Radiology image perception and observer performance: how does expertise and clinical information alter interpretation? Stroke detection explored through eye-tracking.

Lindsey Cooper∗a, Alastair Galea, Iain Darker, Andoni Tomsb & Janak Saada
aApplied Vision Research Centre, Loughborough University, UK
bNorwich Radiology Academy, Norwich, UK
cNorfolk and Norwich University Hospitals NHS Trust, Norwich, UK

ABSTRACT

Historically, radiology research has been dominated by chest and breast screening. Few studies have examined complex interpretative tasks such as the reading of multidimensional brain CT or MRI scans. Additionally, no studies at the time of writing have explored the interpretation of stroke images; from novices through to experienced practitioners using eye movement analysis. Finally, there appears a lack of evidence on the clinical effects of radiology reports and their influence on image appraisal and clinical diagnosis. A computer-based, eye-tracking study was designed to assess diagnostic accuracy and interpretation in stroke CT and MR imagery. Eight predetermined clinical cases, five images per case, were presented to participants (novices, trainee, and radiologists; n=8). The presence or absence of abnormalities was rated on a five-point Likert scale and their locations reported. Half cases of the cases were accompanied by clinical information; half were not, to assess the impact of information on observer performance. Results highlight differences in visual search patterns amongst novice, trainee and expert observers; the most marked differences occurred between novice readers and experts. Experts spent more time in challenging areas of interest (AOI) than novices and trainee, and were more confident unless a lesion was large and obvious. The time to first AOI fixation differed by size, shape and clarity of lesion. ‘Time to lesion’ dropped significantly when recognition appeared to occur between slices. The influence of clinical information was minimal.

Keywords: Observer performance, Image Perception, Stroke, Eye-Tracking, Expertise, Clinical Information, CT, MRI.

INTRODUCTION

1.1 Medical image perception

Radiographic images are produced to examine the human musculoskeletal system as well as organ and metabolic function and are frequently used to confirm or refute a queried diagnosis. Medical image acquisition, accuracy and interpretation have an important role to play in contributing to patient safety, yet diagnostic errors in the interpretation of medical images have been reported since the 1940s and abnormalities have been either missed, or over-read, with various rates across numerous experimental studies. Since the 1960’s, research in visual attention, perception and cognition had dominated chest (e.g. Thomas & Lansdown, 196323; Kundel, 197811, Gale et al., 19796; Krupinski, 199310, Manning, 200412) and breast screening (Mello-Thoms, 200215, Scott et al., 200722) in many countries, to identify the presence or absence of malignant or benign nodules.

Radiology is now an information rich specialty, which is almost completely digital producing varying 2D, 3D and 4D images with imaging procedures available today including ultrasound, Computed Tomography (CT), Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI) and more. Despite this, much of the actual research into factors affecting the interpretation of radiological imaging tends not to encompass such multidimensional imaging and largely neglects other important visual inspection tasks such as head and neck imaging.
1.2 Neuroradiological and cardiovascular diseases

Since the 1940’s new diseases and conditions have challenged health, medical research and healthcare provider practice. With an ageing population, neurodegenerative (e.g. Alzheimer’s’ Disease and Dementia) and cardiovascular diseases (Myocardial Infarction and Stroke) have become more prevalent, impacting heavily upon older adult well-being. Neurological conditions occur as a direct result of disease or damage to the nervous system (ABN, 1996)\(^1\) and are caused by a number of factors that can arise from within an individual (e.g. immunological, heredity and lifestyle factors) or as a result of an external influence (e.g. domestic violence or car accident etc). This group of disorders and their related symptoms can be due to the result of a ‘one-off’ clinical event (e.g. stroke), or a permanent, intermittent and/or degenerative disease such as Dementia, Parkinson’s, Alzheimer’s or Huntingdon’s diseases.

In 2005, The Department of Health reported that approximately 10 million people across the UK had suffered, or were suffering, from a neurological condition, accounting for 20% of acute hospital admissions, with 350,000 people requiring daily living assistance and over 850,000 people providing care for those affected (DoH, 2005)\(^5\). By 2006, the Department of Health reported that worldwide, cardiovascular disease was taking more lives than AIDS, TB and Malaria put together. In the United Kingdom, cardiovascular disease currently kills more people than all of the cancers combined, with Stroke affecting 150,000 people per annum (DoH, 2006)\(^4\).

1.3 Neuroradiology and eye-tracking

Considering the widespread and pervasive nature of the disease, advanced diagnostic imaging methods and sufficient expertise to identify lesions and abnormalities is vitally important. Both Computed Tomography (CT) and Magnetic Resonance (MR) imaging modalities are frequently used in the prevention, identification, diagnosis and treatment of people who are predisposed to, or have suffered an acute or chronic neurological deficit. Despite this, few studies have examined complex interpretative tasks such as the reading of CT or MR images, specifically in the area of neuroradiology. Anatomical changes and abnormalities characteristic of neurodegenerative and cardiovascular diseases tend to be highly variable between individuals and can present themselves quite differently (localised or generalised) in an imaging scan, posing different accuracy and interpretation problems for the observer. In addition, medical image acquisition and interpretation is generally used to confirm a queried diagnosis, rather than act as a preventive measure, and thus it has been reported these conditions had received less attention than treatable cancers by radiologists in the past (Weiner, M. 2008)\(^27\).

Whilst acute stroke has received attention from advanced imaging methods such as Diffusion-Weighted MRI (van Everdingen, 1998)\(^26\), visual search in examining stroke images has not been explored. Eye-tracking methodologies can be used to gain an objective insight into the underlying factors concerning the visual examination of such images that allow cognitive and perceptual implications about the observer’s behaviour to be drawn. Until very recently eye movement analysis across multidimensional images in radiology has not been reported. In 2005, Phillips et al. developed a software framework to track scan patterns across 3D, multi-slice images in differing modalities; including CT and MRI\(^11\). However, neither CT nor MR images have yet been explored to examine clinical conditions and observer performance using eye-tracking. These imaging modalities combine perfectly the investigation of complex digital imagery, with the examination of the accuracy and interpretation of neuroradiological images e.g. Stroke and will be explored in the current study.

1.4 The development of expertise in medical imaging and eye-tracking

In order to understand how neuroradiologists appraise complex medical imagery, it is important to investigate the development of expertise and how experts differ from novice readers and radiology trainees. Radiology studies using eye tracking have demonstrated novice and expert differences in location and duration of gaze, with research suggesting that novice readers scan more of an image and may not be able to generate enough ‘hypotheses’ from the visual imagery to make a clinical decision; whilst experts focus on specific clinically relevant areas of the scan, can generate hypotheses and make complex decisions based upon mental schema of salient image features (Garlatti & Sharples, 1998)\(^7\), drawing upon problem-solving skills and synthesizing biomedical knowledge to support clinical reasoning (Rogers, 1995)\(^25\).

Experts may fall short and miss critical abnormalities when they cease to recognise the limitations of medical image acquisition and/ or become satisfied with their primary visual search, a phenomenon known as ‘Satisfaction of Search’
(Berbaum, 1990). Eye-tracking research has demonstrated that lesions in chest radiographs are looked at, but not reported when satisfaction of search has occurred (Samuel et al, 1995). In addition, experts may become so engrossed in their respective specialism that they discount irregularities in the scan that do not fit with their prior experience, inadvertently alienating themselves from perceiving and engaging with new possibilities and learning new imaging techniques (Donovan, 2006). Nodine and Krupinski (1998) reported that much of what differentiates radiographers from lay people are their training, analytical mindset and their past experience with multiple images.

Nodine and Kundel (1990) reported the minimal dwell time for abnormality detection of lung nodules in chest radiographs to occur in clinical studies is 0.9s. In a series of studies by Manning et al. (2004), dwell time appeared to have implications for confidence in decision-making processes; missed lung nodules were actually dwelt on for an average time of 3.1s in chest radiographs, indicating that they were recognised but not interpreted cognitively, therefore, perception maybe the key to fundamental errors. Dwell time appeared to have links with observer expertise; length of dwell time appeared to be positively correlated with inexperience on nodule detection tasks. Subsequently, Manning (2005) found that correct negative decisions are made quickly and are positively associated with expertise (50% of their true negative decisions were made within 1000ms of gaze duration), and incorrect negative decisions were image features that are fixated on for much longer, especially with decreasing levels of expertise. Experienced radiologists ruled the presence of a possible nodule out within 2 seconds of foveal fixation. Whereas all naive decisions made after 3s were incorrect; once over the 4.75s fixation time point, no negative decisions (true or false) were attributed by radiologists.

Manning (2006) investigated novice, postgraduate and experienced film readers’ detection of chest nodules and found that statistically there was no difference between pre-training radiographers and novice readers, with the more novice the reader, then significantly more fixations per film were made compared to experienced readers. Fixations per film reduced significantly as training continued. Experienced radiographers not only examined the image in ‘longer, sweeping eye movements’ but they also focused in on a smaller number of areas or ‘zones’ for a more detailed inspection than did novice observers. Interestingly, as visual coverage and saccadic amplitude altered through experience, so did speed and accuracy: experts appeared to make better decisions in much less time. These findings overall demonstrate that training and caseload experience differentiate the ability of novice, postgraduate and experienced radiographers to be able to identify chest nodules. In addition, when the radiologists and postgraduate radiologists were grouped together, they performed significantly better than the less experienced groups.

1.5 The effect of clinical information on medical image appraisal

Medical images are rarely viewed in isolation from other influences that affect decision-making processes. The availability of clinical information is reported to have mixed results on diagnostic accuracy; some studies report increasing true positive ratings following presentation of clinical information (Schriever, 1963) whereas others report no significant increase (Good et al., 1990). Tudor et al. (1997) measured the performance of five consultant radiologists and although findings did not reach statistical significance, clinical information improved the accuracy of lesion detection from 77% to 80% in plain radiographs. Moreover, clinical information led observers to make false-positive diagnoses, which appeared consistent between observer ratings. Norman, Brooks, Coblenz and Babcock (1992) investigated the overall effects of brief clinical history on the diagnosis and feature detection of bronchilitis and uncovered that both experts and novice readers were influenced by clinical information, although novices made more false-positive reports than experts on this feature detection task. Hobby et al. (2000) proposed that image assessments are heavily influenced by radiological reports and owing to mixed findings and lack of evidence surrounding the effect of information on image appraisal, confirmed clinical diagnosis and treatment outcome for the patient (e.g. Donovan, 2006). Based on the mixed reported findings above, the influence of clinical information on image appraisal will also be assessed in the present study.

**OBJECTIVE**

From the foregoing, an initial investigation into the visual interpretation of multidimensional stroke medical images was undertaken in relation to observer performance. Observers with differing levels of radiological knowledge were presented with a series of CT and MRI brain images and their performance examined through examining their visual search behaviour. The difference between levels of expertise in relation to neuroradiological image appraisal was examined, including the influence of clinical information on image perception, interpretation, and diagnostic accuracy.
METHOD

2.1 Participants

Eight participants were recruited, namely: a) four novice participants, with no prior knowledge or experience of radiological images were recruited from staff and research students at Loughborough University; b) one second year radiology trainee based at the Norwich Radiology Academy, and c) three experienced readers employed by Norfolk and Norwich University Hospitals NHS Trust.

2.2 Design

A computer-based study was designed to assess diagnostic accuracy and interpretation in stroke CT and MR imagery. Eight predetermined clinical cases were selected from a bank of clinical cases; four acute stroke cases; i) ‘AST’: typical acute stroke, ii) ‘ASF’: acute stroke with fluent aphasia), iii) ‘ASA’: acute stroke (alexia without agraphia), and iv) ‘ASS’: acute stroke with speech arrest), two sub-acute stroke cases; i) ‘SSA’: subacute with aphasia, and ii) ‘SSS’: subacute with loss of sensation), one fatal chronic case ‘CSF’ and one normal control ‘NSN’ representing normal ageing. Five slices per case were selected and rated by an experienced radiologist to represent between 75 & 100% of the lesion, totalling forty images overall. For each case, key descriptive statements were extracted by the research team from the relevant radiology reports, which had already been condensed by expert radiologists to accompany the clinical cases; however, information only accompanied half the clinical cases and was withheld for the other half.

The independent variables of availability of clinical information (offered, withheld) and case severity (acute, subacute, chronic or normal aging control) were assessed in a within-participants design. The orders of presentation of the availability of clinical information and case severity conditions were counterbalanced within participants. The dependent variable considered in the present analysis was inspection strategy.

2.3 Procedure

Visual search behaviour of each participant was monitored using a Tobii X50 remote eye tracker, mounted below the computer monitor, which permits unobtrusive recording of saccadic eye movements. Images were viewed on an LG Flatron L2000, black LCD monitor (Display area: horizontal; 16.8”, vertical; 12.0”, diagonal; 20.1”). Each participant was first calibrated on the eye movement system. All participants were shown a short initial presentation concerning medical these types of images and the clinical features of stroke together with a short lesion identification training exercise. Each participant was then instructed to gaze at a white cross in the centre of the screen (to regulate the initial gaze point between participants) and then examine the eight clinical cases thoroughly and reach a decision regarding the presence or absence of an abnormality. Each participant saw the image slices for each case in the same order and could only scroll down through these slices. Participants were asked to rate each case on a five-point Likert scale, namely; 1) abnormality definitely present, 2) abnormality probably present, 3) Unsure, 4) abnormality probably absent, 5) abnormality definitely absent. If an abnormality was considered present, participants were required to confirm the location of the infarct on a separate brain atlas. After examination of all eight cases, participants were debriefed.

RESULTS

3.1 Image analysis results

Case study 1. Acute stroke: The following gaze-tracker images highlight the differences between readers’ (novice, trainee and expert) visual inspection strategies when appraising images of acute stroke (AST);
Figures 1-3, highlight the fixation patterns of novice (fig.1), trainee (fig.2) and expert (fig.3) readers. Figure 1 demonstrates how novice readers appraise normal, ventricular anatomy to the neglect of cortical tissue. In this image, the trainee reader perceived the AOI (located in the left patient hemisphere) with the 34th fixation and spends a total of 50 fixations on the image, yet the expert fixates upon the AOI with the first fixation from a total of 20 with a structured visual search pattern.

Case study 2. Subacute stroke. The following gaze-tracker images highlight the differences between readers’ (novice, trainee and expert) visual inspection strategies when appraising images of subacute stroke with aphasia (SSA).

Figure 4 highlights the novice unstructured search pattern over 28 fixations; the reader appears to ‘fixate’ upon the abnormality once detected and spends a large amount of time inspecting the left eye. Here the trainee fixates upon the image 17 times and ‘hits’ the lesion by the 12th, whereas the expert ‘hits’ the lesion by the second fixation. Although the expert fixates upon the image 20 times, a more thorough visual search pattern is apparent compared with figures 4 and 5.

Case study 3. Chronic stroke (CSF): The following gaze-tracker images highlight the differences between readers’ (novice, trainee and expert) visual inspection strategies when appraising images of chronic stroke.
This case study highlights interesting features of visual search whereby the expert concentrates upon affected tissue having ‘seen’ the abnormality upon the first fixation and attention then moves towards other affected tissue on the left hemisphere (fig.9). The trainee, as represented by figure 8, also spends more time appraising the affected hemisphere, yet the midline is not examined as thoroughly compared with the expert. The novice only spends a short amount of time in the primary affected area before rating its presence (fig.7).

3.2 Confidence rating data: quantitative analysis

The confidence rating data revealed that the novices detected 65% of abnormalities, the trainee 87% and the experts 96% as being broadly present or absent. As chronicity of case and size of lesions increased then abnormalities were more likely to be reported as present. Overall, participants were most confident about acute case 2 (ASF) and the chronic case, with all participants reporting true positive results. Participants were most unsure about the first acute case (AST, 3 false negative reports, 1 ‘unsure’ rating) and the control case (NSN; 1 false positive, 2 ‘unsure’ ratings). Expertise was significantly positively correlated with judgements on acute case ASA (p=.03) and negatively correlated with the control; NSN (p=.05). Novice participants reported the most ‘unsure’ and false negative responses.

3.3 Location data: brain atlas task

When the location data were examined, the accurate response rate fell by 24%, 23% and 20% respectively for each level of expertise. As the table below highlights, the tertiary lesion in the second subacute case (SSS) was most challenging to detect, followed by the secondary lesion in the same case, primary lesion in acute case, AST and secondary lesion in acute case, ASA. The primary lesion in the chronic case was the easiest to locate. Secondary and tertiary abnormalities in each of the three cases were either not cognitively processed or reported by some experts; this finding appears to imply there were simple processing errors or they reached a satisfaction of search (Berbaum, 1990). One naïve participant correctly located an abnormality on the brain atlas task, but did not report its presence, whereas most naïve participants appeared to examine and rate normal anatomy as abnormal i.e. ventricular anatomy.

<table>
<thead>
<tr>
<th>Case</th>
<th>AST</th>
<th>ASF</th>
<th>ASA</th>
<th>ASS</th>
<th>SSA</th>
<th>SSS</th>
<th>CSF</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesion number</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Expert</td>
<td>67</td>
<td>100</td>
<td>100</td>
<td>67</td>
<td>67</td>
<td>100</td>
<td>100</td>
<td>33</td>
</tr>
<tr>
<td>Trainee</td>
<td>0</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Novice</td>
<td>0</td>
<td>75</td>
<td>75</td>
<td>0</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>0</td>
</tr>
<tr>
<td>Lesion detection</td>
<td>25</td>
<td>85.5</td>
<td>62.5</td>
<td>25</td>
<td>75</td>
<td>87.5</td>
<td>75</td>
<td>25</td>
</tr>
</tbody>
</table>

Table 1. Correctly reported and located lesions (in percentages), by case and participant group.
3.4 Diagnostic accuracy and clinical information

When the trainees’ results were combined with the experts and examined with a 2-way ANOVA, there was a highly significant difference between novices and experts on accuracy tests, with a large effect size of experience ($df_1 = 1, p = .007, F = 16.2, \text{Partial Eta Squared} .73$). Unsurprisingly, experts were the optimal performers on both tasks, yet errors were still made on the location task, in particular the subacute stroke cases.

Overall, the presence of clinical information did not improve performance, either within-subjects or between them; the within-subject effect was small-moderate (Partial Eta Squared .043) but did not reach statistical significance ($p = .620, F = .273$). There was a large effect when we examined the interaction of case severity, expertise and information processing on accuracy, which was approaching significance ($df_1 = 1, p = .071, F = 4.8, \text{Partial Eta Squared} .44$).

3.5 Eye-movement data: quantitative analysis.

The following figures represent the mean time (ms) for each participant group to fixate the Area of Interest (AOI) around each primary lesion across each image ‘slice’ for each of the cases. The expert group is represented by a black line/squares; the trainee individual by a grey line/triangles and the novice group by a light grey line/circles.

Figures 7-14 represent mean time (ms) to primary lesion across each image by participant group for AST-NSN cases. N.B. Trainee slice 4 was excluded (29.3ms) for ease of graph interpretation for case ASA.
Experts were more consistent both within and between cases than novices and trainees; experts operated within a smaller time range than novice participants in every case. Interestingly, novices took twice the amount of time to fixate upon a lesion in half of the cases than did the experts. For AST & SSA cases, the time to fixate the primary lesion was significantly more for novices than experts (t-test AST: \( p = 0.04 \), SSA: \( p = 0.02 \)). The trainee was most inconsistent and the range was largest across four of the seven cases, trainee scores were excluded from t-test analysis.

There appeared a sharp ‘drop-off in time to lesion, at which point recognition may occur. Experts appear to detect the lesion quickly in the first slice, whereas recognition appears to occur in the second or third slice for novices and trainees. Time to primary lesion across slices appeared to exhibit an inverted ‘U’ effect for experienced observers who detected the lesion, followed its size and shape through the image scans, and then examined the final slice for additional affected areas before returning to the lesion and rating its presence.

### 3.6 Number of eye fixations in area of interest (AOI).

<table>
<thead>
<tr>
<th>Case</th>
<th>AST</th>
<th>ASF</th>
<th>ASA</th>
<th>ASS</th>
<th>SSA</th>
<th>SSS</th>
<th>CSF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesion number</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Expert</td>
<td>AOI</td>
<td>12.4</td>
<td>3.2</td>
<td>1.9</td>
<td>1.3</td>
<td>5.1</td>
<td>9.5</td>
</tr>
<tr>
<td>Total</td>
<td>21.1</td>
<td>9.1</td>
<td>19.4</td>
<td>-</td>
<td>9.3</td>
<td>15.2</td>
<td>20.1</td>
</tr>
<tr>
<td>AOI%</td>
<td>58.8</td>
<td>35.2</td>
<td>16.5</td>
<td>-</td>
<td>54.8</td>
<td>62.5</td>
<td>46.3</td>
</tr>
<tr>
<td>Trainee</td>
<td>AOI</td>
<td>11.8</td>
<td>3.4</td>
<td>2.4</td>
<td>3.2</td>
<td>0.0</td>
<td>1.6</td>
</tr>
<tr>
<td>Total</td>
<td>71.4</td>
<td>13.4</td>
<td>55.4</td>
<td>-</td>
<td>7.0</td>
<td>5.8</td>
<td>17.6</td>
</tr>
<tr>
<td>AOI%</td>
<td>16.5</td>
<td>25.4</td>
<td>10.1</td>
<td>-</td>
<td>0.0</td>
<td>27.6</td>
<td>31.8</td>
</tr>
<tr>
<td>Novice</td>
<td>AOI</td>
<td>3.9</td>
<td>3.0</td>
<td>1.2</td>
<td>0.9</td>
<td>3.6</td>
<td>2.9</td>
</tr>
<tr>
<td>Total</td>
<td>16.2</td>
<td>6.7</td>
<td>13.6</td>
<td>-</td>
<td>12.5</td>
<td>8.8</td>
<td>14.8</td>
</tr>
<tr>
<td>AOI%</td>
<td>24.1</td>
<td>44.8</td>
<td>15.4</td>
<td>-</td>
<td>28.8</td>
<td>33.0</td>
<td>52.7</td>
</tr>
</tbody>
</table>

Table 2. The mean number of fixations within an AOI, mean total fixations per case and percentage of fixations in the AOI.
Experts fixated more in the AOI than trainees and novices, especially in cases AST and SSA which proved highly statistically significant (t-test, \( p=.00 \)). However, this trend was reversed where the lesion was large and obvious; in these cases novices remain fixated on the primary lesion site to the neglect of additional affected tissue i.e. secondary and tertiary abnormalities. There was a trend towards the trainee fixating upon secondary lesions more than experts and novices, which may indicate a degree of uncertainty.

In terms of percentage data, experts spent more time in the AOI’s than novices and trainees, with the exception of cases ASF, SSS and CSF. In the latter two cases, the trend of novices fixating on the primary lesion was again observed.

### 3.7 Time in AOI versus total slice time.

On average, experts spent more time fixating within the AOI than trainees and novices, especially in cases ASA and ASS, which proved significant (\( p=.03 \& .05 \) respectively) and AST and SSA cases, which both proved highly statistically significant (t-test, \( p=.00 \)). However, this was reversed where the case lesion was large and obvious e.g. novices remain fixated on the primary to the neglect of additional affected tissue such as secondary and tertiary lesions.

<table>
<thead>
<tr>
<th>Case</th>
<th>AST</th>
<th>ASF</th>
<th>ASA</th>
<th>ASS</th>
<th>SSA</th>
<th>SSS</th>
<th>CSF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesion number</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Expert</td>
<td>AOI</td>
<td>3.7</td>
<td>1.1</td>
<td>6.5</td>
<td>0.6</td>
<td>1.4</td>
<td>2.6</td>
</tr>
<tr>
<td>Total</td>
<td>7.3</td>
<td>3.0</td>
<td>4.7</td>
<td>3.4</td>
<td>6.3</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>AOI%</td>
<td>50.8</td>
<td>35.3</td>
<td>16.4</td>
<td>41.9</td>
<td>55.7</td>
<td>34.9</td>
<td>55.7</td>
</tr>
<tr>
<td>Trainee</td>
<td>AOI</td>
<td>1.0</td>
<td>0.3</td>
<td>0.2</td>
<td>0.4</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>Total</td>
<td>15.6</td>
<td>3.6</td>
<td>16.3</td>
<td>3.6</td>
<td>5.4</td>
<td>2.8</td>
<td>-</td>
</tr>
<tr>
<td>AOI%</td>
<td>6.7</td>
<td>8.4</td>
<td>4.1</td>
<td>0.0</td>
<td>2.1</td>
<td>19.2</td>
<td>-</td>
</tr>
<tr>
<td>Novice</td>
<td>AOI</td>
<td>0.8</td>
<td>0.7</td>
<td>0.1</td>
<td>0.7</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>Total</td>
<td>6.7</td>
<td>2.6</td>
<td>5.5</td>
<td>5.2</td>
<td>4.0</td>
<td>4.9</td>
<td>-</td>
</tr>
<tr>
<td>AOI%</td>
<td>11.5</td>
<td>26.7</td>
<td>5.0</td>
<td>13.2</td>
<td>13.9</td>
<td>33.5</td>
<td>62.0</td>
</tr>
</tbody>
</table>

Table 3. The mean time (seconds) spent fixating within an AOI, mean total fixations per case and percentage of time in AOI.

In terms of percentage data, experts spent more time in the AOI’s (versus time on the slice itself) than novices and trainees, with the exception of case CSF. In the latter case, the trend of fixating on the primary lesion was again observed. When total time was considered, the trainee spent the most time appraising each case; 34.8 seconds, on average, per case, followed by experts (25.3 seconds) and novices (24.8 seconds).

### 3.8 Eye-movements and reported abnormalities

Two experts rated the first acute case (AST) as a true positive and time to first fixation (on the first image) was within 0.7s, correct novice readers took up to 3s. Incorrect (false negative) trainee and expert false negative ratings were both over 2.6s. All rated the second acute case, ASF, a true positive with experts and trainee fixating on the first AOI within 1.5s of seeing the case, novices were within 2.3s to foveal fixation. The trainee and experts rated the third acute case, ASA, a true positive and all fixated upon both AOI’s within 5s. Correct novices took up to 10s to ‘see’ both AOI’s. Correct ASA decisions were accompanied by fixations to the AOI within 2.2s, novices were within 4.2s. Fixation time to subacute (SSA) lesion was 0.8 for experts, 1.5 for trainee and up to 3.5 for novices.

Experts took up to 10s to spot all three AOI’s in subacute case, SSS, whilst 3 of 4 novices reported this case was a true positive, not all lesions were identified on the atlas task and times cannot be accurately compared with expert data. Novice readers fixated upon the true positive, primary chronic lesion within 0.9s, yet the secondary abnormality was barely fixated upon and not reported. Experts and trainee also perceived the abnormality within 0.9 seconds and took up to 2.8s more to appraise the secondary. Overall, it appears that novices took twice as much time to reach an AOI than expert, unless the lesion was large and/or unambiguous. For the control case, participants who reported a lesion was not present (true negative), spent on average 5s per image and 26s per case. Participants who reported they were ‘unsure’ or rated the case as ‘positive’, spent much less time appraising the image (1.9s & 1018ms) and case (9.3s & 5.1s) than those previous.
Overall, experts detected a primary lesion within the first four fixations on an image (mean fixation time to lesion: 1.5s & average dwell time: 240ms) a secondary within 13 fixations (mean fixation time to lesion: 4.1s & average dwell time: 261ms), and tertiary lesions within the first 18 (mean fixation time to lesion: 4.7s & average dwell time: 146ms). Conversely, novices perceived a primary lesion within the first seven fixations on an image (mean fixation time to lesion: 1.9ms & average dwell time: 324ms). In addition, foveal fixations in an AOI were twice more likely to be recorded on the second slice than the first. Only one novice reported a secondary lesion on the chronic case and this was by the second fixation, on the third slice, within 0.7 seconds. Collectively, experts ‘saw’ but did not report the location of 8 (out of a possible 36 across individual experiments) lesions and fixated upon these areas for an average of 219ms. Where recognition did occur and was reported, the first fixation within an AOI was slightly higher at 242ms.

DISCUSSION

4.1 Image Analysis and Eye Movements.

As supported by the image analysis and eye-movements results, it was apparent that experts employed a strategic visual search pattern when identifying abnormalities in these stroke medical images; they also spent more time looking in the area of interest surrounding a lesion than novices and trainees, unless the abnormality was large and unambiguous. Experts detected an area of interest as being abnormal, directed their attention and quickly moved on to appraise other affected areas of tissue, the midline and anatomical symmetry, as also reported by Manning (2006) in chest radiograph images (2006)14. Thus, experts appeared to operate a system of deduction; ruling out certain areas very quickly to economise effort and enabling visual attention to be redirected to more interesting clinical features.

Novices on the other hand, spent more time visually examining normal anatomy such as ventricles and/ or primary lesions following recognition. These findings appear to support previous research by Rogers (1995)25 and Garlatti and Sharples (1998)7 that novices may not generate enough clinical hypotheses to consider complex decisions owing to a lack of biomedical knowledge and clinical problem-solving. The radiology trainee appeared to be less structured than the experts and spent more time on secondary abnormalities than novices, which may indicate an uncertainty to make a decision, also owing to a lack of caseload experience. These novice and trainee findings regarding the influence of caseload experience also appear to be inline with previous research by Nodine and Krupinski (1998)17.

The differences in visual scanning between novice and experts may not only be influenced by direct experience and repeated exposure to clinical images. It may be that experts have developed complex 'cognitive maps' about anatomical features and may hold a bank of acquired information surrounding disease manifestations and progression. For instance, different types of ‘bleed’ in a stroke attack pervade different vascular routes within the brain due to known cerebrovascular pathways. Experts should have preconceptions, not only about anatomy, but how and where another stroke might develop, what patient abilities might have become impaired post-infarct, and how to identify current and future problems from image data.

4.2 Diagnostic Accuracy, Location Data and Clinical Information Processing.

It is unsurprising that experts were the optimal performers on decision-making and the ‘brain atlas’ location tasks. Experts were more confident in their decisions to refute the presence of abnormalities in the control case and more able to detect challenging abnormalities in acute cases than novices and trainees. Novices also made more false positive decisions and these findings are in agreement with much previous research. An interesting finding emerged as one novice participant located a lesion on the brain atlas task, however reported that the abnormality was unlikely to be present. This case appears to demonstrate that perception and cognition have occurred yet a lack of confidence meant the abnormality was not reported.

A few errors were made by experts on the lesion detection tasks, with secondary and tertiary abnormalities in scans being missed or not reported. This finding could potentially be linked with the satisfaction of search phenomenon (Berbaum, 1990)2, and will be explored in more detail in future studies with larger numbers of patient cases. In addition, there was a trend towards clinical information affecting decision-making although this was not statistically significant and supporting research by Good (1990)8 and Tudor (1997)24 the effect of clinical information and will be explored in future studies with improved statistical power.
4.3 Expertise and Eye-movements.

Overall, experts were quicker to detect the primary lesion, and with exception of the challenging ASA case, foveal fixation occurred within 2s on the first slice of each case image ‘stack’. For this acute case, time to lesion dropped to 2.7ms on the second slice from 3.9ms, and to 1.2ms following appraisal of the third. Experts were more consistent and operated within a smaller time range than novices and trainee. It was interesting to evaluate the ‘shoulder effect’ that occurred through the five image slices and between participants, in terms of time to lesion dropping and appeared to level out after recognition of the lesion may be inferred. Such observed trends will be examined in more detail within and across different types of stroke cases in future studies.

The number of foveal fixations in a predefined area of interest increased with level of expertise. Experts also had more fixations per case than novices and the trainee (see table 2), however, the total number of fixations per slice was variable between cases, owing to differences in lesion type, size and number and therefore, a clear conclusion regarding an association between the number of fixations per image and level of experience could not be drawn in this study. Novices spent less time in an AOI, per image and less time overall than trainee and experts. The trainee spent the most time appraising all patient cases, followed by the experts, and finally novices.

True positive decisions are marked by a quick time to lesion in the earliest image, with experts being twice as fast as novice readers until AOI’s become larger and less ambiguous. Multiple lesions take more time to see and be processed cognitively processed. If the image is not fully appraised or the reader is inexperienced, they may be missed altogether or go unrecognised. True negative decisions appear to be associated with a thorough image appraisal, confidence and expertise.

This was an exploratory study investigating aspects of neuroradiology, particularly concerning taking a visual search perspective to observer performance. In doing so it is recognised that certain limitations apply; namely only a small number of experienced and naïve observers were studied as they examined a set of carefully selected images. In normal clinical practice radiologists would scroll up and down image stacks for each case which was not permitted here for experimental reasons.

CONCLUSIONS

In this initial study of observer performance in examining CT and MR stroke images significant differences in visual search behaviour and performance were found. This included differences in search patterns and image coverage between novice, trainee, and expert observers including differences in foveal fixations, dwell and overall case time in the detection and interpretation of these exemplar stroke images. In addition, the study examines observer behaviour through multiple images per patient case rather than appraising single images. To-date few studies have explored such imaging and our ongoing research is further examining neuroradiological imaging interpretation performance.

ACKNOWLEDGEMENTS

CT/ MRI images reproduced with permission from The Whole Brain Atlas; http://www.med.harvard.edu/AANLIB/home.html

REFERENCES

[5] Donovan, T., and Manning, D.J. “Successful reporting by non-medical practitioners such as radiographers, will always be task-specific and limited in scope,” Radiography, 12, 7-12 (2006).