

# The effect of feedback on performance in a fracture detection task

T. Donovan<sup>\*a</sup>, D.J. Manning<sup>a</sup>, P.W. Philips<sup>b</sup>, S. Higham<sup>b</sup>, T. Crawford<sup>b</sup>

<sup>a</sup>Dept. of Medical Imaging Sciences, St Martin's College, Lancaster, LA1 3JD, UK.

<sup>b</sup>Psychology Department, Lancaster University, LA1 4YF, UK

## Abstract

Four observer groups with different levels of expertise were tested to determine the effect of feedback on eye movements and accuracy whilst performing a simple radiological task. The observer groups were 8 experts, 9 year 1 radiography students, 9 year 3 radiography students, and 10 naïve observers (psychology students). The task was fracture detection in the wrist. A test bank of 32 films was compiled with 14 normals, 6 grade 1 fractures (subtle appearance), 6 grade 2 fractures, and 6 grade 3 fractures (obvious appearance). Eye tracking was carried out on all observers to demonstrate differences in visual activity. Observers were asked to rate their confidence in their decision on a ten point scale. Feedback was presented to the observers in the form of circles displayed on the film where fixations had occurred, the size of which was proportional to the length of fixation. Observers were asked to repeat their decision rating. Accuracy was determined by ROC analysis and the area under the curve (AUC). In two groups, the novices and first year radiography students, the feedback resulted in no significant difference in the AUC. In the other two groups, experts ( $p = 0.002$ ) and second year radiography students ( $p = 0.031$ ), feedback had a negative effect on performance. The eye tracking parameters were measured for all subjects and compared. This is work in progress, but initial analysis of the data suggests that in a simple radiological task such as fracture detection, where search is very limited, feedback by encouraging observers to look harder at the image can have a negative effect on image interpretation performance, however for the novice feedback is beneficial as post feedback eye-tracking parameters measured more closely matched those of the experts.

**Keywords:** perceptual feedback, expertise, eye-tracking, fracture detection, performance

## 1. Introduction

Radiological expertise is based on two kinds of skill, the swift and accurate processing of normal appearances and the ability to distinguish disease from normal variation in appearance [1]. These skills need to be learned, and training does result in distinct differences between the way experts and non-experts interpret medical images with different visual scanning strategies [2]. One way of improving detection and interpretation errors is to bring to the observers attention features in an image that are relevant to achieving an accurate diagnosis. One method of computer aided perception does this by identifying areas in an image where a radiologist has looked at for more than a pre-determined length of time, and overlaying the image with these highlighted areas [3]. Visual dwell has for some time been recognised as a predictor of target recognition, particularly in the recognition of missed lung nodules in chest radiographs. Over 70% of missed lung nodules are fixated and processed by fixation durations that are long enough to recognise the abnormality, but are rejected as positive nodules [4]. Nodine et al [4] developed the computed assisted visual search system (CAVS) that highlighted areas that received prolonged visual dwell, and re-presented the areas to the observer. This cuing resulted in a 14% increase in sensitivity. A further study showed an average 16% improvement as measured by the alternative free response operating characteristic (AFROC) analysis [5]. The type of cue is important; Krupinski [6] demonstrated that the cuing was most effective with a complete circle insulating the region of interest (ROI) from outside distracters.

In other domains there has been very little research on perceptual feedback apart from one study where visual feedback of the scan patterns of airframe structural inspectors has been shown to improve performance [7].

The potential applications of using eye-tracking for training and quality assurance, especially with the current climate of clinical governance in medicine, are extensive [8]. Eye-tracking has a role in the design of decision support systems particularly as explicit domain knowledge representation used in such systems often overlooks those factors that are subconsciously applied during visual recognition, such as the way information from the image is filtered. Human visual

pattern recognition is different from general reasoning [9]. The effect of a poorly performing computer-assisted detection cuing environment has been demonstrated to significantly degrade observer performance [10]. In medical imaging, apart from the work by Nodine et al [4] on the detection of lung nodules, there has been little research into the effect of perceptual feedback on visual search and decision strategies of observers and assumptions about the transfer of strategies across tasks cannot always be made. The aim of this study is to use a relatively simple radiological task, provide immediate visual feedback and determine the effects on performance and scan paths.

## 2. Materials and methods

### 2.1 Observer groups

Four groups of observers were compared in the effect perceptual feedback had on their performance and eye movements. The groups were:

1. 8 experienced image reporters, consisting of 6 reporting radiographers and 2 radiologists. The radiographers and radiologists were not separated as they are performing the same radiological task in the work place.
2. 9 third year undergraduate radiography students.
3. 9 first year undergraduate radiography students.
4. 10 naive undergraduate psychology students.

### 2.2 Film test bank

The test bank was compiled of 32 verified digitised wrist images made up of 14 normal cases and 18 cases with a fracture of the wrist. The cases ranged from obvious fractures to quite subtle appearances (figure 1).



Figure 1: Three of the images from the test bank demonstrating the fracture appearances from subtle to obvious.

### 2.3 Observation procedure

Observers were informed that for each of the wrist images they had to decide if a fracture was present or not. It was briefly explained to the novice group, with the aid of a line diagram, what the appearances of a typical fracture were. For each case observers were asked to indicate their diagnostic decision on a 10 point rating scale on the display monitor, using a computer mouse. Each case image was then removed and re-displayed with information on their eye fixation points as circles. The diameter of the circle was in pixels the same as the number of 60Hz samples in the fixation. Observers were asked to review the image and once again indicate their decision on the rating scale. For this study the eye-tracker code used 6 samples as the minimum for a fixation, which at 60Hz is 100ms. The gaze deviation threshold,

or distance that a gaze point may vary from the average fixation point and still be considered part of the fixation, was set internally at 6.35mm.

At the start and end of the study and after every 7 cases the calibration procedure was carried out.

Viewing conditions were controlled and the viewing distance to the monitor was fixed at 54cm. A chin rest was used to support the head. The study took approximately 1 hour. Observers were offered a rest half way through the study to reduce fatigue.

## 2.4 Observer performance measurement

Receiver operating characteristic (ROC) analysis was used in the measurement of performance pre and post feedback. The data was processed through AccuROC 2.5 non parametric analysis (Accumetric Corporation, Canada). The analysis of the ROC curve used the method of Delong, Delong and Clarke-Reason when calculating the area under the ROC curve (AUC). One set of data was discarded due to the observer using an inadequate number of points on the decision rating scale.

## 2.5 Eye-tracking

Eye-tracking data was acquired with a 60Hz LC eyegaze system (LC Technologies, Virginia, USA). Regions of interest (ROI) were defined on all images covering the fracture sites.

Parameters measured from the eye tracking pre and post feedback were:

1. Mean number of fixations in the ROIs,
2. Mean cumulative dwell time in the ROIs,
3. The time taken to the first fixation in the ROI for the true positives.

## 3. Results

### 3.1 Observer performance pre and post feedback.

Figure 2 shows the overall effect of the feedback on the AUC for each observer group. For the experts and the year 3 radiography students the AUC decreased, so perceptual feedback had a negative effect on performance. The difference was significant,  $p = 0.002$  for the experts and  $p = 0.031$  for the year 3 students. For the naïve observers feedback increased the AUC, however this was not significant at  $p = 0.324$  for the naïves and  $p = 0.175$  for the year 1 students.

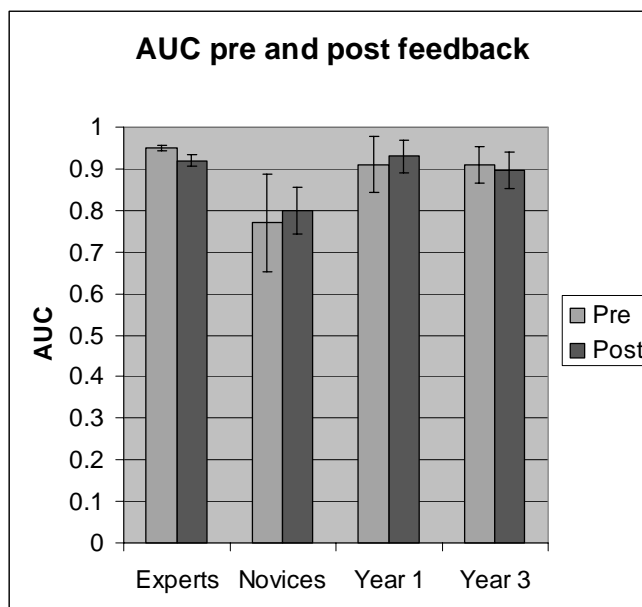


Figure 2: The overall effect of the feedback on the AUC for each observer group.

### 3.2 The effect of feedback on the number of fixations and dwell time.

At the time of going to press only data for the naïve observers and the experts has been analysed.

Figure 3 and figure 4 show the effect of feedback on the mean number of fixations and dwell time respectively in the ROIs of the cases with fractures. For the experts the number of fixations went down by 32% ( $p = 0.007$ ) and the dwell time by 44% ( $p = 0.025$ ). For the naïve observers the effect of feedback was more marked with the fixations reducing by 55% ( $p = 0.003$ ) and the dwell time by 50% ( $p = 0.004$ ).

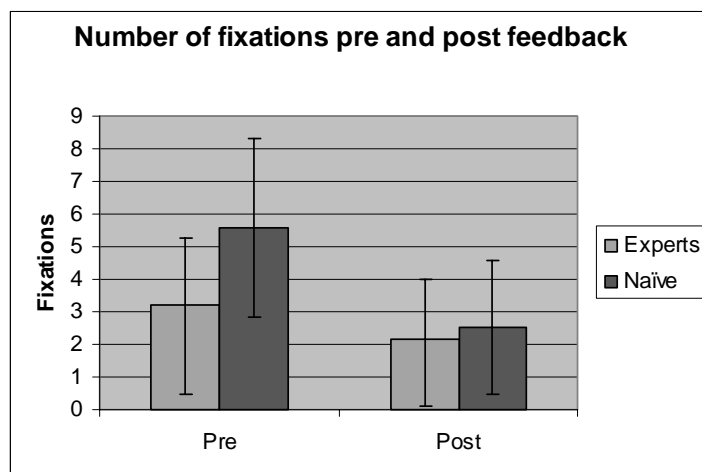


Figure 3: The effect of feedback on the mean number of fixations for the expert and naïve observer groups.

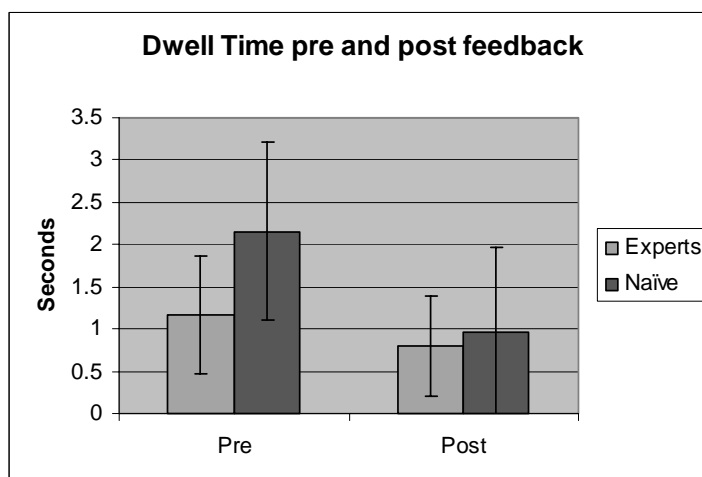


Figure 4: The effect of feedback on the dwell time for the expert and naïve observer groups.

### 3.3 Time to fixate.

The time taken to fixate a fracture and the effect of feedback is shown in figure 5. The mean fixation time for the experts fell by 16% after feedback, this was not significant, whereas for the naïve observers the fall was 40% ( $p = 0.067$ ).

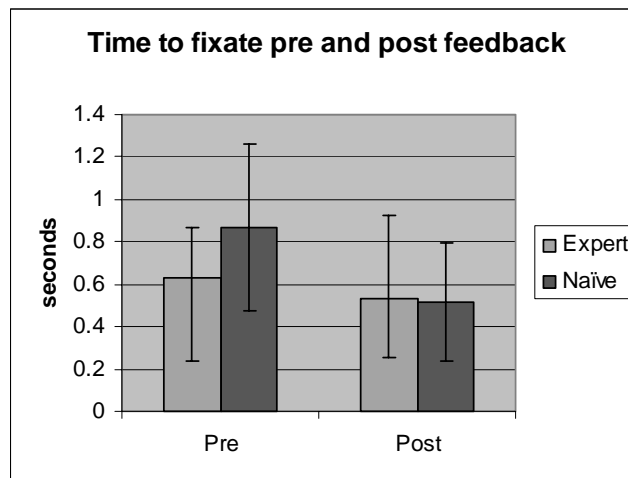


Figure 5: The time taken to fixate in the fracture ROI for each observer group pre and post feedback.

### 3.4 Decisions made with no fixations.

Figure 5 shows a graph indicating the number of fracture cases where a correct true positive decision was made yet no fixation occurred in the ROI containing the fracture site. For the experts this was 6 before feedback and 8 after feedback ( $p = 0.028$ ). For the naïve observers it was 2 before feedback and 5 after feedback ( $p = 0.067$ ).

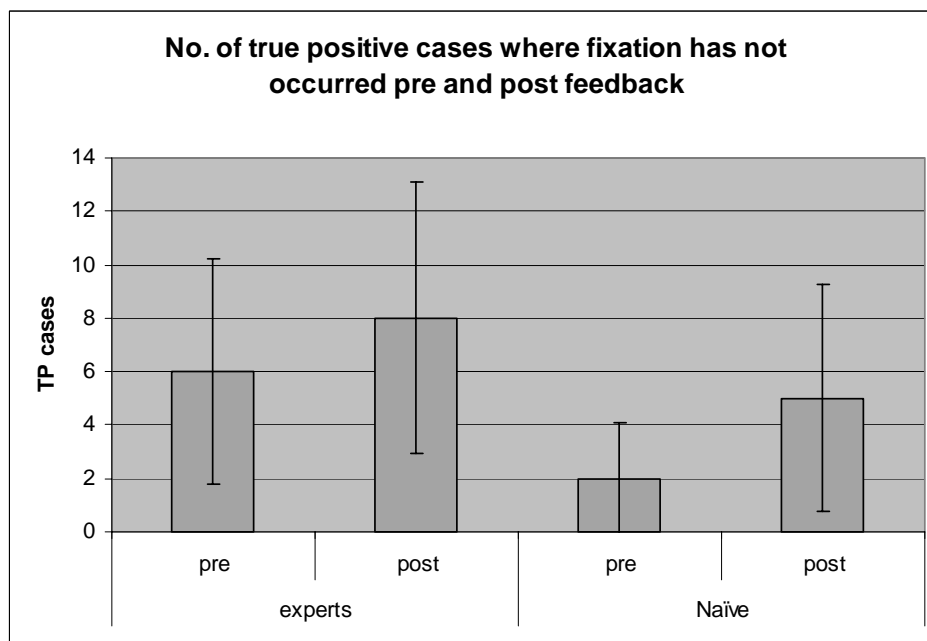


Figure 6: A graph indicating the number of fracture cases where a correct true positive decision was made with no fixation occurring in the fracture ROI for the expert and naïve observer groups.

## 4. Discussion

The aim of the experiment was to determine the effect of feedback on performance and on visual search strategy. The data from the AUC results indicate that the performance of experts and third year radiography students significantly deteriorates following feedback. This is interesting; it appears that forcing a second look makes experienced viewers more uncertain about their diagnosis. For experienced observers this is a relatively simple task, the problem is not in detecting pathology but in deciding whether the negatives are true negatives and detecting subtle variants in anatomy, or variable appearances due to positioning and other technical factors.

There was a significant number of cases where no fixations were recorded for true positives in ROIs over the fracture site. Careful checking of the calibration data was done to ensure the data are correct, our definition for a fixation of 100 ms is widely used in eye-tracking studies. This raises a number of questions:

Does a fixation need to occur before attention takes place? And are fixations a genuine invariable criterion of attention? Is 100ms a realistic threshold?

It is 20 years since Kundel and Nodine conducted a study [11] interpreting chest radiographs without visual search, proposing that observers in a model of the search process start with a global impression or gestalt formed by a pre-attentive global response. A more recent study on the effect of reporting speed on plain film reporting errors [12] determined that increasing reporting speed had no significant overall effect upon accuracy but did lead to fewer false positive reports. Giving experts less time to view an image can be beneficial; conversely the provision of feedback in this task can be detrimental. One possible explanation is that beyond a certain level of expertise observers go beyond analytical reasoning and use holistic recognition [13], which can occur very quickly. There were a number of cases where the experts simply fixated at the centre of the image, as though searching was simply not necessary.

Does this indicate top down or bottom up processing, parallel or serial visual search?

It can be difficult to apply psychophysical models to domain specific image interpretation tasks but in this study it appears that the task is simple enough for experts that a glance or peripheral vision provides enough information to make a decision. Nodine and Kundel's global focal model acknowledges the dominant role of peripheral vision in the global impression stage of their model where obvious pathology is identified that is different from the observers normal prototypic impression. The lack of fixations in many of the experts scan path data suggests the search is parallel, the obvious fracture or target just pops out. Klein and Farrell [14] suggest eye movements are not necessary for parallel search. The concept of the visual lobe may be appropriate in this case, similar to the concept of visual span in reading [15], the size of which varies with the specific task situation. The decline in visual abilities away from central vision is gradual. Similar to Kundel's global impression is the importance of pre-attentive information in the guided search model of Wolfe [17], this is a guide to the deployment of attention through bottom up or top down mechanisms. It is possible that once feedback is generated the task changes to one that involves interpretation of the feedback information and therefore is no longer simply a pop-out bottom up task.

Overall perceptual feedback does seem to result in a reduced number of fixations, a reduced dwell time and reduced time to fixate in the ROIs of positive images. This indicates that the decision making process is more efficient as observers are making decisions in a reduced number of fixations. This effect is much more noticeable with novices. This may be because providing visual feedback is a way of encouraging perceptual learning which is why it is more noticeable in novices. Perceptual learning occurs much earlier in the course of learning than the learning of cognitive processes associated with inference [17]. Therefore one hypothesis is that the experienced observers will be using a top down reasoning strategy, whereas naïve observers will be using a bottom up strategy, feedback for the experienced observers is likely to complicate a straightforward task.

## 5. Conclusion

The initial results from our work suggest that feedback in a relatively simple fracture detection task can compromise the performance of experts. There was relatively little effect on the eye-tracking parameters measured, however performance did decline as measured by the AUC as experts were forced to take on a more pedantic search. The effect

on the naïve observers was more marked with the post feedback eye-tracking parameters measured more closely matching those of the experts. This suggests rapid perceptual learning as a result of the feedback. Further work using more complicated radiological images will need to be undertaken to gain a greater understanding of the effect of perceptual feedback on diagnostic performance.

## References

1. Myles-Worsley M, Johnston WA, Simons M. The influence of expertise on x-ray image processing, *J Exp Psychol learn Mem Cogn*. 1988; **14**: 553-557.
2. Gale AG. *Human Response to Visual Stimuli*. in Hendee WR, Wells PNT (Eds.), The perception of visual information, 1997 2<sup>nd</sup> edition Springer NY.
3. Manning DJ, Ethell SC, Donovan T, Detection or decision errors? Missed lung cancer from the posteroanterior chest radiograph, *Br J Radiol*, 2004, **77**(915): 231-5.
4. Nodine CF, Kundel HL. A visual dwell algorithm can aid search and recognition of missed lung nodules in chest radiographs. In Brogen D (Ed.), *Visual search* 1990. 1<sup>st</sup> edition Taylor S Francis, London.
5. Kundel HL, Nodine CF, Krupinski EA, Computer-displayed eye position as a visual aid to pulmonary nodule interpretation, *Investigative Radiology* 1990; **25**: 8, 890-896.
6. Krupinski EA, Nodine CF, Kundel HL. Perceptual enhancement of tumour targets in chest x-ray images, *Perception and Psychophysics* 1993; **53**(5): 519-526.
7. Gramopadhye AK, Drury CG, Sharit J. Feedback strategies for visual search in airframe structural inspection. *International Journal of Industrial Ergonomics* 1997; **19**: 333-344.
8. Yang G-Z, Dempere-Marco L, Hu X-P, Rowe A, Visual search: psychophysical models and practical applications, *Image and Vision Computing* 2002; **20**: 291-305.
9. Hu X-P, Dempere-Marco L, Yang G-Z, 2003, Implicit feature mapping with visual search for diagnostic decision support in medical imaging, proceedings of ECEM 11 2003, Dundee.
10. Zheng B, Ganott MA, Britton CA, Hakim CM, Hardesty LA, Chang TS, Rockette HE, Gur D. Soft-copy mammographic readings with different computer-assisted detection cuing environments: preliminary findings, *Radiology* 2001; **221**: 633-640.
11. Kundel HL, Nodine CF, Interpreting Chest radiographs without visual search, *Radiology* 1975; **116**: 3, 527 – 532.
12. Edwards AJ, Ricketts C, Dubbins PA, Roobottom CA, Wells IP, The effect of reporting speed on plain film reporting errors, *Clinical Radiology* 2003; **58**: 971-979.
13. Dreyfus HL, Dreyfus SE, Mind over machine: the power of human intuition and expertise in the era of the computer. New York: The Free Press, 1986.
14. Klein R.M., Farrell M, Search performance without eye-movements, *Perception and Psychophysics* 1989; **46**: 476-482.
15. Courtney AJ, Chan HS, Visual lobe dimensions and search performance for targets on a competing homogenous background, *Perception and Psychophysics* 1986; **40**: 39-44.
16. Wolfe JM, Moving toward solutions to some enduring controversies in visual search, *Trends in Cognitive Sciences* 2003; **7**: No 2, 70-75.
17. Sowden PT, Davies IRL, Roling P, Perceptual learning of the detection of features in x-ray images: a functional role for improvements in adults' visual sensitivity? *Journal of Experimental Psychology: Human Perception and Performance*.2000, **26** (1): 379-390.

\* email: t.donovan@ucsm.ac.uk; phone 00 44 1524 384491