Computer-Based Learning to Improve Breast Cancer Detection Skills

Yan Chen[1], Alastair Gale[1], Hazel Scott[1], Andrew Evans[2], and Jonathan James[2]

[1] Applied Vision Research Centre, Department of Computer Science, Loughborough University, United Kingdom
[2] Nottingham Breast Institute, Nottingham City Hospital, Nottingham, United Kingdom
{Y.Chen3,A.G.Gale,H.Scott}@lboro.ac.uk

Abstract. In breast cancer screening it is important both to improve and maintain cancer detection skills at their highest levels. The introduction of digital imaging enables computer-based learning to be undertaken outside breast screening centres using a range of different devices. The potential for providing computer-based interpretation training using low-cost devices is detailed. The results demonstrated that naive observers can be trained to recognise certain key breast cancer appearances using a low cost display monitor along with a range of HCI techniques.

Keywords: mammogram interpretation, training, eye movement, visualization, Human-Computer Interaction (HCI).

1 Introduction

Breast cancer, the most common type of cancer amongst women, is responsible for 13% of all deaths globally and is the leading cause of death in women [1]. To improve the early detection of this disease, the National Health Service Breast Screening Programme (NHSBSP) was initiated by the Department of Health in 1988 across the UK and this was the first nationwide scheme of its kind in the world. With the purpose of facilitating the early detection of breast cancer and improving treatment, the scheme provides free breast screening every three years for all women in the UK aged 50 to 70 years [2].

Typically four mammographic views of a woman are taken, two views of each breast, i.e., Cranio Caudal (CC): a vertical view through the breast; Media Lateral Oblique (MLO): an angular view which include the glands under the arm. Interpreting these mammograms is a difficult task, even by experienced and specially trained personnel. To aid screening personnel in appreciating different presentations of difficult abnormal appearances all UK breast screeners annually interpret sets of known difficult mammographic cases as a mean of self-assessing their breast screening interpretation skills; this is known as the PERFORMS scheme [3]. Results from this scheme over several years demonstrate that mean cancer detection performance on these exemplar cases can range from 81 -95% with a skewed distribution [4]. The scheme can help identify for each individual where they have particular difficulties in detecting or recognising known abnormal appearances in these cases as compared to their peers

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2009, LNCS 5613, pp. 49–57, 2009.

nationally. Additional training for individuals which targets such known discrepancies can then be proposed which should help in maintaining skill levels of cancer detection.

Currently most training in the interpretation of screening mammograms needs to be undertaken where there is a mammo-alternator (a large viewing device, where several hundred cases can be loaded, which comprises a back-illuminated surface on which numerous mammographic cases can be presented simultaneously for examination) or other suitable light box (where only one single mammographic case can be presented at a time) on which to view the mammograms. This then limits the locations and availability where training can take place.

However, digital breast screening, where images of the breast are digitally acquired and the resultant images examined using very high resolution and expensive dedicated workstations in breast screening centres is gradually being introduced nationally. Typically these workstations have a dual monitor set up where each has a resolution of 5 Mega pixels (2048 x 2560) and is capable of displaying 10-bit greyscale images with a high contrast ratio. Each monitor can display various combinations of mammographic views and is used primarily to display a single mammogram which can then be zoomed into and manipulated. The widespread adoption of such digital technology renders additional training opportunities using other less expensive and lower resolution displays and in various locations possible for certain purposes if suitable Human-Computer Interaction (HCI) techniques can be derived which are acceptable to radiologists. Primarily when examining digital mammographic images for diagnostic purposes then an appropriate workstation is fundamentally required. However it may be possible to target training of key mammographic appearances using other lower resolution displays which would make web delivered training using a range of viewing devices both potentially feasible as well as acceptable.

This research investigates the utility of employing low-cost devices to provide individualised training with the support of HCI techniques and then determines what types of image interaction and manipulation techniques are required by potential end users. It was hypothesised that using a lower resolution display would hamper the identification of small signs of breast cancer (e.g. microcalcifications) but that larger signs (e.g. masses) would still be visible. Different HCI approaches were used as an initial approach to training based on facets of typical routine clinical training or reporting. In particular, the possibility of using aspects of visual search (as an HCI technique) to assist in mammographic interpretation training is explored.

2 Methods

Firstly, a questionnaire was sent to all (601) UK screening personnel to determine their satisfaction with current training opportunities and their views on the potential for individualised training as well as on the use of other image display devices for examining the high resolution mammographic images. This included details of current screening (e.g. profession; experience of digital mammography) and the individual's current usage of mammographic interpretation training (e.g. forms of training available; the amount of training opportunities; any difficulties of current training; advantages and disadvantages of current training).

Secondly, as a 'worst case' scenario investigation twenty experienced breast screening radiologists were shown a series of mammograms on a small PDA (screen size 3.5") and questioned as to whether they thought such displays could ever be used for training purposes in breast screening and if possible then what types of HCI image manipulation techniques would they envisage requiring (Figure 1).

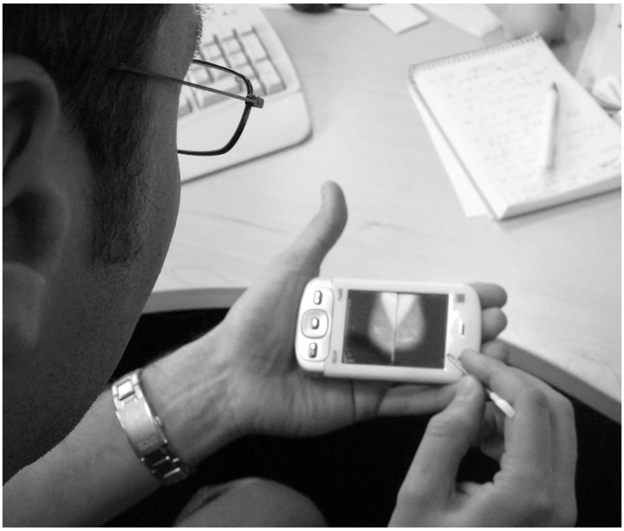


Fig. 1. A radiologist views mammograms (MLO views) on a 3.5" screen PDA

An experimental investigation was then undertaken which examined different aspects of computer based training for breast screening interpretation. Initially, both the eye movements and the audio description of an expert breast screening radiologist were recorded whilst a series of recent screening cases were examined. From these eye movement data key fixations were identified (based on visual dwell measures) which fell on both abnormal areas as well as other normal image areas which attracted visual attention.

Twenty naïve participants were then first given a short standardised introductory computer-based presentation about breast cancer; mammograms and the appearances of two key mammographic features; masses and calcifications. These two features were deliberately chosen as whilst both can be difficult to detect on a mammogram, masses are generally relatively large and calcifications are fairly small irregular abnormalities which can appear singly or in clumps and of various size. This was then followed by an image examination exercise.

During this, each participant was required to examine 21 recent mammographic screening cases (7 normal cases, 7 cases containing a mass and 7 cases containing calcifications) on two occasions. Their performance in identifying abnormal appearance was measured as they first examined these cases, then they either undertook a short training exercise or were assigned to a control condition and then re-examined the cases. On both occasions the participants were asked to identify whether each case contained either of the two features. If they thought an abnormality was present then they also had to indicate its location. Participants' eye-movements were recorded throughout.

The different forms of training used were: whole mammogram presented but with regions of interest highlighted; only regions of interest (where only a magnified portion of the mammographic image around a potential abnormality site was presented); playback of the expert's visual search behaviour, and playback of the expert's audio instruction on each case.

- Participants

Twenty naive observers participated (seventeen research students in various subjects, and three university employees) were involved in the study. None of them had any experience in mammography reading. Participants were split into four experimental groups and a control group (without undertaking any training) with the experimental groups undertaking different forms of computer based training.

- Materials

Visual stimuli: 21 sets of recent digital mammographic images were used. Each image set comprised both the Medio Lateral Oblique (MLO) views and Cranio Caudal (CC) views of both breasts. Fourteen of the pairs featured a specific abnormality which had been grouped into two types (namely: Mass and Calcification) with the abnormality visible on either one or both views. Seven of the sets featured no abnormality.

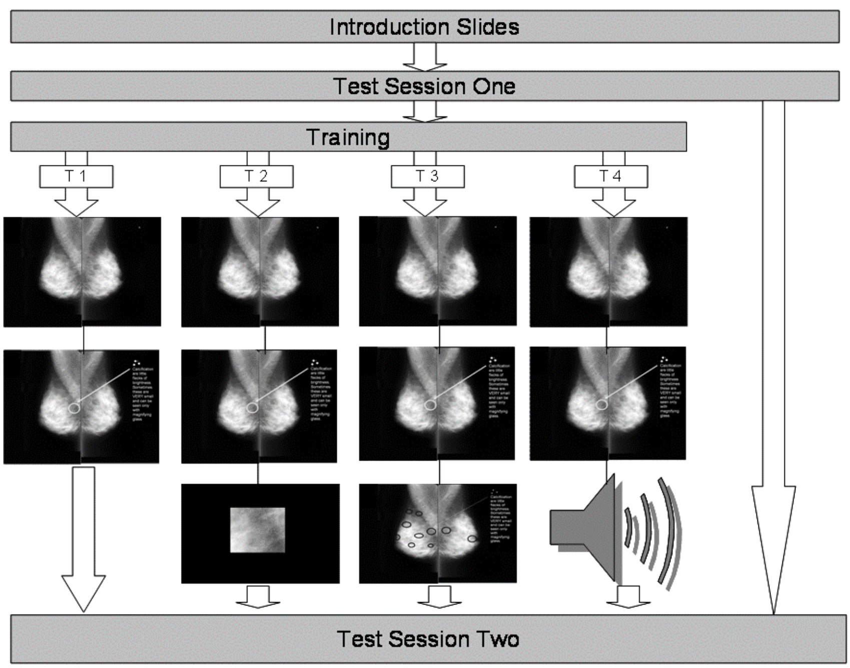


Fig. 2. Schematic of the different training and control approaches

Hardware: The experiment was run on a PC with a LCD monitor (display size: 21.5”; resolution: 1680 x 1050) for displaying images; eye movements were recorded using a remote oculometer – a Tobii X50 eye-tracker. This records eye movements with no attachment to the participant and allows them some degree of free head movement (within 30 x 16 x 20 cm ; W x H x D) at 60 cm from the device. It has a reported accuracy of 0.5-0.7 degrees of visual angle.

▪ Design

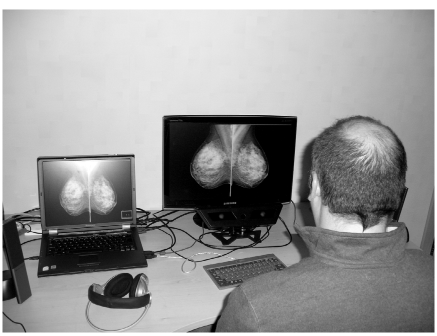
Training sets: each training set included the 21 cases comprising both MLO and CC views of each case but were presented in four different formats (Figure 2).

These formats were:

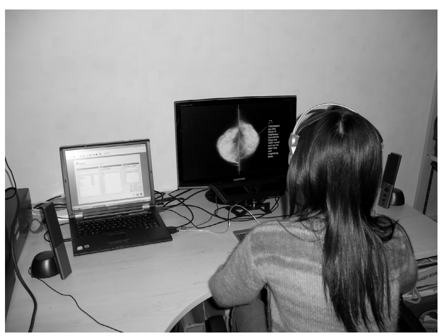
1. T1 – whole image: the MLO views of both breasts were presented then, where appropriate, the feature area was highlighted by a circle with descriptive text, followed similarly by the CC views.
2. T2 – image with portion: the MLO views were presented (as in format T1) but then the area of interest around the abnormality was highlighted followed by this area being magnified and shown alone (size: 8”x 8”). This was followed similarly by the CC views.
3. T3 - eye movements: the MLO views were shown and then overlaid with annotated fixation locations of where the initial expert radiologist had looked and in the order in which visual search had been performed. Also, the area of abnormality was highlighted. This was followed similarly by the CC views.
4. T4 - comments: This was similar to format T1 above but with the addition of the expert’s audio descriptions concerning the case.
5. control: a control activity was undertaken which took the same length of time as the other training sessions.

▪ Procedure

For each participant the eye tracker was first calibrated. Each was then given a short introduction concerning breast cancer, mammogram images, and were familiarised



A



B

Fig. 3. A shows a participant taking the test viewing the MLO mammograms of a case.; B shows a participant undertaking the audio training whilst viewing the CC images of a case.

with the appearance of the two different key breast cancer features in such images (masses and calcification). They then visually examined two practice cases. Subsequently they completed a computer-based image examination task whilst their eye movements were discretely recorded (Figure 3). During this task the participants first fixated on a small centrally presented fixation cross and then this was replaced with the MLO views of the case which in turn was followed by the CC views. Participants were asked visually to examine each case view, identify whether the view was normal or contained an abnormality. If the latter, they also had to specify the feature type (i.e. mass or calcification) and its location.

3 Results

For the questionnaire study some 273 questionnaires were returned (a 45% response rate) which covered the main professions in breast-screening film reading, e.g. 152 consultant radiologists, 78 advanced practitioners (in the UK these are specially trained technologists), and others. These all indicated that screeners considered that whilst existing training was adequate they would like to have individualised and targeted training which is only really possible using digital images. The responses concerning the most useful mammographic interpretation training type is shown in Figure 4.

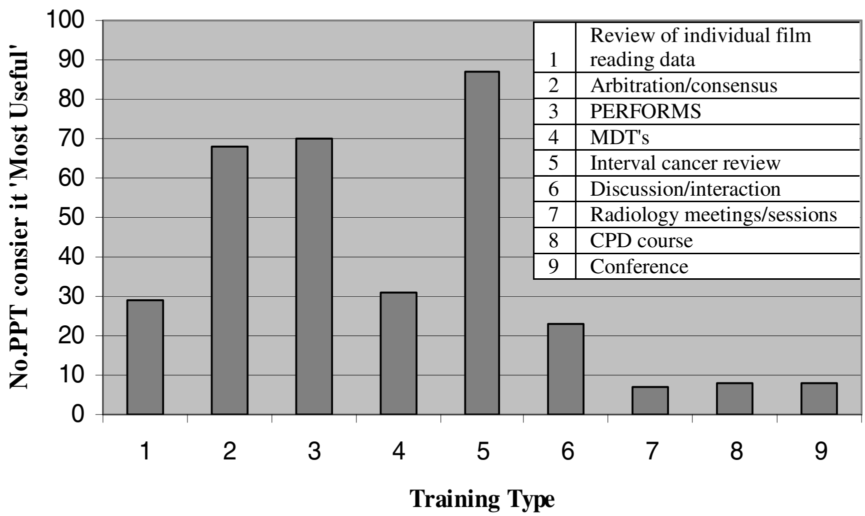


Fig. 4. Participant's response on 'The most useful training type'

'Interval cancer review', 'Arbitration/consensus' and 'PERFORMS' were considered to be the fist three most useful. According to the characteristic of these three training types, it shows that being able to read a large amount of images followed by other forms of information, such as, expert's opinions on the case, other people's feedback on the case, etc. were considered to be the key useful training approaches.

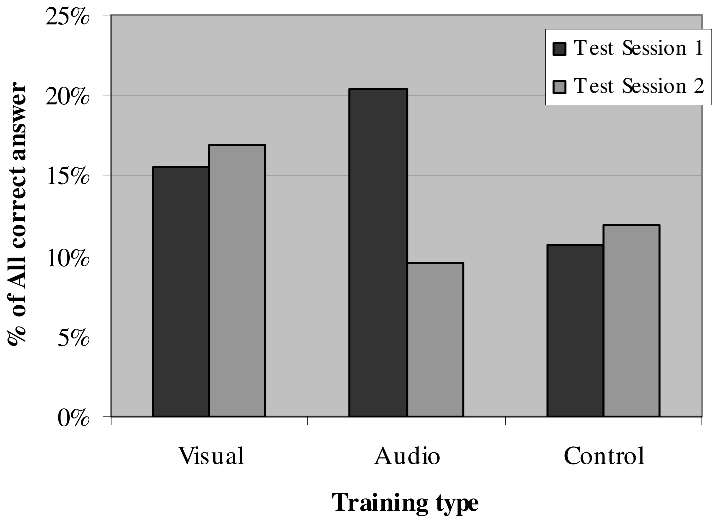


Fig. 5. Percentage of both correct answers (feature & location) of the training types (both test sessions)

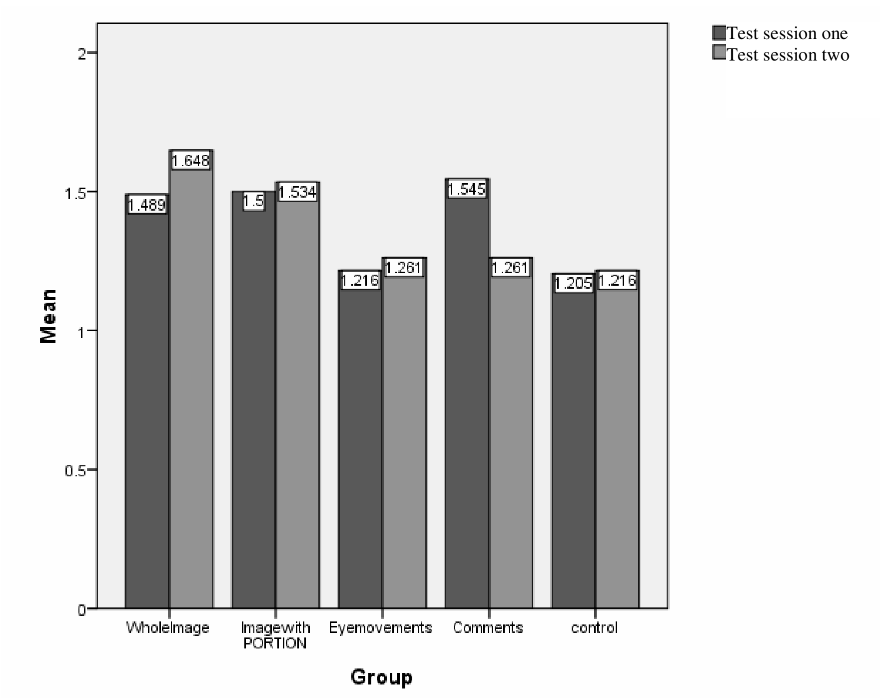


Fig. 6. Mean performance values of each participant group (both test sessions)

With regard to viewing images on the PDA, initially all participants were highly sceptical as the display had approximately 1/10th of the resolution of a digital mammographic workstation but it was determined that such a display could be useful for training (but not for diagnostic) purposes if suitable simple HCI techniques were available which would enable rapid image pan and zoom with contrast level manipulation.

In the experimental study of naïve observers some 880 responses were collected and of these only 124 correctly identified both features and locations on both views of a case. Initially the training types were considered as either visual (whole image, portion, eye movements), audio (comments) or control - details are shown in Figure 5. In audio, performance decreased after training whereas for both the visual and control conditions there was a slight, non significant, increase.

The performance of these participants in the first and second test was then attributed points concerning whether they correctly identified location and features on both views (Figure 6) where a score of 4 indicates identifying both the correct location and feature on both views. Most participants scored 1-1.5 indicating that they only correctly scored either location or feature in either view. In the three visual conditions there was a slight increase in performance in test 2, in the audio condition performance decreased and in the control condition there was little variation.

In all conditions participants were significantly better in identifying features correctly than in identifying correct location ($p<.05$). The performance in identifying features between test 1 and test 2 significantly improved ($p<.05$). In terms of mammographic features then masses were better identified after any visual training and only slightly increased after audio training. Calcifications were detected worse after all three types of visual training as well as the audio training. In the control condition calcification identification improved. Normal cases were reported worse on every second trial.

In the control group the correct identification of normal cases dropped on the second test but mass and calcification detection increased. In the audio condition calcification and normal identification fell on the second test but mass identification improved slightly. In the three visual conditions only mass identification improved.

4 Discussion

From the returned questionnaires, these all indicated that screeners considered that whilst the existing training was adequate they would like to have individualised targeted training [5]. With regard to viewing images on the PDA then for this to be at all useful, easy to use HCI techniques which would allow rapid image manipulation were clearly required [6]. Using a PDA for some training purposes would be possible if this was done in a appropriately lit environment with a low ambient light level.

The experimental study reported here was an initial investigation utilising a single low cost monitor to deliver mammographic interpretation training as compared to using one or two high resolution workstation monitors. Different training regimes were developed which presented observers with images enhanced using different HCI approaches which were considered suitable in this domain. The approaches used were firstly simply showing the full views of both breast simultaneously which mimicked the display on a workstation but at a much lower resolution. Secondly the full view was show followed by a magnified view of the area of interest to mimic basic image manipulation on a workstation. The third approach taken was to utilise areas of interest which attracted the visual attention of an expert together with the scanning path which were overlaid on the image. The fourth approach tried to replicate an expert describing how they examine a case for abnormalities which is a commonly adopted approach in real screening as shown in the questionnaire survey.

The study set out to use short training approaches to examine their effects on naïve observers. This demonstrated that such observers can be trained to recognise certain key breast cancer appearances using a low cost display monitor along with a range of HCI techniques. Two mammographic appearances were studied; small calcifications because these can be difficult to detect in routine breast screening and larger masses. Calcifications were not detected well presumably due to the shortage of image manipulation techniques used here. Naïve observers were used in this study to see how they responded to the different training types.

5 Conclusion

Overall, it is argued that these findings taken together indicate that low cost devices can be used for training purposes in digital breast screening with appropriate HCI techniques. These then extend the opportunity for training beyond the clinical workplace. Ongoing research is further investigating the delivery of computer-based training using a variety of HCI methods to breast screeners.

References

1. World Health Organization "Cancer", Fact sheet 297 (2006) (Accessed: 01 March, 2009)
2. National Statistics Online, Breast Cancer: Incidence rises while deaths continue to fall, Office for National Statistics (2007) (Accessed: 01 March, 2009)
3. Gale, A.G.: PERFORMS-a self assessment scheme for radiologists in breast screening. Seminars in Breast Disease 6(3), 148–152 (2003)
4. Gale, A.G., Scott, H.J.: Measuring Radiology Performance. In: Michell, M. (ed.) Breast Screening. Contemporary Issues in Cancer Imaging – Breast Cancer. Cambridge University Press, Cambridge (in press)
5. Chen, Y., Gale, A.G., Scott, H.J.: Mammographic interpretation training profile in the UK: current difficulties and future outlook. In: Manning, D., Sahiner, B. (eds.) Image Perception, Observer Performance, and Technology Assessment. Proceedings of SPIE (in press)
6. Chen, Y., Gale, A.G., Scott, H.J.: Mammographic interpretation training: how useful is handheld technology? In: Manning, D., Sahiner, B. (eds.) Image Perception, Observer Performance, and Technology Assessment. Proceedings of SPIE, vol. 6917, pp. 691712–691712-10 (2008)