Is image manipulation necessary to interpret digital mammographic images efficiently?

Yan Chen[a], Alastair Gale[a], Anne Turnbull[b], Jonathan James[c]
[a] Applied Vision Research Centre, Loughborough University, Loughborough, UK;
[b] Breast Unit, Derby Royal Hospital, Derby, UK;
[c] Nottingham Breast Institute, Nottingham City Hospital, Nottingham, UK

ABSTRACT

With the introduction of digital breast screening across the UK, screeners need to learn how best to inspect these images. A key advantage over mammographic film is the facility to use workstation image manipulation tools. Forty two-view FFDM screening cases, representing malignant, normal and benign appearances were examined by fourteen radiologists and advanced practitioners from two UK screening centres. For half the cases, the mammography workstation image manipulation tools could be employed and for the other half these were not used. Participants classified each case and indicated whether an abnormality was present. Throughout the study the participants' visual search behaviour as well as their image manipulations was recorded. Whether or not image manipulation tools were used made very little difference to overall performance (t-test, $p > .05$) as confirmed by JAFROC analysis Figure-Of-Merit values of 0.816 and 0.838 (with and without tools respectively); performance not using tools was better. However, using tools significantly increased inspection time ($p < 0.5$) as well as participants' confidence. Detailed examination of participants' image inspection behaviour elicited that the average time on each case in the different viewing conditions differed significantly between the high experienced readers and low experienced readers. The visual data analysis revealed that the participants made similar overall pattern of errors on both modalities. The visual search behaviour on both modalities are surprisingly similar.

Keywords: Mammographic interpretation training, low resolution devices, eye movements, HCI

1. INTRODUCTION

Breast screening has been undertaken across the UK for over 20 years using mammographic film as the imaging medium [1]. Recent developments will see the age range being increased to encompass woman aged 47- 73 years [2]. To cope with the additional workload Full Field Digital Mammographic imaging is being rolled out so that soon all 110 screening centres will have some digital imaging ability with nationwide full digital imaging ensuing. The change to digital allows current screening personnel to manipulate images to optimise appearance. However, there has been little research to show that the image manipulation tools support better performance and also whether it is an efficient way of examining mammographic images. At the same time, clinical workstations are in great demand and if other less expensive monitors could be employed for training, or other non-clinical purposes then this would be very useful.

As reported previously [3] a long and detailed study was performed examining screening performance on a workstation, standard computer monitor and an iPhone where participants either were able to manipulate the displayed images using typical DICOM viewing software or else were not permitted to manipulate the images in such a manner. Here, more detailed data regarding the comparison of performance and the related underlying factors on the workstation and monitor are presented. In particular the observers' visual search and interaction behaviour whilst examining the images on these displays is concentrated upon, particularly where false negative responses were made and where the participants manipulated the images. The fundamental purpose of the overall study was to examine how screeners' performance on different display devices varied and whether non-clinical workstation performance was at all at a level to imply that these could be used instead of workstations to offer training (potentially outside the radiology department). Here the purpose of the investigation is to report in detail how performance on a workstation and a monitor varied.

 A group of screeners was presented with sets of recent difficult FFDM screening cases on a mammographic workstation and asked to examine these images either with or without using any image manipulation functions. Their performance

and visual search behaviour were measured to determine how using these functions affected their case reading behaviour and abilities.

Three UK screening centres were selected which each had had GE digital mammography equipment for at least four years and all screeners were familiar with the appearance of images generated by these systems. Therefore, nineteen screeners from these three major breast screening centres were approached with detailed study-related information. For logistical reasons (the large amount of experimental equipment required to be transported from centre to centre), it eventually transpired that to conduct the study at one of the centres was not practical. Consequently, fourteen participants from two of the major UK breast screening centres took part.

The participants were divided into two groups according to the different screening centres where they primarily worked.

2. METHOD

2.1. Experimental Design

__Participants__: fourteen experienced screeners (including nine consultant radiologists and five advanced practitioners [i.e. technologists]) from two major UK breast screening centres volunteered to undertake the experiments.

__Visual Stimuli__: two sets of 20 challenging digital screening cases (MLO and CC views) which demonstrated difficult examples of normal, benign and malignant appearances were selected and matched by feature type (i.e. masses, calcifications and architectural distortions) and difficulty.

__Hardware__: (1) GE digital mammography workstations (with 5 megapixel dual monitors; resolution 2,048 x 2,560 pixels each); (2) a standard LCD monitor (images were shown using a DICOM viewer running on a laptop, screen size: 21.5", resolution: 1,050 x1,680).

__Eye Tracking Device__: A head mounted eye tracker (ASL 504) was used to monitor the visual search behaviour of the experienced screeners in this study.

__Procedure__: each participant examined the two sets of images with/without using the image manipulation tool (HCI) on both digital workstation and standard monitor, however with a minimum two month gap in between. The participants' behaviour whilst examining each image was recorded for subsequent detailed analyses. The procedure has been reported elsewhere in details [3].

2.2 Behaviour Data Analysis

Detailed examination of the sequence of the image manipulation (HCI) data was then carried out. For each participant, their behavioural data for each study round was recorded on two separate videotapes; one for the case set viewed with the support of the image manipulation tools and the other for the case set viewed without the support of such tools. Each tape was transferred to computer disk and saved as an *.avi* file for data analysis. Each of the.*avi* files was then reviewed and coded using Captiv L-2100 software. Figure 1 illustrates the video analysis process. This software allows the experimenter to replay the video record of each participant's behaviour and code various predetermined actions as and when they occur during the participant's visual examination of each mammographic case. For instance the start and end of examining each mammographic view, use of image zoom and pan etc. are all recorded. These data are then exported to an Excel spreadsheet for further detailed analyses.

Also, to check on any potential subjective interpretation of the video records, one video recording was first analysed using the Captiv software and then the same video was reviewed again independently by another experimenter using the software. The results showed no individual differences in identifying the time sequencing of the actions/events and consequently it was decided that all of the tapes would be analysed by one individual.
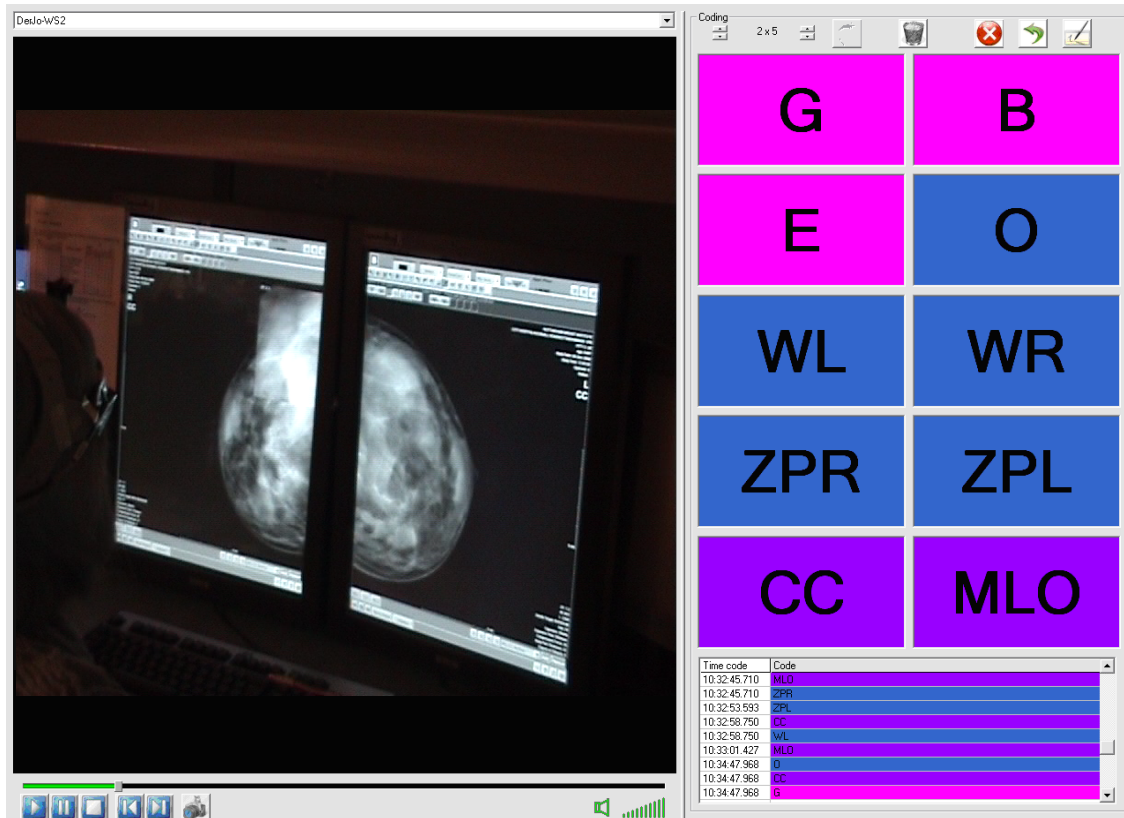
Figure 1. Screen shot of analysis of the imaging interpretation behaviour data using Captiv L-2100 software. The symbol codes on the right were used to identify specific observer actions (e.g. image zooming).

2.3. Visual Data Analysis

Participants' eye movements were recorded throughout the study. A key part of the interaction with the images was how individuals visually examined and interacted with the cases when examining images on the diagnostic workstation and the standard monitor when the image manipulation tools were used. The raw eye movement data were clustered into fixations using criteria (e.g. number of data samples to define a fixation, the spatial separation of each data sample, and a minimum fixation duration of 200ms) adapted from the ASL eye tracking 'Eyenal' software.

Of particular interest was the eye fixation data on all the false negative cases made when participants were examining cases on the workstation using image manipulation and also examining cases on the standard monitor again while using image manipulation. The eye movement data were used to classify the false negative responses into the following three categories:

Search error: neither the location of the abnormality, nor the area of interest, attracted any visual attention;

Detection error: the location of the abnormality was not reported and the area of interest only attracted visual attention for a short (*<1000ms*) period of time;

Interpretation error: the location of the abnormality was not reported even though the area of interest attracted visual attention for longer than 1000ms.

Other information was also collected and analysed, such as:

- 'Time to first hit', which is how long and also how many fixations it took from when the image appeared to first hit the area-of-interest (AOI);

- 'Dwell time' corresponding to the total amount of time spent in the area-of-interest.

3. RESULTS

Overall performance between the digital workstation, whether or not image manipulation tools were used, and the standard monitor, when image manipulation tools were used, was not significantly different ($p>.05$) as confirmed by JAFROC analysis Figure-Of-Merit values of 0.816, 0.838 and 0.827. These performance data have been reported elsewhere [3].

The key interest here is the participants' behaviour data and visual search data analyses. Participants' data were also considered as grouped into high (where mammography experience was 10 years and above) and low (i.e. mammography experience was less than 10 years) experience.

3.1. Behaviour Data

Data were examined across all participants with regard to the average examination time spent on each case. t-tests showed non-significant differences ($p=n.s.$) between the time spent examining cases on the standard monitor, using image manipulation tools, and on the workstation using, or not using, the image manipulation tools. See figure 2 for details.
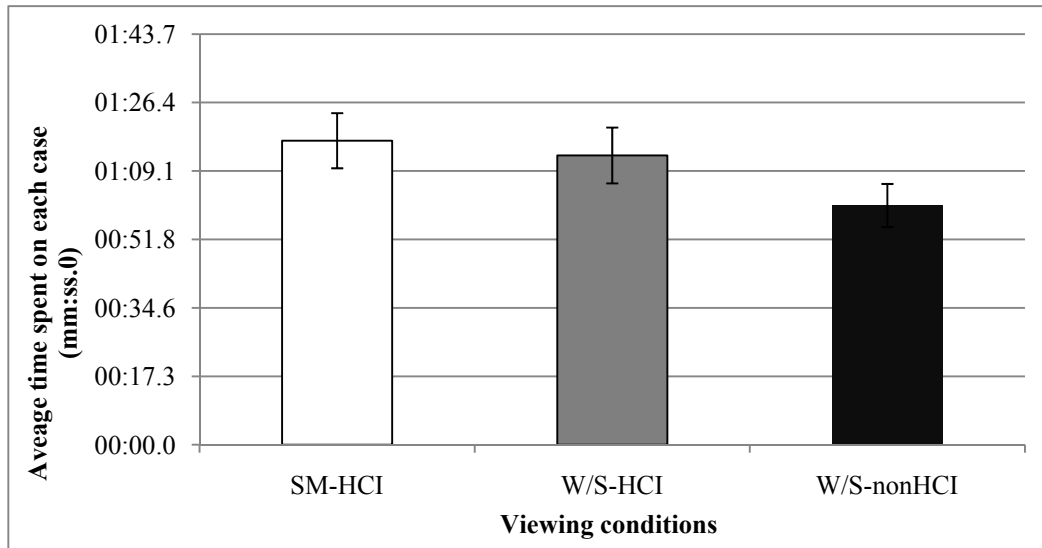
Figure 2. Average time spent per case comparison between the three viewing conditions.

A 2 x 3 mixed design ANOVA revealed that there was a non-significant main effect of mammography experience on the average time spent on each case [$F_{(1, 10)}=4.17$, $p=n.s.$, $r = .54$] and a significant main effect of viewing conditions on the average time spent on each case, [$F_{(2, 20)}=3.9$, $p < .05$, $r = .40$]. Contrast revealed that the average time spent on each case while examining on the standard monitor with image manipulation tools was significant longer than if examining on the workstation without using the image manipulation tools [$F_{(2, 20)}=6.39$, $p < .05$, $r = .49$].

There was a significant interaction effect between viewing conditions and participants' mammography experience level, [$F_{(2, 20)}=3.9$, $p < .05$, $r = .40$]. This indicates that the average time on each case in the different viewing conditions differed significantly between the high experienced readers and low experienced readers. To break down the interaction, contrast was performed which revealed significant interactions when comparing high and low experienced observers' average time on each case:

- while examining images on the standard monitor with image manipulation tools as compared to examining images on the workstation without image manipulation tools, [$F_{(1, 10)}=7.84$, $p < .05$, $r = .66$]

- and whilst examining images on the workstation with image manipulation tools compared to examining images on workstation without image manipulation tools [$F(1, 10)$=14.937, $p < .05$, $r = .77$]. Details are shown in figure 3.
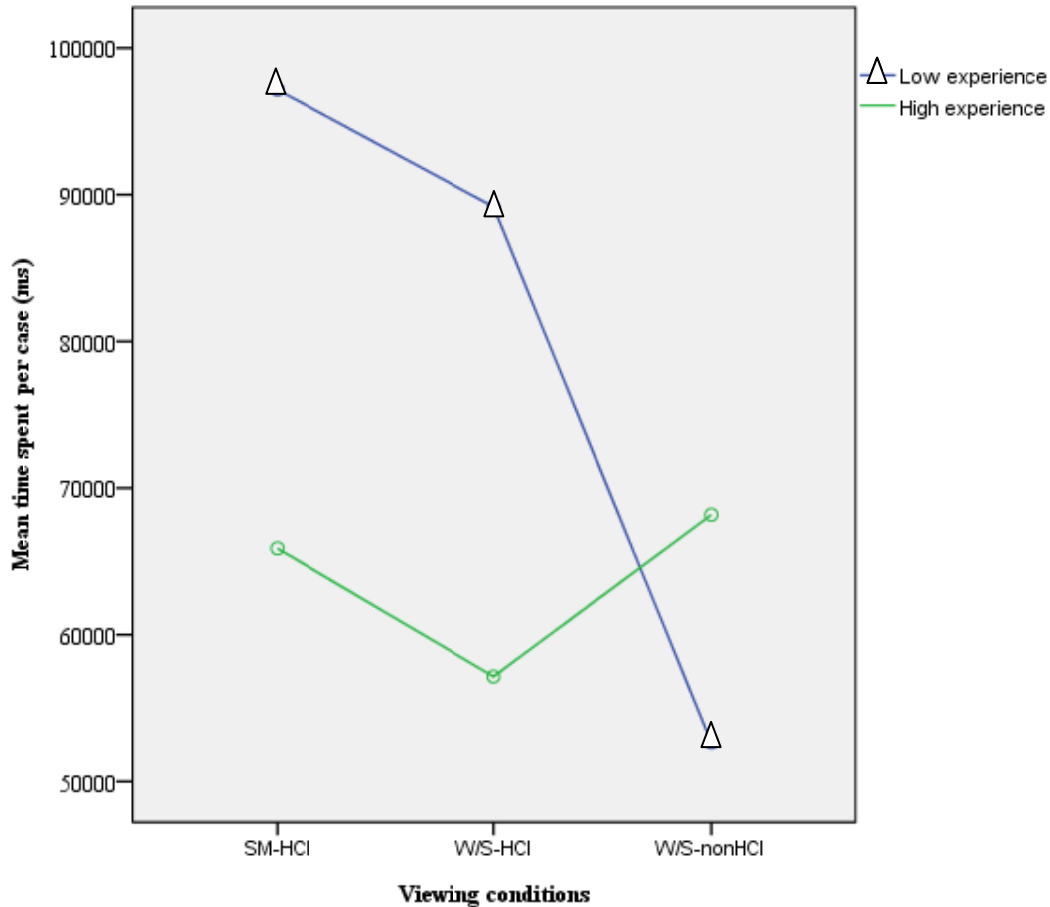

Figure 3. Average time spent per case comparison between the three conditions x mammography experience

3.2. Eye Movement Data

A two-way repeated-measures ANOVA revealed that there was a non-significant effect of viewing modality on false negative error percentage, $F(1, 14) = 3.5$, $p= n.s.$, $r = .13$; there was also a non-significant effect of visual error classification on false negative error percentage, $F(2, 28) = 1.05$, $p= n.s$, $r = .19$.

Visual search error was the main false negative error (48% of false negative errors were search errors on the standard monitor; 31% of false negative errors were search errors on the digital workstation). The percentages of detection and interpretation errors were similar (32.2% of detection error and 19.6% of interpretation error on the standard monitor; 15% of detection error and 33.3% of interpretation error on the workstation). There was a non-significant interaction effect between the types of viewing modality and the types of visual errors, $F(2, 28) = 1.11$, $p= n.s$, $r = .19$. The percentages of each false negative error type on the two different viewing modalities are shown in figures 4 & 5.
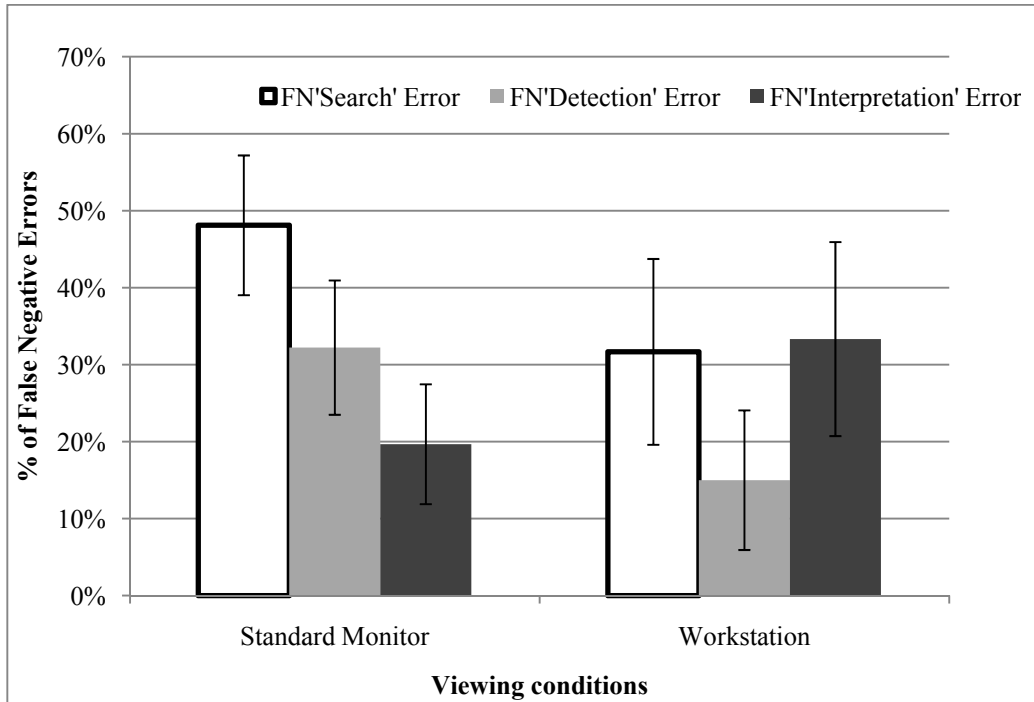
Figure 4. False negative errors made on each modality
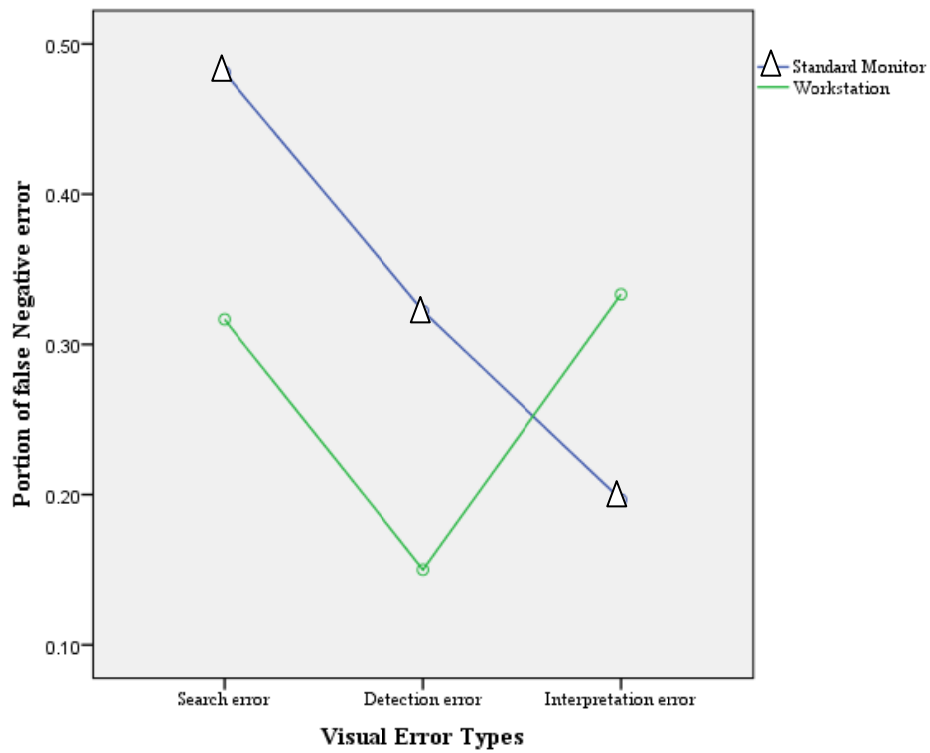

Figure 5. Proportion of FN errors split by error type

On average, the number of times (approximately 4-5 times) that participants fixated within the AOI while examining images on a standard monitor (M = 4.6, SE = 1.5) was not significantly different from examining on the workstation (M = 5.2, SE = 3.1, $t(12)$= -.162, $p=n.s.$, r = .05). Similarly, on average, participants' eye dwell time within the AOI while examining images on the standard monitor (M = 4.82, SE=2.88) was not significantly different from examining on the workstation (M = 3.27, SE=1.86, $t(12)$= .448, $p=n.s.$, r = .13) – although in terms of mean values they 'dwelt' within the AOI for about *2s* longer on the monitor than the workstation. Details are shown in figure 6 & figure 7.
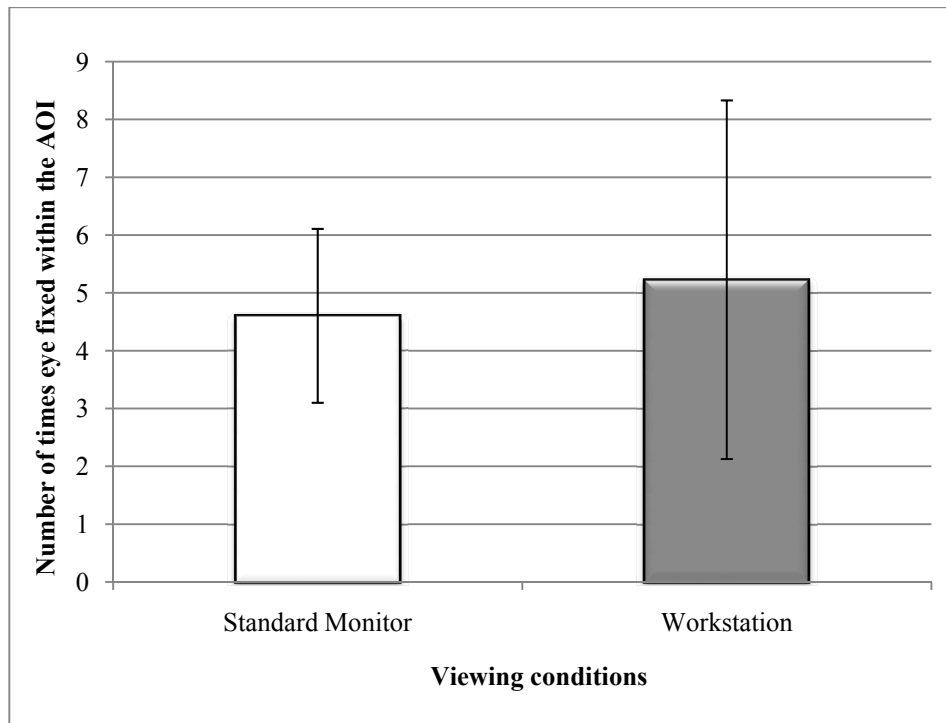

Figure 6. Number of times participants fixated within the AOI while examining images on the standard monitor and the workstation

On average, the number of fixations before the eye first fixated within the AOI while examining images on the standard monitor (M = 87.09, SE = 26.01) was not significantly different from examining on the workstation (M = 175.57, SE= 79.08, t(6)=-1.023, p=n.s., r=.39). Similarly, on average, the participants' average duration between image onset to first hit the AOI while examining images on the standard monitor (M = 38.46, SE = 13.20) was not significantly different from examining on the workstation (M = 95.33, SE = 59.4, t(6)=-.897, p=n.s., r=.34). Details are shown in figures 8 & 9.
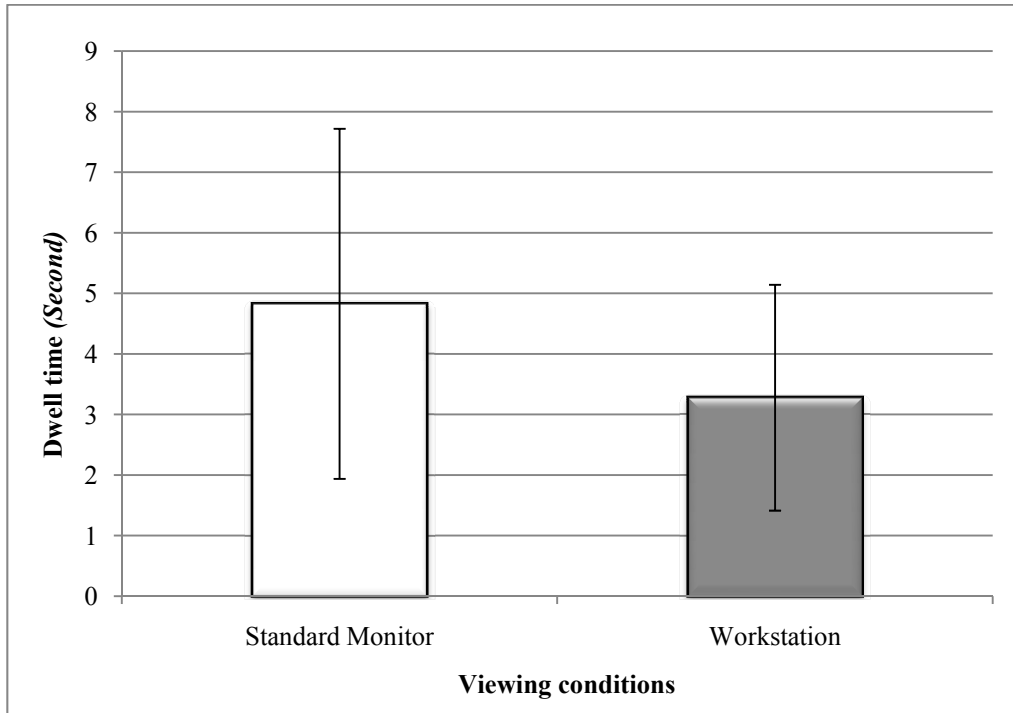
Figure 7. Dwell time comparison between examining images on the standard monitor and the workstation
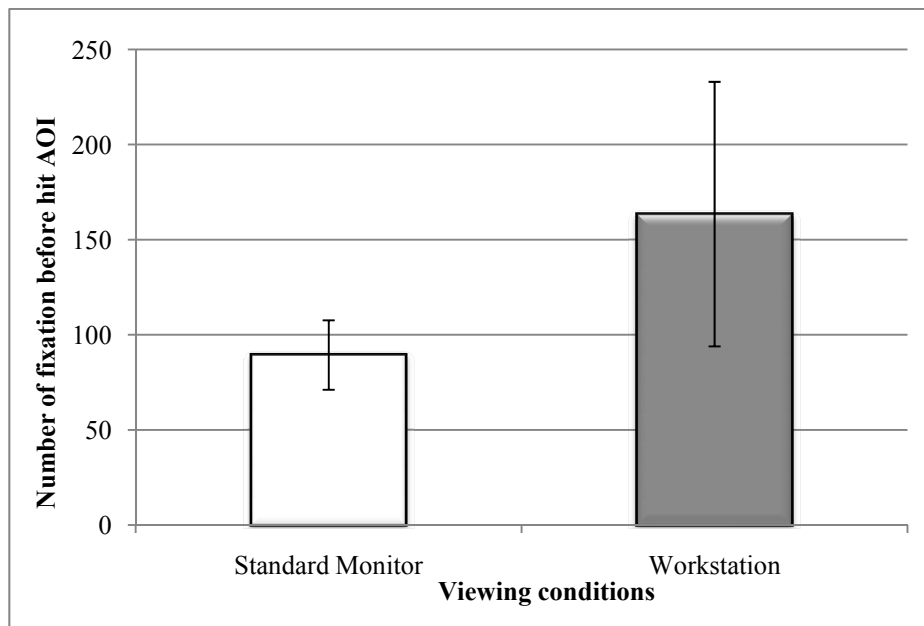

Figure 8. Comparison of number of fixations before the eye first fixated within the AOI while examining images on the standard monitor and workstation
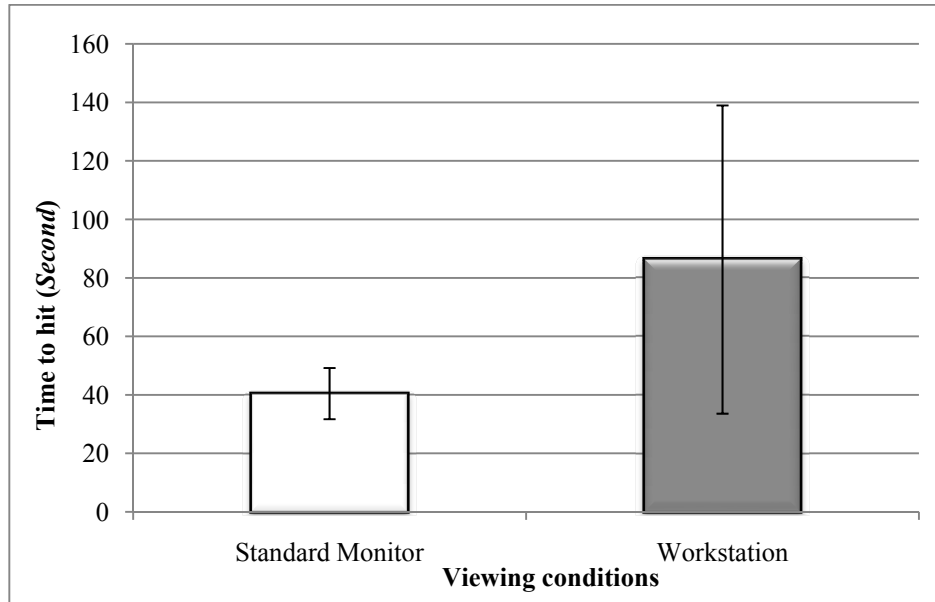
Figure 9. Comparison of time to hit between examining images on standard monitor and workstation

4. DISCUSSION

This study examined how radiologists and advanced practitioner radiographers (technologists) performed when examining sets of difficult recent screening cases on different modalities. The research interest here is the difference between participants' manipulation behaviour under different viewing conditions; also the comparison of participants' visual search behaviour under the different conditions.

Participants' image manipulation behaviour on the workstation and the standard monitor was compared. The important finding was that the average time on each case in the different viewing conditions differed significantly between the high experienced readers and low experienced readers. Compared with the low experienced participants, high experienced ones spent less time on the standard monitor and workstation while the image manipulation tools were used. However, high experienced participants spent more time than the inexperienced participants while the image manipulation tools were not used. This could be because the experienced readers became more careful and therefore spent a longer time examining images while the normal image manipulation tools were not available.

The visual data analysis revealed that the participants made similar overall pattern of errors on both modalities. There were no significant differences between the three different visual errors on either modality. Furthermore, there was no significant difference between modalities in terms of these error types.

On average, the number of times participants fixated within the AOI and participants' eye dwell time within the AOI while examining images on a standard monitor was not significantly different from examining images on the workstation. Similarly, the number of fixations made before the eye first fixated within the AOI and the average duration to first hit the AOI while examining images on the standard monitor was not significantly different from examining images on the workstation. These findings are really somewhat surprising given the differences in monitor and workstation display sizes and resolutions.

This points to similarity in how they inspected images on the two displays. Participants were overall faster on the monitor to hit the AOI and made fewer eye fixations before they hit the AOI. However, once within the AOI then they spent more time dwelling there on the monitor. This may well reflect that the abnormalities were able to be seen on either display type but that once identified then it took longer to examine in detail on the monitor.

5. CONCLUSION

A standard monitor with the support of image manipulation achieved similar performance as on the digital workstation. Visual data analysis on the false negative errors suggested that participants' visual examination behaviour on the standard monitor was very similar to their behaviour on the digital workstation. Also, their screen interaction behaviour on these two modalities was shown to be comparable. This implies that the standard monitor can potentially be considered to be a good alternative to a clinical workstation for the purpose of delivering mammographic interpretation training if image manipulation tools are made available and the training is carefully planned.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Department of Health, "Breast Cancer Screening Report to the Health Ministers of England, Wales, Scotland and Northern Ireland by a Working Group Chaired by Sir Patrick Forrest," London, HMSO, 1986.

[2]   Department of Health, "Cancer Reform Strategy," London, 2007.

[3]   Chen, Y., Gale, A.G., James, J., Evans, A. & Turnbull, A., "Breast Screening: visual search as an aid for digital mammographic interpretation training," In B. Sahiner & D.J. Manning (Eds.) Proceeding of SPIE Medical Imaging 2010: Image Perception, Observer Performance, and Technology Assessment, Vol 7627, 2010.