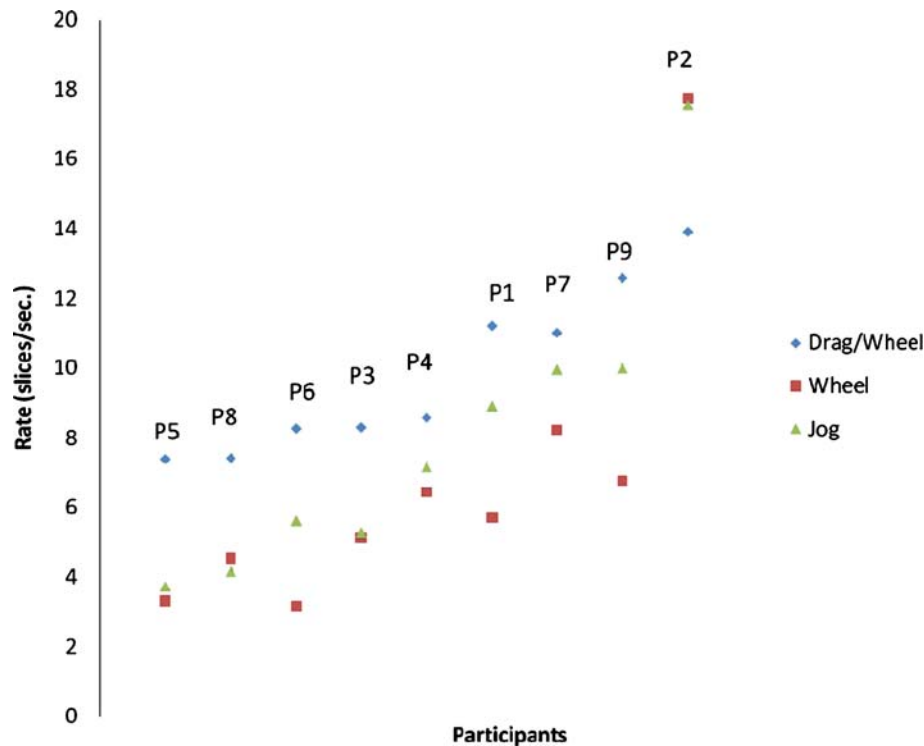

Fig 7. Mean rate of forward navigation in slices/second, for each participant. Sorted by mean rate.

Figure 7 shows that P5, P8, and P6 are very slow in their rate of forward navigation using jog and wheel, and fastest using drag/wheel. These are not just task-learning effects; although for both P5 and P8, their fastest technique (drag/wheel) was the last one they performed, P6's fastest technique (drag/wheel) was the first technique. Were these users somehow constrained by the speed of scrolling in the wheel and jog techniques?

Figure 8 shows a navigation chart for P6 using wheel in a typical trial (Trial 5). Note that the forwards scroll is very slow and smooth, whereas the backwards scroll is much faster, scrolling in bursts of four or five images before the user moves the finger to make the next scroll. However, the burstiness does not appear to limit the user in the speed of scrolling, as this user can scroll backwards at 6.7 slices/s. P8 is similar. The slow navigators are not constrained by the biomechanics of the interaction technique.

The slow navigators were not limited by their ability to spot the anomalies, either. All three participants achieved substantially faster rates using the drag/wheel technique, with the same perfect accuracy. It is not clear, then, what made these three radiologists so much slower on the wheel and jog techniques.

An Unusually Fast Navigator

Figure 7 shows that P2 navigates forwards at a much faster rate than the others, although P2's thinking time is the second-highest (Fig. 6). P2 did drag/wheel first, then wheel, then jog. As P2 achieved the same 100% accuracy as the other radiologists, it is interesting to see how this fast rate of navigation was achieved. The navigation charts for P2 indicate that he used a much fuller range of scroll-wheel movement before moving the finger, as the bursts occur in six or seven image slices, seen in Figure 9.

P2 quickly learnt that for this task a review (backwards) scan was often not necessary, so several of P2s trial charts such as this one for the tenth trial, show no review navigation backwards. Instead, P2 was able to note the presence of an anomaly as he navigated swiftly through the image slices, and note the slice where the anomaly ended (slice 33 in this trial). P2 then navigated slowly backwards to confirm the slice where the anomaly first appeared (slice 28) and then navigated slowly forward through the slices while the anomaly was present, before speeding up again to detect the next anomaly in slices 71–75.

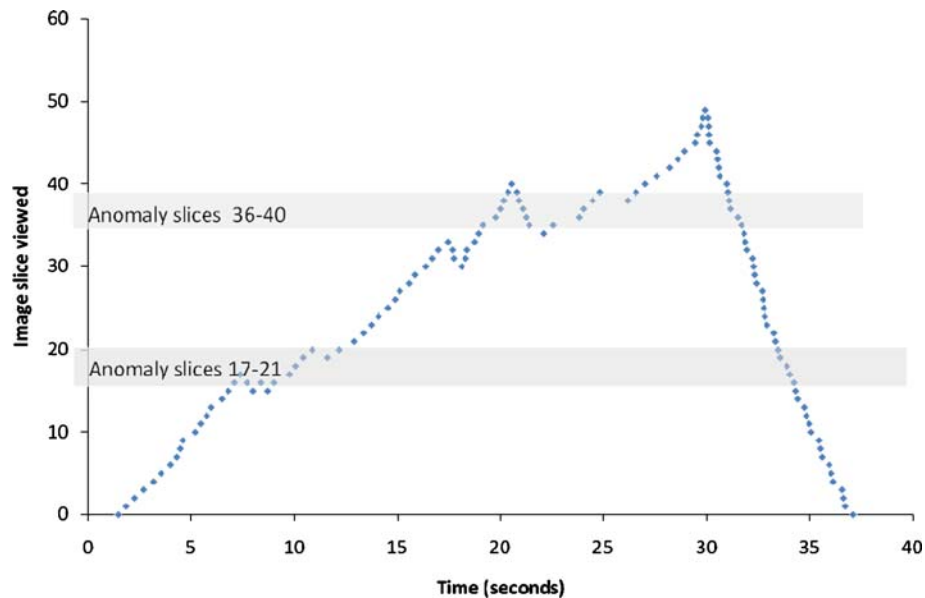Figure 10 shows P2 using the jog, again very fast and smoothly. Again, he performed the same

Fig 8. Slow navigator P6 on Trial 5 using the wheel technique. Forward rate = 3.2 slices/s, backward rate = 6.7 slices/s.
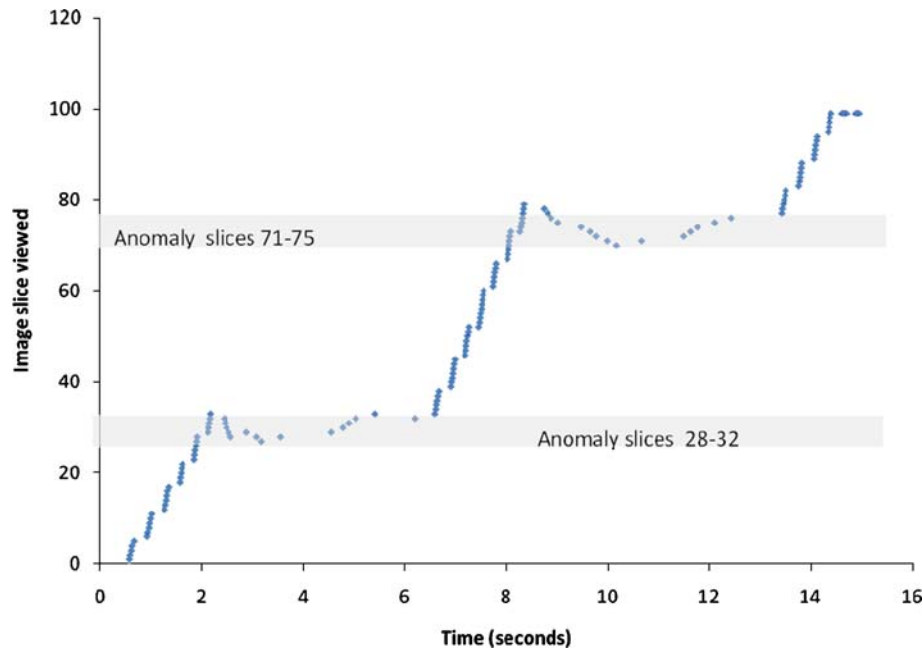
Fig 9. Fast navigator P2 using wheel, Trial 10. Rate of forward navigation = 17.7 slices/s. Note the absence of a review pass.

technique of navigating to the end of the anomaly before backing up finding the start (anomaly is in slices 26–30 inclusive). Note that in this later (11th) trial, P2 did a review pass backwards through all the images.

We conclude that the source of P2's much faster rate was his use of a different strategy from the other eight radiologists. P2 would navigate very fast through the slices, recognize when an anomaly appeared, and then back up only when the anomaly
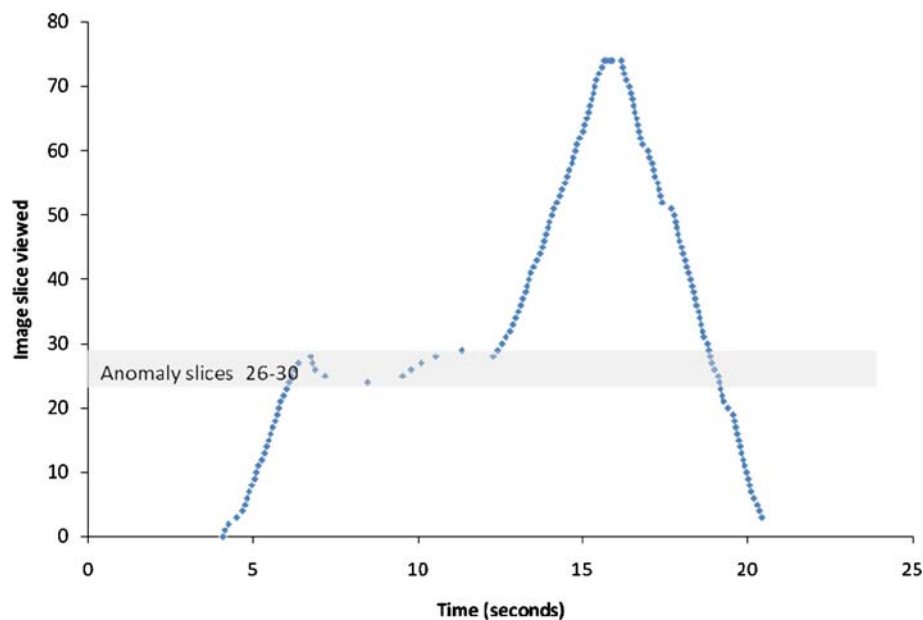

Fig 10. Fast navigator P2 using jog, Trial 11. Rate of forward navigation = 13.0 slices/s, backward navigation = 16.5 slices/s.

was no longer seen. This is an interesting strategy, which may possibly be good for other situations. This strategy may have been particularly suited to the stimuli used in our study. The optimal speed for combining timely image review and accuracy has to our knowledge not been researched, but no doubt depends in part on the organ system, the contrast of the abnormality against background ('contrast resolution') and also the observer and viewing environment. In the case of the simulated lesions presented in this experiment, contrast resolution was high, somewhat like lung nodules against normal lung background, but most unlike many liver lesions against the background of normal liver. We surmise that this high contrast allowed P2 to adopt a strategy of deliberately overshooting the anomalies.

Qualitative Results (from questionnaire and comments)

In the post-session questionnaire, all nine radiologists thought that the task was realistic.

User Preferences

All nine users were accustomed to using drag/wheel. A total of six users preferred the wheel technique, two preferred the drag/wheel technique, and one user preferred the jog wheel.

Comments on Jog Technique

Two users said they disliked the jog wheel because they were not familiar with it. Both preferred the pure wheel technique in this study, but use the drag/wheel technique on PACS. One user thought the rotational movement of the finger on the jog might become tiring. This user preferred the drag/wheel technique in the study, and uses that technique on PACS.

Comments on the Wheel Technique

One user chose the wheel technique because it was the best for the small data sets (<100 slices). Two users thought the scroll wheel required too much finger movement, and one user said it caused strain on the finger. Both of these users use drag/wheel on PACS. Two more users thought the scroll wheel was too slow. We observed that some users had difficulty controlling the rate of scrolling, causing overshooting.

Comments on the Drag/Wheel Technique

We observed that when using this technique, all the radiologists used their index finger for both dragging the mouse while holding the button, and for using the scroll wheel for fine adjustments. Consequently, when shifting from gross movement to fine adjustment, they had to move their finger from the mouse button to the scroll wheel. The effect of this was noted in their preferences, one person stating they disliked moving their fingers back and forth from the scroll wheel to the button, even though they preferred the scroll wheel for this task. However, it did not appear to slow them down.

Results for Student Participants

We piloted the study with eight student volunteers from neighboring computer science labs who performed a similar experimental task, using four interaction techniques. The jog and the wheel techniques were identical to those used by the radiologists, but the students also tested a drag/wheel technique that went at a fixed speed, and a rate-based mouse technique, where the user had to click the right or left button and hold it, to navigate through the images at a fixed speed. The stimuli and total number of slices per trial were different for the students, who had a large lung CT image which took time to load, which was not used for the radiologists. The students performed four trials in each condition (vs. the radiologists' five trials), with an average of 1.5 anomalies per trial (vs. the radiologists' 1.6 anomalies per trial). The software was also slightly different, as the images were not pre-loaded, causing delays for some trials. The viewing conditions differed from those of the radiologists, as the students had a 17″ CRT display in a darkened room. Hence, the trial times for students and radiologists are only approximately comparable.

Some students were inaccurate: three out of eight students missed an anomaly, leading to one unrealistically fast trial by one particular student. Removing that student's trial time from the data, the student mean times were 20.8 s for the pure scroll and 21.3 s for the jog. These times compare well with the radiologists' locate pass times of 23.8 s for the pure scroll, and 24.3 s for the jog. These results corroborate our hypothesis that students can stand in for performance evaluations of radiologists on this task.

The jog and wheel were the most preferred interactions with three preferences each; one student

preferred the drag/wheel, and one preferred the "click and hold" interaction. It is interesting that three out of eight students preferred the jog, suggesting that training may influence preferences.

CONCLUSION

We compared three interaction methods for stack mode image navigation during a radiology diagnosis task. One method introduced a new device, the jog wheel. The task had external validity, as the radiologists considered the task realistic, albeit substantially easier than clinical practice. We should however qualify that our image sets are substantially smaller than the larger CT volumes often found in current clinical practice. We might well see different scrolling behavior (and in turn different relative performance of the techniques) for the longer scrolling required in bigger data sets.

Comparison of the techniques found no significant differences in total trial time or the time to perform the first pass through the images. However, the rates at which the radiologists scrolled forward through the slices in the first (locate) pass varied significantly with technique, with the drag/wheel technique 2.5 slices/s faster than the jog technique, and the jog technique 1.5 slices/s faster than the wheel technique.

Only one radiologist preferred the jog over the mouse-based techniques. Radiologists were usually fastest on the techniques they used in clinical practice. The preference of the radiologists for the mouse-based techniques may result from their extensive experience with the mouse.

Students (who had no familiarity with stack mode image navigation) were fastest on the novel jog wheel, although, like the radiologists, the large individual differences obscured the smaller differences between the techniques. Several students preferred the new device over the mouse-based techniques.

All the radiologists were 100% accurate, whereas three students missed a target (a false negative error).

All the radiologists (except the fastest) used their index finger for mouse wheel scrolling *and* for the click and move; so for fine adjustments in the click and move technique, they had to move their finger from the button to the scroll wheel. The questionnaire revealed that radiologists did not like moving their finger this way. Furthermore, several stated that they thought it would be stressful to use the scroll wheel for scrolling more image slices than the 60 used in our experiments.

Mechanics of the techniques were observed through our so-called navigation charts, graphing the number of the image slice currently displayed against time. Comparison of the navigation paths taken by the two user populations showed that most radiologists scanned the data set twice whereas students only scanned once. Students performed their single scan of the data sets in time comparable to the first scan by the radiologists.

We observed that one participant adopted a unique strategy, navigating quickly through the slices until just past an anomaly, then backing up to the anomaly start. This is an interesting strategy, which may possibly be good for other situations. The strategy may be particularly suited to anomalies with high contrast.

Both the radiologists and the students had a very high inter-subject variability in the time to complete the task. The similar trial times of the radiologists' first "locate" pass and the single pass made by the students, corroborated our hypothesis that students can stand in for performance evaluations of radiologists on this task. However, student times only predict the time radiologists will take for their first pass and will not account for the second pass.

These results, showing that the different techniques are broadly comparable in performance, suggest that several techniques should be provided on workstations, allowing individuals to choose their preferred method. It may be worthwhile to introduce new techniques such as the jog wheel to radiologists during training.

The biomechanics and system performance of the different techniques are in fact subtly different. Whether changing specific parameters would substantially improve performance is an open question.

Future work will test if using larger data sets will change the results.

REFERENCES

1. Reiner BI, Siegel EL, Siddiqui K: Evolution of the digital revolution: a radiologist perspective. J Digit Imaging 16:324–330, 2003

2. Krupinski EA, Mincilla R, Sewell P, Steiner E, Widlus D: Value of image motion in detecting stenoses. Proceedings SCAR 10–11, 2006

3. Ellis SM, Hu X, Dempere-Marco L, Yang GZ, Wells AU, Hansell DM: Thin-section CT of the lungs: eye-tracking analysis of the visual approach to reading tiled and stacked display formats. Eur J Radiol 59:257–264, 2006

4. Kim YJ, Han JK, Kim SH, Jeong JY, An SK, Han CJ, Son KR, Lee KH, Lee JM, Choi BI: 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

5. Mathie AG, Strickland NH: Interpretation of CT scans with PACS image display in stack mode. Radiology 203:207–209, 1997

6. Krupinski EA, Kallergi M: Choosing a radiology workstation: technical and clinical considerations. Radiology 242:671–682, 2007

7. van der Heyden JE, Inkpen KM, Atkins MS, Carpendale MST: Exploring presentation methods for tomographic medical image viewing. Artif Intell Med 22:89–109, 2001

8. Moise A, Atkins MS: Interaction techniques for radiology workstations: impact on users' productivity. Proc SPIE Medical Imaging 5371:16–22, 2004

9. Atkins MS, Kirkpatrick AE, Knight A, Forster B: Evaluating user interfaces for stack mode viewing. Proc SPIE Medical Imaging 6515:65150A1–A10, 2007

10. Teistler M, Breiman RS, Lison T, Bott OJ, Pretschner DP, Aziz A, Nowinski WL: Simplifying the exploration of volumetric Images: development of a 3D user interface for the radiologist's workplace. J Digit Imaging, doi:10.1007/s10278-007-9025-8, Mar 27, 2007

11. Wang XH, Durick JE, Lu A, Herbert DL, Golla SK, Foley K, Piracha CS, Shinde DD, Shindel BE, Fuhrman CR, Britton CA, Strollo DC, Shang SS, Lacomis JM, Good WF: Characterization of radiologists' search strategies for lung nodule detection: slice-based versus volumetric displays. J Digit Imaging, doi:10.1007/s10278-007-9076-x, September 15, 2007

12. Sherbondy AJ, Homlund D, Rubin GD, Schraedley PK, Winograd T, Napel S: Alternative input devices for efficient navigation of large CT angiography data sets. Radiology 234:391–398, 2005

13. Weiss DL, Siddiqui KM, Scopelliti J: Radiologist assessment of PACS user interface devices. J Am Coll Radiol 3:265–273, 2006

14. Moise A, Atkins MS, Rohling R: Evaluating different radiology workstation interaction techniques with radiologists and laypersons. J Digit Imaging 18:116–130, 2005