The assessment of stroke multidimensional CT and MR imaging using eye movement analysis: does modality preference enhance observer performance?

The assessment of stroke multidimensional CT and MR imaging using eye movement analysis: does modality preference enhance observer performance?

Lindsey Cooper[*a], Alastair Gale[a], Janak Saada[c,] Swamy Gedela[c], Hazel Scott[a] & Andoni Toms[b]

[a]Applied Vision Research Centre, Loughborough University, UK
[b]Norwich Radiology Academy, Norwich, UK
[c]Norfolk and Norwich University Hospitals NHS Trust, Norwich, UK

ABSTRACT

Although CT and MR imaging is now commonplace in the radiology department, few studies have examined complex interpretative tasks such as the reading of multidimensional brain CT or MRI scans from the observer performance perspective, especially with reference to Stroke. Modality performance studies have demonstrated a similar sensitivity of less than 50% for both conventional modalities, with neither modality proving superior to the other in Stroke observer performance tasks (Mohr, 1995; Lansberg, 2000; Wintermark, 2007). Visual search studies have not extensively explored stroke imaging and an in-depth, comparative eye-movement study between CT and MRI has not yet been conducted. A computer-based, eye-tracking study was designed to assess diagnostic accuracy and interpretation in stroke CT and MR imagery. Forty eight predetermined clinical cases, with five images per case, were presented to participants (novices, trainees and radiologists; n=28). The presence or absence of abnormalities was rated on a four-point Likert scale and their locations reported. Results highlight differences in visual search patterns amongst novice, trainee and expert observers; the most marked differences occurred between novice readers and experts. In terms of modality differences; novice and expert readers spent longer appraising CT images than MR, compared with trainees, who spent longer appraising MR than CT images. Image analysis trends did not appear to differ between modalities, but time spent within clinical images, accuracy and relative confidence performing the task did differ between CT and MR reader groups. To-date few studies have explored observer performance in neuroradiology and the present study examines multi-slice image appraisal by comparing matched pairs of CT and MRI Stroke cases.

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

INTRODUCTION

1.1 Cardiovascular disease and Stroke

Worldwide, cardiovascular disease (CVD) is a major health concern. By 2006, cardiovascular disease was reported to have claimed more lives than AIDS, TB and Malaria combined (DoH, 2006). In the UK alone, recent figures state that cardiovascular disease currently kills more people than all cancers combined, with Stroke affecting 150,000 people per annum (DoH, 2006). When considered separately from other CVDs, Stroke is the third leading cause of death in the US and UK, and the largest single cause of severe disability. Severe strokes leave 20% of individuals institutionalized and permanently disabled in a further 15-30%. Some 50-70% of stroke sufferers recover and can lead relatively normal lives, dependent upon the stroke severity and the area of the brain affected (Asplund, Stegmayr & Peltonen, 1998).

In the US, Stroke affects 700,000 people per annum and in 2007 the nationwide cost of Stroke had risen to $62.7 billion (American Heart Association, 2007). Timely and accurate diagnosis is imperative when providing an efficient and

---

[*] Lindsey Cooper: E-mail: L.H.K.Cooper@lboro.ac.uk www.appliedvision.org

efficacious radiology service for all conditions. Due to the adverse nature of patient outcomes and affiliated direct and indirect costs associated with Stroke, a well run service is particularly important to provide quick identification, feedback to physician and treatment of Stroke, which may enable a reduction in the potential impact a cerebral infarction can have on an individual (Gonzalez, 1999; Mullins, 2002; Kloska 2004).

1.2 Modality Research CT versus MR

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. The imaging of neuroradiological conditions has received much attention from radiologists to compare rates of sensitivity and specificity between modalities. In a 1986 study by Haughton *et al*, diagnostic accuracy in neuroradiology between CT and MR modalities was examined. Whilst MR imaging was considered more sensitive for spinal region abnormalities (88% MR sensitivity compared with 85% for CT), CT was considered more sensitive overall with 91% sensitivity rating and 82% for MRI. At that time, specificity was also perceived to be higher for CT (90%) than MR (81%) when 100 cases were viewed by two neuroradiologists and one research fellow.

When Stroke was the only condition being considered, mixed results were yielded. Gonzalez in 1999 compared rates of 18% sensitivity and 100% specificity for MR imaging and 45% sensitivity and 100% specificity for CT when 14 stroke patients were imaged within 6 hours of stroke-like symptoms. In a much larger review of 563 CT and 498 MR patients, Mullins *et al* (2002) uncovered that MR sensitivity had been underrated in the Gonzalez study by 40%, yet specificity remained constant at 100%. These studies and others have demonstrated a similar sensitivity of less than 50% for both conventional modalities, with neither modality proving superior to the other from an observer performance perspective (Mohr, 1995; Lansberg, 2000; Wintermark, 2007).

Recently, studies have reviewed the efficacy of Diffusion-Weighted MR Imaging (DWI) and found it to be superior to conventional CT and MR, owing to superior lesion contrast and its ability to capture the diffusion of water molecules indicative of early physiological shifts (Gonzalez *et al.*, 1999; Lansberg, 2000). Gonzalez *et al.* reported 100% sensitivity and 86% specificity for DWI, yet with a larger sample size, Mullins *et al* (2002) reported a reverse trend of 97% sensitivity and 100% specificity. Mullins also demonstrated that accuracy depended largely upon when the cerebral infarction had occurred and time elapsed from admission to the emergency department. Following 12 hours of infarction onset, DWI accuracy was considered equivalent to CT. More recently, Perfusion CT has been shown to enhance conventional CT from 45% sensitivity to 76.3% and uncover lacunar infarctions previously missed by unenhanced CT readers, but still remains less sensitive than DWI (Kloska *et al.*, 2004). Despite this imperative research, visual search in neuroradiological images has not been extensively explored, highlighting a paucity of research from a human observer perspective in this clinical field.

1.3 Human-Computer interaction in Neuroradiology

Modern imaging methods can now produce hundreds of images, viewed in multiple sequences, offering free scrolling between sagittal, axial, and coronal orientations, even diagonal planes for the reader. As 3D and 4D digital imagery is now commonplace in the clinical setting, accompanied by increasingly flexible viewing software, research must endeavor to examine how multidimensional and multimodal imaging affects accuracy and efficiency in radiology. Each stride in technology impacts heavily upon the amount of human-computer interaction required from radiologists and resultant clinical performance (Phillips *et al.,* 2008). Eye-tracking can be used to gain an objective insight into human-computer interaction and the underlying factors involved in the visual examination of such images. Visual search studies in medical image perception allow conclusions to be drawn about the observers' search behaviour and underlying cognitive processing. Visual search studies in medical imaging have been published since the 1960's (Thomas & Lansdown,1963; Kundel, 1978; Nodine & Kundel, 1987), yet only recently have experimental studies been designed to track eye movements across a number of related medical images in radiology (e.g. Phillips *et al.,* 2005 & 2008; Cooper *et al*, 2009).

1.4 Factors affecting medical image perception

Factors that influence the ability of the observer to efficiently examine and interpret patient images are not only dependent upon image quality and modality, but are also dependent upon the observers' level of expertise accompanied

by cognitive aptitude and caseload experience (Manning *et al.*, 2005). Novice readers, with little or no experience in reading medical images, will undoubtedly differ from expert readers with accredited training and often decades of reading experience. The distinction between trainee performance and novice and expert readers may be less clear. Researchers have dedicated much time to understanding how experts perform their duties with clarity under time pressure and constant patient demand. Whilst consultants are considered the optimum performers in clinical radiology, and the experts in this study, human error has been reliably well documented over the past five decades, specifically in the areas of chest and breast imaging at various rates for misreading or omitting important clinical features (e.g. Thomas & Lansdown, 1963; Kundel, 1978; Gale et al., 1979; Krupinski, 1993; Manning, 2004 & Mello-Thoms, 2002).

1.5 Visual search and observer performance

One of the well established ways of assessing observer performance across differing abnormalities is through tracking visual search behaviour (Kundel, 1978; Mello-Thoms, 2002). Visual search studies in radiology have shown consistent significant differences between novice and expert readers irrespective of the image interpretation task (Donovan, 2006; Cooper *et al.*, 2009). Experts are said to generate hypotheses and make complex decisions based upon mental schema of salient image features (Garlatti & Sharples, 1998), drawing upon problem-solving skills and synthesizing biomedical knowledge to support clinical reasoning (Rogers, 1995). Whilst it is expected expert readers are the optimum performers, studies demonstrate subtle differences in visual search between the transition from novice to trainee and through to expert reader. Studies of radiographers highlighted no significant difference between pre-training radiographers and novice readers (Manning, 2006), but once training was underway or completed, Nodine and Krupinski (1998) suggested that a medically analytical mindset (coupled with experience of viewing multiple images), differentiated these radiographers from lay people.

Visual search behaviour provides clues to cognitive and perceptual processes involved in medical decision-making. Eye movements of expert readers demonstrate "longer, sweeping eye movements", fewer fixations per image and clustering of fixations in clinically relevant areas of the image, when compared with novice readers. In 2006, Manning confirmed that as experience increased so did speed and accuracy; marked by alterations in visual coverage and saccadic amplitude when searching for chest nodules. An earlier study by Manning (2004) highlighted that dwell time in particular appeared to be positively linked with inexperience and lack of 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. In addition true negative decisions were made quickly with 50% of decisions being made within 1000ms of gaze duration, and were positively correlated with expertise. Incorrect decisions are gazed at for much longer (e.g. >3 seconds) indicating cognitive dissonance surrounding the task (Manning, 2005).

OBJECTIVES

Stroke presentation has received attention from researchers to compare modalities in terms of sensitivity and specificity to meet patient requirements (Gonzalez, 1999; Mullins, 2002). Most studies appear to agree that CT and MR offer similar sensitivity and specificity ratings (Mohr, 1995; Lansberg, 2000; Wintermark, 2007). To the authors' knowledge, an in-depth examination of visual search and observer performance analysis between CT and MR multidimensional images in neuroradiology has not yet been reported. This study aims to report any changes in visual search or behaviour through analysis of the following; eye movements, location data with accompanying confidence rating for primary abnormality (if at all), diagnostic accuracy for each stroke case type and modality, and conclude by drawing comparisons between the two modalities. To summarise, the objectives of this paper are as follows;

- Explore observer performance between participant groups when viewing conventional CT and MR images, multi-slice imaging of Stroke.
- Explore visual search behaviour between participant groups when viewing conventional CT and MR images.
- Assess visual search behaviour within and between 'stack' images of CT and MR Stroke cases.
- Re-examine pilot findings, with a larger sample size (n=28) and an increase in patient cases from 8 to 48.

METHOD

2.1 Participants

Twenty eight participants were recruited to the study, namely: a) Ten novice participants, with little or no prior knowledge or experience of radiological image interpretation were recruited from staff and research students at Loughborough University; b) Ten Specialist Registrars with differing levels of experience (years 1-4 radiology trainees) and were based at the Norwich Radiology Academy, and c) Eight Consultant Radiologists with varying years of experience and were 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. Forty-eight predetermined clinical cases were selected from a bank of clinical cases; eight acute, six subacute cases, four chronic cases and six normal controls were matched for each modality. Five axial slices per case were selected and independently rated by two Consultant Neuroradiologists, totalling 240 anonymised images overall. The independent variables of modality (CT, MRI) and case severity (acute, subacute, chronic or normal aging control) were assessed in a within and between participant design. The orders of presentation of the independent variables were counterbalanced within and between participants to control for order effects and the first Area of Interest either appeared on the first or second slice of the case. The dependent variable considered in the present analysis was inspection strategy, reported confidence and diagnostic accuracy.

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 consistent for size and viewed on an LG Flatron L2000, black LCD monitor (Display area: horizontal; 16.8", vertical; 12.0", diagonal; 20.1"). Participants' eye movements were calibrated on a 5-point chart. A short presentation was viewed by all, per modality, to give a basic training on the clinical features of Stroke, as presented by CT and MR imagery, together with a short lesion identification training exercise. Each participant was then instructed to gaze at a fixation point in the centre of the screen (to regulate the initial gaze point between participants) prior to image viewing. Patient axial slices were presented in the same order, and therefore, participants could only scroll down through the 'stack'. Participants were not confined to a time limit by slice or case but were asked to appraise each case thoroughly and reach a decision regarding the presence or absence of an abnormality. Participants rated each case on a four-point Likert scale, namely whether a primary abnormality (i.e. Stroke) was; 1) definitely present, 2) probably present, 3) probably absent, or 4) definitely absent. If an abnormality was considered present, participants were required to confirm the location of the infarct on a separate brain atlas task. After examination of all 48 cases, all participants were informed of the aims and objectives of the study.

RESULTS

3.1 Image analysis results: visual search behaviour between groups and modality

Case study 1. CT Acute stroke: The following gaze-tracker images highlight the differences between readers' (a novice, a trainee and an expert) visual inspection strategies when appraising images of Acute CT Stroke images.

Figure 1	Figure 2	Figure 3

Figures 1-3, highlight the fixation patterns of a novice (fig.1), a trainee (fig.2) and an expert (fig.3) gaze patterns. Figure 1 demonstrates how a novice reader appraised the normal, ventricular anatomy to the neglect of cortical tissue. In the second image, the trainee reader perceived the AOI (located in the left patient hemisphere) with the 4th fixation and spends a total of 28 fixations on the image. The expert fixates upon the AOI with the second fixation from a total of 5, with an apparently efficient visual search pattern accompanied by a true positive rating.

Case study 2. MR Acute stroke: The following gaze-tracker images highlight the differences between readers' (a novice, a trainee and an expert) visual inspection strategies when appraising images of Acute MR Stroke images.
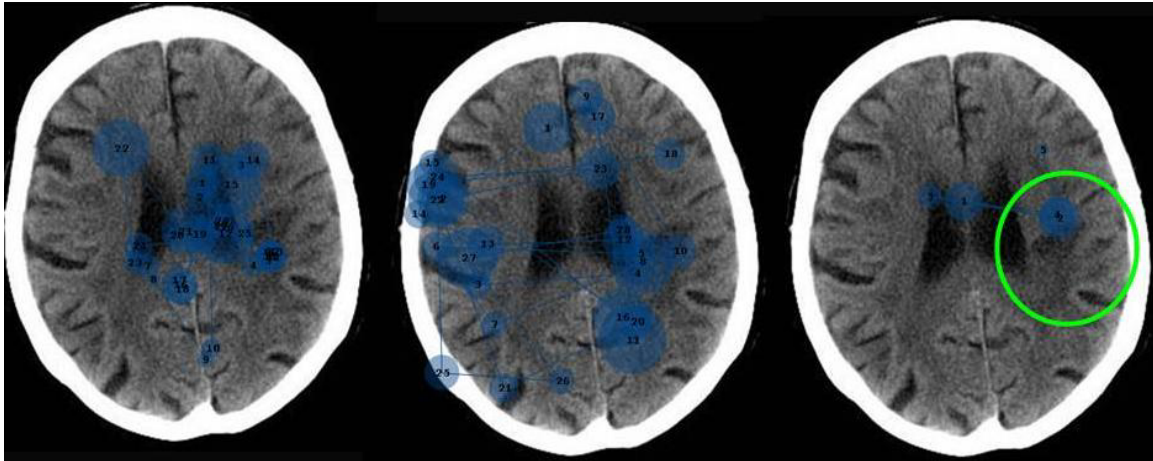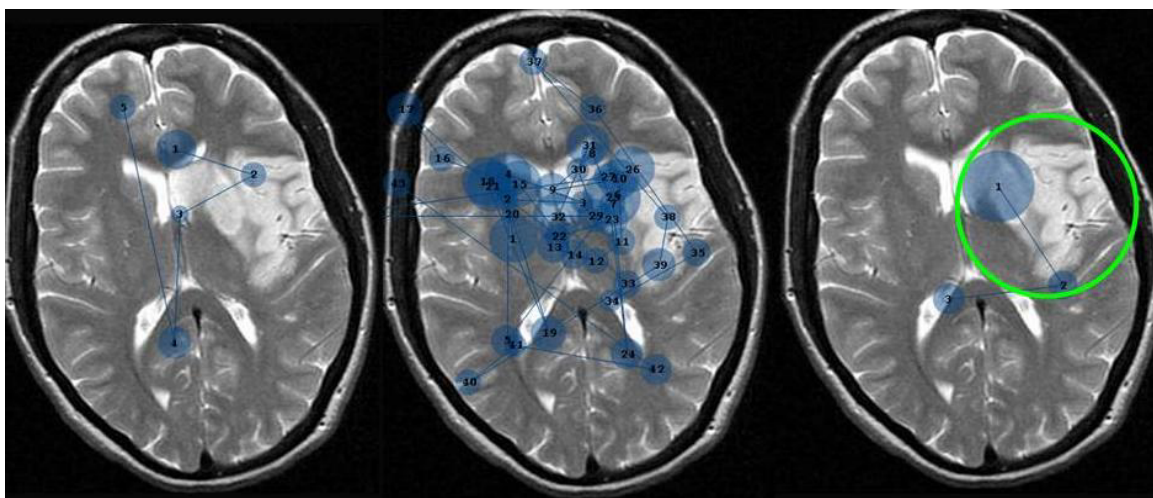

Figure 4	Figure 5	Figure 6

Figure 4 highlights a quick time to 'hit' over 5 fixations; this readers' eye movements are somewhat similar to those of the expert in figure 6, yet novices are known to be unaware of the potentiality of secondary problems and small vessel changes, which may influence patient status. The trainee spends further time within the slice, which appears indicative of a thorough appraisal of peripheral sulci, cortical tissue and ventricles. The trainee fixates upon the abnormality, as indicated by the third fixation, and a confident rating, but s/he also examines the surrounding tissue, unlike the novice. The expert makes few eye saccadic movements and a quick, confident and accurate decision.

Collectively, Figures 1-6 demonstrate the differing visual search behaviour between these exemplar novice, trainee and expert observers and by modality. For instance, although the infarct in the MR patient is slightly larger than the CT

patient, the CT novice reader makes more fixations around the ventricular area than the MR novice reader, who fixates within the AOI immediately denoting quick identification. The trainee readers demonstrate similar search patterns between CT and MR cases; detailed inspection including much cross comparison is evident on both, however there are an increased number of fixations on the MR (43 fixations) image compared with the CT image (28 fixations). This finding may be indicative of uncertainty when searching for the possibility of secondary abnormalities such as small vessel disease or lacunar infarcts, as the primary abnormality (Stroke) appears much less challenging for nearly all participants. Expert eye movements demonstrate consistency between readers and modality and a quick time to fixate the infarct.

3.2 Accuracy and Confidence rating data: quantitative analysis of observer performance

When considering *diagnostic accuracy* in CT, novice participants correctly reported the absence of abnormal locations in 70.83%, trainees 81.25% and experts 90.1% of all patient cases. In MRI, novice participants correctly reported the absence of abnormal locations in 63.8%, trainees 85.4% and experts 96.4% of patient cases. This indicated that novice accuracy was better in CT, but trainee and expert performance was better in the MR condition. When ANOVA tests were performed, there were significant group differences for diagnostic accuracy in both MR (*df* 2, p<.000, F=44.07, Eta Squared=.12) and CT (*df* 2, p<.000, F=13.77, Eta Squared=.04). MR results also demonstrate a large effect size compared with CT. Tukey posthoc comparisons demonstrated that the differences were most evident between expert and novice participant groups for CT (p<.000) and MR (p<.000). Comparisons between trainee and expert readers were significant to the (p<.007) level in MR and also significant (p<.036) in CT. Similar results were found for novice and trainee reader results, which were significant in both MR (p<.000) and CT (p<.010). When reporting participant *confidence* in their decision-making in CT, descriptive statistics show overall percentage of cases correct indicates that acute cases were perceived to be most difficult by all participants (79.1%), followed by controls (79.2%), chronic (93.1%) and then subacute stroke types (96.4%). In MRI, overall percentage of cases correct indicates that control cases were perceived to be most difficult by all participants (78.5%), followed by subacute (80.3%), acute (81%) and then chronic stroke types (91%). Therefore, MR control cases may be considered to be the most difficult cases when all participant results were combined shortly followed by acute and control CT cases.

Confidence scores (%) by Stroke type	Modality	Novice	Trainee	Expert
Normal Control Score	CT	48.8	52.5	63
	MR	51.7	60.4	66.1
Acute	CT	70.6	83.1	92.2
	MR	81.3	95.9	99.6
Subacute	CT	87.5	98.3	98.4
	MR	75.0	91.7	96.4
Chronic	CT	83.3	95	98.4
	MR	90.6	93.1	99.2
Average Confidence Score	**CT**	**73.3%**	**81%**	**81.5%**
	MR	**73.8%**	**80.1%**	**82.3%**

Table 1 represents confidence scores (in percentages) by participant group and stroke type.

Novice descriptive statistics show a higher confidence when rating chronic cases in MR. Trainees show higher confidence levels when rating subacute cases in CT and experts show a higher confidence when rating acute cases in MR. All participants show lower confidence levels when ruling out abnormalities in CT cases. There were significant differences between participant groups for confidence in both MR (*df* 2, p<.001, F=6.85, Eta Squared=.02) and CT (*df* 2, p<.001, F=7.123, Eta Squared=.02) in ANOVA statistical tests. Tukey posthoc comparisons demonstrated differences in confidence between novice and trainee groups (MR p<.013, CT p<.005), and expert and novice groups in both conditions (MR p<.013, CT p<.005). However, there were no significant differences between trainee and expert confidence scores in CT (p<.926) and MR (p<1.00) scores.

3.3 Eye-movement data: quantitative analysis.

3.3.1 Task viewing time per modality

Table 2 below highlights a trend towards novices and experts spending more time appraising CT compared with MR AOI's and experimental images as a whole. The inverse is true for radiology trainees - a trend which continued into mean fixation time per AOI and individual image. For all participant groups, the mean fixation time was larger within the AOI for CT than MR images.

Participant group	Modality	Average fixation time appraising all images (in seconds, 120 per modality)	Mean fixation time per axial image (s)	Mean fixation time in primary AOI (ms)
Novice (n=10)	CT	6087.87	5.07	1050
	MR	4348.64	3.62	580
Trainee (n=10)	CT	5208.34	4.34	710
	MR	5811.52	4.84	530
Expert (n=8)	CT	3525.93	3.67	700
	MR	2984.50	3.11	610

Table 2 demonstrates the average fixation time across all images, mean fixation time per axial image and mean fixation time in primary AOI per group and modality.

3.3.2 Stroke, Expertise, Accuracy and Visual Search.

Table 3 represents a cross comparison of eye movement data i.e. the mean time (in seconds) to 'hit' the first Area of Interest (AOI), mean time spent in the AOI, mean time spent out of the AOI in the same slice and mean fixation duration in the first AOI slice overall by group and modality, for **true and false positive** decisions.

Table 3 demonstrates that, irrespective of Stroke type, novices took longer to fixate an AOI in CT images, and spent more time in and around the AOI, when compared with MR images for *true positive* (TP) decisions. Trainee and expert results were dependent upon stroke type and modality. In acute cases, trainees and experts were quicker to fixate the AOI in MR images. Whilst trainees spent the same time viewing the AOI in CT and MR images, experts spent less time in MR images – indicating quick lesion recognition. Experts spent more time appraising cortical areas out of the AOI in CT, whereas trainees spent marginally more time inspecting the same areas in MR images. Overall trainees spent longer in MR than CT and the opposite was true for experts. Across all acute cases, experts show faster response times and spent less time in and around the AOI, than novices and trainees. In subacute cases, although experts spent the same amount of time within the AOI in both CT and MR, trainees and experts appeared to be quicker to 'hit' an AOI and spent less time in and around the AOI than in MR images. In chronic cases, experts were quickest to spot the lesion (mean time 1.1 seconds) in MR images, but both trainees and experts spent the least amount of time in chronic CT cases than MR*)*.

Once again, trainees spent more time appraising surrounding MR tissue than in CT cases, with experts demonstrating the opposite behaviour; spending more time in CT cases. Across all cases, experts were quickest to identify the lesion when making accurate decisions.

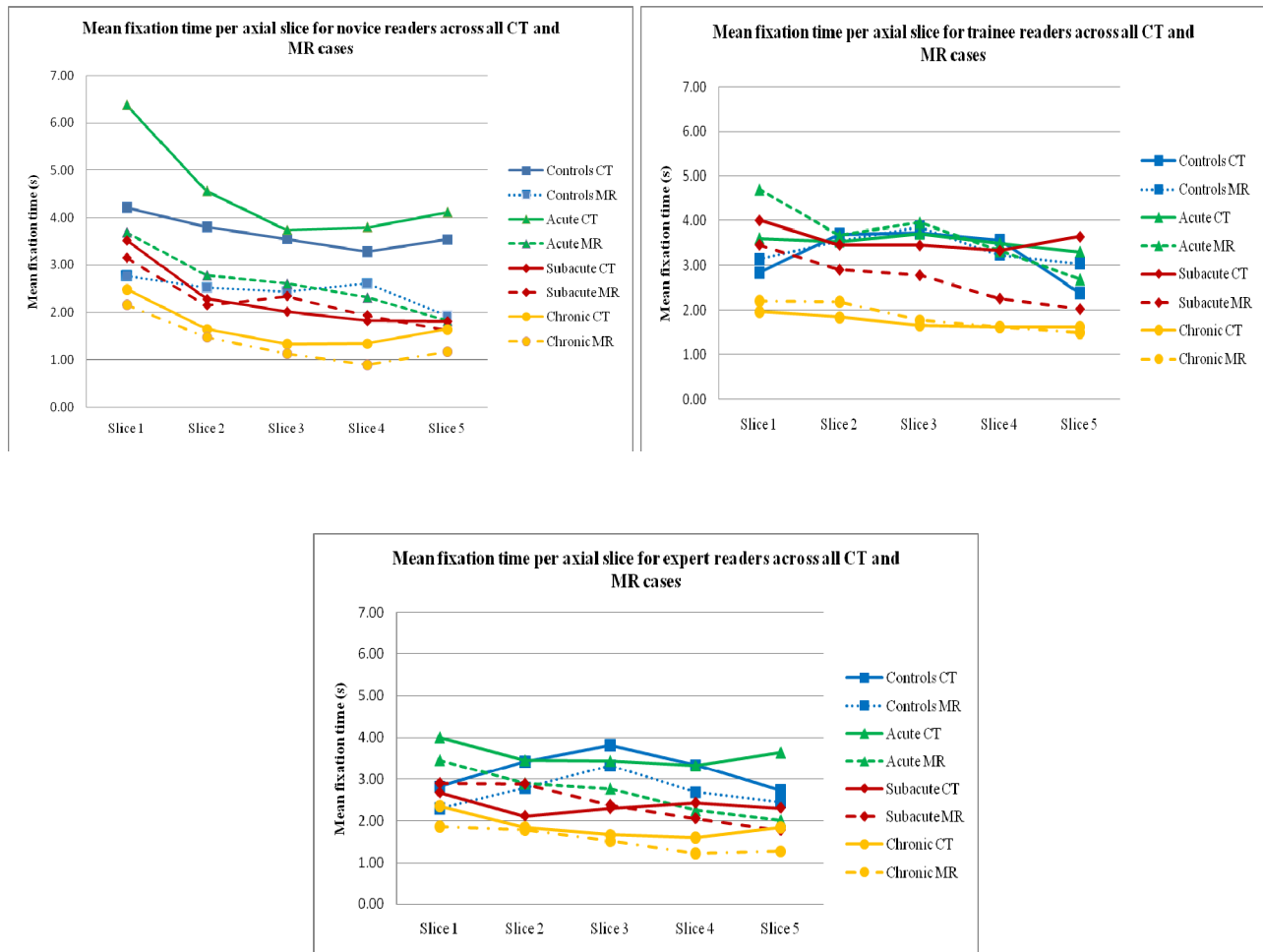| Stroke type, level of expertise and modality			Mean time to hit Primary AOI		Mean time spent in Primary AOI		Mean time out of AOI (in same slice)		Mean fixation duration in first AOI slice overall	
			TP	FP	TP	FP	TP	FP	TP	FP
Acute	Novice	CT	3.1	2.5	1.6	1.4	8.2	6.3	9.6	7.2
		MR	1.2	1.0	1.4	0.5	2.4	4.2	4.1	4.7
	Trainee	CT	1.6	2.3	1.2	0.8	3.8	3.8	5.0	4.6
		MR	1.4	0.4	1.2	0.7	3.9	2.9	5.2	3.6
	Expert	CT	1.4	1.6	1.4	1.4	2.3	2.7	3.9	3.5
		MR	1.1	0.2	1.1	0.8	2.1	0.5	3.2	1.3
Subacute	Novice	CT	2.5	2.5	1.7	2.5	5.0	7.1	6.5	8.6
		MR	0.9	1.5	1.6	0.8	3.4	4.1	4.9	5.1
	Trainee	CT	1.1	1.5	1.4	0.6	2.2	3.5	3.6	4.0
		MR	1.3	4.7	1.5	0.4	4.1	5.4	5.6	5.8
	Expert	CT	0.9	0.0	1.4	0.0	2.3	0.0	3.6	0.0
		MR	2.0	1.8	1.4	0.3	2.6	4.6	4.0	5.0
Chronic	Novice	CT	1.5	0.3	1.4	1.8	4.9	4.6	5.5	5.4
		MR	1.2	0.5	1.1	0.3	2.6	1.8	3.9	3.0
	Trainee	CT	1.5	0.8	0.9	2.2	3.9	11.3	4.5	13.4
		MR	1.5	0.5	1.2	0.3	4.4	2.3	5.7	2.6
	Expert	CT	1.5	6.7	1.1	0.3	3.8	6.9	4.2	7.2
		MR	1.1	0.0	1.2	0.0	2.5	0.0	4.0	0.0

Table 3 represents the mean time (in seconds) to 'hit' the first Area of Interest (AOI), mean time spent in the AOI, mean time spent out of the AOI in the same slice and mean fixation duration in the first AOI slice overall by group and modality, for *true and false positive* decisions.

For *false positive* decisions regarding acute cases, descriptive statistics showed that all participant groups took longer to fixate the AOI and subsequently spent longer fixating within and around the AOI in CT than MR images. In subacute cases, whilst the same trend was true for novices, trainees took longer to fixate upon a lesion in MR images than CT and also spent much more time out of the AOI in MR images than CT. No expert made an incorrect diagnosis for subacute CT cases, nor did they make any incorrect decision regarding chronic MR cases. False positive (FP) decisions appeared to be characterised by more irregularities within the data than true positive decisions.

When comparing TP versus FP decisions between modalities; false positive decisions made by novices regarding acute cases, were accompanied by a quicker time to hit, but less time spent in and around the AOI than TP decisions in CT images. Of the trainees that made incorrect decisions, the time to hit was longer for acute and subacute cases indicating they were not drawn quickly to the abnormality as they had been for TP decisions and they spent less time in the CT AOI. In accordance with our previous study, true negative decisions appear to be associated with a thorough image appraisal, confidence and expertise.

3.3.3 Visual search behaviour throughout the image 'stack'.

The following graphs demonstrate mean time spent in each axial slice throughout the five image 'stack' by case type (control, acute, subacute and chronic), level of expertise (novice, trainee and expert) and modality (CT, MR).

Mean fixation time per axial slice for novice readers across all CT and MR cases

Mean fixation time per axial slice for trainee readers across all CT and MR cases

Mean fixation time per axial slice for expert readers across all CT and MR cases

The first graph demonstrates that novice readers spent the most time appraising control and acute cases overall and the least time on chronic CT and MR images. There appears a trend towards a 'concave' effect in the search behaviour of novices through the stack with novices spending the most time on the first slice with a quick shoulder effect towards slice 3, which sometimes reduced further in slice four and increased when viewing the final slice. It appears a sharp 'drop off' in time in slice may be indicative of lesion recognition, even if a correct clinical decision was not made following the viewing.

Radiology trainee and expert results (displayed in graph 2 and 3), highlighted that visual search patterns throughout the stack appeared to depend upon the Stroke type; with some case groups following the concave line e.g. acute and chronic CT, whereas other data followed the convex trend e.g. control CT and MR. For trainees and experts the CT and MR control cases followed a 'convex' data streams through the stack, where an increasing amount of time was spent within the middle slice. This trend exhibited in normal cases may be indicative of experts examining middle axial sections for additional abnormalities in the absence of an obvious lesion, such as small vessel changes and lacunar infarctions, which frequently appear around the ventricular area. For other cases such as subacute and chronic MR, much time was spent on the first slice, followed by a steady decline of time spent examining the rest of the stack.

Experts, akin to novices, spent the most amount of time in examining acute CT cases as indicated by the descriptive results. In agreement with a prior pilot study, experts appeared to be more consistent both within and between cases than novices and operated within a smaller time range than novice participants in every case (Cooper *et al*, 2009). Although there was slightly more variability between trainees and experts in terms of average time spent in and around AOIs and total slice time, trainees followed a similar visual search pattern to the experts between stroke types i.e. between acute

and control visual search trends, on a case by case basis. The above graphs also demonstrate that in the majority of cases, participants spent more time in CT than MR images.

DISCUSSION

4.1 Diagnostic Accuracy and Confidence by Group and Modality.

This study aimed to explore observer performance between participant groups when viewing conventional CT and MR multidimensional images of Stroke. For both modalities, ruling out the presence of an infarct and detecting small infarctions were perceived to be the most challenging tasks among all Stroke cases. Highly significant differences were observed between groups and modalities when investigating diagnostic accuracy, but the largest effect size observed was for accuracy in MR rather than CT. When considering confidence ratings, novice and expert participants' confidence was higher across MR than CT cases. Conversely, trainee ratings appeared to favour CT over MRI, which may be indicative of a reduced amount of reading time that trainees had previous to the study with MRI, compared with expert readers. Novice readers may simply prefer the enhanced detail of MRI as they had received no formal training for detecting Stroke in either modality.

4.2 Image Analysis, Expertise and Eye Movements.

Image analysis results highlighted a clear qualitative difference between participant groups when viewing all medical images, irrespective of modality; novices spent more time visually examining normal anatomy such as ventricles and/ or spent much time 'fixated' upon a large and unambiguous lesion. Owing to lack of experience when appraising medical images and clinical problem-solving, novices had to make decisions about clinical features with very little training and no feedback on performance throughout the experimental study. Thus novices had to 'build up' their own mental schema regarding what constitutes an abnormality and what might not.

Early trainees outperformed novice readers but also spent a similar amount of time appraising the images, indicating a degree of uncertainty when making a final decision. Particularly in MR, experts were more confident in their ratings than trainee readers. In CT, however, there were no significant differences between trainee and expert readers. Trainee readers spent much more time cross comparing hemispheres than expert readers. With much caseload experience experts were more capable of directing their visual attention to an abnormal clinical feature, make a quick confident decision and promptly move on to appraise the surrounding cerebral tissue. Compared with novice readers, experts and experienced trainees, appeared to operate a system of deduction; ruling out certain areas very quickly to economise time and effort. In addition, experienced trainees and experts appear to have a mental schema of acquired information surrounding disease manifestation and how, due to known cerebrovascular pathways, different stroke types pervade different vascular routes. These qualitative findings are in agreement with literature previously discussed e.g. Rogers (1995), Nodine and Krupinski (1998), Garlatti and Sharples (1998), Manning (2006).

4.3 Diagnostic accuracy, Modality, Expertise and Eye-movements.

Previous studies have not rated conventional CT as superior to conventional MR or vice-versa (Mohr, 1995; Lansberg, 2000; Wintermark, 2007) and unenhanced CT remains the primary modality adopted when patients present with stroke-like symptoms (Kloska, 2004). Yet in this study of forty-eight cases, it appears experts performed this reading task much quicker and with more accuracy in the MR condition over CT. Trainees appeared to prefer reading CT than MR images; a finding correlated with a reduced reading time in CT and higher confidence in clinical decisions. For trainee readers, the enhanced anatomical detail and/or lack of previous reading time with MRI may have led to uncertainty and an increase in task time. Specific reasons for this difference would need to be further explored with trainee radiologists in a separate study.

Overall, experts were quicker to detect the primary infarction in MR for acute (overall mean 1.1s) and chronic stroke types (overall mean 1.1s) but quicker in CT for subacute cases (overall mean 0.9s). Experts were quicker to reach an AOI and spent less time in and around the AOI than novices and trainees in the most challenging cases, appearing to confirm the qualitative image analysis results demonstrating quicker, more accurate decisions and also confirming

findings from our previous pilot study (Cooper *et al.*, 2009). False positive decisions were characterised by more inconsistencies within and between groups in terms of eye-movements; either taking an unusually long time to first fixate within an AOI (e.g. 6.7 seconds for an expert FP decision in chronic CT) or not enough time when within the AOI itself (e.g. a novice spent only 300 milliseconds on average between chronic CT case AOI's).

In this study, radiology trainee and expert visual search patterns appeared to be characteristic of a particular case type i.e. control, acute, subacute and chronic, when viewing the small cross-section of clinical images, so indicating that images from each case type in Stroke are examined in much the same way, even between modalities. Whilst this study suggested that expert reading time is reduced and accuracy enhanced by adopting MR over CT imagery, resources and patient condition at time of imaging may not permit its increased use. Patient status aside, future research could conduct a cost-based analysis into the time spent viewing CT and MR images in a typical radiology department per annum to establish whether one modality might be more cost effective in Stroke detection, or indeed whether time taken to perform clinical tasks altered considerably when examined over a long period of time and patient cases.

In this study, it is recognised that certain limitations were applied; in normal clinical practice radiologists would scroll up and down image stacks, and use pan and zoom functions at will. For experimental design reasons this could not be permitted here. In addition, radiologists would not normally report a 3D image on a 2D reporting sheet and would have the opportunity to write reports, which accompany their observations; this was not permitted either. Future studies will endeavour to surpass these limitations to gather more in-depth and ecologically valid analysis of expert observers' behaviour in neuroradiology. In addition, many published papers have favourably reviewed the accuracy of DW imaging, a future study comparing visual search behaviour between conventional MR and DWI would be of interest. .

CONCLUSIONS

To-date few studies have explored observer performance in neuroradiology and the present study examines multi-slice image appraisal by comparing matched pairs of CT and MRI Stroke cases. Novice and expert readers spent longer appraising CT images than MR, compared with trainees where the inverse was true. Diagnostic accuracy and confidence rating followed the same pattern with trainees being more confident and accurate in CT rather than MR imaging and vice-versa for novice and expert readers. Differences were observed between novice, trainee and expert visual search behaviour. In particular, trainee and expert visual search appeared to depend upon stroke type being scrutinised. Image analysis trends did not appear to differ between modalities, but time spent within clinical images, accuracy and relative confidence performing the task did differ between CT and MR reader groups. Eye movement results suggest that true positive decisions were marked by a quick time to fixate the infarct and a consistency within reader group, whereas false positive decisions were more inconsistent both within and between groups.

ACKNOWLEDGEMENTS

REFERENCES

[1] Asplund, K., Stegmayr, B. and Peltonen, M., "From the Twentieth to the Twenty-First Century: A Public Health Perspective on Stroke." (1998) In: Ginsberg, MD, Bogousslavsky J, eds. Cerebrovascular Disease Pathophysiology,Diagnosis, and Management. Vol. 2. Malden, Mass: Blackwell Science; Chapter 64.
[2] American Heart Association,"Heart Disease and Stroke Statistics—2007 Update: A Report From the American Heart Association Statistics Committee and Stroke Statistics Subcommittee," Circulation, American Heart Association. Dallas, TX (2007)
[3] Cooper, L.H.K., Gale, A.G., Darker, I.T., Toms, A. and Saada, J., "Radiology image perception and observer performance; how does expertise and clinical information alter interpretation? Stroke detection explored through eye-tracking," Proc. SPIE 7263-19 (2009).
[4] Department of Health, "Mending Hearts and Brains," Department of Health, London (2006).

[5] Dorfman, D.D., Berbaum, K.S. and Metz, C.E., "Receiver operating characteristic rating analysis: generalization to the population of readers and patients with the jackknife method," Investigative Radiology; 27: 723–731 (1992).

[6] Gonzalez, R.G., Schaefer, P.W., Buonanno, F.S., Schwamm, L.H., Budzik, R.F., Rordorf, G. Wang, B., Sorensen, A.G. and Koroshetz, W.J. Diffusion-weighted MR Imaging: Diagnostic Accuracy in Patients Imaged within 6 Hours of Stroke Symptom Onset. Radiology; 210:155-162 (1999).

[7] Hobby, J.L., Tom, B.D.M., Todd, C., Bearcroft, P.W.P. and Dixon, A.K. "Communication of doubt and certainty in radiological reports," Br J Radiol, 73, 999-1001 (2000).

[8] Haughton, V.M., Rimm, A.A., Sobocinski, K.A., Papke, R.A., Daniels, D.L., Williams, A.L., Lynch, R. and Levine, R., "A Blinded Clinical Comparison of MR Imaging and CT in Neuroradiology," Neuroradiology; 160:751-755 (1986).

[9] Kloska, S.P., Nabavi, D.G., Gaus, C., Nam, E., Klotz, E., Bernd Ringelstein, E. and Heindel, W., "Acute Stroke Assessment with CT: Do We Need Multimodal Evaluation?" Radiology; 233: 79-86 (2004).

[10] Kundel, H.L. "Visual Scanning, Pattern Recognition and Decision-making in Pulmonary Nodule Detection," Investigative Radiology (1978).

[11] Krupinski, E.A., Nodine, C.F., and Kundel, H.L., "A perceptually based method for enhancing pulmonary nodule recognition," Investigative Radiology; 28: 289-294 (1993).

[12] Lansberg, M.G., Gregory, W.A., Beaulieu, C. and Marks, M.P., "Comparison of diffusion-weighted MRI and CT in acute stroke," American Academy of Neurology; 54:1557-1561 (2000).

[13] Manning, D.J., Ethell, S.C., and Donovan, T., "Detection or decision errors? Missed lung cancer from the posteroanterior chest radiograph," The British Journal of Radiology; 77:231-235 (2004).

[14] Manning, D.J., Barker-Mill, S.C., Donovan, T. and Crawford, T., "Time-Dependent Observer errors in pulmonary nodule detection," The British Journal of Radiology;78:1-5 (2005).

[15] Manning, D.J., Ethell, S.C., Donovan, T. and Crawford, T., "How do Radiologists do it? The influence of experience and training on searching for chest nodules," Radiography; 12, 134-142 (2006).

[16] Mello-Thoms, C. "What attracts the eye to the location of missed and reported breast cancers?" ACM;58113(1):111-117 (2002).

[17] Mohr, J.P., Biller, J., Hilal, S.K., Yuh, W.T.C., Tatemichi, T.K., Hedges, S.D., Tali, E., Nguyen, H., Mun, I., Adams, H.P., Grimsman, K. and Marler, J.R., "Magnetic Resonance Versus Computed Tomographic Imaging in Acute Stroke," Stroke; 26: 807-812 (1995).

[18] Mullins, M.E., Schaefer, P.W., Sorensen, A.G., Halpern, E.F., Ay, H., He, J. Koroshetz, W.J. and Gonzalez, R.G., "CT and Conventional and Diffusion-weighted MR Imaging in Acute Stroke: Study in 691 Patients at Presentation to the Emergency Department," Neuroradiology; 224:353-360 (2002).

[19] Nodine, C.F. and Kundel, H.L. "Using eye movements to study visual search and to improve tumor detection," RadioGRaphs; 7:1241-1250 (1987).

[20] Nodine, C.F. and Krupinski, E.A. "Perceptual skill, radiology expertise, and visual test performance with NINA and WALDO," Acad Radiol, 5, 603–612 (1998).

[21] Phillips, P.W., Manning, D.J., Donovan, T., Crawford, T. and Higham, S., "A Software Framework for Diagnostic Medical Image Perception with Feedback, and a Novel Perception Visualisation Technique," Proc. SPIE 5749: 572-580 (2005).

[22] Phillips, P.W., Manning, D.J., Crawford, T. Burling, D. Tam, C.H. and Taylor, A., "Searching in Axial and 3D CT Visualisations," Proc. SPIE 6917 (2008).

[23] Thomas, E.L. and Lansdown, E.L., "Visual Search Patterns of Radiologists in Training," Radiology; 81: 288-91 (1963).

[24] Wintermark, M., Meuli, R., Browaeys, P., Reichhart, M., Bogousslavsky, J., Schnyder, P. and Michel, P., "Comparison of CT perfusion and angiography and MRI in selecting stroke patients for acute treatment," Neurology; 68:694-697 (2007).