AN ANALYSIS OF RADIOLOGISTS' ASSIMILATION OF COMPUTED TOMOGRAPHIC EXAMINATIONS OF THE LUNGS

Stephen M Ellis

A thesis submitted to the
University of London
For the degree of
Doctor of Medicine

Department of Radiology
Royal Brompton Hospital
London
2002

BRITISH LIBRARY

IMAGING SERVICES NORTH

Boston Spa, Wetherby
West Yorkshire, LS23 7BQ
www.bl.uk

BEST COPY AVAILABLE.

VARIABLE PRINT QUALITY

BRITISH LIBRARY

IMAGING SERVICES NORTH

Boston Spa, Wetherby
West Yorkshire, LS23 7BQ
www.bl.uk

ORIGINAL COPY TIGHTLY BOUND

ABSTRACT

The acquisition of experience is a time consuming process resulting from exposure to multiple eventualities and deriving the most appropriate and efficient way of dealing with them. In radiology, the recognition of visual patterns, an appreciation of their significance, and relationship to one another help the radiologist to interpret an image. The aim of this thesis was to develop techniques to analyse the massive amount of data acquired when the eye-movements of a radiologist reading HRCT examinations are recorded with a view to understanding ~~the~~ some of the processes behind the assimilation of images.

Initially, the radiologists' approach to HRCT examinations presented in different viewing formats was assessed. It was found that in terms of time taken, number of fixation points and consistency of approach between examinations of similar disease processes, that images displayed at full screen size, one at a time was the more efficient presentation format.

Subsequent experiments were aimed at assessing the cognitive processes being employed by the observers in terms of their eye-movements. The approach of experienced observers was recorded and incorporated into a teaching presentation. The eye-movements of novice radiologists were then compared to those of the experienced observers, in terms of their visuospatial and feature-based characteristics, before and after training. The novice group was divided into two sub-groups one of which was given specific feedback as to how the experienced observers approached the examinations. This feedback did not result in a significant difference between the two groups of novices and any post training improvement in approach seen was more evident in the group of novices not give the extra information regarding the experienced observers' approach.

The thesis details techniques for analysing the approach of observers to HRCT examinations and the investigations demonstrated a fair degree of consistency among experienced observers. However, there was no discernible improvement in the approach of the novices following a limited training protocol, which was based on the behaviour of the experienced observers.

ACKNOWLEDGMENTS

I would like to acknowledge the help and support I have received from the computer department at Imperial College of Science and Medicine. In particular Laura Dempere-Marco and Xiaopeng Hu. Laura for her programming of the hierarchical distortion algorithm into a form that I could then apply to the eye-tracking data from the experiment. Xiaopeng for his adaptation of the eye-tracking camera set-up for use in the experiments conducted in this thesis, the development of the script based presentation and eye-tracked data recording program, the image pre-processing required for the presentation of lung parenchymal images to the hierarchical distortion algorithm program and the development of software for feature marking and analysis of eye-tracking data with respect to marked features. I would like to thank Professor Guang-Zong Yang, Director of the Department of Computing, Imperial College for Science, Technology and Medicine, University of London, for his support, advice and liaison with Laura and Xiaopeng.

I am indebted to Dr Athol Wells for his patience and assistance in the development of statistical techniques to analyse difficult data sets, his clarity of though helped to clear some very muddy waters. I am very grateful to the experienced observers, David Hansell (DMH), Michael Rubens (MBR), Sujal Desai (SRD) and Simon Padley (SPP) and the novices from University College Hospital and King's College Hospital.

I have saved particular thanks for my supervisor Professor David Hansell who has been a constant source of support and advice and without whom this thesis would not have been possible.

TABLE OF CONTENTS

TITLE PAGE i

ABSTRACT ii

ACKNOWLEDGEMENTS iii

TABLE OF CONTENTS iv

LIST OF FIGURES AND TABLES x

LIST OF ABBREVIATIONS xxii

CHAPTER 1: GENERAL INTRODUCTION.

1.1 PULMONARY ANATOMY **1**

1.1.1 AIRWAYS, PULMONARY ARTERIES AND VEINS 1
1.1.2 THE SECONDARY PULMONARY LOBULE 4
1.1.3 ACINI AND ALVEOLI. 4
1.1.4 INTERSTITIUM 6
1.1.5 CORTICAL AND MEDULLARY LUNG 6
1.1.6 DENSITY CHARACTERISTICS ON HRCT OF THORACIC ANATOMY. 7

1.2 OPTICS OF THE EYE **9**

1.2.1 FOCUSING 9
1.2.2 PUPILLARY SIZE VARIATION 10
1.2.3 VISUAL ACUITY 10

1.3 FUNCTIONAL ANATOMY OF THE EYE **11**

1.3.1 GROSS ANATOMY OF THE GLOBE 11
1.3.2 THE RETINA 14
1.3.3 CONTRAST PERCEPTION 16
1.3.4 THE EXTRINSIC EYE MUSCLES 16
1.3.5 EYE MOVEMENTS 18

1.4 EYE TRACKING **20**

1.4.1 HISTORY AND DEVELOPMENT 20
1.4.2 TECHNIQUES BASED ON REFLECTED LIGHT 20
1.4.3 A TECHNIQUE BASED ON ELECTRIC SKIN POTENTIAL 25
1.4.4 TECHNIQUES BASED ON CONTACT LENSES 25
1.4.5 APPLICATIONS OF EYE TRACKING 26

1.5 VISUAL PERCEPTION **28**

1.6 COMPUTED TOMOGRAPHY OF THE LUNGS 30

1.6.1 PRINCIPLES OF CT 30
1.6.2 HIGH-RESOLUTION COMPUTED TOMOGRAPHY 31

1.7 IMAGE DISPLAY 36

1.7.1 TRADITIONAL DISPLAY OF DIAGNOSTIC IMAGES. 36
1.7.2 CURRENT DISPLAY TECHNOLOGIES 36
1.7.3 "SOFT COPY" REPORTING VARIABLES 37

1.8 SELECTED PULMONARY DISEASE 40

1.8.1 DIFFUSE INFILTRATIVE LUNG DISEASE 40
1.8.2 OBSTRUCTIVE LUNG DISEASE. 44

1.9 CONSIDERATION OF RADIOLOGIST PERFORMANCE 52

1.9.1 TRAINING 52
1.9.2 REDUCING THE SOURCES OF ERROR IN RADIOLOGICAL INTERPRETATION 53
1.9.3 ASSESSMENT 54
1.9.4 IMPLICATIONS OF QUALITY ASSURANCE 55

1.10 RADIOLOGICAL DIAGNOSIS 56

1.10.1 COMPONENTS OF A "RADIOLOGICAL DIAGNOSIS" 56
1.10.2 ASPECTS OF AN HRCT IMAGE THAT ARE USED TO REACH A DIAGNOSIS 56
1.10.3 HIERARCHICAL HYPOTHESES THEORY FOR COMPUTER-AIDED DIAGNOSIS. 59
1.10.4 DIAGNOSTIC ACCURACY OF HRCT 61

1.11 COMPUTER AIDED DIAGNOSIS IN RADIOLOGY 62

1.11.1 PRINCIPLES OF CAD 62
1.11.2 IMAGE PROCESSING AND INCREASING CONSPICUITY. 64
1.11.3 THE BASIC COMPONENTS OF A CAD SCHEME 65
1.11.4 CURRENT RESEARCH 68

1.12 AIM OF THE THESIS 70

CHAPTER 2: GENERAL METHODOLOGY

2.1 EXPERIMENT SET UP. 71

2.1.1 EYE-TRACKING CAMERA AND SOFTWARE. 71
2.1.2 CALIBRATION 76

2.2 COMPUTER DISPLAY SOFTWARE DESIGN. 78

2.3 DERIVING A NORMALISED LUNG 81

2.3.1 NORMAL IMAGE SELECTION 81
2.3.2 PARENCHYMAL SEPARATION 83
2.3.3 HIERARCHICAL IMAGE REGISTRATION 85
2.3.4 DEFORMATION GRIDS 85
2.3.5 SELECTION OF THE "NORMAL" OUTLINE 87
2.3.6 DEFINITION OF THE NORMALISATION DEFORMATION GRID. 87
2.3.7 LIMITATIONS 87

2.4 ANALYSIS OF UNPROCESSED (RAW) EYE-TRACKING DATA 90

2.4.1 TIME (UNITS 10^{-4}S) 90
2.4.2 DISTANCE 91
2.4.3 VALIDITY 91
2.4.4 TREMOR, SHORT SACCADE OR LONG SACCADE 91
2.4.5 FIXATION 92
2.4.6 FIXATION DURATION 92
2.4.7 SHORT SACCADE LEADING TO FIXATION 93
2.4.8 LONG SACCADE LEADING TO FIXATION 93
2.4.9 MOVEMENT BETWEEN IMAGES 93

2.5 STATISTICAL METHODS USED FOR ANALYSIS 94

2.5.1 MEASURES OF OBSERVER AGREEMENT 94
2.5.2 MEASURES OF OBSERVER DIFFERENCES 97

CHAPTER 3: COMPARISON OF THE APPROACH TO READING HRCT SCANS DISPLAYED IN DIFFERENT VIEWING FORMATS.

3.1 INTRODUCTION 101

3.2 EXPERIMENT SET-UP 104

3.3 MATERIALS AND METHODS (EXPERIMENT 1) 104

3.4 RESULTS (EXPERIMENT 1) 109

3.5 MATERIALS AND METHODS (EXPERIMENT 2) 111

3.6 STATISTICAL ANALYSIS (EXPERIMENT 2) 116

3.6.1 DIFFERENTIATION OF SACCADE TYPE 116
3.6.2 EYE-MOVEMENT FEATURES ANALYSED 116
3.6.3 SCORING OF VERBAL REPORTS 118

3.7 RESULTS (EXPERIMENT 2) 118

3.8 INTERPRETATION OF RESULTS (EXPERIMENT 2) 120

3.9 DISCUSSION 126

3.10 SUMMARY 129

CHAPTER 4: A VISUOSPATIAL ANALYSIS OF THE APPROACH TO READING HRCT SCANS DISPLAYED IN TWO DIFFERENT VIEWING FORMATS.

4.1 INTRODUCTION 130

4.2 MATERIALS AND METHODS 131

4.2.1 EXTRACTING THE LUNG PARENCHYMA. 131
4.2.2 ANATOMICAL MATCHING 132
4.2.3 DISTORTION GRID CONSTRUCTION 132
4.2.4 SEPARATION OF THE EYE-TRACKING DATA 133
4.2.5 EYE-TRACKING DATA POINT MAPPING. 135

4.3 STATISTICAL ANALYSIS 139

4.3.1 POST-PROCESSING OF DATA 139
4.3.2 MASKING AND COARSENING OF DATA 139
4.3.3 COMPARISON OF DATA-SETS 140
4.3.4 INTER-OBSERVER ANALYSIS 143
4.3.5 WITHIN OBSERVER ANALYSIS. 143

4.4 RESULTS 144

4.5 DISCUSSION 147

4.6 SUMMARY 150

CHAPTER 5: ANALYSIS OF THE STRATEGY USED BY EXPERIENCED RADIOLOGISTS ASSIMILATING HRCTs.

5.1 INTRODUCTION 151

5.2 MATERIALS AND METHODS 152

5.2.1 CASE DISTRIBUTION. 152
5.2.2 CASE DISPLAY 152
5.2.3 EXPERIMENT SEQUENCE 155
5.2.4 REVIEW OF THE EYE-TRACKING DATA. 155

5.3 RESULTS **158**
5.3.1 STRATEGY WITH CASES OF INTERSTITIAL LUNG DISEASE: 158
5.3.2 STRATEGY WITH CASES OF ABNORMAL LOW ATTENUATION LUNG: 159
5.3.3 FURTHER ANALYSIS 159
5.4 INTERPRETATION OF THE RESULTS. **162**
5.5 DISCUSSION **165**
5.6 SUMMARY **169**

CHAPTER 6: VISUOSPATIAL ANALYSIS OF NOVICE'S APPROACH TO HRCT IMAGES COMPARED TO EXPERIENCED OBSERVERS.

6.1 INTRODUCTION **170**
6.2 MATERIALS AND METHODS **172**
6.2.1 TRAINEE SELECTION 172
6.2.2 FEATURE RECOGNITION 172
6.2.3 CASE SELECTION 173
6.2.4 TRAINING 180
6.3 STATISTICAL ANALYSIS **181**
6.3.1 VISUOSPATIAL DATA 181
6.3.2 FEATURE IDENTIFICATION SCORING 185
6.4 RESULTS **185**
6.4.1 OBSERVER CONSISTENCY 185
6.4.2 INTER-OBSERVER COMPARISONS. 186
6.4.3 FEATURE RECOGNITION SCORE 189
6.5 INTERPRETATION OF THE RESULTS **190**
6.6 DISCUSSION **197**
6.7 SUMMARY **202**

CHAPTER 7: ANALYSIS OF EXPERIENCED AND NOVICE OBSERVERS' FEATURE-LED ASSIMILATION OF HRCT.

7.1 INTRODUCTION **203**

7.2 MATERIALS AND METHODS **204**

7.2.1 FEATURE MAPPING 204
7.2.2 DATA ANALYSIS 205

7.3 STATISTICAL ANALYSIS **210**

7.4 RESULTS **211**

7.4.1 EXPERT AGREEMENT 211
7.4.2 NOVICE PERFORMANCE 213
7.4.3 AVERAGE NORMALISED TIME DENSITY BY FEATURE 216
7.4.4 TIME SEQUENCE RESULTS 217

7.5 INTERPRETATION OF THE RESULTS **222**

7.6 DISCUSSION **224**

7.7 SUMMARY **228**

CHAPTER 8: CONCLUSION **229**

APPENDIX **234**

REFERENCES **240**

LIST OF FIGURES AND TABLES

Chapter 1

		Page
Figure 1.1:	Anatomy of the bronchial tree distal to the segmental bronchi	3
Figure 1.2:	Schematic of an open lung biopsy: Defines the relative position of the anatomical structures that comprise the secondary pulmonary lobule	5
Figure 1.3:	Gross anatomy of the eye	13
Figure 1.4:	Cellular anatomy of the retina	13
Figure 1.5:	Anatomy of the ocular muscles. Note the course of the superior oblique muscle as its tendon passes through a sling before inserting onto the globe.	17
Figure 1.6:	Source of the 4 Purkinje images: Light is reflected at both surfaces of the cornea and lens, the strongest, and therefore easiest to identify, reflection is that from the external surface of the cornea and is termed the glint (Purkinje 1 reflection)	24
Figure 1.7:	Pupil and corneal reflections (the glint): As the eye moves the position of the glint alters in relation to the pupil. With calibration and a stationary illuminating source, the relative position of the glint and pupil can indicate the direction of vision.	24
Figure 1.8:	Centrilobular distribution: **(a)** and **(b)** diagrammatic representations with abnormal tissue depicted in black and **(c)** an HRCT of a patient with pneumoconiosis demonstrating centrilobular nodularity.	43
Figure 1.9:	Lymphatic distribution: Nodular infiltrates **(a)** with HRCT **(b)** showing micro-nodularity and diffuse infiltrates **(c)** with thickening of the interlobular septa on HRCT **(d).**	43

		page
Figure 1.10:	A peripheral acinar distribution: **(a)** subpleural scarring, **(b)** more severe involvement of entire lobule with destruction and honeycombing. **(c)** HRCT showing a honeycomb pattern and peripheral acinar shadowing in adjacent tissue	46
Figure 1.11:	Parenchymal consolidation: **(a)** patchy [**(b)** HRCT] and **(c)** diffuse [**(d)** HRCT]	46
Figure 1.12:	Emphysematous destruction: Normal lung **(a)** is compared with emphysematous lung **(b)** where the alveolar sacs have been damaged with destruction of the alveolar walls.	46
Figure 1.13:	HRCT image demonstrating **(a)** mosaic attenuation due to air trapping and **(b)** due to chronic thrombo-embolic disease causing reduced perfusion. The darker lung is abnormal in both cases and the vessels in the abnormal lung are attenuated.	51
Figure 1.14:	HRCT image demonstrating a mosaic attenuation pattern secondary to parenchymal infiltrates. Note there is no discrepancy in the vasculature of the different attenuation regions of lung.	51
Figure 1.15:	Schematic diagram of a simplified hierarchical hypothesis mode for constructing a diagnosis from an HRCT examination. Regions encompassed by computer aided diagnosis and automated diagnosis are shaded.	60

Chapter 2

Figure 2.1:	The ASL-504 eye-tracking device: The light emitting diodes arranged in circle around the central camera lens emit infra-red light to which the camera is sensitive. The camera mounting enables rotation in two orthogonal planes and therefore permits a wide range of movement.	72

		page
Figure 2.2:	Schematic of the experiment set-up	73
Figure 2.3:	Screen capture of the eye-tracker video image: The software associated with the eye-tracking camera identifies the pupil (white circle), calculates the centre of the pupil (white cross) and identifies the glint (black cross). Note the circle and crosses are displayed with a constant displacement to aid the manual checking of the automated detection.	75
Figure 2.4:	Calibration screen: The crosses were designated with the numbers one to nine and the observer asked to look at each cross in turn thereby calibrating the eye-tracking system over the entire screen.	77
Figure 2.5:	An example of a case presented in the 2 x 2 tiled viewing format: The green square indicates the image that can be enlarged and, when moved off the visible screen area, stimulates the repositioning of the displayed images to allow scrolling through the images in the case.	79
Figure 2.6:	Explanation of the data acquired during eye-tracking recording.	80
Figure 2.7:	15 levels, based on mediastinal structures, selected for calculation of the normalised lung	82
Figure 2.8:	The above twenty-four images demonstrate the extracted parenchyma of the left lung, at level seven for the 'normal' lung, from each of the normal HRCT cases. Through a series of re-scaling and distortion on each image, an 'average' shape was derived from these images.	84

		page
Figure 2.9:	Deformation grids at progressively finer detail levels: The bold line represent the shape to be distorted. At detail level 0 the margins of the shape are deformed to match those of the final shape. At the next level, the initial grid is subdivided into four and each smaller grid element is deformed to match the final grid at that detail level. The subdivision process continues until no further deformation is required.	86
Figure 2.10:	**a)** and **b)** are contour maps depicting the overlapping of the lung parenchymal images from level seven, the red end of the spectrum on the colour map or the whiter areas on the grey scale map depict the areas of greater overlap. **c)** and **d)** are probability contour maps (**d** is a simplified version) each contour defines the probability that a given point within that contour will map onto one of the reference images. **e)** and **f)** show the probability contours superimposed upon the grey scale probability map.	89

Chapter 3

		page
Figure 3.1	Image designed to determine the efficacy of peripheral vision in the identification of small detail, in this case mild bronchiectasis present in all but the bottom left quadrant. The green circle defines the area in the centre of the image into which the eye must be tracked for the image to be visible. The distance used in the results section is indicated by the green double headed arrow marking the distance from the edge of the central circle to the outer aspect of the bronchiectatic airway in that quadrant, the entire airway needs to be visualised to make decision as to whether it is bronchiectatic.	106
Figure 3.2	Image designed to determine the efficacy of peripheral vision in the identification of shades of grey, in this case mosaic attenuation present in all but the top right quadrant. The green circle represents the area as defined above. The distance used in the results section is again demonstrated as a double headed arrow this time extending from the edge of the central circle to the nearest margin of mosaic attenuation in that quadrant.	107

		Page
Figure 3.3	Flow chart demonstrating the sequence of actions that comprised experiment 1	108
Figure 3.4	Four images displayed at once: The green box identifies the current selected image.	113
Figure 3.5	**a)** moving the green selection box to the left, off the screen, from the left upper quadrant results in **b)** a single step through the sequence of images in the case. **c)** and **d)** A single step in the opposite direction occurs if the green selection box is moved to the right from the bottom right quadrant.	113
Figure 3.6	Flow chart demonstrating the sequence of experiment 2	115
Table 3.1	The distance from the outer edge of the central fixation circle on the image to the outer limit of a bronchiectatic airway (images 1-3) or the nearest grey / black interface of mosaic attenuation (images 4-5). The shaded cells corresponds to quadrants without an abnormal feature	109
Table 3.2	Responses from the four observers for experiment 1, only positive replies are indicated, the shaded cells correspond to quadrants in which no abnormal features were seen.	110
Table 3.3	The radius of useful peripheral vision, in pixels, calculated from the responses in experiment 1.	111
Table 3.4	A summary of the technical differences between the two formats used in this experiment.	114
Table 3.5	Table of results from experiment 2. P values from Wilcoxon's signed rank test. All p values < 0.05 have been highlighted in red and mean values recorded in blue. The presented data represents the accumulated results from all 4 observers.	119

Chapter 4

		page
Figure 4.1	Summary of data splicing and concatenation of the raw eye-tracking data using Observer 3 prior to and following re-mapping onto the normalised lung via distortion grids.	134
Figure 4.2	Description of the mapping of eye-tracking data points by defining each as a circular area projected by a 2° cone with density weighting determined by a Gaussian distribution. **(a)** demonstrates an isolated circular area and the variation of weighting depicted by colour and then demonstrated on a 3-dimensional plot with weighting on the y axis. **(b)** shows eye-tracking data mapped as individual points and **(c)** demonstrates the appearances when each point is plotted as a circular area following a Gaussian distribution for weighting	136
Figure 4.3	Demonstration of the distortion of the circular area projected by a cone when angle off the central axis; 'D' is the distance from the eye to the screen, 'r' is the radius of the circle of projection of the cone with the vertex at the eye and aligned along the orthogonal axis to the plane of the screen. 'w' and 'h' represent the width and length of the oval of projection of the cone when at an angle α from the orthogonal axis.	138
Figure 4.4	Plot of correlation coefficient against anatomical level for a comparison of the eye-tracking derived time density data for DMH and SRD for case 1. Comparisons of 32x32 grid,16x16 and 8x8 grids are plotted; reduction in data fluctuation is evident in the 8x8 grid data. The trend lines indicate higher correlation at the more cranial or caudal levels.	141
Figure 4.5	Histogram of normalised time density from the eye-tracking data: Threshold points for the calculation of ranks are shown and correspond to points at which the histogram demonstrates a departure a smooth curve, rank 1 was taken as zero, the limits for rank 2 were >0 and ≤0.00358, for rank 3 >0.00358 and ≤0.00716 and for rank 4 >0.00716.	141

		Page
Figure 4.6	Explanation of the processing of eye-tracking data prior to analysis. **(a)** is the mask derived from the normalised lung, **(b)** is the plot of the time density distribution for the eye-tracking data at the same anatomical level following grid distortion to match the normal lung, **(c)** displays the time density distribution overlaid on the normal lung mask and **(d)** displays the ranked assignment to the grid squares dependent on the total time density found within them.	142
Table 4.1	Summary of inter-observer kappa values comparing tiled format with stack format. The kappa values for stacked format were significantly greater than those for tiled format ($p<0.05$).	144
Table 4.2	Summary of inter-observer kappa values comparing cases of different disease type. The kappa values for ILD cases were significantly greater than those for NORM cases ($p<0.05$) but no other statistically significant differences were seen.	145
Table 4.3	Definition of observer ranking based on mean kappa values	145
Table 4.4	Definition of format comparison ranking based on mean kappa values. Tiled corresponds to format A and stack to format B.	146
Table 4.5	Definition of disease comparison ranking based on mean kappa values. ILD – interstitial lung disease, OLD - obstructive lung disease, NORM - no disease.	146
Table 4.6	Summary of the result of stepwise linear regression analysis on the within observer data. $R^2 = 0.25$. There is a significant positive correlation between ranking of observer, format comparison and disease comparison with the calculated kappa values.	147

Chapter 5

		page
Figure 5.1	Flow diagram demonstrating the sequence of case reading and calibration followed in the experiment. x = case number.	156
Figure 5.2	An example of the feedback questions that were given to the observers. The feedback was performed via a PowerPoint (Microsoft Corp, USA) presentation and the observers given the option to see the image in question with and without the marking circles and arrows.	157
Table 5.1	Summary of features in the cases used in the experiment	154
Table 5.2	2 x 2 contingency table of whether abnormal low attenuation lung was present and whether the low attenuation lung was viewed first (Chi squared test, $p<0.0001$).	161
Table 5.3	2 x 2 contingency table of whether abnormal high attenuation lung was present and whether the high attenuation lung was viewed first (Chi squared test, $p<0.0001$).	161
Table 5.4	2 x 2 contingency table of whether a peripheral abnormality was present and whether the periphery was scrutinised (Chi squared test, $p<0.0001$).	161
Table 5.5	2 x 2 contingency table of whether a comparison of the bronchovascular bundles was made between the left and right lung and whether abnormal low attenuation lung was present (BL) (Chi squared test, $p<0.0001$).	161

Chapter 6

		page
Figure 6.1	The five images comprising Case 2 with regions of low and high attenuation lung marked to demonstrate the mosaic attenuation pattern	175
Figure 6.2	The five images comprising Case 3 with nodules and reticulation marked in this case of sarcoidosis.	176
Figure 6.3	The five images comprising Case 7 with ground glass opacity and reticulation marked in this case of non-specific interstitial pneumonia.	177
Figure 6.4	The five images comprising Case 9 with ground glass opacity and reticulation marked in this case of non-specific interstitial pneumonia.	178
Figure 6.5	The five images comprising Case 19 with ground glass opacity, nodules, bronchiectasis and mosaic attenuation marked.	179
Figure 6.6	Histogram of normalised time density from the eye-tracking data: Threshold points for the calculation of ranks are shown and correspond to points at which the histogram demonstrates a departure a smooth curve, rank 1 was taken as zero, the limits for rank 2 were >0 and ≤0.053, for rank 3 >0.053 and ≤0.105 and for rank 4 >0.105.	183
Figure 6.7	Visual projection of the native eye-tracking data for case 7 from the two experts with the greatest level of agreement (kappa = 0.79). **a)** the original images with eye-tracking density data for DMH overlaid and **b)** the original image with eye-tracking density data for MBR overlaid Density ranges from blue (least) through green and yellow to red (greatest).	192
Figure 6.8	Visual projection of the native eye-tracking data for case 7 from the expert and novice with the greatest level of agreement (kappa = 0.80). **a)** the original images with eye-tracking density data for DMH overlaid and **b)** the original image with eye-tracking density data for Novice B overlaid Density ranges from blue (least) through green and yellow to red (greatest).	193

		Page
Figure 6.9	Visual projection of the native eye-tracking data for case 2 from the two novices with the least level of agreement (kappa = 0.12). **a)** the original images with eye-tracking density data for Novice A overlaid and **b)** the original image with eye-tracking density data for Novice D overlaid Density ranges from blue (least) through green and yellow to red (greatest).	194
Figure 6.10	Visual projection of the native eye-tracking data for case 2 from the novice who showed the greatest elevation in kappa value between readings (kappa from 0.43 to 0.67) and from the experienced observer with whom this change occurred. **a)** the original images with eye-tracking density data for Novice D overlaid and **b)** the original image with eye-tracking density data for DMH overlaid and **c)** the original images with eye-tracking density data for Novice D post-training overlaid. Density ranges from blue (least) through green and yellow to red (greatest).	195
Table 6.1	Scoring system based on the number of features identified on the HRCT images presented to the novices. The recording of features not present was marked by reducing the score for that case by one but zero remained the minimum score.	173
Table 6.2	Summary of features found in the cases used for the visuospatial analysis; ILD = interstitial lung disease, OLD = obstructive lung disease	174
Table 6.3	Summary of the three ranges of inter-observer comparisons performed in the analysis of results. Experienced = experienced observers (DMH, MBR and SRD), King's novices were labelled A to C and UCH novices D to F	182
Table 6.4	Ranks and mean kappa values for the comparisons between the experienced observers and novice groups as a whole.	186
Table 6.5	Ranks and mean kappa values for the comparisons between the expert and novice groups separated by training centre	186
Table 6.6	Ranks and mean kappa values for the comparisons between individual experts and individual novices.	187

		Page
Table 6.7	Summary of the total scores for feature recognition from the novices for each of the two experiments.	189

Chapter 7

Figure 7.1	Sample case of subtle mosaic attenuation in a subject with small airways disease. **a)** the total region within which features were defined is displayed, note the mediastinum is included but the chest wall is excluded. **b)** regions of abnormal low attenuation lung are overlaid with blue, **c)** the bronchovascular bundles are marked in red and **d)** both features are combined on the one image. In **e)** an eye-tracked data point is superimposed on the images and feature maps to demonstrate the recording of visual duration on two features from only one data point and the impact of overall feature area on the resulting total feature duration.	207
Figure 7.2	Graph of average normalised time density for each feature averaged from the results from the experienced observers for each case (different colours).	218
Figure 7.3	Graph of average normalised time density for each feature averaged from the results from the novice observers prior to training for each case (different colours).	219
Figure 7.4	Graph of average normalised time density for each feature averaged from the results from the novice observers, after training, for each case (different colours).	220
Figure 7.5	Graph of KLIC against the percentage of the total time taken for reading a case, for each case read by novice C after training.	221
Table 7.1	Summary of the features marked for analysis and the cases in which the respective features were found.	205
Table 7.2	Summary of comparisons between the experienced observer's results for each case with significant correlation identified by bold type.	212

		page
Table 7.3	Case 9 Spearman rank correlation test p values for the pre-training results from novices A, B, C, E and F compared with the results from the three experienced observers (no significant correlation was found with novice D).	214
Table 7.4	Spearman rank correlation test p values for the case 19 pre and post-training results from novices A, C, D, E and F compared with the results from the three experienced observers (no significant correlation was found with novice B).	216

LIST OF ABBREVIATIONS

CT	Computed tomography
HRCT	High resolution computed tomography
ms	millisecond
mA	milliampere
kVp	Peak kilovoltage
HU	Hounsfield unit
CAD	Computer aided / assisted diagnosis
PACS	Picture archive and communication system
κ_w	Weighted kappa coefficient
Ho	Null hypothesis
H_A	Alternative hypothesis
W	Wilcoxon's signed rank test statistic
T	Wilcoxon's ranksum test statistic
r_s	Spearman rank correlation coefficient
KLIC	Kullback-Leibler information criterion
MRI	Magnetic resonance imaging
DICOM	Digital Imaging and Communications in Medicine
ILD	Interstitial lung disease
OLD	Obstructive lung disease
ANOVA	Analysis of variance
BX	Bronchiectasis

CHAPTER 1

GENERAL INTRODUCTION

The experiments undertaken in this thesis involve the interpretation of high-resolution computed tomography (HRCT) images and the recording of eye movements. In order to understand the interpretation of HRCT images a core knowledge of pulmonary anatomy is presented and the patho-radiologic correlation of disease processes discussed. Furthermore, assertions as to the type and sequence of eye-movements require background knowledge with respect to the anatomy and physiology of the eye and the bio-physiology and psychology of the visual system. Detail regarding the intra-cerebral mechanisms of vision are beyond the scope of this introduction

1.1 PULMONARY ANATOMY

1.1.1 Airways, pulmonary arteries and veins

Air is conducted into and out of the lungs by the bronchial tree. The trachea divides into right and left primary bronchi, which subsequently divide into secondary bronchi each entering a lobe and hence there are three on the right and two on the left. Each secondary, or lobar, bronchus then divides into tertiary, or segmental, bronchi each supplying an individual bronchopulmonary segment. The segmental bronchi divide further into sub-segmental bronchi. The branching pattern is one of asymmetric dichotomy, i.e. each bronchus gives rise to two daughter bronchi of different lengths. All the airways to this

point have cartilage in their walls and do not contribute to gas exchange. An artery accompanies each bronchi; the diameter of the bronchus and its accompanying artery gradually reduce toward the periphery. Bronchi that are of greater diameter than their adjacent arteries are taken to be abnormally dilated (or bronchiectatic). However, bronchi may, on occasion, appear larger than adjacent arteries in normal subjects (Lynch et al. 1993) and in dependent regions of the lung the artery is often of larger calibre than the neighbouring bronchus.

In the lung periphery the pulmonary veins are found at the margins of the secondary pulmonary lobules and the arteries at the centre; the secondary pulmonary lobule is described in more detail later. Beyond the sub-segmental divisions, the bronchioles are classified as terminal (membranous) bronchioles or respiratory bronchioles (Wright et al. 1992). The terminal bronchioles represent the most distal components of the conducting airways (as they are characterised by a complete fibromuscular wall and have no capacity for oxygen exchange). The terminal bronchioles vary in length from 0.8 to 2.5mm and have an internal diameter of approximately 0.6mm, which appears to be relatively uniform along their length (Hansen et al. 1975). A transitional zone downstream of the terminal bronchioles gives rise to the respiratory bronchioles and represents a change in function of the airways from gas conduction to gas exchange. The respiratory bronchioles have alveoli arising from their walls and contribute to gas exchange. Each terminal bronchiole supplies a gas exchange unit, the pulmonary acinus (Fig.1.1).

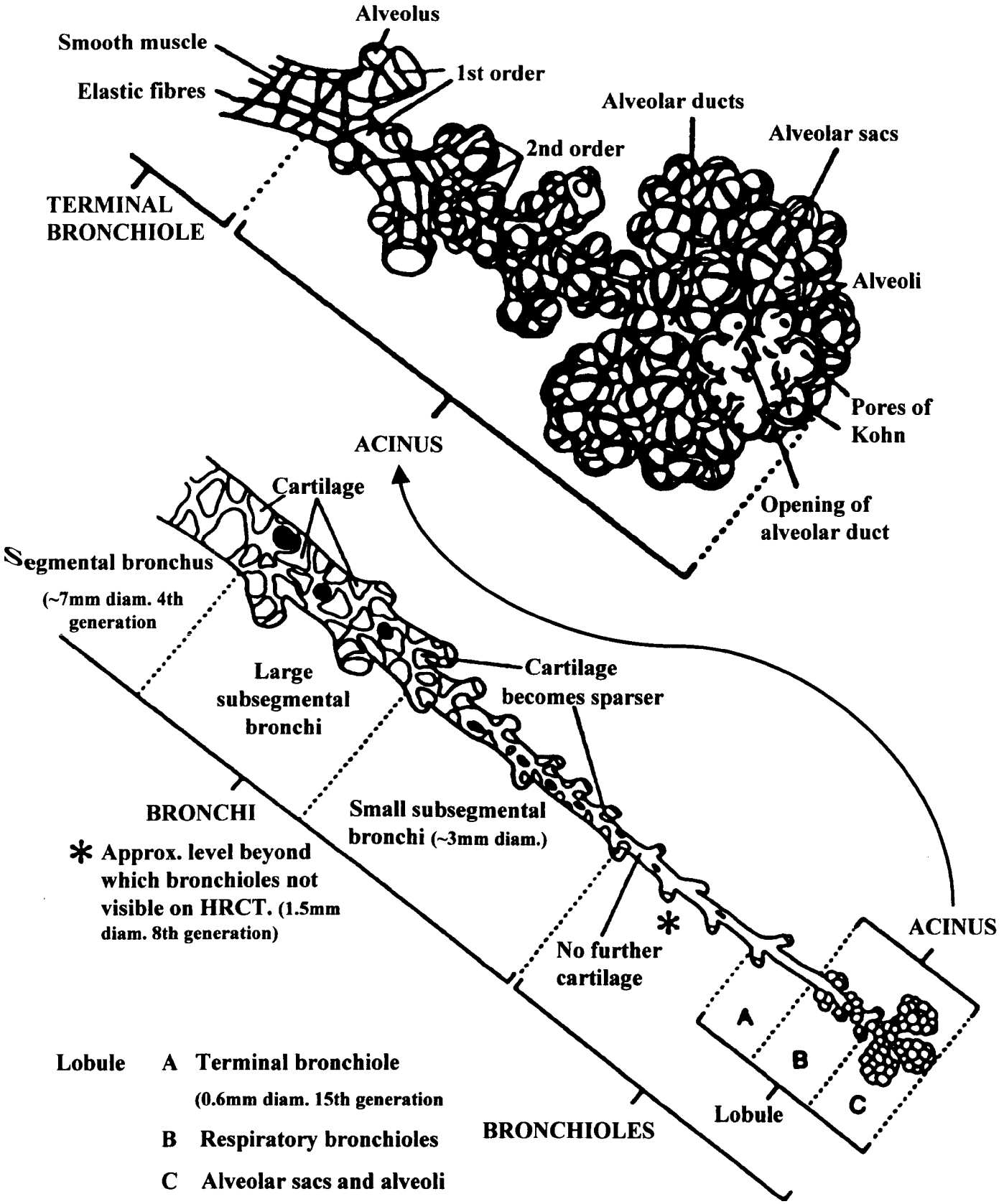


Figure 1.1: Anatomy of the bronchial tree distal to the segmental bronchi. From (Hansell 1999)

1.1.2 The secondary pulmonary lobule

The secondary pulmonary lobules, the smallest units of lung structure marginated by connective tissue septa, are variable in size (between 1 and 1.5 cm in diameter) and have an irregular polyhedral shape. Each secondary lobule is supplied by a small bronchiole and pulmonary artery and variably marginated by interlobular septa containing veins and lymphatics. The number of acini found within a given secondary pulmonary lobule is variable (3 to 24). The superficial secondary pulmonary lobules are more clearly defined than the deeper lobules where the interlobular septa are less developed and thinner. The central region of the lobule is referred to as the centrilobular region and contains the supplying pulmonary artery and bronchiole along with supporting connective tissue, the centrilobular core structures (Fig. 1.2).

1.1.3 Acini and alveoli.

The acinus is a 6 to 10 mm diameter air-space enclosed by a unicellular respiratory epithelium facilitating gas exchange with the blood cells traversing the respiratory capillaries in the walls of the acinus. Communication with the respiratory bronchiole is either direct or via other acini through the pores of Köhn, which also facilitate communication between acini in adjacent secondary pulmonary lobules. The alveoli are small sacs lined by respiratory endothelium with intraluminal capillaries and represent the smallest functional units in an acinus. A single acinus contains one terminal bronchiole, approximately fourteen respiratory bronchioles, 1,200 to 1,500 ducts, 2,500 to 4,500 sacs and 14,000 to 20,000 alveoli (Hansen, Ampaya, Bryant, & Navin 1975).

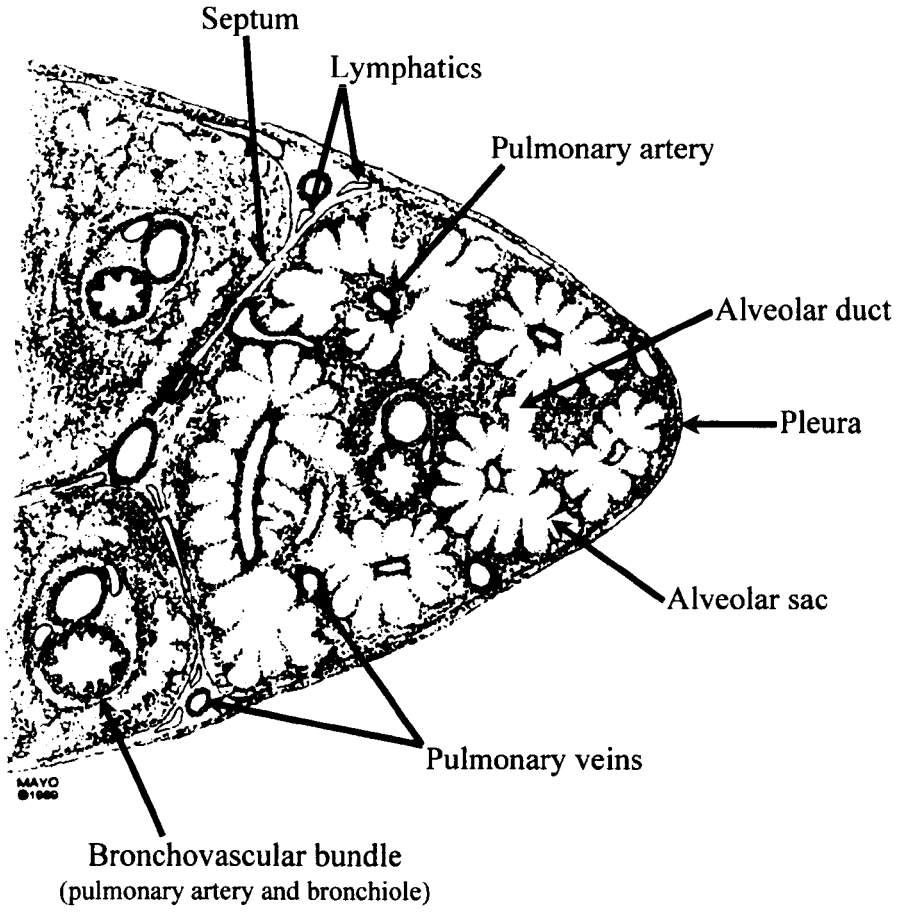


Figure 1.2: Schematic of an open lung biopsy.

Defines the relative position of the anatomical structures that comprise the secondary pulmonary lobule

(from Colby TV, Lombard CM, Yousem SA, Kitaichi M. Atlas of pulmonary surgical pathology. Philadelphia: WB Saunders, 1991)

1.1.4 Interstitium

A network of connective tissue fibres termed the lung interstitium supports the lung. The peribronchovascular interstitium, in the peri-hilar region, forms a strong connective tissue sheath surrounding the large bronchi and arteries, which becomes progressively less substantial toward the periphery as it invests the distal small centrilobular bronchioles and arteries (the centrilobular interstitium). The peribronchovascular and perilobular interstitium, which extends to the level of the alveolar ducts, together form the "axial fibre system" (Weibel 1979). The subpleural interstitium, located beneath the visceral pleura, envelopes the lung in a fibrous capsule from which connective tissue septa penetrate into the lung parenchyma forming Weibel's "peripheral fibre system". A fine connective tissue mesh of thin fibres in the walls of the alveoli forms the intralobular interstitium, which bridges the gap between the interlobular septa and sub pleural interstitium in the lobular periphery and the centrilobular interstitium in the centre of the lobules. The important conceptual point is that the intralobular, peribronchovascular, centrilobular and subpleural interstitium together with the interlobular septa form a continuous fibre skeleton for the lung.

1.1.5 Cortical and medullary lung

The division of the lung parenchyma into a peripheral cortical lung and more central medullary lung is based, at least in part, on differences in lobular anatomy. Cortical lung consists of 2-3 rows of well-defined and organised secondary pulmonary lobules forming a 3-4 cm thick layer at the lung periphery, including the lung surfaces adjacent to the

fissures. The pulmonary lobules in cortical lung are larger than those in medullary lung and marginated by thicker better defined interlobular septa. The small arteries and veins in cortical lung can be identified on HRCT but the small airways are not usually discernible. By contrast the larger airways in medullary lung are readily identified on normal HRCT examinations. The secondary pulmonary lobules in medullary lung are smaller and more irregular in shape than those in cortical lung, and are marginated by thinner less well defined interlobular septa. The lobules on cross sectional inspection appear polygonal but are not usually visible on normal HRCT examinations.

The distinction between cortical and medullary lung is not usually brought to bear in refining an HRCT diagnosis (in contrast to the attention paid to disease distribution in terms of the secondary pulmonary lobule). Nevertheless, there are some anatomical (and possibly physiological) differences between peripheral and central lung, which can be useful in predicting and explaining the HRCT distribution of some diseases (Gurney 1991).

1.1.6 Density characteristics on HRCT of thoracic anatomy.

Normal lung parenchyma on a CT scan varies in density. An antero-posterior density gradient exists in normal parenchyma due to the presence of mobile fluid in the interstitium, which, with the patient supine in the CT scanner, is drawn posteriorly by gravity; the dependent increase in parenchymal density is most apparent at the major fissures where the parenchyma anterior to the fissure is of increased density due to an increase in interstitial fluid but the parenchyma posterior to the fissure is of reduced

density due to the movement of interstitial fluid to the more dependent posterior regions of the lung. There is also a density gradient in the cranio-caudal direction due to the increase in perfusion of the caudal areas of the lung; the increase in vascular flow increases the amount of blood within the capillary bed and the amount and level of turnover of interstitial fluid (more engorged lymphatic channels). All these factors increase the lung parenchymal density but the structures involved are too small to be resolved individually on even an HRCT scan; the result is an overall increase in density. The use of quantitative CT data to reach decisions with regard to lung function or disease processes (Goldin 2002) requires an appreciation of the normal physiological changes in lung parenchymal density and attempts to create a normal template of parenchymal density have been made (Fromson & Denison 1988).

1.2 OPTICS OF THE EYE

1.2.1 Focusing.

The focusing of light on the macula is dependent on the cornea and the lens. The refraction of light at the cornea is considerably greater than that at the lens. The degree of light refraction at an interface between two media is related to the angle of incidence and the difference in the refractive indices of the two media, medium1 and medium2, given by the following expression based on Snell's law:

$$\frac{\sin\Theta_1}{\sin\Theta_2} = \frac{I_{medium2}}{I_{medium1}}$$

where Θ_1 and Θ_2 represent the angles of incidence and refraction respectively and $I_{medium1}$ and $I_{medium2}$ represent the refractive indices of the two media at the interface. The difference between the refractive indices of air and cornea is much greater than that between those of aqueous humour and the lens. Any increase in curvature of the lens, increasing the angle of incidence, is insignificant compared to the impact of the discrepancy in size of the refractive index differences. However, corneal refraction is fixed and it is the lens, with its variable convexity, that fine-tunes the focusing of an image on the macula. The lens shape is held in a flattened form when the ciliary muscles are relaxed and contraction of the ciliary muscles results in an increase in lens curvature due to its inherent elasticity (Guyton & Hall 1996a).

1.2.2 Pupillary size variation

Pupillary constriction accompanying accommodation prevents the most divergent light rays, which cannot be focused and therefore introduce blurring, from entering the eye. The role of varying pupillary size in enabling the eye to function over a wide range of light intensities is a minor one. The pupil area changes over a ratio of only 16:1 but the eye works efficiently over a light intensity range of 1,000,000:1. The majority of light adaptation occurs in the retina (Guyton & Hall 1996a).

1.2.3 Visual acuity

Theoretically, light from a distant point source, when focused on the retina, should produce an infinitely small retinal spot. However, the eye is not optically perfect and such a retinal spot actually has a minimum diameter of 11 micrometres, brightest at its centre gradually diminishing toward the edges. The average diameter of cones in the fovea is 1.5 micrometers, one-seventh the diameter of the retinal spot, but the intensity profile of the retinal spot enables the distinction between two retinal spots only 2 micrometres apart. The normal visual acuity of the human eye for discriminating between two point sources of light is around twenty-five seconds of arc. This means that a person with normal acuity looking at two bright pinpoint spots of light ten metres away can barely distinguish the spots as entities when they are 1.5 to 2 mm apart. The fovea subtends approximately two degrees of the visual field outside of which visual acuity progressively decreases due to a reduction in density of light sensitive cells, an increase in their size (transition from all cone to all rods) and the connection of many rods and cones to the same nerve fibre (Guyton & Hall 1996a).

1.3 FUNCTIONAL ANATOMY OF THE EYE

1.3.1 Gross anatomy of the globe

The globe of the eye is approximately spherical, divided into anterior and posterior segments by the lens with a wall formed by three tunics (Fig. 1.3). The outermost fibrous tunic, composed of dense avascular connective tissue, has two very different components: the sclera and the cornea. The sclera forms the posterior portion and the bulk of the fibrous tissue. It is tough, opaque and the site of attachment of the extrinsic eye muscles. Posteriorly, where the sclera is pierced by the optic nerve, it is in continuity with the dura mater of the brain. The anterior sixth of this fibrous capsule is modified to form the transparent cornea, which bulges anteriorly from its junction with the sclera. The cornea has no blood supply and obtains its nutrition from the aqueous humour, which fills the anterior segment of the eye and is continually being renewed every three to four hours. Within the sclera lies the vascular tunic or uvea, which has three distinct regions, the choroid, the iris and the ciliary body. The iris bounds the pupil, an anterior circular defect in the uvea. The iris is pigmented, rendering it opaque, and contains muscle fibres which on contracting and relaxing alter the size of the pupil. The choroid is a highly vascular dark brown membrane that forms the posterior five sixths of the uvea; the pigmentation prevents reflection or scattering of light within the eye. The uvea is continuous anteriorly with the ciliary body to which the suspensory ligament of the lens, the zonula, is attached. The lens, a crystalline structure isolated from the blood stream, like the cornea, lies immediately behind the iris and is suspended under tension by the zonula. Contraction of the ciliary muscle releases the tension in the zonula and increases the convexity of the

lens. The innermost tunic of the globe is the retina comprising two layers. The outer pigmented layer, a single-cell-thick lining, abuts the choroid and extends anteriorly to cover the ciliary body and the posterior surface of the iris. The transparent inner neural layer extends anteriorly only as far as the junction between the ciliary body and the choroid. The posterior segment of the globe is filled with a clear gel (the vitreous humour), which transmits light, supports the posterior surface of the lens, holds the neural retina firmly against the pigmented layer and contributes to the intraocular pressure, which helps to counteract the pulling forces of the extrinsic eye muscles (Berry et al. 1995).

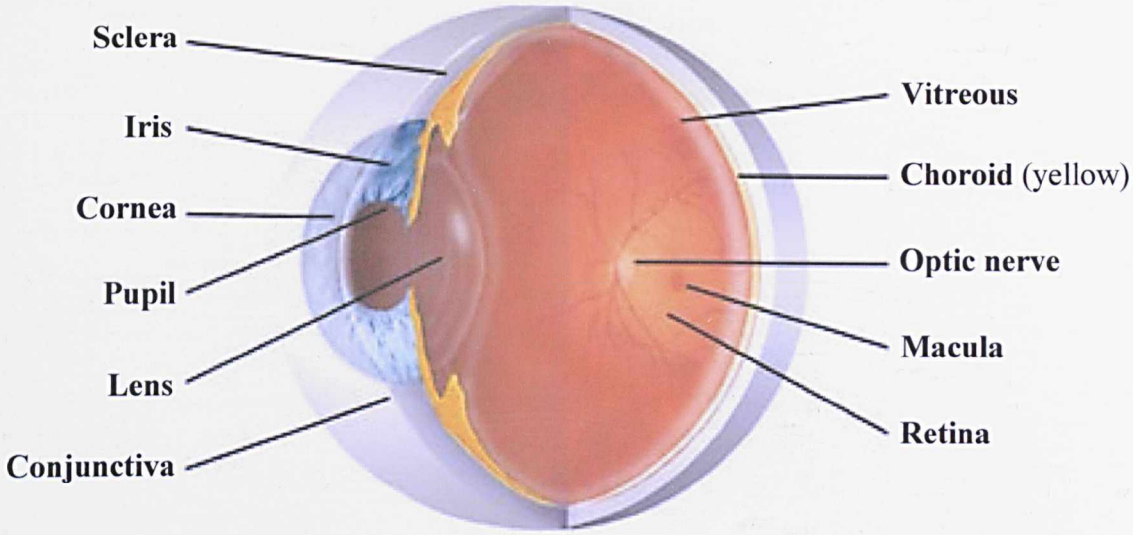


Figure 1.3: Gross anatomy of the eye

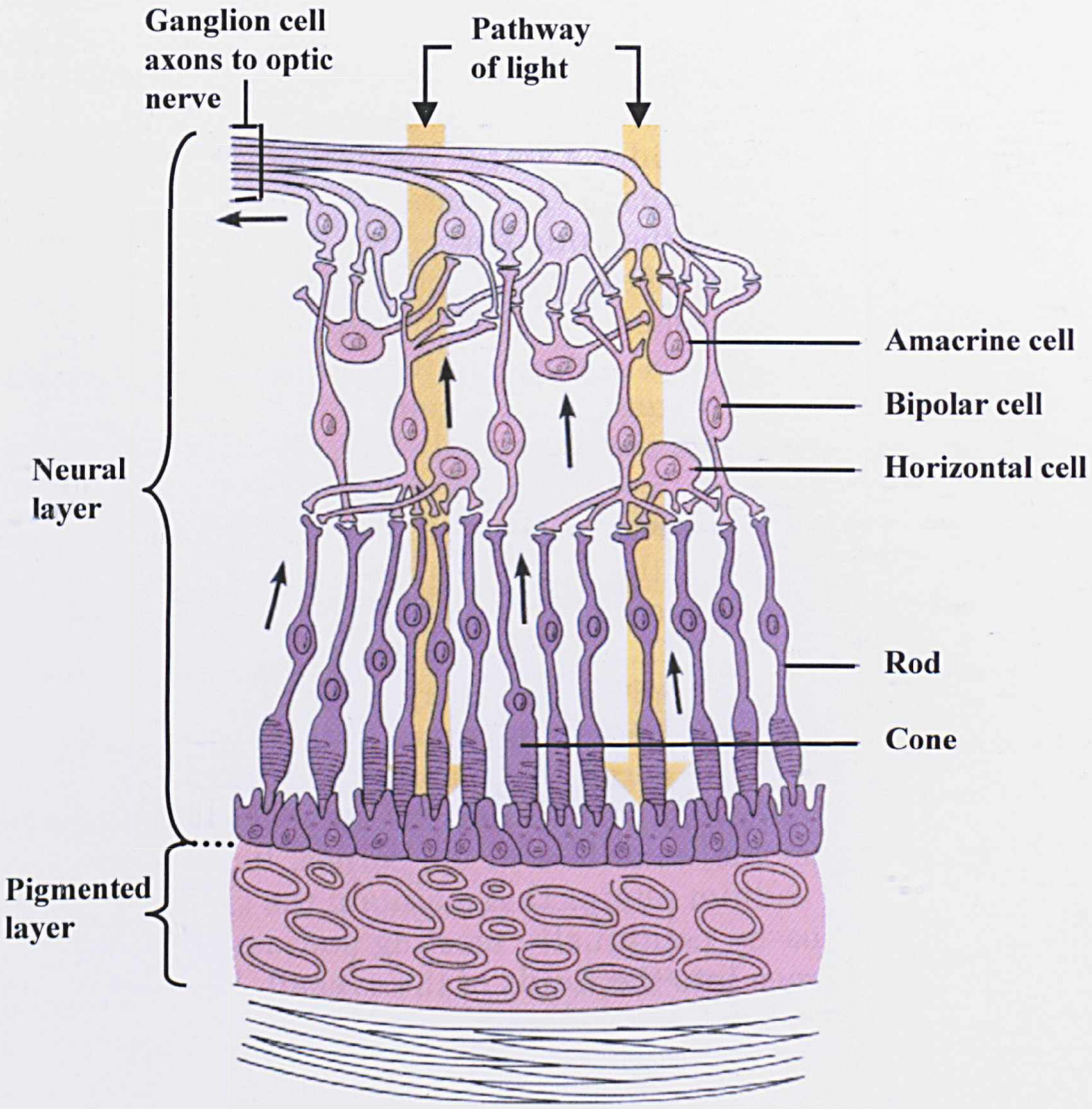


Figure 1.4: Cellular anatomy of the retina

1.3.2 The retina

1.3.2.1 Overall structure

The retina forms the inner surface of the eye and is divided into outer pigmented and inner neural layers (Guyton & Hall 1996b). The retina can be considered to be anatomically and functionally inside out; the embryological development of the retina is as an out-pouching of the brain that has become sensitive to light. Light must travel through a network of nerves and vessels before reaching the light sensitive cells. The functional deficit that would result from this arrangement is to some extent offset by the nerve fibres from the periphery skirting around and avoiding the central foveal region where the density of light receptors is greatest and the best resolution is obtained. The light receptor cells are linked to ganglion cells via bipolar cells. Each bipolar cells receives signals from multiple adjacent light receptor cells and their outputs are then further integrated by ganglion cells from which the final nerve fibres that leave the eye are derived (Fig.1.4). This hierarchy of nerve cells reduces the input from $120x10^6$ receptors to an output of $1x10^6$ nerve fibres, reducing the required size of the optic nerve 120 fold . There are two types of light sensitive receptor in the retina: those that respond to different wavelengths of light (i.e. therefore colour) called cones, and those that respond to intensity of light called rods. This nomenclature is based on the histological appearance of the receptors.

1.3.2.2 Rods and Cones

The photoreceptor cells contain light sensitive pigments, which are "bleached" on exposure to light setting up a chain reaction that results in the firing of the stimulated receptor. The pigment found in rods, rhodopsin, is deep purple in colour and absorbs light throughout the visual spectrum providing no information as to the wavelength of incident light. Cone cells contain one of three different pigments, each sensitive to a specific range of wavelengths of light, making them colour specific. The three types of cone cell can distinguish the three optical primary colours, red, green and blue; combinations of these primary colours can produce any colour in the visual spectrum. The threshold of light intensity at which a cone will fire is much greater than that for a rod; rods provide most of the visual information at low light intensities and cones become dominant at greater light intensities.

1.3.2.3 Fovea centralis and macula lutea

The distribution of photoreceptor cells in the retina reflects the greater importance attached to the higher spatial resolution, colour discriminating, cone cells in visual perception. The macula is an oval shaped condensation of, primarily, cone cells. At the centre of the macula, at the focal point of the lens, is a 0.4mm area termed the fovea, containing exclusively cone cells at the greatest concentration in the retina. The density of cone cells gradually declines from the edge of the macula to the retinal periphery where only rod cells are found. The cone density required for detailed colour vision is only found at the fovea. Fixation of an object results in the projection of that object onto the fovea.

1.3.3 Contrast perception

A given region generally looks brighter the darker the surrounding area is. Perception of grey scale is not absolute but relative. This is due to the cross-connections of retinal receptors and the importance placed by retinal processing on the perception of boundaries. The brain requires little information with regard to regions of uniform intensity and this is largely extrapolated from one border to the next, cutting down on visual processing (Gregory 1998). Therefore the most significant aspect of contrast sensitivity is in the perception of contrast boundaries. A gradual alteration of intensity may be completely ignored if the boundaries at each end fail to highlight the change. The contrast threshold for a static edge at photopic luminances is 1% (Draper 1997).

1.3.4 The extrinsic eye muscles

Each eye has six extrinsic eye muscles that originate at the bony orbit and insert into the outer surface of the globe (Fig. 1.5). They allow the eyes to follow a moving object and provide external support that help to maintain the shape of the eyeball and hold it in the orbit. Four of the muscles originate at a common tendinous ring, the annular ring, at the back of the orbit and run directly to their insertion on the globe. Their names, superior, inferior, lateral and medial rectus muscles are anatomically and functionally self-explanatory. The remaining two muscles are the superior and inferior oblique muscles. The superior oblique muscle rotates the eye downward and laterally; it originates in common with the rectus muscles and runs along the medial wall of the orbit and inserts on the superolateral aspect of the globe having passed through a fibrocartilaginous loop. The superior oblique muscle. The inferior oblique muscle rotates the eye up and laterally;

it originates from the medial orbit surface and runs lateral and obliquely to insert on the inferolateral eye surface.

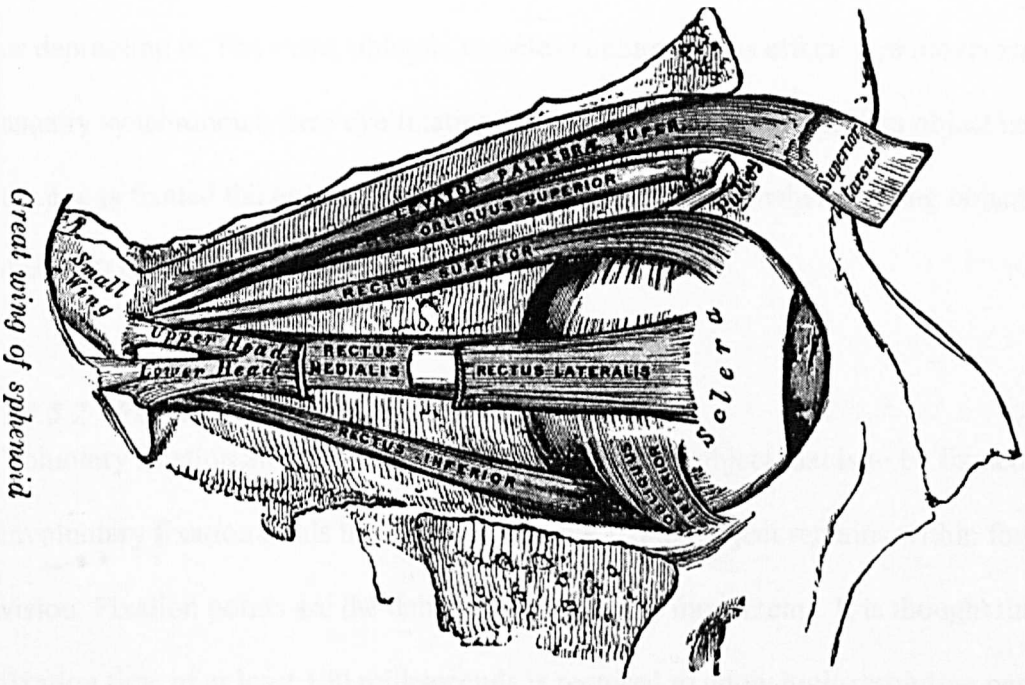


Figure 1.5: Anatomy of the ocular muscles.
Note the course of the superior oblique muscle as its tendon passes through a sling before inserting onto the globe. *Taken from Gray's Anatomy ©1995 Churchill Livingstone.*

1.3.5 Eye movements

1.3.5.1 Muscular control

The abducens and trochlear nerves innervate the lateral rectus muscle and the superior oblique muscle respectively. The oculomotor nerves innervate the remaining extrinsic eye muscles. The need for six muscles when four should be sufficient is a result of the mechanics of the inferior and superior recti, which rotate the eye medially while elevating or depressing it. The extra, oblique, muscles counteract this effect. Eye movements are usually synchronous, each eye fixating the same object. However, if an object nearer to the eye is fixated the eyes converge and conversely diverge when fixating objects more distant (Guyton & Hall 1996c).

1.3.5.2 Fixation movements

Voluntary fixation allows the eyes to move to find the object that is to be fixated on; involuntary fixation holds the eye such that the fixated object remains within foveal vision. Fixation points are the delimiters of saccadic movements. It is thought that a fixation time of at least 100 milliseconds is required to allow high-resolution perception from the macula (Guyton & Hall 1996c).

1.3.5.3 Voluntary fixation and saccadic movements

Voluntary fixation movements are controlled by a small cortical field in the pre-motor cortical regions of the frontal lobe, distant from the visual cortex and brainstem, and override the involuntary fixation movements when an alternate fixation point is to be found. Voluntary fixation movements are saccadic; jerky movements that quickly move the eye from one fixation point to another, allowing rapid foveal interrogation of multiple

sites in the field of view. Saccadic movements occupy only 10 per cent of the total time spent in moving the eyes, the remaining 90 per cent is allocated to fixation movements. During saccadic movement the visual signals from the retinal are suppressed such the movement is not perceived visually, resulting in a stable field of view (Guyton & Hall 1996c).

1.3.5.4 Involuntary fixation

Involuntary fixation is regulated by a negative feedback mechanism that prevents the object of attention from leaving the foveal portion of the retina. The eyes normally have three types of continuous but almost imperceptible movements; continuous tremor at a rate of 30 to 80 Hz caused by successive contractions of motor units in the ocular muscles, a slow drift of the eyeballs and sudden flicking movements controlled by the involuntary fixation mechanism. Tremor movements do not usually affect fixation but the drifting movement can move the object of fixation outside the foveal region at which point there is an automatic flicking movement that returns the fixated object's image to the centre of the fovea (Guyton & Hall 1996c).

1.3.5.5 Pursuit movements

Pursuit movements enable the eye to remain fixed on a moving object. A highly developed cortical mechanism automatically detects the course of movement of an object and then gradually develops a similar course of movement for the eyes. The greater the length of time of pursuit the smoother the pursuit movement becomes (Guyton & Hall 1996c).

1.4 EYE TRACKING

1.4.1 History and development

As early as 1936, Mowrer succeeded in making automatic recordings of the orientation of the eye in the head, and thus the direction of gaze (Glenstrup & Engell-Nielson 1995). The techniques for making these recordings improved only gradually during the 20th century, but recently they have become sufficiently non-intrusive to allow investigation outside the laboratory. Eye tracking allows testers to identify what participants look at, providing an insight into the cognitive processes being employed in image or scene assimilation. Eye tracking equipment may use several different technologies, including skin electrodes, marked contact lenses, image processing cameras, and reflector trackers. All techniques require some form of calibration before usage, as well as periodic re-calibration during use.

1.4.2 Techniques based on reflected light

The video-based eye movement measurement technique, designated as video-oculography (VOG), was first developed by Dr. Clarke and his co-workers at the Vestibular Research Laboratory of the Freie Universität Berlin. His work was largely funded by the German Space Agency to fulfil the requirements of clinical and research studies of eye movements related to the vestibular and oculomotor systems (Clarke & Baartz 1998). The technique lends itself seamlessly, following calibration, to eye-tracking recording and as such has developed into a clinical tool.

1.4.2.1 Limbus tracking

The *limbus* is the boundary between the white sclera and the dark iris of the eye. Due to the fact that the sclera is (normally) white and the iris is darker, this boundary can easily be optically detected and tracked. This technique is based on the position and shape of the limbus *relative to the head*, so either the head must be held quite still or the apparatus must be fixed to the user's head.

Due to the more or less occasional covering of the top and bottom of the limbus by the eyelids, "it is probably fair to regard limbus tracking as suitable for precise horizontal tracking only" (Glenstrup & Engell-Nielson 2002).

1.4.2.2 Pupil tracking

Similar to limbus tracking, pupil tracking uses the still smaller boundary between the pupil and the iris and the apparatus must be held completely still in relation to the head. The advantages of this technique over limbus tracking is that the pupil is far less covered by the eyelids than the limbus, enabling vertical tracking and the border of the pupil is often sharper than that of the limbus yielding a higher resolution. However, the difference in contrast between the pupil and iris is lower than that between the iris and sclera, making the border detection more difficult (Glenstrup & Engell-Nielson 2002).

1.4.2.3 Corneal reflection and eye image using an artificial neural network

This technique is based on a more *wide-angled* digitised video image of the user, so that the entire head is in the field of view of the camera. A stationary light is placed in front of the subject, and the system starts by finding the right eye of the subject by searching the

video image for the reflection of this light (the glint), distinguished by being a small, very bright point surrounded by a darker region. It then extracts a smaller, rectangular part of the video image (typically only 40 by 15 pixels) centred at the glint, and feeds this to an Artificial Neural Network (ANN) . The output of the ANN is a set of display co-ordinates. The ANN requires more than the simple calibration that is required by the other techniques; it must be trained by gathering images of the user's eye and head for at least three minutes while the user visually tracks a moving cursor on the display. This is followed by an automatic training session that uses the stored images lasting approximately 30 minutes using the current technology, but then the system should not require re-calibration on the next encounter.

The accuracy of the ANN-based system is not as good as for the other techniques; it can be improved slightly by augmenting the corneal/pupil-based calculations with a calculation based on the position of the glint in the eye socket, but it is still limited to an accuracy of about 1.5-2°. The great advantage of this technique is that the wide angle of the base image allows up to 30 cm of head movement, making this an attractive technique for situations in which head movement is inevitable, and where a high degree of accuracy is not essential (Glenstrup & Engell-Nielson 2002).

1.4.2.4 Purkinje Image tracking

The first and fourth Purkinje images can be used for tracking the direction of gaze by the *Dual-Purkinje Image* technique, which uses the relative positions of these reflections to calculate the direction. The Dual-Purkinje-Image technique is generally more accurate

than the other techniques, and the sampling frequency is high, up to 4000Hz. The disadvantage with this technique is that the fourth Purkinje image is rather weak, so the surrounding lighting must be closely controlled (Glenstrup & Engell-Nielson 2002).

1.4.2.5 Corneal and pupil reflection relationship

When infrared light is shone into the user's eye, several reflections occur on the boundaries of the lens and cornea, the so-called *Purkinje images* (Fig. 1.6). The first Purkinje image is also called the *glint*, and this together with the reflection of light off the retina, the so-called *bright-eye*, can be video-recorded using an infrared sensitive camera as a very bright spot and a less bright disc, respectively. When the eye is panned horizontally or vertically, the *relative* positioning of the glint and the centre of the bright-eye change accordingly, and the direction of gaze can be calculated from these relative positions (Fig. 1.7) (Hall 1983).

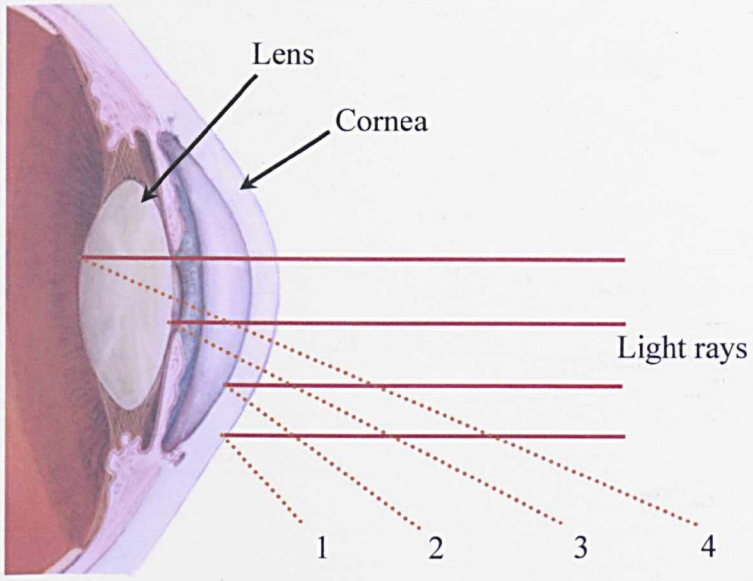


Figure 1.6: Source of the 4 Purkinje images

Light is reflected at both surfaces of the cornea and lens, the strongest, and therefore easiest to identify, reflection is that from the external surface of the cornea and is termed the glint (Purkinje 1 reflection)

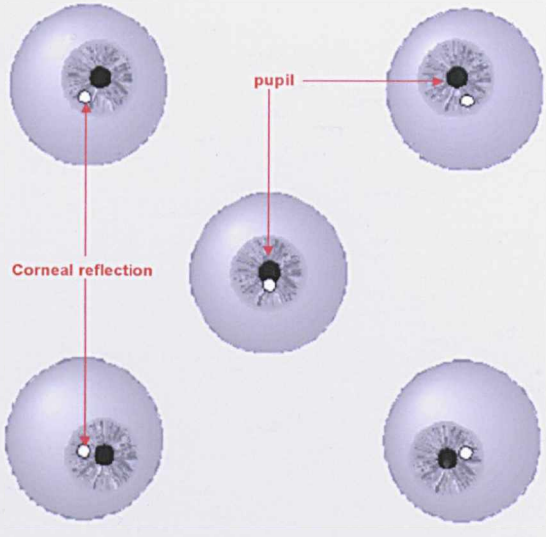


Figure 1.7: Pupil and corneal reflections (the glint).

As the eye moves the position of the glint alters in relation to the pupil. With calibration and a stationary illuminating source, the relative position of the glint and pupil can indicate the direction of vision.

The eye tracking equipment used in the experiments described in this thesis is based upon this particular technique. The problems associated with the technique are primarily those of establishing an unimpeded view of the eye; lateral head movement can put the video image of the eye out of focus, or even make the image of the eye fall out of view of the camera. In general, spectacles do not pose a problem unless there is significant glare from the lenses which obscures the corneal or pupillary reflections (Glenstrup & Engell-Nielson 2002).

1.4.3 A technique based on electric skin potential

The *electro-oculography* tracking technique is based on the fact that an electrostatic field rotates along with the eye. The position of the eye can be detected by recording quite small differences in the skin potential around the eye using electrodes placed on the skin around the eye. This technique does not require a clear view of the eye, which results in a large dynamic range of approximately ±70° but accuracy is poor (Glenstrup & Engell-Nielson 2002).

1.4.4 Techniques based on contact lenses

By making the user wear a special contact lens, it is possible to make quite accurate recordings of the direction of gaze. Two lens techniques exist: by engraving one or more plane mirror surfaces on the lens, reflections of light beams can be used to calculate the position of the eye. And by implanting a tiny induction coil into the lens, the exact positioning of the lens can be recorded through the use of high-frequency electromagnetic fields placed around the user's head.

These techniques are inappropriate for routine use: the health issues concerning high-frequency electromagnetic fields have not yet been resolved, let alone the requirement that the user must wear special contact lenses, possibly connected to wires, which makes the technique very cumbersome outside a laboratory setting (Glenstrup & Engell-Nielson 2002).

1.4.5 Applications of eye tracking

1.4.5.1 General

The applications of eye-tracking are far reaching. Eye movement data has been used to analyse the neurophysiology of vision with clinical applications including the identification of schizophrenia and improvements in the accuracy of ocular surgery. Many tasks require the assimilation of visual stimuli and in some instances the time frame allowed is very short. For example, pilots flying supersonic jets will travel a distance beyond their visual range in the time it takes them to look at a cockpit dial. Although head up displays circumvent this problem, analysis of eye tracking behaviour helps identify potentially disastrous distracting visual stimuli. This form of analysis has lent itself to the assessment of many visually demanding tasks and to the improvement of human-display interfaces. Eye tracking enables actions to be instigated by eye movements when in a suitably automated environment; applications for such technology would be relevant where an operators hands are otherwise occupied.

1.4.5.2 In radiology

As a primarily "visual" speciality, radiology lends itself to the analysis eye movement. To date the visual tracking of radiologists' approaches to a variety of diagnostic images has been assessed, primarily to identify reasons for false positive and false negative reports concentrating on perceptual errors (Kundel, Nodine, & Carmody 1978;Kundel, Nodine, & Krupinski 1989) but also to compare different viewing techniques such as "hard" and "soft" copy reading of CT imaging (Beard et al. 1990;Beard et al. 1994b), different display technologies in mammography (Beard et al. 1997) and determining the best viewing formats for magnetic resonance imaging studies (Niimi et al. 1997). Most work to date has been performed in relation to chest radiographs (e.g. the enhancement of nodule detection by identifying areas of prolonged visual dwell time (Krupinski, Nodine, & Kundel 1993a;Kundel, Nodine, & Krupinski 1990)), mammograms (Krupinski 1996;Krupinski & Nishikawa 1997) and radiographs for trauma (Berbaum et al. 2001).

1.5 VISUAL PERCEPTION

There are conflicting accounts of perception, which divide broadly into active and passive (Gregory 1998). The passive school of thought is that there is a passive acceptance of what is out there, as a window facing the world. However, a passive account fails to explain how we perceive objects from very sketchy basic components such as vague lines or dots. The favoured account, based on the assertions attributed to Hermann von Helmholtx (1821-1894), describes perception as 'unconscious inferences' from sensory data to what may be out there. The Helmholtzian theory permits the incorporation of dynamic brain processes, employing knowledge stored from the past, to see the present and predict the immediate future.

The 'intelligent eye' philosophy, largely derived from Helmholtzian theory, states that visual and other perception is intelligent decision-taking from limited sensory evidence. Essentially, sensory signals are considered inadequate for direct or certain perceptions, so intelligent guessing is needed for seeing objects. Perceptions are therefore predictive, never certain, hypotheses of what may be out there and the brain must decide what interpretation to make regarding a retinal image, which could be produced by an infinity of sizes, shapes and distances of objects. It is remarkable that the brain only rarely presents an ambiguous interpretation and, at the same time, worrying as there is no facility to consciously over-ride an unambiguous assertion. Numerous theories and models have been proposed to explain the mechanisms of human visual perception but none provides a complete answer (Ginsburg 1988). In the field of radiology there are

numerous examples of an abnormality overlooked on one occasion but readily identified by the same observer subsequently, the "how did I miss that?" scenario. Understanding the impact of pattern learning, visual cues and scanning paths used in the interpretation of an image may help to reduce the errors caused by the 'intelligent eye'.

1.6 COMPUTED TOMOGRAPHY OF THE LUNGS

1.6.1 Principles of CT

Computed tomography depends upon the scattering of x-rays as they pass through the subject being scanned; the amount of scatter is proportional to the tissue density and will therefore vary depending on the part of the subject being traversed by the x-rays. An x-ray beam with a restricted thickness (collimated) traverses the subject being scanned and the proportion of the x-ray beam still lying within the predetermined finite thickness (i.e. that which has not been scattered or absorbed) on exiting the subject is detected by sensors arranged in a curved array on the opposite side of the CT gantry to the x-ray source. The "view" of the sensors is restricted to the thickness of the original collimated x-ray beam. In a single position this technique merely provides a measure of the average tissue density of the subject lying within the defined x-ray beam thickness. However, by repeating the procedure at multiple different angles around the subject but in the same axial plane, a data set of tissue densities corresponding to the entire 'slice' mapped out by the rotation of the x-ray beam is collected. Using back projection (a mathematical technique) an image of the scanned slice is created with each picture element, or pixel, assigned a value corresponding to the average density of tissue in a small block of tissue. As each pixel represents a three-dimensional block, or volume, of tissue (the third dimension being the slice thickness) they are referred to as voxels. The value assigned is measured in Hounsfield units (HU), named after Godfrey Hounsfield the pioneer of CT scanning, with water as reference assigned a value of zero. The theoretical HU value of a tissue is defined by its linear attenuation coefficient (μ) by the following formula.

$$\mathbf{HU} = \frac{(\mu_{tissue} - \mu_{water}) \times 1000}{\mu_{water}}$$

The two-dimensional data set is then displayed as an image with the brightness of each pixel corresponding to the Hounsfield unit number of the relevant voxel.

1.6.2 High-resolution computed tomography

As the use of CT has developed in clinical practice and CT technology has improved, refinements have been made with scanning protocols tailored to specific imaging tasks. In 1985, 11 years after the first clinical CT scanners were developed, the concept of high-resolution CT (HRCT) was reported (Webb, Müller, & Naidich 2001d).

1.6.2.1 HRCT technique

The aim of HRCT is to optimise the demonstration of lung anatomy, particularly the smaller structures. The most important modifications to conventional CT technique in acquiring an HRCT image are thin collimation and image reconstruction with a high spatial frequency (sharp) algorithm (Griffin & Primack 2001).

1.6.2.2 Scan collimation.

In practice collimation of 1.5 mm or less is considered ideal for HRCT imaging (Webb, Müller, & Naidich 2001d) affording greater contrast between small vessels and surrounding lung parenchyma and detection of subtle changes in parenchymal density that are associated with early interstitial lung disease. The number of slices and radiation

dose resulting from contiguous HRCT (very thin) sections through the whole lung are prohibitively high. A compromise is therefore employed whereby discontinuous HRCT slices are separated by varying amounts of non-imaged lung. The disease processes typically examined with an HRCT examination, i.e. interstitial lung disease, are diffuse by nature and inter-spaced sections are sufficient to identify relevant features.

1.6.2.3 Window width and centre

The HU values of the tissues in the chest range from −1000 for air to +1000 for bone. The human eye can only detect up to one hundred shades of grey, which imposes limits on how much contrast information can be displayed in a useful form at any one time. The assignment of pixel brightness (shade of grey) to an HU number is determined by setting a window width and centre. The window width defines the range of HU numbers that will be represented by shades of grey and the window centre defines the HU value that is assigned the mid tone value of grey. Therefore, any pixels with HU values below the range defined by the window width (window centre value ± (window width)/ 2) are displayed as black and those above as white. The range of HU values found in chest CT requires the displaying of the image data twice, with window levels for soft tissue (width 400, centre 50) and for lung parenchyma (approximately width 1500, centre −500).

1.6.2.4 Reconstruction algorithm.

The inherent or maximum spatial resolution of a CT scanner is determined by the geometry of the data collecting system and the frequency at which scan data are sampled during the scan sequence (Webb, Müller, & Naidich 2001d). In HRCT scanning the

reconstruction algorithm used, the matrix size, and the field of view which affect the pixel size are optimised to increase the spatial resolution of the image. However, image resolution is always less than the inherent resolution. Conventional CT scans are reconstructed with smoothing algorithms that reduce visible image noise and enhance, to some extent, contrast resolution. A high-resolution (high spatial frequency) algorithm reduces image smoothing, increasing spatial resolution and making structures appear sharper.

1.6.2.5 Radiation exposure in HRCT

The amount of radiation exposure from HRCT scanning is dependent on the number of slices, the time taken to scan each slice and the amount of radiation present in the incident x-ray beam. Imaging with ionising radiation always involves a trade-off between image quality and exposure dose. The combination of thin collimation and a high spatial frequency reconstruction algorithm increase the visibility image noise, which represents the local statistical fluctuation in the CT numbers of individual picture elements of a homogeneous region of interest. The impact of noise is reduced by increasing the amount of "true" data through an increase in the number of photons that contribute to the "true" image data. The number of x-ray photons used to construct the image is dependent on the peak potential and current employed in the x-ray source and the time of exposure. Peak potential, measured in kilovolts, affects the energy profile and photon density of the x-ray beam. The greater the proportion of high-energy photons, the greater the number that traverse the tissue and reach the detectors. However, a proportion of photons must be absorbed or scattered outside the beam collimation limits in order to record information

about the tissue traversed and this is determined by the energy profile of the x-ray beam. The current represents the rate at which electrons are propelled toward the anode in an x-ray source and has a direct effect on the number of photons produced. The peak voltage and current can be readily increased when obtaining an HRCT, reducing image noise. However, with modern scanners this is rarely necessary and typical values are a peak voltage of 120-140 kVp and a tube current of 240 mA with a 1 second scan time per slice. The radiation dose resulting from HRCT is 10-20% of the radiation dose of conventional CT (Mayo, Jackson, & Müller 1993).

1.6.2.6 Artefacts in HRCT

Quantum mottle

Quantum mottle in HRCT is the variation of HU number resulting from the random spatial distribution of the x-ray quanta (an x-ray beam is not uniform but composed of multiple quanta of energy) and is most apparent when the amount of "real" information is at a minimum. Quantum mottle is the main contributor to image noise.

Beam hardening.

As an x-ray beam passes through a partially absorbing medium, the lower energy photons are preferentially absorbed. The result is a greater proportion of higher energy photons in the x-ray beam, which becomes more penetrating as a result. The proportion of the x-ray beam absorbed by a tissue prior to hardening would be greater than that absorbed by the same tissue after hardening. The non-uniformity of the x-ray beam energy profile will

result in erroneously low HU numbers for tissues following beam hardening; the effect of beam hardening is most evident in regions adjacent to highly attenuating tissue, such as bone, e.g. beneath a rib.

Motion artefact

The acquisition of data for any single slice takes a finite amount of time. The reconstruction process assumes that the tissues in the scanned slice remain stationary during data acquisition. Any movement results in misregistration of image data giving a blurred appearance. Respiratory motion can be reduced by breath-holding, and the number of slices imaged in a single breath-hold can be tailored to the patients ability to hold their breath. Cardiac motion artefact causes images distortion in the lung parenchyma close to the heart and can only be reduced by decreasing the time taken to capture each section of a scan. Effects of oesophageal motility are largely insignificant.

Aliasing artefacts

Aliasing artefacts are the result of under-sampling of data and / or insufficient projections performed on the data. The result may be a fine banding artefact or moiré pattern that may mimic lung disease.

1.7 IMAGE DISPLAY

1.7.1 Traditional display of diagnostic images.

The first radiographic images were displayed dynamically on phosphorescent screens, before the hazards of ionising radiation were known. Subsequent images were produced on film coated with x-ray sensitive emulsion. The radiation dose required for these early radiographs far exceeded what is now considered an acceptable level.

1.7.2 Current display technologies

1.7.2.1 "Hard copy"

The creation of "hard copy" x-ray images on film has undergone many improvements, the most significant of which was the introduction of screens. A screen lies next to the film. Comprised of high atomic number elements which phosphoresce on exposure to x-rays, the screen produces quanta of light in response to incident x-rays. The film, coated with emulsion sensitive to the wavelengths of light emitted by the screen records the image. The use of screen film combinations dramatically reduced the radiation dose associated with plain radiographs. The screen absorbs 90-95% of the incident x-rays and each incident x-ray creates multiple quanta of light resulting in a lower exposure requirement to achieve an image. However, the trade-off for the reduction in dose is the reduction in image quality; there is an inherent unsharpness resulting from film screen combinations and lower x-ray exposures increase image noise.

The advent of CT scanning, and then magnetic resonance imaging, created the need for converting digital images to "hard copy" film. A film coated with light sensitive emulsion is scanned in rows by a laser beam; the intensity of the laser at any given point, hence the amount of exposure, is defined by the brightness of the picture element being printed at that point. The resolution of the laser printer easily exceeds that of the original image.

1.7.2.2 "Soft copy"

The increase in the availability of rapidly advancing computer hardware has heralded the adoption of "soft copy" reporting. Images captured digitally are viewed and reported on via computer monitors. The implementation of digital or computed radiography lends itself to "soft copy" reporting and eliminates, in many cases, the need for "hard copy". A delayed phosphorescence initiated by incident x-rays and subsequently stimulated by a scanning laser beam produces sequential scintillation that is recorded by light sensitive circuitry and stored as a digital image. Like CT images, the digital plain film images have contrast resolution beyond the limits of the human eye, but windowing similar to that used in CT can enhance image interpretation. The diagnostic accuracy of viewing "as soft copy" CT images has been shown to be at least equivalent to viewing "hard copy" images (Krupinski et al. 1994;Reiner et al. 2001).

1.7.3 "Soft copy" reporting variables

1.7.3.1 Display formats

Comparison of multiple images, whether chronological sequences of plain images, different views of the same part of the body or multiple adjacent slices as in CT can be achieved by dividing the monitor screen into multiple smaller images and / or by having

more than one monitor. The viewing of multiple sequential CT images can be enhanced by viewing a single image at a time but having the capacity to rapidly change the viewed image to an adjacent slice, stacked mode, enabling the following of anatomical structures, in real time, through a three dimensional data set (Mathie & Strickland 1997).

1.7.3.2 Image windowing

The human eye can distinguish a maximum of one hundred shades of grey and computer monitor can typically only display up to two hundred and fifty-six different shades of grey. Therefore the range of values that may be assigned to an individual pixel in an HRCT image is approximately ten times that discernible by the human eye. Alteration of window settings enables the observer to visualise all the contrast information available in the image file rather than just the selected range available on a single hard copy. The information that can be gleaned from a given image is enhanced and ambiguity as a result of varying exposure parameters can be overcome. However, the alteration of window levels increases the time taken to read an image and failure to identify an abnormality only apparent on non-standard window settings might, in medico-legal terms, be considered negligent.

1.7.3.3 Monitor display settings

The monitors used for "soft copy" reporting need to have resolution and contrast ranges beyond human visual limits to ensure no loss of image quality when compared to "hard copy". The translation of digital data in the computer to analogue data displayed by the monitor requires digital to analogue conversion. The conversion process will vary

depending on a set of pre-defined weightings that converts the digital information into a picture that the human eye, which has a non-linear perception of tone, will perceive as being formed on a linear scale. The weighting values are expressed as a curve of weighting against grey level; the use of the Barten curve has been shown to be superior to a non-weighting altered display in viewing mammograms (Krupinski & Roehrig 2000); monitor luminance was shown to have a minor effect on performance in the same study. In general the monitor brightness and contrast settings should be standardised and most sophisticated soft copy reporting monitors have automatic luminance adjustment for changes in ambient light.

1.7.3.4 Spatial resolution.

The advent of digital imaging introduces the concept of an objective discernible spatial resolution. Ultimately the image is formed of discrete picture elements, pixels, the size of which determine the image resolution. The resolution limit of analogue images, conventional plain radiographs, is determined by the size of the silver halide crystals within the x-ray film emulsion. The minimum pixel size on a monitor screen is defined by the resolution of the monitor phosphor. High resolution monitors with pixel picture element sizes less than 0.25mm are of satisfactory quality for radiographic soft copy reporting, and the efficacy of using cheaper lower specification display hardware has been questioned (Kido et al. 1999).

1.8 SELECTED PULMONARY DISEASE

HRCT is the chosen imaging modality for analysis in this thesis. HRCT is most commonly used for the assessment of diffuse lung diseases as it provides greater parenchymal detail with high spatial and contrast resolution, compared to standard CT. The theoretical disadvantage of HRCT technique is the non-contiguous nature of the sections captured, effectively leaving 85% of the lung unexamined. The nature of diffuse lung and small airways disease is ideally suited to a non-contiguous approach to imaging as the parenchymal abnormalities detected on imaging are spread throughout the lungs and can be better characterised on HRCT sections.

1.8.1 Diffuse infiltrative lung disease

Diffuse infiltrative lung disease is a broad term that encompasses diseases that result in an increase in abnormal tissue in the lung, predominantly in the interstitium. At one end of the spectrum there is destruction of alveolar walls resulting in non-functional cystic airspaces within a predominantly fibrotic lung parenchyma termed end-stage or honeycomb lung. At the other end of the spectrum an exclusively inflammatory infiltrate, which responds to anti-inflammatory treatment, may resolve with no subsequent long-term lung parenchymal damage. In practice, most diseases causing diffuse interstitial lung disease lie somewhere between these two extremes.

1.8.1.1 HRCT-pathology correlation.

The resolution of histology slides far exceeds that of HRCT. Histologically identified infiltrates that are below the resolving power of HRCT will result in an overall increase in parenchymal attenuation. The result is "ground glass opacity" characterised by an increase in lung parenchymal attenuation with no obscuration of other parenchymal features. However, there are similarities to be found in the gross histological and HRCT appearances in diffuse lung diseases and the anatomical distribution of abnormal tissue provides valuable clues as to the likely aetiology (Colby & Swensen 1996). A schematic of an open lung biopsy is presented in Figure 1.2 with the relevant anatomy labelled. Furthermore, ancillary signs such as dilatation of the airways, traction bronchiectasis or bronchiolectasis, helps differentiate inflammation from fibrosis as the cause of "ground glass" opacity (Remy-Jardin et al. 1993)

Centrilobular distribution

When the pathological infiltrate is predominantly at the bronchiolar lumen and walls, a centrilobular nodular pattern is seen on HRCT (Fig.1.8). Partial filling of the central alveoli or thickening of the alveolar walls will be below the resolving power of HRCT and present as background ground glass opacity.

Lymphatic distribution

Such a distribution results from a disease that shows a predilection for the lymphatics, which are found along the veins and bronchovascular bundles, in the pleura and in the septa. There may be a diffuse infiltration of the lymphatic channels themselves or more

nodular follicular peri-lymphatic infiltrates. The HRCT appearances clearly reflect the expected distribution of lymphatic channels (Fig.1.9).

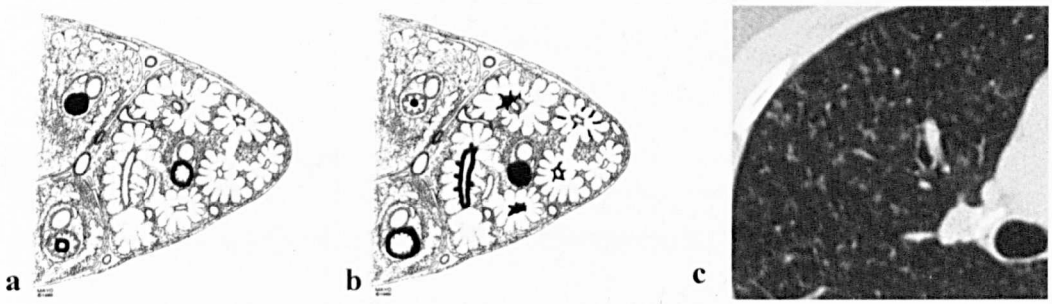


Figure 1.8: Centrilobular distribution.

(a) and **(b)** diagrammatic representations with abnormal tissue depicted in black and **(c)** an HRCT of a patient with pneumoconiosis demonstrating centrilobular nodularity.

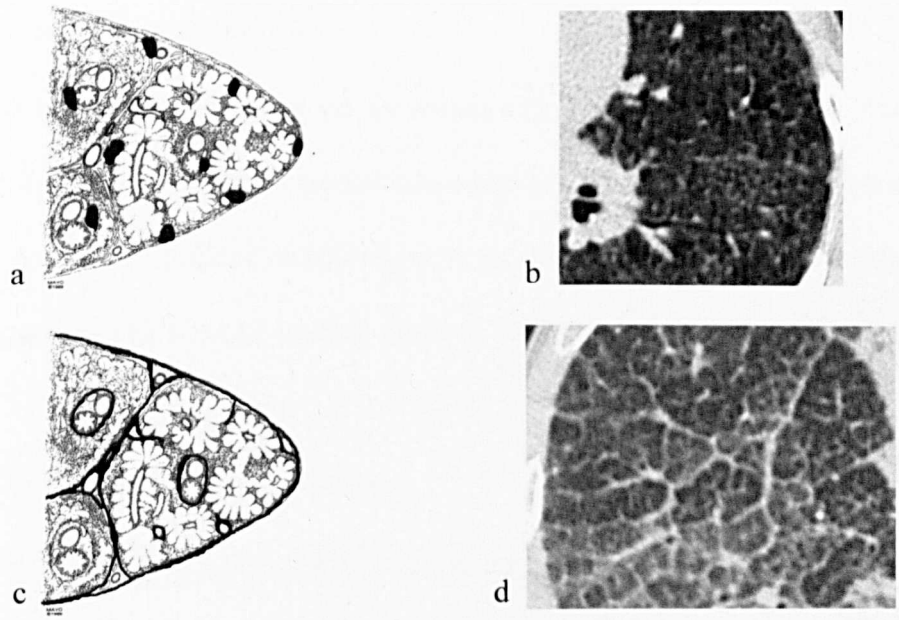


Figure 1.9: Lymphatic distribution

Nodular infiltrates **(a)** with HRCT **(b)** showing micro-nodularity and diffuse infiltrates **(c)** with thickening of the interlobular septa on HRCT **(d).**

Peripheral acinar distribution

The peripheral acini are found adjacent to the pleura, interlobular septa and bronchovascular structures proximal to the respiratory bronchioles. Therefore a peripheral acinar distribution may superficially resemble a lymphatic distribution. Often below the spatial resolution of HRCT a peripheral acinar distribution of disease if identifiable will be most obvious in the regions where the peripheral acini are more numerous, in the periphery of the lung (Fig.1.10).

Parenchymal consolidation

Complete filling of the alveolar air spaces will result in consolidation. The appearances on HRCT are of increased attenuation obscuring the vessels (Fig.1.11) but accentuating the unaffected air filled conducting airways which remain of very low attenuation contrasting with the surrounding increase in attenuation (air bronchograms).

1.8.2 Obstructive lung disease.

During inspiration a negative pressure is created within the thoracic cage drawing air into the lungs. The negative intra-parenchymal pressure relative to the pressure in the conducting airways helps to maintain the airway patency. However, during expiration the intrathoracic pressure is increased, the pressure in the lung parenchyma will exceed that within the conducting airways, which then have a tendency to collapse. The walls of the larger conducting airways contain cartilage and so retain their patency during expiration

but the smaller non-cartilaginous bronchioles have no cartilage in their walls and rely on the elasticity of the surrounding lung parenchyma to remain patent. There is a general reduction in calibre of the airways during normal expiration and pre-existing airways narrowing increases the likelihood of this physiological reduction in calibre causing significant airflow limitation. Therefore airflow obstruction or limitation may result from the destruction of the supporting elastic parenchyma or pathological narrowing of the airways (Stulbarg & Frank 1998).

1.8.2.1 Emphysema

Emphysema is a disease characterised by dilatation of the alveolar spaces and destruction of the alveolar walls (Fig. 1.12), resulting in the loss of much of the elastic recoil of the lung (Gurney 1998).

Compliance of the lung in emphysema is significantly above normal; the lung becomes easy to distend but empties slowly. This results in a chronically overinflated lung (high total lung capacity, functional residual capacity, and residual volume), which lessens the curvature of the diaphragm, making it less efficient in generating even the small swings in pleural pressure, necessary for optimal respiration.



Figure 1.10: A peripheral acinar distribution

(a) subpleural scarring, **(b)** more severe involvement of entire lobule with destruction and honeycombing. **(c)** HRCT showing a honeycomb pattern and peripheral acinar shadowing in adjacent tissue

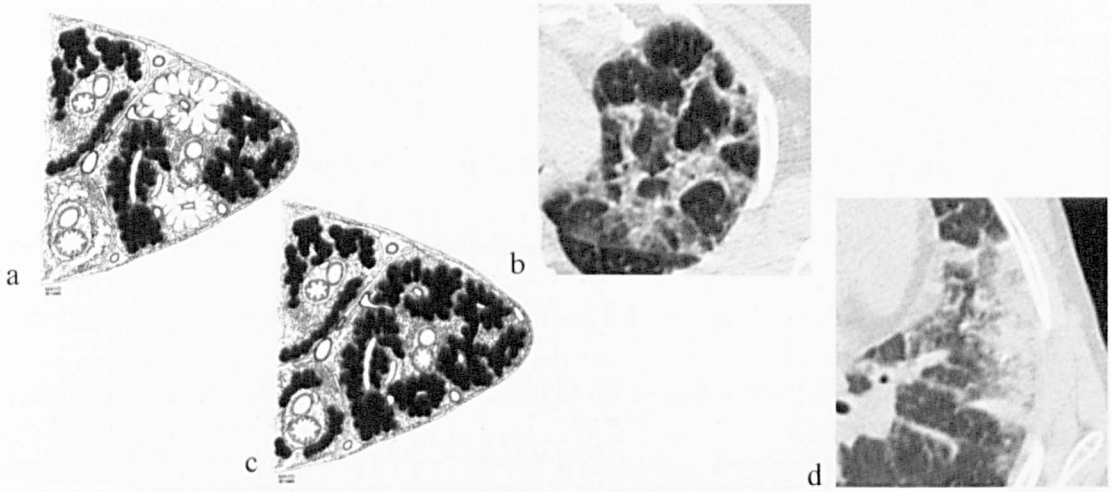


Figure 1.11: Parenchymal consolidation

(a) patchy [**(b)** HRCT] and **(c)** diffuse [**(d)** HRCT]

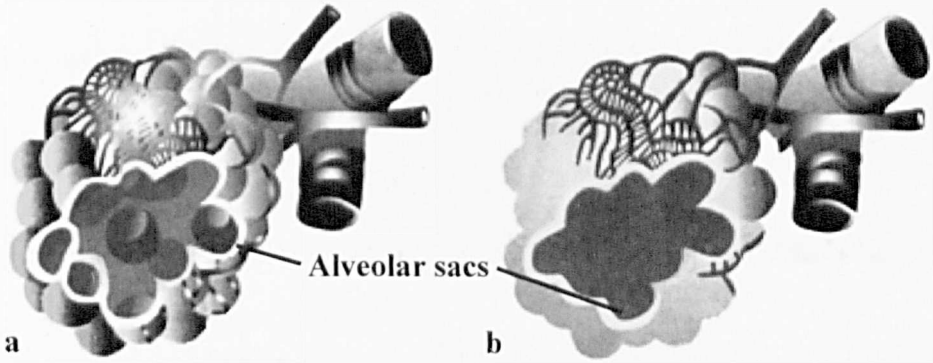


Figure 1.12: Emphysematous destruction

Normal lung **(a)** is compared with emphysematous lung **(b)** where the alveolar sacs have been damaged with destruction of the alveolar walls.

1.8.2.2 Small airways disease

The small non-alveolated conducting airways, terminal bronchioles, contribute only 20% of the overall airways resistance (Hogg, Macklem, & Thurlbeck 1968) and even large changes in the airflow conductance at this level in the bronchial tree can go undetected on lung function testing. Abnormalities of the small airways are found in a wide variety of disorders and the increasing array of reported causes and associations prohibits a manageable clinical classification. The histopathological sub-types described do not necessarily have clinical or imaging correlates and a simplification of the classification is needed. A division of small airways disease (bronchiolitis) into "proliferative" and "constrictive" is associated with recognisable and distinctive clinical and imaging features (Hansell 1999). The small airways in "proliferative" bronchiolitis are filled with polyps of granulation tissue that may obstruct the airway and go on to fill the distal air-spaces. In "constrictive" bronchiolitis the airway is narrowed, due to peribronchiolar fibrosis resulting from some form of bronchiolar injury, and may progress to complete obliteration. Subjects with apparently normal appearances on HRCT images taken in suspended inspiration may reveal significant small airways disease, manifest as air trapping, on images taken in suspended expiration (Arakawa & Webb 1998;Stern & Frank 1994)

1.8.2.3 Asthma and chronic bronchitis

Asthma and chronic bronchitis are both characterised by inflammation of the bronchi. In asthma there is a hypersensitivity reaction causing bronchospasm resulting in airflow

limitation. The severity of airflow limitation fluctuates with prolonged periods of minimal or no bronchospasm. Chronic bronchitis is a chronic inflammatory process of the bronchi, most often caused by cigarette smoking. The airflow limitation is less labile by comparison with asthma but insults to the airways such as infection can cause a worsening of the inflammation and exacerbate airflow limitation

1.8.2.4 HRCT-pathological correlation.

Emphysema

The gross histopathological finding in emphysema is destruction of the alveolar walls leaving conglomerate but disorganised air-spaces. On HRCT the result is a reduction in lung attenuation in those areas affected. Emphysema may be confined to the subpleural areas (paraseptal), the secondary lobules (centrilobular), more extensive areas of the lung parenchyma (pan-lobular) or form large cyst like air spaces (bullous). In paraseptal emphysema bubble like air spaces are seen along the subpleural surfaces. In centrilobular emphysema discrete areas of low attenuation, corresponding to the affected lobules, can be seen and give a somewhat moth-eaten appearance; the parenchymal vessels become attenuated and distorted due to the tissue destruction. Pan-lobular emphysema results in a widespread low attenuation with little lung parenchyma remaining; pulmonary vessels are attenuated but not as obviously distorted as in centrilobular emphysema. (Gurney 1998;Takasugi & Godwin 1998).

Small airway disease

In normal lung the distal small airways are too small to identify on HRCT. Regions of lung affected by small airways disease will have reduced flux of oxygenated air into the airspaces compromising the functional usefulness of the affected lung. In response, a physiological shunting of blood to more functional areas of lung occurs. On HRCT areas of decreased attenuation of the lung parenchyma, usually with poorly defined margins, correspond to regions of reduced blood flow (the ratio of air to blood in each voxel is greater) and adjacent areas into which the blood is shunted are of a slightly higher attenuation. The result is a patchy variation in lung attenuation for which the general term *mosaic attenuation* has been coined from the initial description of *mosaic oligaemia* (Eber, Stark, & Bertozzi 1993). The relative reduction in perfusion in the areas of decreased attenuation is initially due to hypoxic vasoconstriction (Gückel et al. 1999) but in the long term, vascular remodelling results in irreversible occlusion. A mosaic attenuation pattern due to small airways disease can be accentuated by obtaining an HRCT with the patient breath holding in expiration, as opposed to inspiration used in conventional HRCT (Arakawa & Webb 1998;Stern & Frank 1994). In expiration the expulsion of gas from normal lung parenchyma results in an increase in lung attenuation. Regions of lung affected by small airways disease will retain a greater proportion of gas due to trapping. As a result the increase in attenuation of lung affected by small airways disease, if present at all, will be less than that of unaffected lung enhancing the contrast in attenuation between the two and exaggerating the mosaic pattern (Hansell 1999;Worthy & Müller 1998)

Mosaic attenuation pattern

Regional inhomogeneity of the density of the lung parenchyma on HRCT has been described as a mosaic attenuation pattern and is a non-specific finding. A mosaic attenuation pattern can reflect the presence of occlusive vascular disease, airway abnormalities (Eber, Stark, & Bertozzi 1993) or parenchymal infiltration (Stern et al. 1995b;Stern et al. 1995a). It is possible in most cases to distinguish between the basic causes of mosaicism on CT scanning (Worthy et al. 1997). Quantitative measurement of the HU number of different regions of mosaic lung attenuation on conventional inspiratory scanning is not generally helpful, as lung attenuation shows substantial inter and intra-subject variability (Webb et al. 1993). However, certain extremes in lung attenuation can be helpful in distinguishing primary airway and vascular disease from an infiltrative process. Normal lung attenuation in adults ranges from –850 HU to –700 HU. Regions of lung that trap air or are under-perfused, can measure between –920 HU and –800 HU, whereas true infiltrative lung disease or hyperaemia can measure between –750 HU and ≥ -400 HU (Stern, Webb, & Gamsu 1994). Furthermore, in small airways and vascular diseases the vessels in the low attenuation lung are attenuated in number and size (Fig. 1.13) compared to those in the more opaque lung; such a difference is not found in infiltrative lung disease (Fig. 1.14). Distinguishing small airways disease from primary vascular disease may be aided by scrutiny of paired inspiratory and expiratory CT images (Arakawa et al. 2000) recommended as routine practice by some in the investigation of diffuse lung disease (Arakawa et al. 2000).However, recent evidence suggests air-trapping, and thus an accentuated mosaicism, is not uncommon in pulmonary embolic disease (Arakawa et al. 2002).

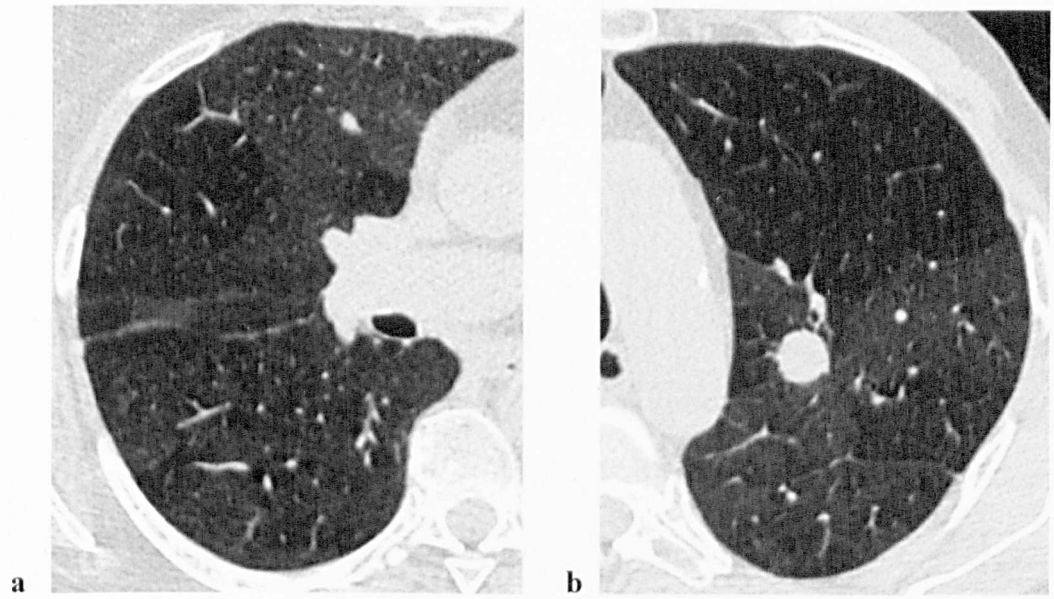


Figure 1.13: HRCT image demonstrating **(a)** mosaic attenuation due to air trapping and **(b)** due to chronic thrombo-embolic disease causing reduced perfusion. The darker lung is abnormal in both cases and the vessels in the abnormal lung are attenuated.

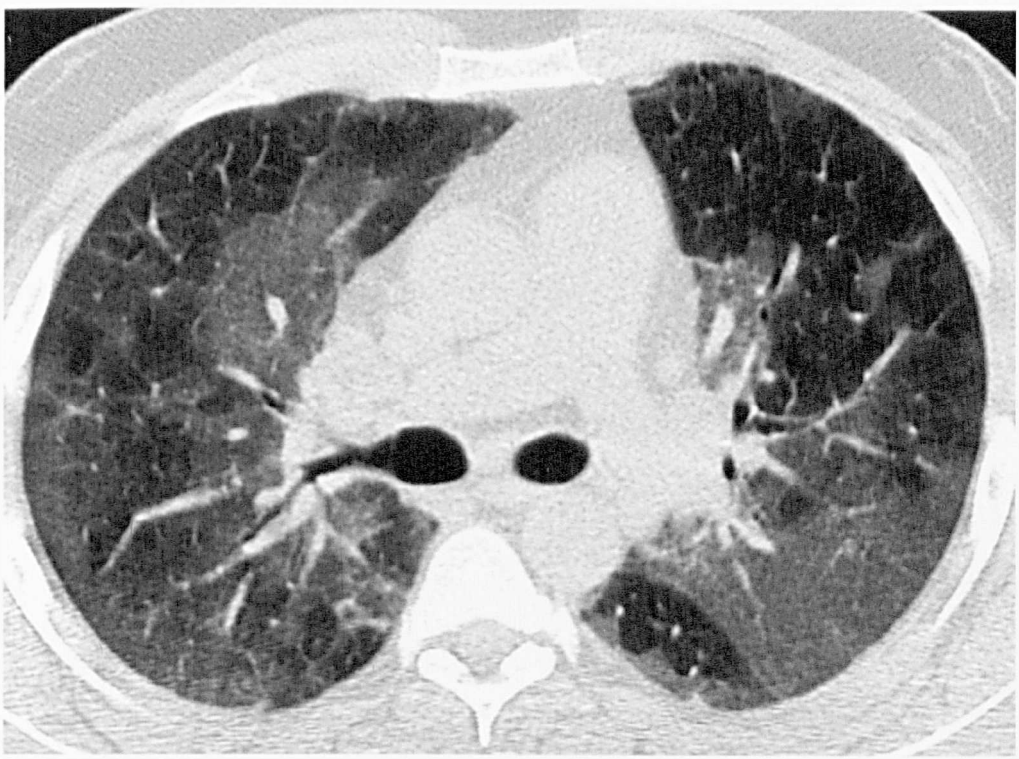


Figure 1.14: HRCT image demonstrating a mosaic attenuation pattern secondary to parenchymal infiltrates. Note there is no discrepancy in the vasculature of the different attenuation regions of lung

1.9 CONSIDERATION OF RADIOLOGIST PERFORMANCE

1.9.1 Training

Current training in radiology in the UK demands five years of practical experience and the passing of formal examinations set by the Royal College of Radiologists covering normal anatomy, radiography, radiation physics, diagnostic radiology and practical image viewing. The majority of the examinations are in the form of multiple choice question papers that are designed to test overall knowledge. The final exam concentrates on film reading ability and is arguably the most relevant to subsequent radiological practice. During training the trainees are exposed to a mixture of hands-on and observational teaching. The reading of CT imaging is learned primarily by observing consultants during reporting sessions and writing preliminary reports on which there is consultant feedback. Structured teaching sessions in preparation for the final exam are usual. Radiological interpretation has been taught in this fashion for many years but the number and complexity of radiological images with which the radiologist must become familiar is continuously increasing and requires the development of more streamlined teaching techniques. A systematic approach to image interpretation, taught for example by a computer program, introduced into the training scheme at an early stage would enable the trainee to develop practical interpretative skills rapidly. The learning by observation should remain a vital element of training but an automated learning process would enhance overall learning.

1.9.2 Reducing the sources of error in radiological interpretation

Errors in interpretation may be cognitive or perceptual. A cognitive error is when an observer incorrectly interprets what the observed abnormality is, due to faulty reasoning (Berlin 1996a). Perceptual errors occur when the observer fails to identify the abnormality (Berlin 1996b). Cognitive errors tend to result from a failure to consider a diagnosis rather than lack of knowledge (Berlin 1996a). Therefore cognitive errors can be reduced by obtaining as much clinical information as possible, reviewing previous imaging and reports (White, Berbaum, & Smith 1994), consulting with colleagues and making a concerted effort to consider alternative interpretations of abnormal findings observed before submitting a report. The blame and shame culture, that has developed within medicine, usually fails to identify, and therefore reduce the risk of, errors. The current emphasis in error reduction is to take a team approach in order to identify where the error managed to avoid contingencies designed to trap it. The cause may still be entirely attributed to a single observer's error but other failures in the system should also be sought and rectified (Fitzgerald 2001). Perceptual errors are harder to address and numerous factors such as image quality, viewing conditions, the nature of the lesion (i.e. its conspicuity) and inadequate search have been implicated (Brogdon, Kelsey, & Moseley, Jr. 1983). A possible cause of an inadequate search has been coined "satisfaction of search" where the identification of a significant lesion prematurely terminates a search resulting in the failure to identify a second lesion (Berbaum et al. 1990). However, the concept of satisfaction of search is not universally accepted (Nodine et al. 1992). In addition, it may be difficult to decide in retrospect whether or not a "missed" lesion truly represents an error (Berbaum 1995). Computer assisted diagnostic

systems have the potential to reduce the frequency of perceptual misses (Robinson 1997) and image processing such as digital subtraction masking with previous images (temporal subtraction) or dual energy subtraction radiography have been shown to be of value (MacMahon 2000).

1.9.3 Assessment

There is an increasing emphasis on the quality of care provided by the health service (Milburn 2002). The advent of clinical governance requires the setting of goals in all aspects of patient care, instituting the necessary processes to ensure those standards are met and implementing appropriate changes if they are not. In radiology the accuracy of image interpretation, the speed with which a radiological examination is undertaken and the communication of the radiological findings to the requesting physician are all aspects that fall within the remit of clinical governance (Ayers 2001). The advent of digital imaging and patient archive systems have a definite impact upon the communication of results to clinicians, and potentially increase the efficiency with which examinations are performed by the radiographers, no longer burdened with the printing and packaging of conventional film images.

The setting and achieving of goals in the delivery of an imaging service can be readily assessed and will usually involve altering manpower and technical resources. However, the assessment of cognitive and interpretative skills, which should inevitably become a part of the whole process of clinical governance in the setting of a radiology department, has significant implications with regard to the teaching methods employed in radiology

and the mechanisms by which standards are maintained. Methods for evaluating the competence of radiologists interpretative skills are being assessed (Cascade et al. 2001) and are likely to be increasingly used to ensure an acceptable level of competence.

1.9.4 Implications of quality assurance

The quality of radiographic images is unlikely to undergo significant improvement and most technical research is based around increasing the amount of image data without increasing the radiation dose incurred. As a result the interpretation of CT imaging will become more time consuming and complex with multiplanar reconstructions becoming the norm. As improvements in the images acquisition stage of the radiological service diminish and the communication with clinicians is optimised by picture archive and communication systems (PACS) the performance of the radiologist will increasingly become the focus of attention in attempts to improve the overall quality of the radiology service.

Demands to improve efficiency and accuracy in the face of increasing complexity of imaging and inadequate manpower provision creates an increasing need to optimise radiologists' performance. The adoption of computer aided diagnostic software to assist in the interpretation of radiological images has considerable appeal but improvement in training is necessary. Furthermore steps to analyse a radiologist's approach to images with a view to identifying correctable errors in visual search technique may become feasible; this is potentially the most controversial and intrusive consequence of quality assurance in radiology.

1.10 RADIOLOGICAL DIAGNOSIS

1.10.1 Components of a "radiological diagnosis"

A clinical diagnosis is a hypothesis based upon a number of assertions, the sources of which range from clinical history and examination to complex investigations. Each assertion can in turn be considered a hypothesis based on further data or assertions and an iterative hierarchy can be generated until the raw non-interpreted data is reached. A radiological diagnosis is usually a first order hypothesis, one of the first branches in a diagnostic hierarchical tree from the final diagnosis. Other first order hypotheses would include the clinical examination, blood results and lung function tests.

The information that contributes to a radiological diagnosis may include multiple imaging modalities, each of which will have its own hierarchy of hypotheses with the iterative processes ultimately ending with basic visual information. For this discussion the hypothesis based on an HRCT examination will be considered.

1.10.2 Aspects of an HRCT image that are used to reach a diagnosis

1.10.2.1 Basic visual information

Pixels

An HRCT image is constructed from pixels, each of which represents a voxel, the average density of which determines its HU number and thus the pixels grey level. HRCT images are not analysed at pixel level but areas as small as 4x4 pixels may impart useful information such as micronodularity.

Structure

Small picture elements such as lines, nodules and thickened airway walls may contribute to the diagnostic hypothesis. Structural elements are assigned as normal or abnormal to varying degrees and in combination may form an overall recognisable pattern.

Density

Alluded to above is the visual information imparted by pixel brightness; focal formed areas of increased density may be recognised as specific structures and considered in those terms. However, an overall impression of density, whether increased or decreased, provides diagnostic information such as found in a mosaic pattern of attenuation or emphysematous lung.

1.10.2.2 Constructing diagnostic hypothesis from an HRCT examination.

Analysis of picture elements

The basic visual information is considered "en masse" to construct an overall impression of the findings on the HRCT image. At this stage recognisable patterns are sought and picture elements that are likely to be related are identified, e.g. interstitial lines and ground glass opacity.

Construction of a working hypothesis

A working hypothesis, based on the visual patterns and other findings, is constructed using the knowledge base gleaned from training and experience.

Consideration of other available information

The working hypothesis may be altered or refined by the inclusion of information from other sources such as other imaging modalities, other investigations, from the patient or the clinician requesting the examination.

Weighting of diagnostic importance

Each item of information down to the most basic elements will to some degree require weighting as to the relevance and significance of the finding with regard to the working hypothesis. The weighting may therefore change if the working hypothesis is altered. Weighting will also include a consideration of the reliability of a finding. Every stage in the diagnostic hierarchy is subject to weighting and, as the number of contributing hypotheses to a given hypothesis increases, the reliability of that hypothesis becomes an increasingly significant feature. Ultimately the clinical value of the final radiological diagnosis will depend on the relevance and reliability of the investigation (Ker et al. 1988).

Relative importance of positive and negative findings

Findings that fail to support a working diagnosis are extremely valuable. There is a tendency to attempt to support a hypothesis by seeking information that can only confirm it, when findings that would disprove, or at least cast doubt on, the hypothesis would be more powerful. In this respect positive findings are necessary to formulate a hypothesis but negative findings will strengthen the final diagnostic hypothesis (Richmond 2002).

1.10.3 Hierarchical hypotheses theory for computer-aided diagnosis.

A simplified hierarchical tree is demonstrated in Figure 1.15. A computer-aided diagnostic system would have to start at the lowest level of the hierarchical tree and feed information to the radiologist at the level of pattern recognition or, if provided with a database of possible diagnostic alternatives, at the level of constructing a working hypothesis. The next stage would be automated diagnosis. This would require the computer-based algorithm to incorporate all the information in the hypothesis tree up to the level of a final diagnostic hypothesis. The complexity of the process at this stage should not be under-estimated (Piraino et al. 1989) and requires the formulation, adjustment and re-formulation of multiple contributory hypotheses all of which will have variable weighting depending on reliability, relevance and degree of positivity or negativity. The computer processes involved require a high degree of parallel processing, assimilating different data elements independently at the same time. The human brain is easily the most powerful parallel-processing computer available and computer technology is unlikely to come up with a feasible competitor for the brain for some time, if ever. Nevertheless, the field of computer-aided diagnosis is stimulating much research.

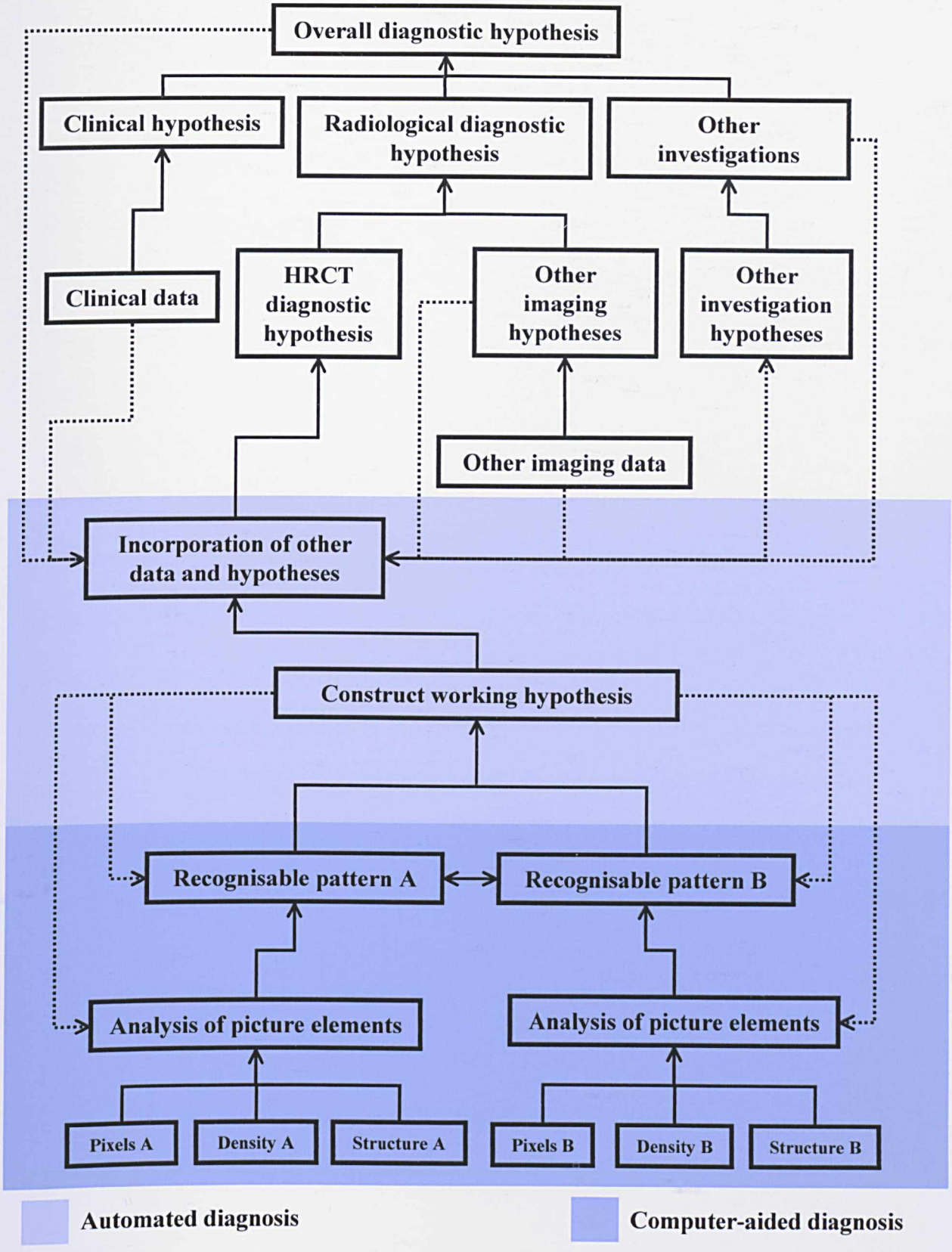


Figure 1.15: Schematic diagram of a simplified hierarchical hypothesis mode for constructing a diagnosis from an HRCT examination. Regions encompassed by computer aided diagnosis and automated diagnosis are shaded.

1.10.4 Diagnostic accuracy of HRCT

The greater spatial resolution of HRCT images is achieved at the expense of a reduction in anatomical coverage. Inter-spaced sections are scanned to reduced the overall scanning time (more of an issue ten years ago than now) and reduce the dose of ionising radiation associated with the examination (still an issue today). Early studies demonstrated the superiority of HRCT over chest radiography in the diagnosis of diffuse lung disease (Nishimura et al. 1993;Padley et al. 1991) and at least equivalence with conventional "thick" section CT (Leung, Staples, & Müller 1991). HRCT appearances have now been described in numerous pulmonary diseases (Swensen et al. 1992) including small airway diseases and thromboembolic disease; a 94-100% sensitivity and 96-98% specificity is reported for the accurate diagnosis of chronic thromboembolic disease (Bergin et al. 1996). The greater resolution of HRCT has enabled the correlation of HRCT appearances with histological findings (Colby & Swensen 1996), taken in most instances as the "gold standard", and thus provides a marker for the accuracy of the HRCT examination. In cases of diffuse interstitial lung disease the ability to depict disease distribution on HRCT enables a more accurate assessment of the histologic appearances with the two investigative modalities working synergistically.

1.11 COMPUTER AIDED DIAGNOSIS IN RADIOLOGY

Computer-aided diagnosis (CAD) may be defined as a diagnosis made by a radiologist who takes into account the results of a computer output as a "second opinion". The last ten to fifteen years have seen many attempts to develop computerised schemes for automated detection and quantitative analysis of abnormalities in radiographic images. The development of such CAD systems encompasses the fields of diagnostic radiology, medical physics, computer sciences and electrical engineering (Doi 1999).

1.11.1 Principles of CAD

The computer is used as a tool to provide additional information to the radiologist based on quantitative analysis of a diagnostic radiologic image; the final decision as to the diagnosis is determined by the radiologist taking into account the output from the computer. The concept of "automated diagnosis", investigated in the early 1960s and 1970s, is different to the principles behind CAD.

The aim of CAD is to improve the diagnostic accuracy and consistency of a radiologist's image interpretation. Inter and intra-observer variations in image interpretation (Robinson 1997) and the rate of false negative reporting are a long-standing cause of concern in radiology (Birdwell et al. 2001;Greene 1992), particularly in screening programmes (Warren Burhenne et al. 2000) where the likelihood of a positive diagnosis is small and the volume of work is large. Studies have identified numerous possible reasons for false

negative findings, such as overlap of anatomical structures, and eye-tracking experiments have demonstrated that lesions are not usually overlooked due to faulty scanning patterns but are scrutinised and discarded (Krupinski, Nodine, & Kundel 1993a;Tuddenham 1969).

The identification of regions requiring further attention can improve diagnostic sensitivity (Krupinski, Nodine, & Kundel 1993a;Krupinski, Nodine, & Kundel 1993b) without necessarily compromising specificity (Birdwell, Ikeda, O'Shaughnessy, & Sickles 2001;Warren Burhenne et al. 2000) but a beneficial effect will depend on the accuracy of the CAD scheme (Ikeda, Shigaki, & Yamauchi 2000). In terms of nodule detection on plain chest radiographs, a field that dominated the early work in perceptual analysis, it was found that longer dwell times correlated with regions that had an abnormality whether or not the abnormality was actually identified (Kundel, Nodine, & Krupinski 1989). The highlighting of regions on a chest radiograph subject to longer dwell times can reduce the number of false negative reports (Krupinski, Nodine, & Kundel 1993a) but requires the recording of eye-tracking data, and therefore is only useful as research tool. The principle of highlighting regions of particular interest to improve diagnostic accuracy becomes practicable when a readily available technique for identify these regions is available. This forms the core application of computer assisted image interpretation.

A CAD system may merely concentrate on increasing the conspicuity of potential abnormalities by image processing or in more complicated scenarios offer a possible identification.

1.11.2 Image processing and increasing conspicuity.

As a result of the increasing use of digital imaging in radiology, computer algorithms prior to, or during, interpretation can readily manipulate the imaging data. The simplest form of image processing can be performed by the radiologist manually, i.e. altering "window" levels, but more complex processes are becoming available particularly with regard to plain radiography.

1.11.2.1 Dual energy subtraction chest radiography

Dual-energy subtraction chest radiography exploits the selective attenuation of low energy photons by calcium containing structures. Two x-ray images of the chest are obtained representing exposures by incident x-rays at two fundamentally different energies. The most intuitive technique is to take two exposures with different kilovolt peak (kVp) settings (the determinant of x-ray energy produced). An alternative technique uses a single exposure of two phosphor plates separated by a copper filter. The first plate is exposed by the full spectrum of x-rays traversing the patient but absorption of the low energy photons by this first plate and the copper filter results in only the higher energy portion of the x-ray spectrum exposing the second plate. A digital subtraction of the two images results in the filtering out of calcium containing structures, such as ribs, increasing the conspicuity of soft tissue structures. The technique is already commercially

available and has been used with success (Kido et al. 2000;MacMahon 2000), but is not in general clinical use.

1.11.3 The basic components of a CAD scheme

Computerised schemes for CAD generally include three basic components based on three different technologies.

1.11.3.1 Image processing or enhancement

Image processing is employed to enhance and extract lesions. The processes are designed to facilitate computer analysis, not that of humans, and identify initial candidates for lesions and suspicious patterns. Subsequent analysis is then concentrated on these candidates. Image processing techniques include Fourier analysis (Panchangam et al. 2001), wavelet transformation (Yoshida et al. 1996), morphological filtering (Thiran & Macq 1996), difference image technique and artificial neural networks (Wu, Doi, & Giger 1995).

1.11.3.2 Quantitation of image features

The candidates selected by image processing are quantified such as by size, contrast and shape. It is possible to define numerous features by mathematical formulae that may not be readily understood by human observers, but in the early stages of CAD the development of imaging features already recognised and described subjectively by the radiologist is prudent. The most important development is the identification of a unique feature that can distinguish reliably between a lesion and normal anatomical structures.

1.11.3.3 Data processing

The distinction of normal from abnormal patterns based on the features obtained from quantitation requires data processing. The simplest method is based on rules established from an understanding of lesions and other normal patterns, which can be refined to improve specificity. Other methods include discriminant analysis, which takes into account the different variables of an object and works out which group the object most likely belongs to, Bayes' theorem, and artificial neural networks (Penedo et al. 1998)

Artificial neural networks

Also referred to as connectionist architectures, parallel distributed processing, and neuromorphic systems, an artificial neural network (ANN) is inspired by the way the densely interconnected, parallel structure of the mammalian brain processes information. Artificial neural networks are collections of mathematical models that emulate some of the observed properties of biological nervous systems and draw on the analogies of adaptive biological learning. The key element is the novel structure of the information processing system composed of a large number of highly interconnected processing elements, analogous to neurones, tied together with weighted connections, analogous to synapses. ANNs learn by example through training, the weighting of the connections between each processing element is altered dependent on how close to the truth the result is. Ultimately the connection weights store the knowledge necessary to solve specific problems. ANNs are good pattern recognition engines and robust classifiers, with the ability to generalise in making decisions about imprecise input data. They are often good

at solving problems that are too complex for conventional technologies, which do not have an algorithmic solution or for which an algorithmic solution is too complex to be found. The primary disadvantage of ANNs is the extent to which they are reliant on training, which must be exhaustive and is very time consuming.

Bayesian analysis

Thomas Bayes, born in 1702 in Tunbridge Wells, was a mathematician who first used probability inductively and established a mathematical basis for probability inference (a means of calculating, from the number of times an event has not occurred, the probability that it will occur in future) (Anon, 2001). The theorem allows new information to be used to update the conditional probability of an event and is expressed as

$$P(G_i|D) = \frac{P(D|G_i)P(G_i)}{\sum_{i=1}^{n} P(D|G_i)P(G_i)}$$

where G_i is group I (I = 1, 2, 3, etc); D is the value of the discriminant function; $P(G_i)$ is the probability that a case (for example a feature defined on an image) is in group I, where no other information is available and is termed the *prior probability*; $P(D|G_i)$ is the probability of getting a given discriminant function value if a case is in a given group, and is termed the *conditional probability*; and $P(G_i|D)$ is the probability that a case is in group i given a discriminant function D. In application the discriminant D is calculated and $P(G_i|D)$ is sought (Anthony 1999).

Bayesian theory compliments the hierarchical hypotheses theory of diagnostic formulation with weighting variables accommodated by changes in the conditional probability.

1.11.4 Current research

CAD remains far from being widespread in routine clinical practice but a wide range of possible applications are currently under development from automated measurement on orthopaedic radiographs (He, Zhang, & Zhang 2000;Rajnics et al. 2001;Wigderowitz et al. 2001) to identifying interstitial lung disease on chest radiographs (Ishida et al. 1997) or identifying pulmonary emboli on contrast enhanced CT (Wildberger et al. 2001) . However, the main fields in radiology in which CAD research is advancing relate to the techniques used in screening programmes.

1.11.4.1 CAD in breast screening

The availability of CAD in the analysis of screening mammograms could greatly enhance radiologist performance in a high volume, low yield process and studies have demonstrated the potential utility of CAD in this context (Birdwell, Ikeda, O'Shaughnessy, & Sickles 2001;Warren Burhenne et al. 2000) although, to some extent dependent on the experience of the radiologist in reading mammograms (Brem & Schoonjans 2001). Of CAD systems in development those for breast screening are at the most advanced stage and closest to clinical use.

1.11.4.2 Nodule detection

The increasing pressure to screen for lung cancer, whether by chest radiographs or chest CT, has stimulated the development of CAD in nodule detection. Up to 30% of lung nodules on chest radiographs are missed by radiologists (Greene 1992), which can be reduced by the implementation of CAD schemes (Doi 1999;Kobayashi et al. 1996;MacMahon et al. 1999). The characterisation of lung nodules discovered during screening can be improved using CAD to determine size changes (Ko & Betke 2001) and variations in form (Nakamura et al. 2000;Reeves & Kostis 2000a), with recent efforts concentrating on the three-dimensional morphology of small nodules to distinguish between benign and malignant lesions (Reeves & Kostis 2000b).

1.12 AIM OF THE THESIS

1) Design analytical techniques to interpret the massive data sets recorded from eye-tracking experiments of observers reading HRCT cases.
2) Determine the optimum viewing format, based on eye-movement data, for analysing HRCT images of diffuse lung disease and airways disease
3) Create a training presentation based on the experienced observers approach to HRCT examinations
4) Compare experienced radiologists approach to HRCT examinations with that of inexperienced radiologists, based on eye-movement data.
5) Determine the impact on the inexperienced observers' approach to HRCT examinations as a result of training that includes insights derived from the more experienced approach.

CHAPTER 2

GENERAL METHODOLOGY

2.1 EXPERIMENT SET UP.

2.1.1 Eye-tracking camera and software.

A commercially available ASL-504 Eye Tracking System [Applied Science Laboratories, Massachusetts, USA] (Fig. 2.1), based on the corneal-pupil reflection relationship, was used in the following investigations. The system is unobtrusive but head movement needed to be constrained to a degree, necessitating the use of a padded chin rest to inhibit the observer's head movement during data acquisition. A large blacked out enclosure, open only at one end, was built and a 20" screen diagonal, flat tube, high resolution monitor (MultiSync FP1370, NEC-Mitsubishi, Japan) was placed at the enclosed end. The monitor selected for the study was the same as those used on personal computer-based image viewing stations at the Royal Brompton Hospital, London.

The distance between the observers' eye and the monitor screen was 70cm with the 41 x31 cm (landscape format) monitor occupying 32° of horizontal visual field and 25° of vertical visual field. The ASL-504 camera was interfaced with a personal computer (PC) running proprietary eye-track processing software [Applied Science Laboratories, Massachusetts, USA], this in turn was linked to another PC running the custom designed DICOM viewing software from which the eye-tracking data were recorded along with information pertaining to the images being viewed at the time (Fig. 2.2).



Figure 2.1: The ASL-504 eye-tracking device. The light emitting diodes arranged in circle around the central camera lens emit infrared light to which the camera is sensitive. The camera mounting enables rotation in two orthogonal planes and therefore permits a wide range of movement.

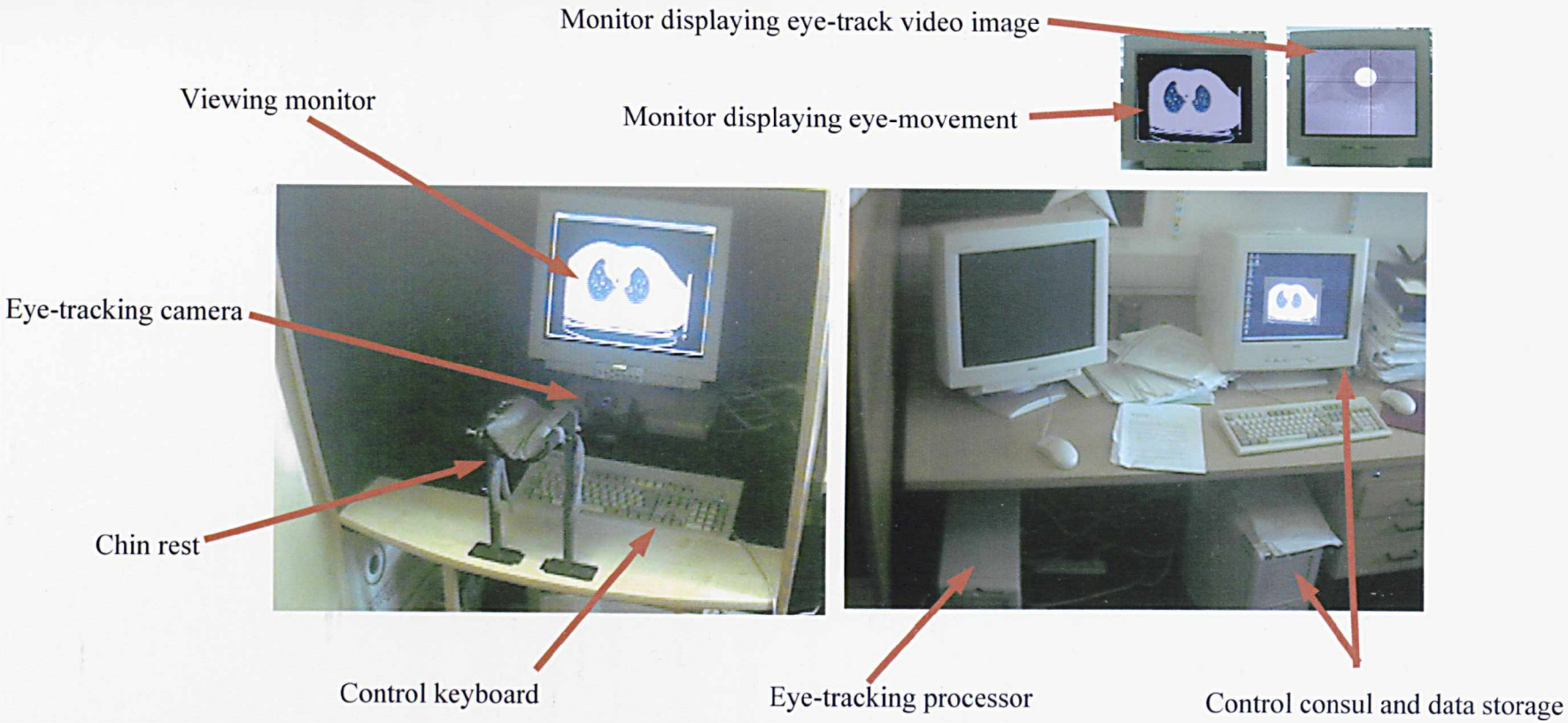


Figure 2.2: Schematic of the experiment set-up

Two small monitors formed part of the eye-tracking targeting system. One displayed the view from the eye-tracking camera enabling the orientation of the camera to be adjusted by remote control to establish an optimal image of the left eye tracked in this study; superimposed on the camera image of the eye was a circle outlining the pupil circumference, a white cross-hair identifying the centre of the pupil and a black cross-hair identifying the Purkinje 1 light reflex or glint (Fig. 2.3). The stability of the marked pupillary outline and glint was assessed on this monitor and any potentially confounding features such as eye-watering, contact lenses and sub-optimal head orientation were identified. The second small monitor displayed the image being viewed with a cross super-imposed to identify the position on the computer screen at which the eye was looking as calculated by the eye-tracking system. Two technicians were employed during the experiments. One operated the display software and camera calibration programs from a PC on which both the image being viewed by the observer and the projected eye position were displayed. Two monitors in the set-up displayed this information providing an early warning of drift in the eye-tracking calibration indicated by the consistent tracking of eye movements to unlikely areas of interest. A second technician taught the observer now to manipulate the images, recorded observer responses and recorded any untoward events that occurred during the experiment that may have an influence on the results. In the event of problems with the automatic tracking of the eye by the eye-tracking camera, a remote control device was used to ensure the camera maintained a clear view of the eye.

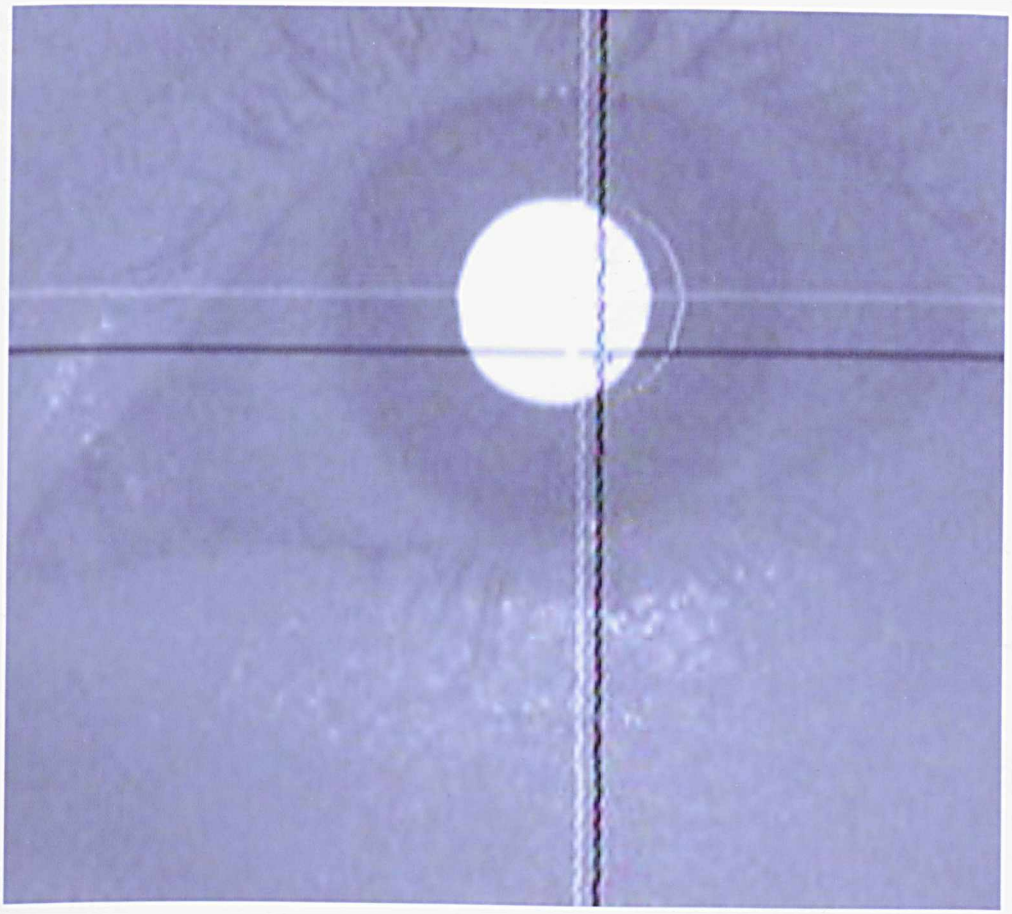


Figure 2.3: Screen capture of the eye-tracker video image.

The software associated with the eye-tracking camera identifies the pupil (white circle), calculates the centre of the pupil (white cross) and identifies the glint (black cross). Note the circle and crosses are displayed with a constant displacement to aid the manual checking of the automated detection.

The eye-tracking camera was positioned within the enclosure directly in front of the left eye of the observer, at the base of the viewing monitor with elevation of the camera to direct its lens at the eye, this is the recommended set-up according to the manufacturers of the eye-tracking camera. The use of spectacles or fixed contact lenses for vision correction caused no problems provided the glare from the lens did not interfere with the detection of the glint. Mobile contact lenses caused tracking inaccuracies and one individual with mobile contact lenses used spectacles in preference. For the experiments that required observer manipulation of the displayed images a computer keyboard was placed in front of the observer.

2.1.2 Calibration

Full screen size was used for the calibration display, which consisted of a black screen with nine crosses displayed in a three by three grid. The crosses were designated the numbers one to nine (Fig 2.4). The observer was asked to fixate crosses, identified by their number, while their eye movements were tracked. When the scale of eye movements was at a minimum and concentrated in a localised area, the eye was considered to be fixating on the desired cross and the relative position of the glint and the centre of the pupil were recorded within the calibration software. The process was repeated for each of the nine crosses (fixation targets) to calibrate the eye-tracking system over the entire area of the image display. The calibration was then checked by asking the observer to fixate on each cross in turn and observing the proximity of the calculated eye movements to the known fixation target. Clusters of eye movements around the target were accepted but a

consistent tracking of eye movements to a point offset from the target initiated re-calibration. Calibration was not considered complete until satisfactory tracking to each of the nine targets in succession was achieved.

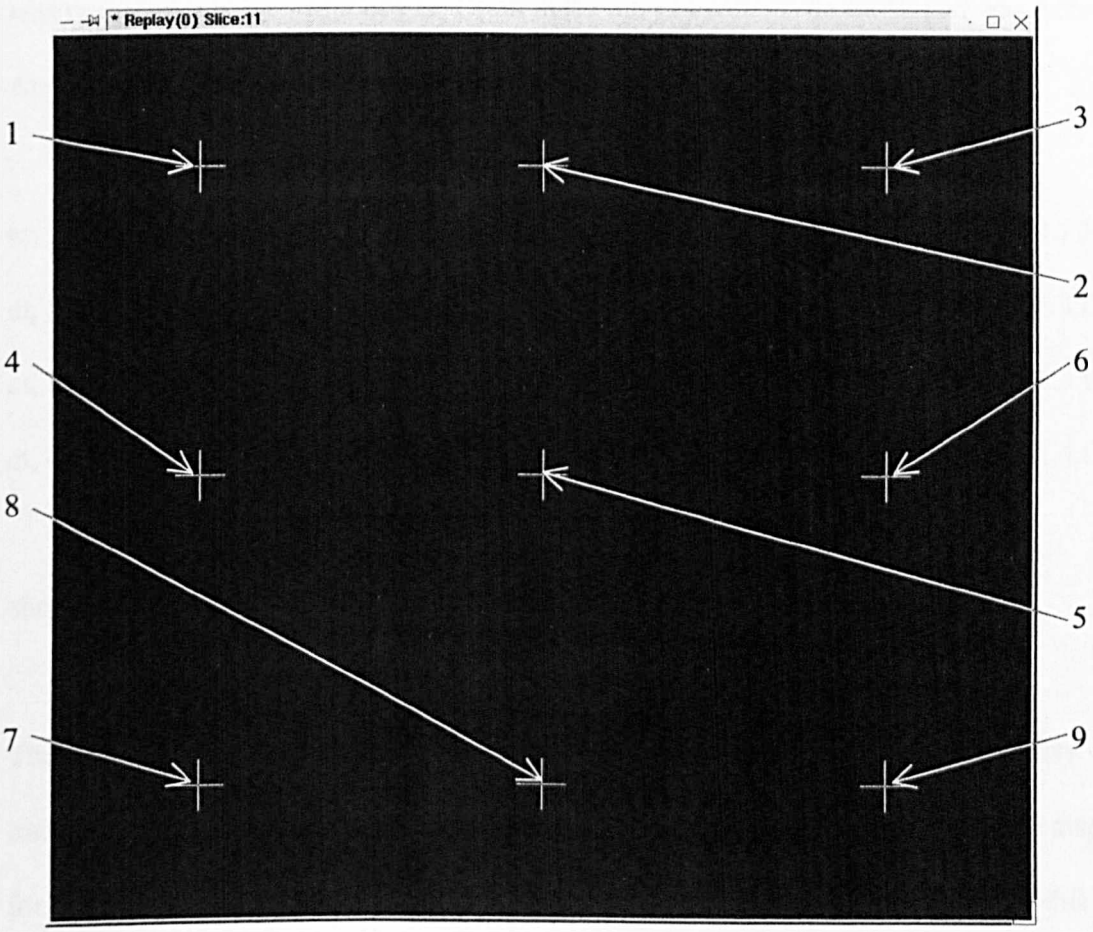


Figure 2.4: Calibration screen
The crosses were designated with the numbers one to nine and the observer asked to look at each cross in turn thereby calibrating the eye-tracking system over the entire screen.

2.2 COMPUTER DISPLAY SOFTWARE DESIGN.

A script file was defined for each experiment whereby the location and name of images was recorded and the sequence in which these images were to be displayed was defined. An example script file is shown below.

ct, 1, 0, /mnt/windows/Md/format/format/1/BUTLER, RAYMOND.CT.1.000001.1.DCM

ct, 2, 0, /mnt/windows/Md/format/format/1/BUTLER, RAYMOND.CT.1.000001.2.DCM

ct, 3, 0, /mnt/windows/Md/format/format/1/BUTLER, RAYMOND.CT.1.000001.3.DCM

ct, 4, 0, /mnt/windows/Md/format/format/1/BUTLER, RAYMOND.CT.1.000001.4.DCM

.....

show_ct, 0, 30, **1, 30**, 101, 1, *2, 2*

The first four lines define the source of the first four images of a case that has thirty images in total. The line beginning "show_ct" defines the first and last images to display for that line of script (in bold) and the number of images to display (in italics) in this case two by two (Fig. 2.5). In this way each script line beginning "show_ct" defines the display of an entire case. The script file instructions were then interpreted by a custom designed DICOM viewer, which also recorded the eye-tracking data from the eye-tracking camera software. A breakdown of the data recorded for each eye-tracked data point is presented in Figure 2.6. At a sampling frequency of 50Hz and an average case reading time of three minutes, nine thousand data points were recorded on average for each observer for each case viewed.

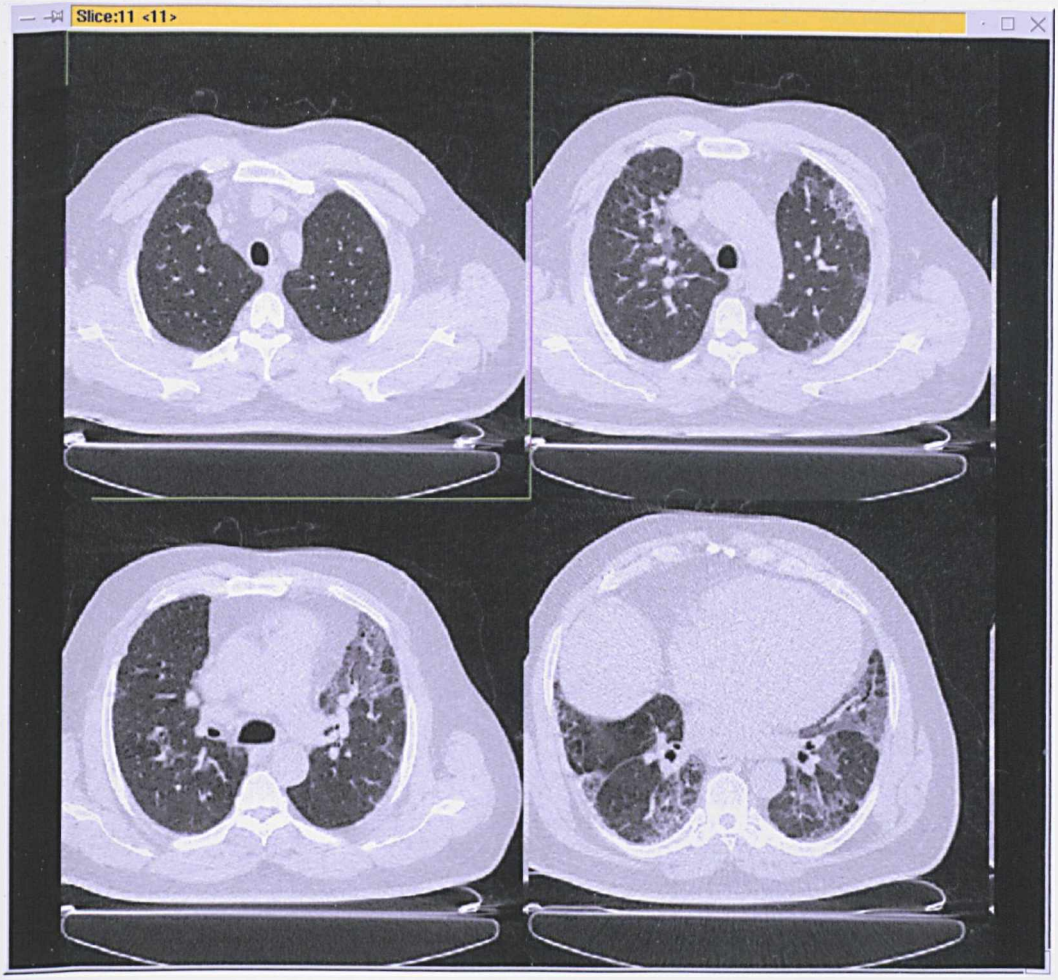


Figure 2.5: **An example of a case presented in the 2 x 2 tiled viewing format**. The green square indicates the image that can be enlarged and, when moved off the visible screen area, stimulates the repositioning of the displayed images to allow scrolling through the images in the case.

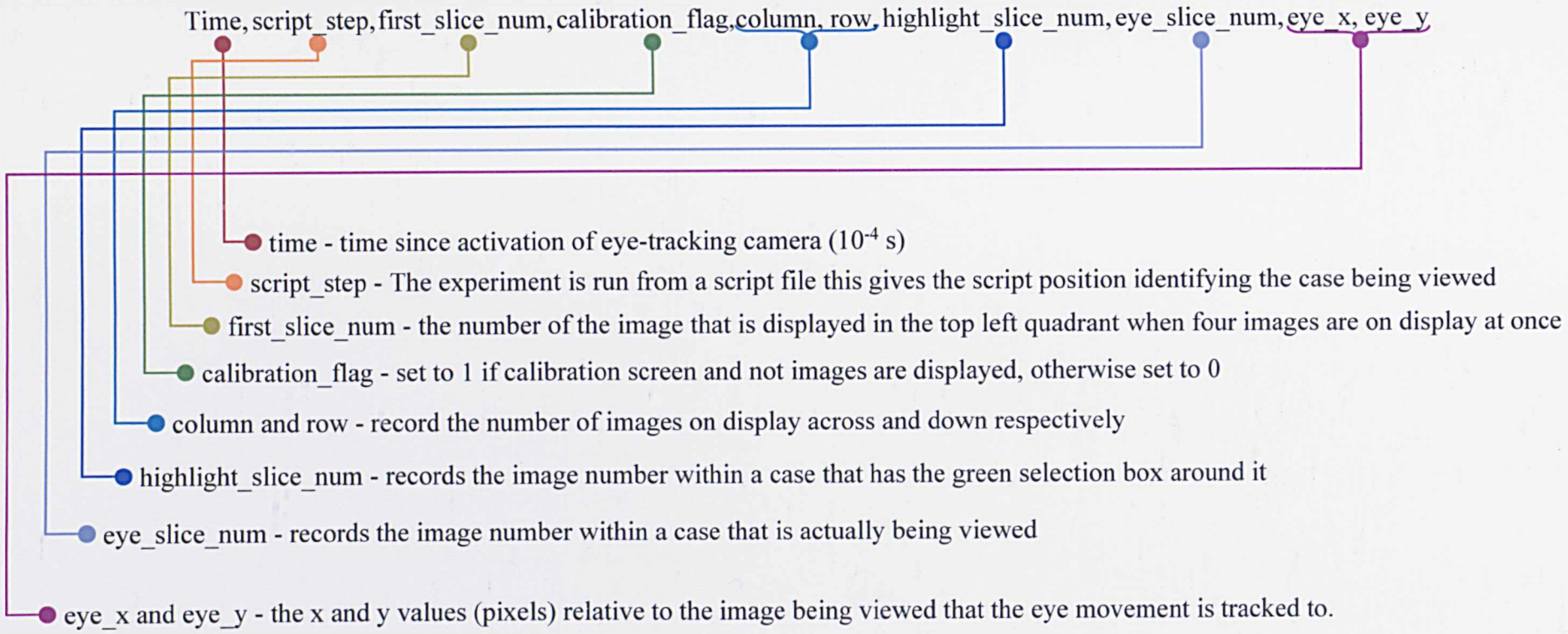


Figure 2.6: Explanation of the data acquired during eye-tracking recording.

2.3 DERIVING A NORMALISED LUNG.

In order to directly compare the eye position and movement between different cases it was necessary to define a 'normal' lung template onto which the eye tracking data from each case could be mapped permitting direct comparison. A selection of normal HRCT scans was taken and using deformation algorithms an average lung parenchymal shape for each of fifteen defined anatomical levels was calculated. Onto these average lung shapes the eye-tracking data could be mapped by using further distortion algorithms based on matching the shape of the lung parenchyma from the case being viewed to the average shape from the 'normal lung'.

2.3.1 Normal image selection

Twenty-four HRCT examinations of individuals with no abnormality detected on HRCT and normal lung function tests were selected as a sample group of normal adult HRCT examinations of the lungs. Fifteen anatomical levels were predefined to cover the length of the chest and corresponding levels from each of the HRCT examinations were selected (Fig. 2.7).

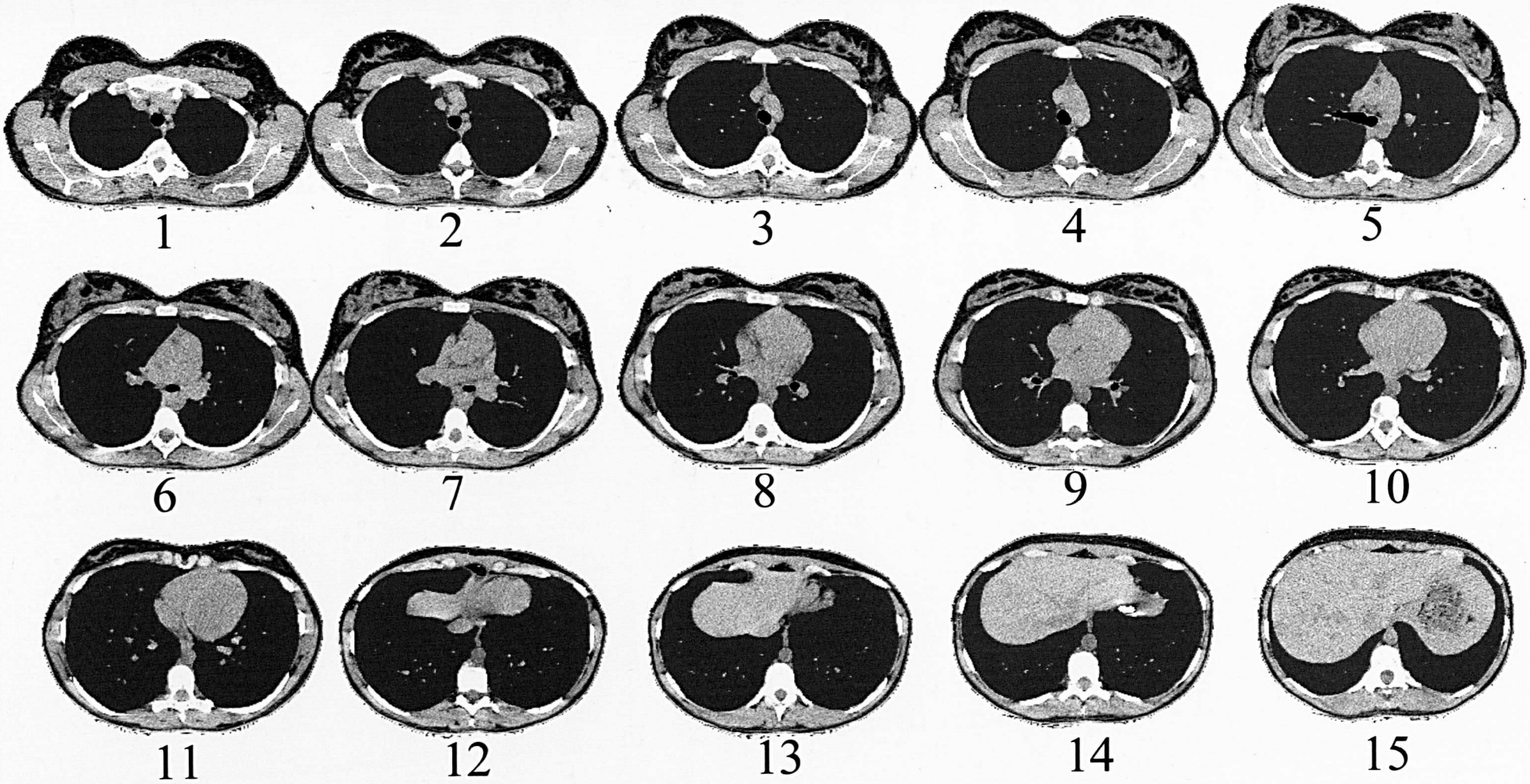


Figure 2.7: 15 levels, based on mediastinal structures, selected for calculation of the normalised lung

2.3.2 Parenchymal separation

The left and right lung parenchyma were extracted from each image using a threshold analysis algorithm based on the Hounsfield unit (HU) numbers represented by the pixel densities. All areas of the image with HU numbers of less than –700 were assigned as lung parenchyma if completely surrounded by an area comprising HU numbers greater than -300 (indicating chest wall). The algorithm for the extraction of the lung parenchyma assumed each image to have separate left and right lungs. In some instances the automatic parenchymal detection failed due to an inability to separate the two lungs, for example when the anterior junctional line was too narrow to be considered as a bounding area ('chest wall'). In such cases the division between the two lungs was artificially exaggerated by drawing on the image pixels of high HU number at the problematic boundary. Care was taken to ensure that this process did not alter the lung contour. The only other instance in which artificial high HU number pixels were drawn on the image were to avoid the inclusion of the main bronchi and the trachea as part of the lung parenchyma. For each original image, corresponding to a single HRCT slice, two separate images, corresponding to the left and right lung, were obtained. These images retained the data corresponding to any structures, such as vessels, that comprised HU numbers outside the range defined for lung parenchyma but clearly did not represent 'chest wall' (Fig.2.8).



Figure 2.8: The above twenty-four images demonstrate the extracted parenchyma of the left lung, at level seven for the 'normal' lung, from each of the normal HRCT cases. Through a series of re-scaling and distortion on each image, an 'average' shape was derived from these images.

2.3.3 Hierarchical image registration

The principle behind hierarchical image registration is to consider an image at different detail levels. The task is to define an optimised deformation grid that registers a source image onto a reference image. Deformation grids are based on the distortion of an arbitrary quadrilateral defining a region of an image. A two dimensional deformation function incorporates the four deformation processes imposed upon the vertices of the quadrilateral to create the desired alteration in the underlying image. The size of the virtual quadrilateral defines the detail level at which the registration is occurring. The initial deformation grid considers the entire image to be registered. Progressively finer detail features are then considered in the registration process in a recursive fashion such that at any given detail level, misalignments coarser than that level, have already been removed. An example is given in Figure 2.9

2.3.4 Deformation grids

Hierarchical image registration was more successful when the registered image was scaled up. Therefore, at each anatomical level the largest parenchymal image was taken as a temporary reference standard and the other images put through hierarchical image registration to approximate to the reference image. The use of a single image in this way introduced a bias into the overall calculated deformation grid. Therefore, an average deformation was calculated from all the deformations performed and subtracted from the individual deformation grids that had been used to register the images to the reference

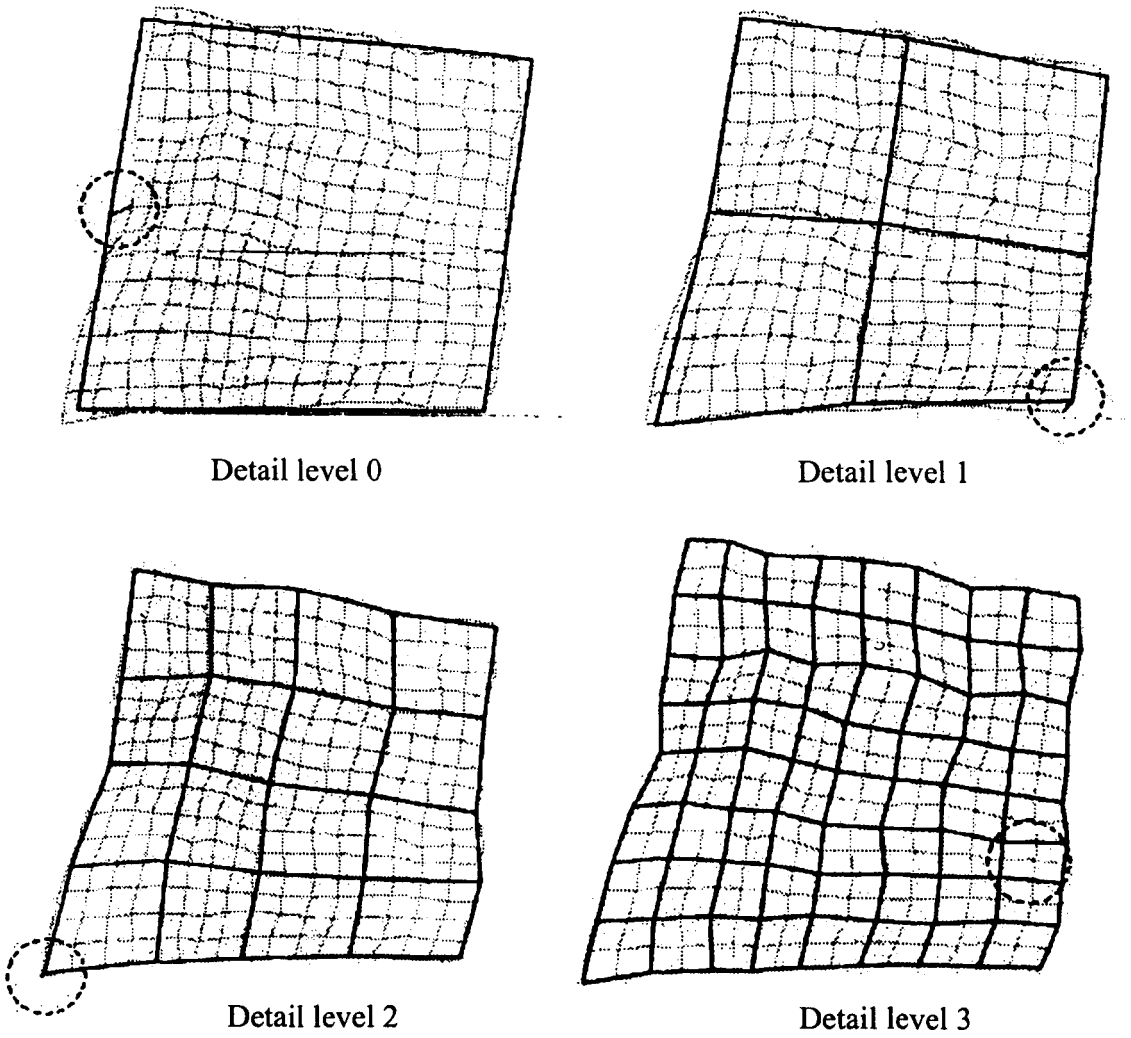


Figure 2.9: **Deformation grids at progressively finer detail levels.**

The bold lines represent the shape to be distorted. At detail level 0 the margins of the shape are deformed to match those of the final shape. At the next level, the initial grid is subdivided into four and each smaller grid element is deformed to match the final grid at that detail level. The subdivision process continues until no further deformation is required.

images. The now normalised deformation grid corresponding to each image was used to perform a registration of that image creating a set of deformed images. At this stage the internal detail of the lung image had been lost and only an outline remained.

2.3.5 Selection of the "normal" outline

The images resulting from the modified deformation grids were superimposed on one another and a probability map generated displaying the number of images onto which a given point could be mapped. The maximum probability threshold that had a reasonably smooth outline was at 50%; within the 50% probability margin any point could be mapped onto at least 50% of the original images (Fig 2.10).

2.3.6 Definition of the normalisation deformation grid.

All the modified deformation grids recorded for each of the images at a specific anatomical level were averaged to determine the deformation grid that would be applied to the eye tracking data, which was then plotted on the standard lung template at that level.

2.3.7 Limitations

The process of deriving a normalised lung discards parenchymal detail in favour of the overall anatomical position. The technique is insensitive in the analysis of feature-led eye movements as the stimulating feature can only be guessed at based on the gross anatomical position. However, the technique elegantly characterises anatomically led eye

movements. During the calculation of the normalised lung a relatively large fluctuation in lung outline was apparent a) adjacent to the mediastinum, where large vessels were entering or exiting the lung, and b) at the bases due to the variable position of the diaphragm.

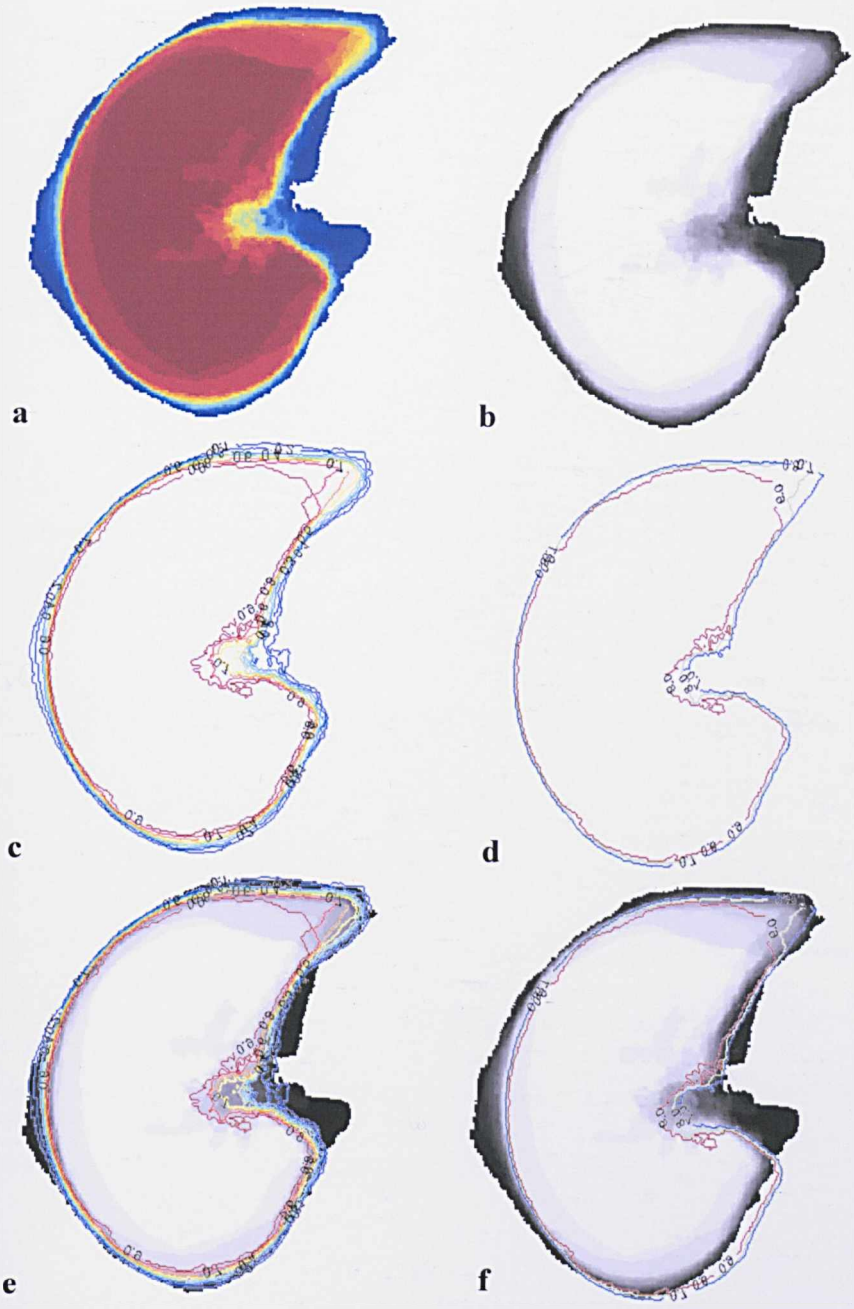


Figure 2.10: a) and **b**) are contour maps depicting the overlapping of the lung parenchymal images from level seven, the red end of the spectrum on the colour map or the whiter areas on the grey scale map depict the areas of greater overlap. **c)** and **d**) are probability contour maps (**d** is a simplified version) each contour defines the probability that a given point within that contour will map onto one of the reference images. **e)** and **f)** show the probability contours superimposed upon the grey scale probability map.

2.4 ANALYSIS OF UNPROCESSED (RAW) EYE-TRACKING DATA

The eye-tracking data consisted of a sequential set of coordinates corresponding to the perceived position of foveal gaze at a specific point in time. Initial analysis dealt with the data in a raw form considering the duration and number of fixation points, the length and number of saccadic movements and interchanges between images. Such an analysis ignores specific spatial information taking no account of the actual areas of the image to which the eye has been tracked.

The raw data analysis was performed using the Excel spreadsheet program (Microsoft Inc., Redmond, USA). The data recorded during an experiment is described in Figure 2.6 and was saved as a text file with each line corresponding to a single data point determined by the eye-tracking camera. The text file was then divided into data points relevant to specific cases by identifying the script_step and each segment of data saved again as a text file. Each text file corresponding to a single script_step was then imported into a separate spreadsheet in a standard comma delimited text file format.

2.4.1 Time (units 10^{-4}s)

The time taken to view a case was calculated by subtracting the 'time' value of the last data line from the 'time' value of the first data line.

2.4.2 Distance

The distance of a saccade, in pixels, from a point x_1,y_1 to a point x_2,y_2 was calculated from Pythagorus' theorem as follows.

Distance = $((x_1-x_2)^2+(y_1-y_2)^2)^{1/2}$

2.4.3 Validity

Calculated distances were only considered valid if the saccadic distance calculated from sequential x and y coordinates greater than zero and less than eight hundred; the maximum possible legitimate saccadic distance on a 512 x 512 pixel image was calculated as 724, therefore the upper limit eight hundred was used. Erroneous data resulted from failure of the eye-tracking camera to retain clear view of the eye, e.g. during blinking or eye watering, the x and y coordinates recorded were invariably very large and clearly off the screen, resulting in a very large value for saccadic distance.

2.4.4 Tremor, short saccade or long saccade

Each validated saccadic distance was classified as tremor, short or long by the following criteria.

Tremor – any validated saccades of less than seven pixels (at full screen resolution) were defined as tremor movements corresponding to eye fixation. The threshold value of seven was calculated from the calibration data corresponding to the upper limit of saccadic

distance recorded during fixation on the calibration targets. This was seen to be uniform across the screen.

Short saccade – any validated saccades of between seven and an observer specific value were defined as short saccades. The observer specific value was taken as that corresponding to the calculated radius of useful peripheral vision for that observer.

Long saccade – any validated saccades that did not meet the criteria for tremor or short saccades.

2.4.5 Fixation

A fixation point was defined as a sequence of at least five sequential tremor saccades corresponding to at least one hundred milliseconds and calculated by progressively incrementing a count value by one for each sequential tremor saccade (technique recommended by the manufacturers of the eye-tracking camera).

2.4.6 Fixation duration

At the end of a sequence of saccades that satisfied the criteria for a fixation point the duration of the fixation (in milliseconds) was calculated by multiplying the count value by twenty, each data line corresponded to twenty milliseconds.

2.4.7 Short saccade leading to fixation

Inspection of the count value, used to define fixation duration, five data lines in advance of a short saccade if equal to five identified those short saccades that were immediately followed by a fixation.

2.4.8 Long saccade leading to fixation.

Calculated as for a short saccade leading to a fixation.

2.4.9 Movement between images

Count – each change in the image being viewed was recorded and the total number of these changes were calculated.

Distance – the image coordinates before and after movement from one image to another were used to calculated the relative distance moved by the eye with respect to the image (rather than screen) coordinates. The calculation took the form of that used to calculate distance as described above with x_1,y_1 corresponding to the last coordinate on one image and x_2,y_2 the first coordinate on the next.

Resulting in fixation – whether a movement between images led directly to a fixation was calculated as for the short and long saccades leading to fixation.

Sequential or non-sequential – a two by two image display permits movements between images that are not adjacent in a sequence. By comparing the number of the images being viewed, either side of a movement between images, a non-sequential movement was recorded if the difference in image values was greater than one.

2.5 STATISTICAL METHODS USED FOR ANALYSIS

2.5.1 Measures of observer agreement

Appropriate measures of observer agreement are required for the assessment of different observers approaches as investigated in the following experiments.

2.5.1.1 The kappa statistic

The eye-tracking data derived from the viewing of an image was mapped onto a masked grid that encompassed all the lung parenchyma displayed on that image. The proportion of time spent viewing the area within each grid square was recorded and ranked according to thresholds calculated from the results from all the grid square values in the overall experiment's data set. The ranked (ordinal) data generated for each viewing of each case was compared using the weighted kappa coefficient (κ_w). This is a derivative of the kappa coefficient.

The kappa coefficient attempts to express the level of agreement between two paired data sets whilst correcting for agreement due to chance alone. The kappa coefficient is defined by the following expression.

$$\kappa = \frac{P_O - P_E}{1 - P_E}$$

P_O = the observed proportion of times the data values agree; P_E = the proportion of times that values are expected to agree.

The advantage of the weighted kappa coefficient is that it can identify large discrepancies in graded observations, which cannot be gleaned from the kappa coefficient. As an example, the kappa statistic would consider the discrepancy between paired grid squares from ranked 1 and 2, based on the proportion of time spent viewing them, as equivalent to that between paired grid squares ranked 1 and 4. The disparity in the latter case is clearly greater and the weighted kappa coefficient would take this into account.

There is currently no consensus for the interpretation of kappa values but kappa values of greater than 0.4 have been shown to represent acceptable observer agreement in a clinical context (Brennan & Silman 1992)

2.5.1.2 Kullback-Leibler information criterion

The Kullback-Leibler information criterion (KLIC) was developed as a ruler to measure the similarity between a statistical model and a true distribution and is predominantly used in statistical research and the assessment of financial prediction models. However, the KLIC can be equally applied to the measurement of similarity between two experimentally derived distributions (personal communication from Dr E Krupinski, University of Arizona, USA).

The KLIC is defined as

$$KLIC(g:f) = \sum_{y=0}^{n} g(y) \times \frac{Ln[g(y)]}{Ln[f(y)]}$$

Where *KLIC(g:f)* is the Kullback-Leibler information criterion for a comparison between the two probability distributions *g(y)* and *f(y)*.

Properties of the KLIC are that as

$$(i).....KLIC(g:f) \geq 0$$
$$(ii).....KLIC(g:f) = 0 \Leftrightarrow g(y) = f(y)$$

Therefore, the closer the two distributions $g(y)$ and $f(y)$ are to each other the smaller the value of the KLIC. By comparing the paired values in the two distributions the degree of distance between the distributions is assessed and not just their overall trends (Kullback & Leibler 1951). There are, as yet, no satisfactory measures of statistical significance of KLIC values limiting its use to comparative analyses.

2.5.2 Measures of observer differences

2.5.2.1 Wilcoxon's signed rank test

The Wilcoxon's signed rank test is a non-parametric test used to test hypotheses about population medians. To use the Wilcoxon's signed rank test the following assumptions must be satisfied: the data is not biased, the observations are independent and there is a reasonable symmetry about the median (less restrictive than an assumption of normality). The test statistic used for the Wilcoxon's signed rank test is W and, when comparing paired data sets with a null hypothesis that the population median difference is equal to some hypothesised value (usually zero), is calculated as follows:

If the observations for the first sample are denoted by x_{1i} and those for the second sample by x_{2i} then the difference between the paired observations is d_i where

$$d_i = x_{1i} - x_{2i}$$

The null hypothesis is

H_O: Population median difference = m_o

where m_o is the median of the sample of values denoted by d_i.

The two sided alternative hypothesis is:

H_A: Population median difference $\neq m_o$

A new sample of observations, n_i, is derived by

$$n_i = d_i - m_o$$

and all zero values of n_i are removed.

The non-zero values of n_i are arranged in ascending order of absolute value and ranked with the lowest rank given to the smallest value. Two or more values that are equal will tie for the same rank in which case they are assigned the average of the ranks the would have had if they had not been tied. Each rank is then given the sign of the observation to which it corresponds. The sum of all the positive ranks is W.

If the null hypothesis holds, the spread of ranks between positive and negative new values (n_i) should be approximately equal and W will equal half of the sum of the absolute values of all the ranks. If W is very large or very small, there is an unequal distribution of ranks between positive and negative difference values suggesting the median difference is not close to m_o. A p value is assigned to the W statistic from look-up tables (Shott 1990).

2.5.2.2 Wilcoxon's ranksum test

The Wilcoxon's ranksum test is a simple non-parametric test to compared the similarity of two samples of unequal size. The observations for both groups are ranked in ascending order of magnitude and any values that are equal their ranks are averaged. The ranks of the smaller sample are than added up to calculate the T statistic that can then be assigned

a p value based on look-up tables with respect to the likelihood that the two samples have different medians (Shott 1990).

2.5.2.3 Spearman's rank correlation

This is a non parametric measure of the degree of association between two numerical variables. The values of each variable are individually ranked (i.e. within sample ranking) and the measure is based on the pairs of ranks of the two variables. If d_i is the difference between paired ranks then the Spearman's rank correlation (r_s) is given by:

$$r_s = 1 - \frac{6\sum_{1}^{n} d_i^2}{n(n^2 - 1)}$$

Where n is the number of paired ranks. The correlation will have a value that varies between –1 and 1 and the significance of any association between the two sets of values is derived from look-up tables (Kirkwood 1988c).

2.5.2.4 Analysis of variance and stepwise linear regression

An analysis of variance and stepwise linear regression, techniques with considerable overlap, are used to compare the means of several groups, assessing how much of the overall variation in the data is attributable to differences in the group means and comparing this with the amount attributable to differences between individual observations in the same group. The data sets analysed by these techniques in the

forthcoming experiments comprised one set of calculated values compared with data sets representing different variables within the experiment. The experiment variables were assigned rank values based on the mean of the calculated values related to that specific variable. This enabled the relative influence of each experiment variable on the calculated values to be assessed. The mathematical calculations involved are complex and are usually performed by dedicated computer software (Kirkwood 1988b;Kirkwood 1988a).

CHAPTER 3

COMPARISON OF THE APPROACH TO READING HRCT SCANS DISPLAYED IN DIFFERENT VIEWING FORMATS.

3.1 INTRODUCTION

The advent of digital imaging has dramatically increased the proportion of radiology reporting that is performed directly from a computer screen (soft copy reporting) as opposed to that performed from film (hard copy reporting) (Arenson et al. 2000). In terms of historical development, soft copy reporting is more readily associated with cross-sectional image reporting (as opposed to projectional radiographic reporting); the imaging data obtained from cross-sectional imaging such as magnetic resonance imaging (MRI) and computed tomography (CT) has always been acquired and displayed in digital form. However, the computing power and monitor quality required for interpreting MRI and CT images has only become widespread in the last decade, necessitating the comparison of reporting efficacy between viewing hard and soft copy images (Beard et al. 1993;Beard et al. 1995). The advantages of soft copy reporting over hard copy reporting are the immediacy of the images, the ability to alter windowing values in real time using the raw DICOM data, and the flexibility in display formats whereby the image size and number of images displayed at any one time can be tailored (Strickland & Allison 1995). This last apparent advantage discloses a disadvantage related to the technology available for soft copy reporting. Printed sheets typically contain 20 images laid out in a 4 x 5 format. The size of quality monitors for soft copy reporting do not

allow the display of so many images at once; with conventional monitors the size of the resulting images are prohibitively small. Even with the use of two monitors the number of images that can be displayed at once remains less than that possible on printed sheets viewed on adjacent viewing boxes. The impact this has on reporting in terms of time taken to report cases (Beard et al. 1993;Foley et al. 1990;Reiner et al. 2001), the display speed of a computer workstation (Beard, Pisano, Denelsbeck, & Johnston 1994b), the number of computer screens required (Beard et al. 1994a) and the comparison of chronologically distinct CT scans (Beard et al. 1995) have been previously investigated . The trend toward filmless departments, on-the-fly windowing and possibly computer-aided diagnosis will establish soft copy reporting as the only option in the near future and will effectively close the soft copy versus hard copy debate (Arenson, Andriole, Avrin, & Gould 2000). The remaining issues are the determination of the best way to view images in terms of workstation design (Beard, Johnston, Toki, & Wilcox 1990) and optimisation of viewing modes (Strickland & Allison 1995;van der Heyden et al. 2001). To a large extent personal preference takes precedence but some indication of which viewing formats provide the most effective presentation of images for reporting is necessary. The assessment of viewing efficacy can be approached from different angles. The simplest criterion would be a comparison of the time taken to read a set of cases in various different formats taking account of the accuracy of any resulting reports. A more rigorous approach, used in the present study, assesses the actual eye movements undertaken when reading cases measuring efficacy in terms of the proportion of eye movements associated with effective visual fixation and identifying the compensatory eye movements that occur when an observer is confronted with different image viewing formats. A similar study

used eye movement analysis to determine the most appropriate method for displaying MRI images of the brain (Niimi et al. 1997). The aim of this experiment was to determine which of two basic, but fundamentally different, viewing formats allows the most efficient assimilation of (non-contiguous) HRCT images and attempt to identify the underlying ergonomic reasons. Eye movements were classified as feature led, where the a fixation results directly from an observation within the observers' range of peripheral vision and anatomically led where the saccadic eye movement is outside the range of the observers useful peripheral vision and is therefore driven by an intention to view a point due to its anatomical position. Analysis of routine anatomically directed eye movements enable the feature led movements, with which this project is most concerned, to be more readily extracted. Comparing the data from two different formats allow the eye movements inherent in the assimilation of a 3-D data set to be distinguished from those that reflect the format in which a 3-D data set is presented.

3.2 EXPERIMENT SET-UP

The ASL500 eye tracking camera and computer monitor were set up as described in Chapter 2. Four consultant radiologists, with a special interest in thoracic imaging and experience ranging from three to 22 years in consultant posts, took part in each of two experiments (DMH, MBR, SRD and SPP). The first experiment was designed to ascertain how much detailed and contrast information could be accrued from the observers peripheral vision. The results of the range of peripheral vision for small detail was used as a threshold for calculating the proportion of saccadic eye movements that fell within this range of peripheral vision and were therefore potentially feature led as opposed to anatomically led. The second experiment aimed to identify the approach of observers to two presentation formats, assessing the trend in the number and length of fixation points and the predominant length of saccadic movements.

3.3 MATERIALS AND METHODS (EXPERIMENT 1)

Five images were created in a painting program (Paint Shop Pro 5.12, JASC, USA) which displayed a uniform background of typical lung parenchyma taken from CT images and superimposed on this were bronchiectatic airways or mosaic attenuation using a clone brush facility (a function that enable the copying of a section of one picture to another). Failure to include a normal parenchymal background would have artificially enhanced the observer's peripheral vision-based detection (Kundel 1975). Three images contained bronchiectatic airways and two represented mosaic attenuation. Each image was divided by a horizontal and a vertical line into four quadrants and the distance from the centre of

the image to the cloned abnormality differed between quadrants. The images were designed to test the extent of useful peripheral vision in identifying small detail (mildly bronchiectatic airways) and density differences (mosaic attenuation) two representative images are displayed in Figures 3.1 and 3.2. The collage images were displayed at a resolution of 1272 x 992 pixels with a circle of radius thirty pixels at the centre. The image was only visible when the eye movements were tracked within the central circle, the moment the eye movements were tracked outside the centre circle the image was replaced by a neutral dark grey background leaving no after-image.

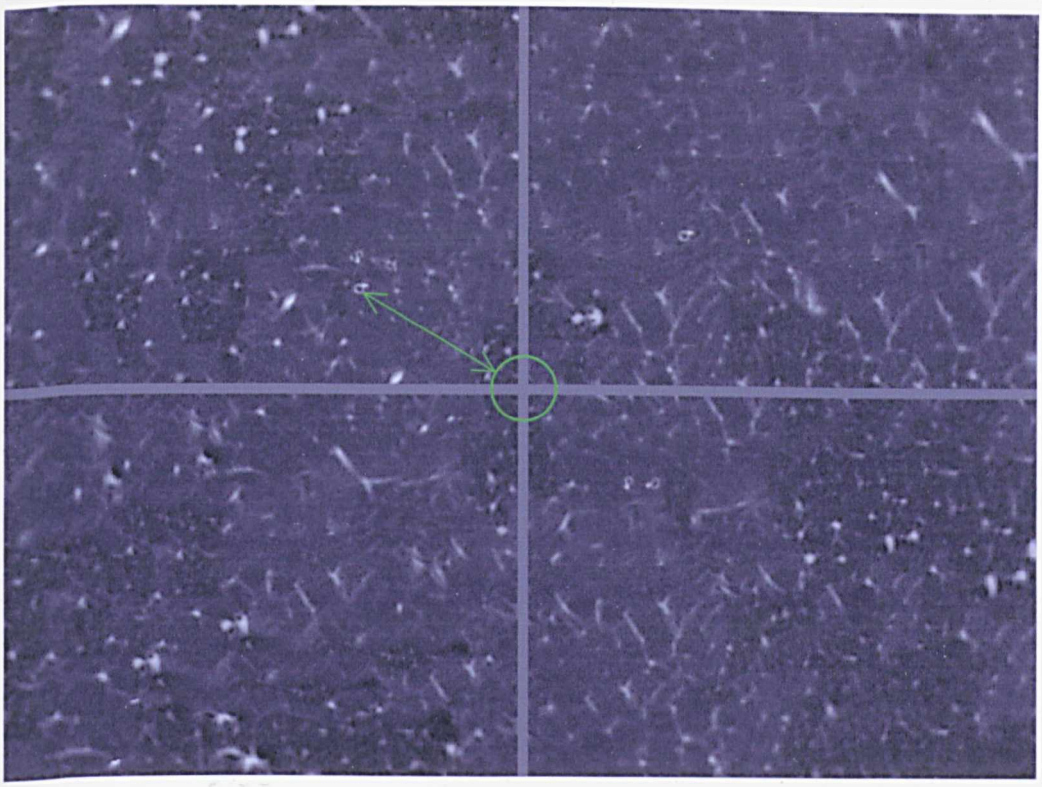


Figure 3.1: Image designed to determine the efficacy of peripheral vision in the identification of small detail, in this case mild bronchiectasis present in all but the bottom left quadrant. The green circle defines the area in the centre of the image into which the eye must be tracked for the image to be visible. The distance used in the results section is indicated by the green double headed arrow marking the distance from the edge of the central circle to the outer aspect of the bronchiectatic airway in that quadrant, the entire airway needs to be visualised to make decision as to whether it is bronchiectatic.

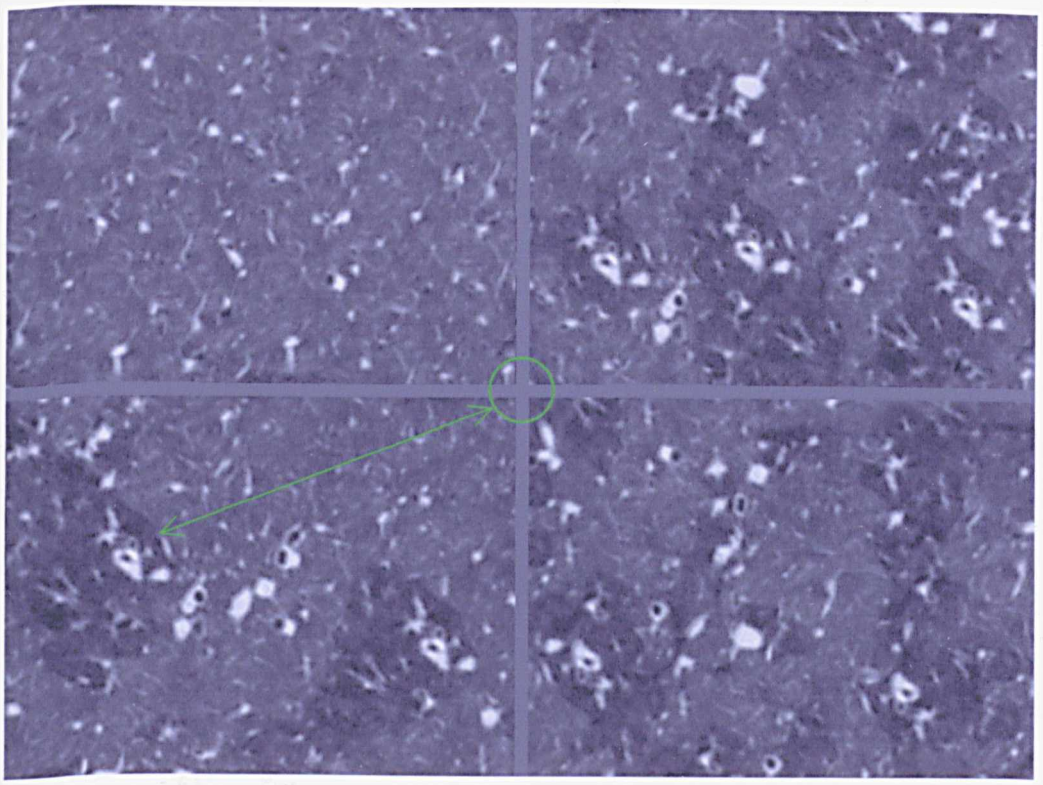


Figure 3.2: Image designed to determine the efficacy of peripheral vision in the identification of shades of grey, in this case mosaic attenuation present in all but the top right quadrant. The green circle represents the area as defined above. The distance used in the results section is again demonstrated as a double headed arrow this time extending from the edge of the central circle to the nearest margin of mosaic attenuation in that quadrant.

The size of the central circle was calculated by trial and error as the smallest possible that allowed prolonged display of the image with central fixation, given a tendency for eye tremor during fixation.

At the outset the monitor brightness and contrast settings were adjusted to the observers' satisfaction. The eye-tracking camera was then calibrated as described in Chapter 2. The images detailed above were displayed sequentially and no time limit was placed on the observation of any of the images. The observer, whilst remaining in position for eye-tracking, indicated verbally for each image when they had had sufficient time to observe the image and in which quadrants they had identified the specified abnormality: bronchiectasis for the first three images and mosaic attenuation for the fourth and fifth images.

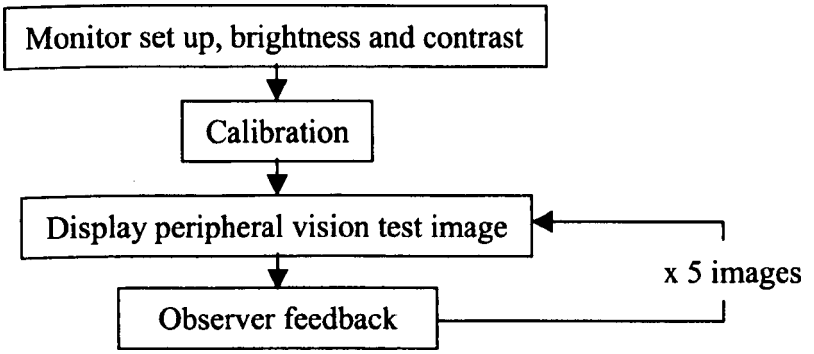


Figure 3.3:
Flow chart demonstrating the sequence of actions that comprised experiment 1

3.4 RESULTS (EXPERIMENT 1)

The results of the peripheral vision experiments are detailed in Table 3.1. The pixel values given relate to the full screen resolution used in subsequent experiments in which pixels are used as the unit of recorded distance. Each observer was able to fixate on the centre of the images such that they were visible for at least 500ms at a time. Observer B had the most difficulty achieving a stable image display which may have resulted in the smaller calculate radius of useful peripheral vision.

Image	Top-left (pixels)	Top-right (pixels)	Bottom-left (pixels)	Bottom-right (pixels)
1	88		39	41
2	146	162		47
3	147	211	202	
4		157	262	187
5	66		315	328

Table 3.1: The distance from the outer edge of the central fixation circle on the image to the outer limit of a bronchiectatic airway (images 1-3) or the nearest grey / black interface of mosaic attenuation (images 4-5). The shaded cells corresponds to quadrants without an abnormal feature.

Observer	Image	Top-left	Top-right	Bottom-left	Bottom-right
DMH	1	YES		YES	YES
	2		YES		YES
	3	YES			
	4		YES	YES	YES
	5	YES		YES	YES
MBR	1		YES	YES	
	2				
	3				
	4		YES	YES	YES
	5	YES		YES	YES
SRD	1		YES	YES	YES
	2	YES			YES
	3		YES		
	4		YES	YES	YES
	5	YES		YES	YES
SPP	1	YES		YES	YES
	2	YES	YES		YES
	3	YES		YES	
	4		YES	YES	YES
	5	YES		YES	YES

Table 3.2: Responses from the four observers for experiment 1, only positive replies are indicated, the shaded cells correspond to quadrants in which no abnormal features were seen.

Observer	Radius of useful peripheral vision (pixels)
DMH	160
MBR	40
SRD	140
SPP	200

Table 3.3: The radius of useful peripheral vision, in pixels, calculated from the responses in experiment 1.

3.5 MATERIALS AND METHODS (EXPERIMENT 2)

Sixteen high-resolution CT cases were retrospectively selected from a digital archive at The Royal Brompton Hospital, London; each comprised 23 to 30 (median 26) images. The cases comprised a mix of obstructive lung disease (OLD)[n=6], interstitial lung disease (ILD)[n=6] and normal [n=4]; all were recorded and displayed from DICOM files in a custom designed DICOM viewing program with fixed windows (centre of –500 HU, width 1500 HU). The two formats used to display the cases were:

Format A – four sequential images displayed in a 2x2 tiled arrangement with the lowest numbered image in the top left quadrant (n), the next image in the top right quadrant (n+1), the next in the bottom left quadrant (n+2) and the fourth image in the bottom right quadrant (n+3) (Fig. 3.4). A green box, moved by the keyboard cursor keys, was used to select individual images from the four on display and pressing the space bar on the keyboard would enlarge the selected image to full screen size. Pressing the space bar again would return the image to its original size and quadrant revealing the other three

images. An attempt to move the green selection box to the left from the top-left image (i.e. off the screen) would cause shuffling of the images with the image number of the top left quadrant decremented by one and the other displayed images altered accordingly (Fig. 3.5). Similarly an attempt to move the selection box to the right from the bottom right quadrant would result in the image number of the top left quadrant to be incremented by one and again all the other images altered accordingly. In this way all the images for a given case could be viewed with four sequential images being displayed at any one time. Pressing the "Alt" key on the keyboard in conjunction with the left and right cursor keys would increment (right cursor) or decrement (left cursor) the image number in the top left quadrant by four and alter the other displayed images accordingly facilitating a more rapid progression through the CT images. Advancing the image sequence beyond the last image would return the sequence to the beginning ("roll-over") and vice versa. The cursor keys were disabled when a selected image was enlarged.

Format B – A single image, at full screen size, was displayed at a time . Sequential movement through the CT images was performed by pressing the left (decremental) or right (incremental) cursor keys. Only single image advancement was possible and a roll-over occurred at the beginning and the end.

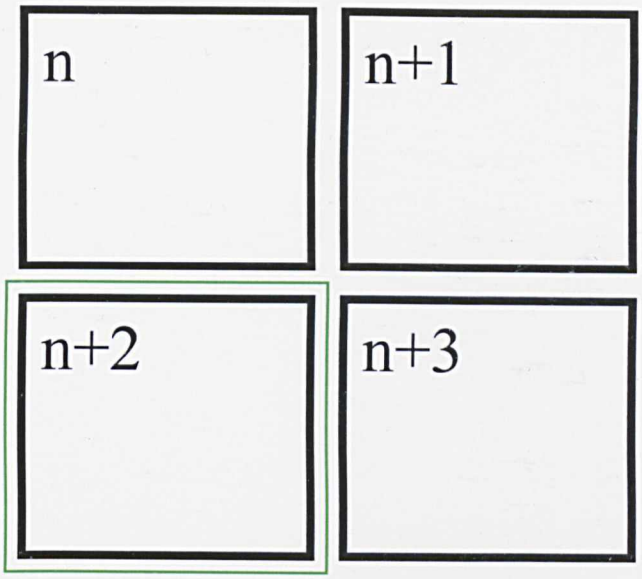


Figure 3.4: Four images displayed at once

The green box identifies the current selected image.



Figure 3.5: a) Moving the green selection box to the left, off the screen, from the left upper quadrant results in **b)** a single step through the sequence of images in the case. **c,d)** A single step in the opposite direction occurs if the green selection box is moved to the right from the bottom right quadrant.

The fundamental differences between the two formats used in this experiment are summarised in Table 3.4.

Format A	Format B
Four images viewed simultaneously	Only a single image viewable at a time
Smaller scale of displayed images	Images at full screen resolution
Keyboard controls requires use of 4 keys	Keyboard control by only two keys
Images selection more complex	Image selection simple.
Rapid progress through case possible	Progress restricted to 1 image at a time.

Table 3.4: A summary of the technical differences between the two formats used in this experiment.

The sixteen cases were divided into two groups of eight each containing normal cases [n=2], cases of ILD lung disease [n=3] and cases of OLD [n=3]. One group was presented in format A and the other in format B. Prior to the exposure to each format a practice case was presented to enable the observer to become familiar with the control keys. There was concern that eye fatigue might have an impact on the results so only five cases were viewed in format A initially followed by five cases in format B. The remaining three cases in each group were then viewed, format A again preceding format B. After each case the observer removed themselves from the eye-tracking setup and dictated a short report onto tape to encourage the observer to approach the case as they would in a clinical setting. The observer would then return to the apparatus and calibration was performed, therefore calibration was performed prior to each case. The dictated reports were transcribed and used to facilitate the analysis of overall scanning patterns.

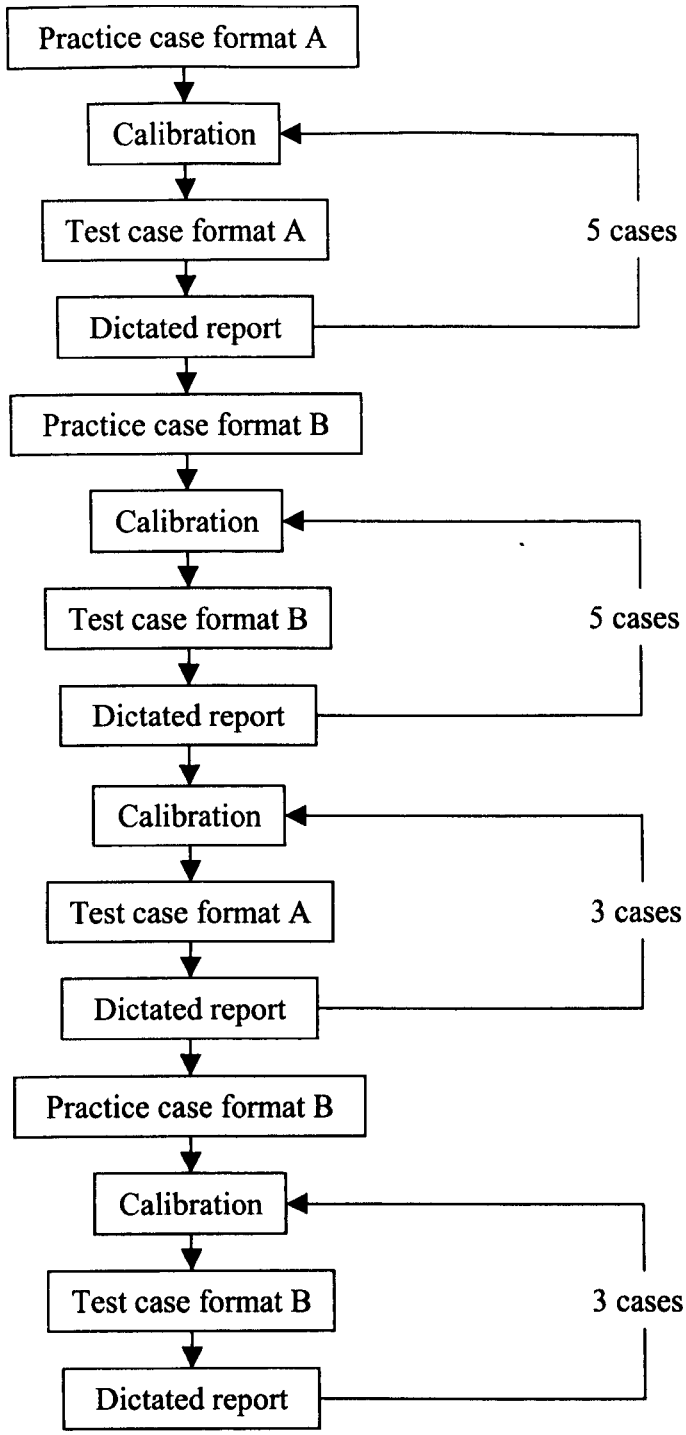


Figure 3.6: Flow chart demonstrating the sequence of experiment 2

3.6 STATISTICAL ANALYSIS (EXPERIMENT 2)

3.6.1 Differentiation of saccade type

Saccadic eye movements were divided into three categories: tremor, short saccades and long saccades based upon the calculations outlined in Chapter 2. The limits for each were defined in terms of full screen resolution pixels and a scale factor of two was used to convert image coordinate values from format A displays to equate to full screen resolution. The upper limit of tremor saccades was defined as the maximum saccadic movement recorded during a steady state phase of fixation on the calibration marks, during calibration, eliminating the larger saccadic movements associated with calibration cross seeking activity; and was consistent for all four observers. The useful peripheral vision calculated for each observer in experiment 1 was used as the upper limit of short saccadic movement for that observer in experiment 2. A long saccade represented all saccades of greater distance to a short saccade.

3.6.2 Eye-movement features analysed

The following eye-movement features were analysed:

Time	The time taken, in seconds, to view the entire case.
Number	The number of saccadic movements recorded
Tremor	The number of saccadic movements classed as tremor
%Short	The percentage of non-tremor saccadic movements classified as short

by reference to the radius of useful peripheral vision.

%Long	The percentage of non-tremor saccades classified as long.
Fix'n time	The average duration in milliseconds of fixation points
No.Fix'n	The total number of fixation points (defined as continous tremor activity for at least 100ms) divided by the total number of saccadic movements recorded as a normalization to allow comparison of the two formats independent of differences in the number of saccadic movements recorded.
%Short-Fix'n	The percentage of short saccadic movements that led directly to a fixation point.
%Long-Fix'n	The percentage of long saccadic movements that led directly to a fixation point.
No.Changes	The number of movements between images.
ChangeDist	The distance in pixels between the image coordinates of the beginning and end points of a saccadic movement traversing two images
%Change-Fix'n	The percentage of movements between images that led directly to a fixation point

%Enlarged time The percentage of time for which an enlarged image was used during the reading of a case under format A

All the data were analysed within Microsoft Excel (Microsoft Corp., Redmond, USA) according to the methods outlined in Chapter 2.

Statistical comparisons were performed using Wilcoxon's signed rank test as the data could not be assumed to follow a normal distribution.

3.6.3 Scoring of verbal reports

Verbal reports for each case were scored for features identified; 0 for failing to identify abnormal features present or identifying abnormalities that were not present, ½ for identifying some of the abnormalities present and 1 for identifying a normal scan as normal or identifying all the significant features.

3.7 RESULTS (EXPERIMENT 2)

All four observers recorded 14½ out of 15 for their verbal reports and there was therefore no difference between viewing formats with respect to accuracy.

Format A v Format B		Time (s)	Number	Tremor	%Short	%Long	Fix'n time (ms)	No.Fix'n	%Short-Fix'n	%Long-Fix'n	No.Change	ChangeDist (pixels)	%Change-Fix'n
Overall	**p value**	0.059	0.003	0.003	0.002	0.002	0.002	<0.001	0.003	0.001	<0.001	<0.001	<0.001
Format A	average	132	4482	3107	93.6%	6.4%	303	0.045	88.9%	4.1%	142	266	10.7%
Format B	average	105	2832	1884	96.6%	3.4%	245	0.050	84.9%	10.5%	51	49	38.2%
Normals	**p value**	0.853	0.561	0.546	0.805	0.805	0.214	0.197	0.006	0.103	0.287	0.061	0.021
Format A	average								86.2%				11.1%
Format B	average								76.8%				37.9%
ILD	**p value**	0.018	0.014	0.037	0.058	0.058	0.097	0.039	0.095	0.070	0.016	0.007	0.005
Format A	average	144	4788	3322				0.046			155	270	10.4%
Format B	average	111	3040	2025				0.050			53	46	38.2%
OLD	**p value**	0.406	0.219	0.203	0.436	0.436	0.077	0.054	0.013	0.031	0.036	0.003	0.012
Format A	average								86.0%	6.7%	136	263	10.7%
Format B	average								75.3%	20.0%	48	54	38.3%

Table 3.5 : Table of results from experiment 2. P values from Wilcoxon's signed rank test. All p values < 0.05 have been highlighted in red and mean values recorded in blue. The presented data represents the accumulated results from all 4 observers. The eye-tracking features analysed were; time - total time taken to view a case; number - number of recorded saccades; tremor - number of saccades attributable to eye tremor; %short - percentage of non-tremor saccades within range of peripheral vision; %long - percentage of non-tremor saccades outside peripheral vision; Fix'n time - average duration of each fixation; No.Fix'n - number of recorded fixation points; %Short-Fix'n - percentage of short saccades leading to a fixation point; %Long-Fix'n - percentage of long saccades leading to a fixation point; No.Change - number of movements between images; ChangeDist - distance moved between images relative to image co-ordinates; Change-Fix'n - number of movements between images leading directly to fixation points.

3.8 INTERPRETATION OF RESULTS (EXPERIMENT 2)

Total time taken to view the case (Time)

The time taken to view the cases tended to be greater when displayed in format A compared to format B but this difference failed to reach significance when all the cases were considered. However, there was significantly greater time spent evaluating the cases of interstitial lung disease in format A when compared to format B implying that interstitial lung disease is more difficult to analyze on smaller scale images, possibly because of the small scale of features such as, fine texture within abnormal lung parenchyma and traction bronchiolectasis.

Total number of saccades recorded for each case (Number)

There was a significantly greater number of saccades recorded from the format A display compared to the format B display, which reflected the trend seen in the time taken to read the cases and was largely attributable to the extra time taken to interpret cases of interstitial lung disease. However, the significant difference overall may in part reflect some loss of saccadic data during transitions between images viewed in format B. If the transitions, in practice taking less than 0.5 seconds, had been faster it is likely that the difference in time taken to view a case would have reached significance.

Percentage of non-tremor saccades within the range of useful peripheral vision (%Short)

The scale of the images viewed in format A were half that in format B. Thus any two features on an image would be closer together when viewed in format A compared to format B. It is therefore likely that a greater proportion of features on a smaller image (format A) could be visualised from a given fixation point compared to the proportion of those same features on an image of twice the size (format B), which are less likely to lie within a single field of useful peripheral vision, necessitating short saccadic movements to enable assimilation of an equivalent proportion of the image. In format B, a sub-analysis demonstrated a greater proportion of short saccadic movements in cases of ILD compared to cases of OLD, reflecting the greater range of useful peripheral vision in identifying mosaicism, the main abnormality in OLD.

Average time per fixation (Fix 'n time.)

The significantly longer fixation times in format A compared to format B most likely relate to the relative difficulty involved in the recording of image data at two different scales. Some features on the smaller scale image would be more difficult to discern and require greater visual processing and hence a longer fixation time. An apparent side effect of this prolongation in fixation time would appear to be an overall reduction in short saccadic movements as described above and implies the presence of parallel processing with peripheral vision providing information in tandem with foveal vision

Number of fixation points (No.Fix'n)

The raw data revealed a greater number of fixation points in format A cases compared to format B. However, the greater number of saccadic recordings in format A cases could explain the difference in the number of fixation points recorded. To eliminate the impact of this, the number of fixation points was normalised by dividing by the number of recorded saccadic movements resulting in a value expressed in terms of number of fixation points per saccadic movement. The result was a reversal in the difference between format A and format B indicating a greater number of fixation points in format B than format A in a given time frame. Nevertheless, as stated above, the average duration of fixation points in format B is less than that in format A.

Percentage of short saccades leading directly to a fixation. (%Short-Fix'n)

A greater proportion of short saccadic movements led directly to fixation points in format A than in format B. This difference was most marked when viewing the cases of obstructive lung disease and normal cases. Appreciation of a mosaic perfusion pattern has been shown to be identifiable by peripheral vision (experiment 1), but determination of whether the darker or lighter lung is abnormal requires regional scrutiny of the vessels and airways. Such small detail as presented in the smaller format A images probably required foveal fixation and a comparison of adjacent areas of dark and light lung resulted in multiple short saccadic movements leading to fixation points. On the larger

scale format B images, vessel calibre is more likely to be discernible with peripheral vision enabling comparison of adjacent regions of lung from a single fixation point, so reducing the number of short saccades leading to fixation. The assimilation of normal cases where abnormal features are not readily identifiable would lead to scrutiny of those features that are readily apparent, the vessels and airways. Again foveal fixation is likely to be required for this in the smaller scale images.

Percentage of long saccades leading directly to a fixation. (%Long-Fix'n)

A greater proportion of long saccades led directly to a fixation point in format B than in format A. This is the converse of the difference seen in the proportion of short saccades leading to a fixation point and was most apparent in cases of obstructive lung disease (OLD). It is possible that this finding is an epiphenomenon of the more frequent comparison of left and right lung performed in the scrutiny of cases of OLD where comparison of equivalent generations of vessels aids the distinction between normal and abnormal lung. Such a comparison in format B would require long saccadic movements but may be more readily appreciated in format A from peripheral vision. However, this should be taken in the context of the %Short-Fix'n finding that implies foveal fixation would be necessary in a left / right comparison, thus use of peripheral vision cannot explain the difference. An alternative explanation is the tendency to non-fixating saccades during thought processing which in terms of absolute scale (i.e. that defined by full screen resolution) would be constant in both formats but in the smaller scale format A data such saccades are more likely to be recorded as a long rather than short decreasing the overall proportion of log saccades leading to fixation.

Number of changes between images (No.Change)

There was a much greater number of movements between images in format A compared to format B. This is not surprising given the comparative ease of moving from one image to another; displaying both images simultaneously enabled a rapid comparison of those images. Comparison between adjacent images was more evident in the cases of ILD and OLD probably reflecting a tendency to greater comparison between images when abnormal findings are present. A proportion of these movements were probably non-functional, the eye wandering into one quadrant before settling as intended in another.

Distance relative to image co-ordinates at each end of a saccade traversing two images (ChangeDist)

The distance moved in terms of image coordinates was considerably less in format B compared to format A implying a comparison of equivalent regions of the lung on adjacent images in format B not evident in format A. The difference was most marked in cases where there were abnormalities and less obvious in normal cases (where a more systematic search is more likely). This behaviour is somewhat unexpected in reading non-contiguous images and may reflect the value of scrutinizing features over more than one image, or may merely result from an ergonomic tendency to maintain vision when a change of image is anticipated. The latter is supported in the finding that in format B alone the ChangeDist values for cases with ILD were lower than for OLD. In non-

contiguous imaging it is unlikely that a particular abnormal feature is scrutinised over multiple images but the eye is more likely to maintain a fixation from one image to the next if it had been concentrating on a regional feature, more prevalent in the cases of ILD.

Percentage of image changes followed immediately by fixation (%Change-Fix'n)

A greater proportion of movements between images led to fixation in format B than format A and was evident in all cases. This data supports the idea that a proportion of perceived movement between images is artefactual and results from random movements around the screen that would be recorded as movement between images in format A. However, in conjunction with a marked difference in ChangeDist between the formats the %Change-Fix'n data also reflects a difference in the ability to rapidly fixate on an image after movement when the image itself has changed without the need to move the eye. The facilitation of sequential fixation of a specific region in multiple sequential images is of more relevance in contiguous imaging.

Percentage of time for which an enlarged image was viewed in format A (%Enlarged time)

In format A full screen viewing was used for, on average, 16.6% of the time indicating the majority of the image interpretation was performed at the lower resolution of the smaller images. The interpretation of the differences between the two formats therefore

appropriately assumes that the vast majority of the data corresponds to the viewing of the smaller images associated with the two by two-tile format.

3.9 DISCUSSION

Experiment 1 demonstrated the large area of peripheral vision in which a mosaic attenuation pattern can be appreciated. This is not surprising given the design of the retina to facilitate contrast definition under low incident intensity in the periphery. The extent of useful peripheral vision in identifying small detail, in this case airway and vascular detail, varied between observers but for one individual it effectively encompassed about a quarter of the screen, a finding previously noted in the detection of nodules on CXR (Kundel, Nodine, & Carmody 1978). Any analysis of eye-tracking needs to take account of the input from peripheral vision, which can be processed in parallel with foveal vision and occurs at a subliminal level. The data presented shows that the vast majority of saccadic movements occur within the ambit of useful peripheral vision. Whether such short saccadic movements are led by peripheral vision processing or merely reflect a tendency to short saccades during a visual search process is a matter of speculation and cannot be ascertained from these experiments but the use of peripheral vision as an adjunct to foveal search has been suggested in relation to nodule detection on chest radiographs (Kundel, Nodine, & Toto 1991).

Previous experiments to determine the efficacy of viewing CT scans on computer screens have mostly made comparisons with hard copy alternatives and provided twelve to twenty images on a single monitor at any one time (Beard, Johnston, Toki, & Wilcox

1990;Reiner, Siegel, Hooper, Pomerantz, Dahlke, & Rallis 2001;Strickland & Allison 1995). The computer hardware available for these studies was relatively limited and the slow speed of computer screen redrawing, relative to current computer processing speed, favoured formats with a minimal number of screen image changes. Faster computer hardware has enabled a departure from the attempted duplication of hard copy presentation on computer monitors and allowed viewing formats not possible on hard copy. The most significant of these has been described as "stack mode" and corresponds to format B in experiment 2. Stack mode viewing has been demonstrated to be superior to tiled images (e.g. format A) in the assimilation of three dimensional structures on contiguous axial images (Hoffman et al. 1983), this can be explained entirely in terms of the physiology of retinal processing. The retina is highly adapted to recognising changes in interface boundaries as would occur when viewing a CT in stack mode (Mathie & Strickland 1997). However, the impact of stack mode imaging on the assimilation of non-contiguous imaging (i.e. a typical HRCT protocol) has not been assessed, where the perceived advantages of stack mode are less relevant. Nevertheless, this experiment suggests that stack mode is preferable. One obvious advantage stack mode has over tiled images in non-contiguous imaging is the greater size of the images, and therefore the image features. If the size of each image in the tiled display is sufficient to allow the scrutiny of small features then the advantage of being able to view multiple images at one time should be overwhelming. The observers spent on average 16.6% of the time in format A with the images in full screen mode implying that most features were discernible at the native smaller scale resolution, with no difference in observer accuracy seen between the two formats.

The data presented in this chapter represents raw values with no attempt to interrogate the temporal or spatial components. A technique described that analyses the direction of saccades (Ponsoda, Scott, & Findlay 1995) was not used as it would translate well to the three-dimensional data sets used in this experiment. Despite this limitation some useful observations about the visual efficacy of the two formats can be made. The analysis of specific saccadic movements and fixation behaviour identifies the adaptive behaviour in terms of eye movements between the two formats and demonstrates how the observers cope with smaller images but maintain the ability to assimilate them. However, the time taken to view the cases in tiled mode tends to be greater than that taken in stacked mode, particularly in cases of interstitial lung disease. The discrepancy may be a result of an increase in time taken to assimilate features on smaller scale images or a manifestation of the slightly more complex manipulation involved in viewing in tiled mode. Although the observers felt able to assimilate the HRCT images at the smaller size an expected improvement in efficiency resulting from the display of four images at once was not apparent. The analysis of the raw eye–tracking data has led to some assertions as to how eye-movements performed in the viewing of HRCT images are tailored to the format in which the images are presented. This data will be taken into consideration when the visuospatial data is analysed in the following chapter and has led to the adoption of format B for subsequent experiments.

3.10 SUMMARY

1. ***Some aspects of the visual approach to the soft copy reporting of HRCT scans can be objectively elucidated from eye-tracking data***
2. ***Stack mode presentation, considered advantageous in viewing contiguous axial images, facilitated more efficient image interpretation of non-contiguous HRCT scans due to the larger size of each axial image.***
3. ***The presented image format did not appear to have an impact on reporting accuracy.***

CHAPTER 4

A VISUOSPATIAL ANALYSIS OF THE APPROACH TO READING HRCT SCANS DISPLAYED IN TWO DIFFERENT VIEWING FORMATS.

4.1 INTRODUCTION

In Chapter 3, the analysis of eye movements of experienced radiologists, when confronted with two different display formats, demonstrated some fundamental differences. However, the analysis did not take account of spatial information and considered only how the eyes moved and which image was being observed. The analysis is now extended to include the spatial information recorded during the experiment, i.e. the coordinates on the images to which the eye was tracked. It has been established that the observers were able to formulate valid interpretations of all the CT cases and that this was more rapidly achieved in the stack format (format B) compared to a tiled format (format A). The following analysis was aimed at determining whether there were any differences in the regions of the HRCT images covered during case reading between the two format types, and whether there was a difference in the visuospatial approach to the cases that paralleled the previously described differences in eye movement behaviour.

4.2 MATERIALS AND METHODS

The experimental design is described in Chapter 3. In order to compare spatial data from different cases, which differ in terms of anatomy, a standardised "normal lung" was created from HRCT data. The "normal lung" construct comprised fifteen axial CT sections at anatomical levels defined by mediastinal structures. At each level the lung parenchyma for the left and right lungs was separated from the CT image, its outline defined and an average shape for the lung at that level was calculated using hierarchical distortion (see Chapter 2). For each case used in the main experiment a distortion grid was generated to match the shape of the lung in the experimental case to that of the "normal lung", at an equivalent anatomical level; the distortion grid calculated was then used to map the eye-tracking information onto the "normal lung" template.

4.2.1 Extracting the lung parenchyma.

For every image section in each case the left and right lung parenchyma were separated from each other and the surrounding tissues, and stored as individual graphic files. The process was identical to that used in the extraction of lung parenchyma from the normal cases in the creation of the "normal lung" (see Chapter 2).

4.2.2 Anatomical matching

The fifteen anatomical levels of the "normal lung" were defined by obvious mediastinal structural landmarks so that representative axial images from each of the normal HRCT cases were obtained. The actual number of axial CT images containing lung parenchyma on an HRCT scan varies from twenty to more than thirty, depending on the habitus of the individual. The selection of fewer anatomical levels than the minimum that would be acquired in an HRCT examination of an adult, facilitated anatomical matching of sections with more than one section being mapped to a given "normal lung" section.

All of the HRCT cases used in the experiment had more than fifteen axial images displaying lung parenchyma. The images that did not correspond to one of the fifteen anatomical layers were assigned to the nearest match. In practice such images were matched to the nearest "normal lung" level in the caudal direction.

4.2.3 Distortion grid construction

The same hierarchical distortion algorithm employed in the creation of the "normal lung" was used to distort the experiment cases sections to the appropriate "normal lung" section. The lung parenchyma from the axial images was extracted at each level and an outline generated for each side of the lung at each axial level. The hierarchical distortion algorithm required some two dimensional detail within this outline. Therefore, the lung outline and the equivalent level "normal lung" outline were filled with random noise.

The lung parenchyma, now filled in, from the experiment cases was distorted to match the equivalent, filled in, "normal lung". The distortion was performed and recorded separately for each side (left and right) for each section within each case. The distortion grid required to perform this shape matching for each side was recorded. Therefore, there were two distortion grids recorded for each image section at each level for each case; one grid for each side.

4.2.4 Separation of the eye-tracking data

The eye-tracking data for each observer was recorded as a continuous text file comprising a single line for each element of eye-tracking data (see Chapter 2). For the process of mapping the eye-tracking data for a case onto the normal lung, the text file was first divided into individual files containing the data for each individual case. Each of the resulting case files was then divided into individual files corresponding to individual sections within a case. The resulting section files were then divided into the data corresponding to the left and right lung separately using the lung parenchyma mask employed in the distortion grid calculation. The process of splicing the eye-tracking data created data files that matched the anatomical source from which the distortion grids were calculated for each case (Fig. 4.1).

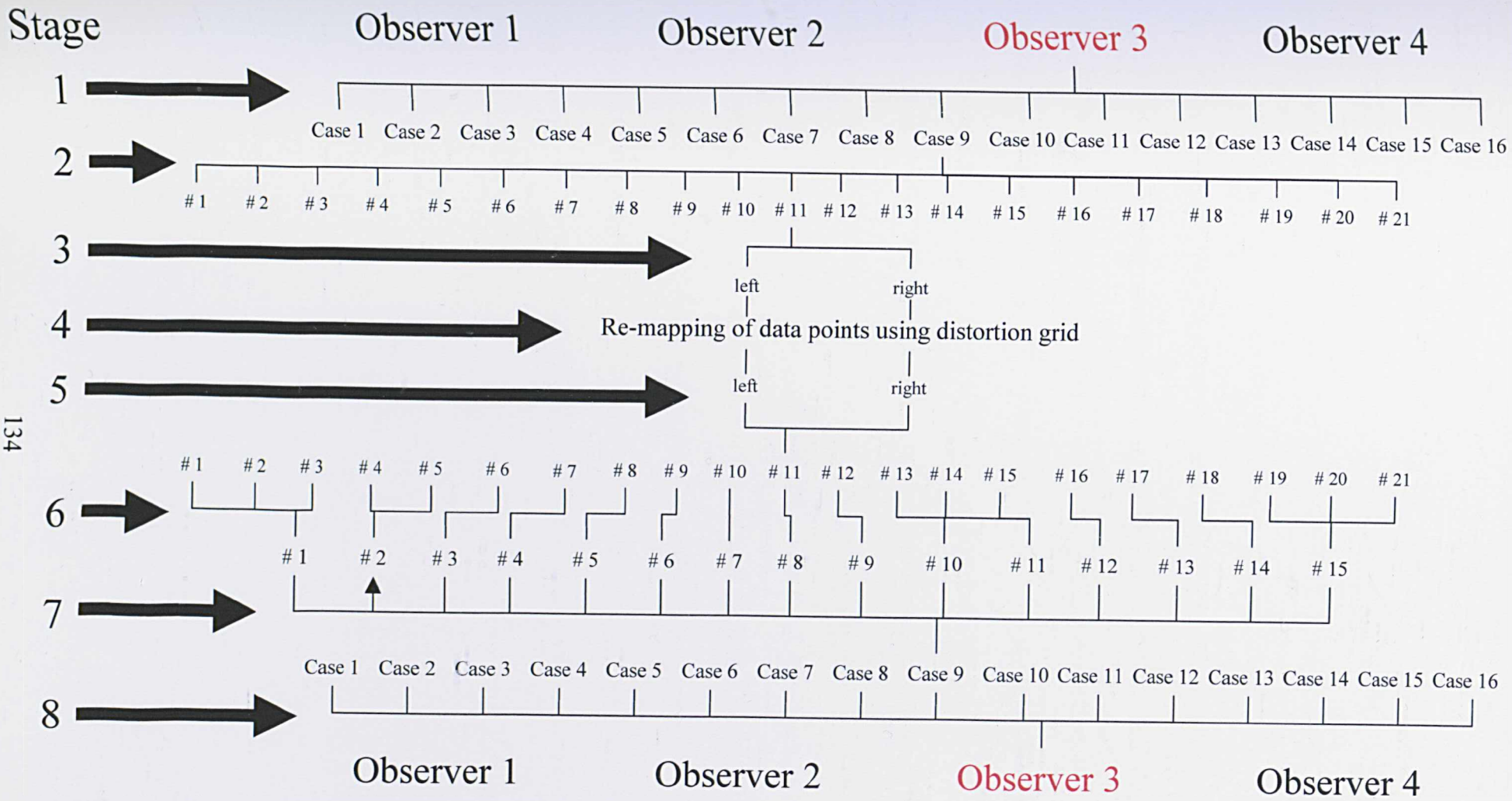


Figure 4.1: Summary of data splicing and concatenation of the raw eye-tracking data using Observer 3 prior to and following re-mapping onto the normalised lung via distortion grids.

Stage 1 - splicing of raw data set for each observer into files corresponding to each case; **stage 2** - splicing of case data into data for each individual section; **stage 3** - splicing section data into left and right halves; **stage 4** - data re-mapped using distortion grids defined by normal lung level for each section; **stage 5** - concatenation of left and right data sets to form a single data set for each section; **stage 6** - sections assigned to normalised lung levels and data averaged if more than one section assigned to any specific level; **stage 7** - the data for the 15 resulting levels concatenated to form a single case; **stage 8** - the cases for each observer concatenated to form a single file for each observer. # = section number.

4.2.5 Eye-tracking data point mapping.

Each line of eye-tracking data contained co-ordinate values in two-dimensional space corresponding to the point on the image being viewed to which the eye was tracked. Following alteration of the co-ordinates, for each eye-tracking data point, through the relevant distortion grid, each point was represented by a circle on a 512 x 512 pixel matrix; the circle diameter corresponded to the equivalent of two degrees of visual angle An accepted figure for the foveal region of the retina). A time value of twenty milliseconds (the eye-tracking device recorded fifty points per second) was added to the value of the pixels at the centre of the circle and smaller values added to the pixels further from the centre. The smaller values were calculated based on the imposition of a Gaussian distribution of values with a peak of twenty milliseconds at the centre (Fig. 4.2). The photoreceptor density of the fovea follows a Gaussian distribution peaking at the centre (Osterberg 1935). The values assigned to each pixel were cumulative and the resulting 512 x 512 array of values represented a two-dimensional plot of the time spent viewing areas of the image (Fig. 4.2). Where more than one section was assigned to a given "normal lung" section, the final 512 x 512 matrix comprised the mean values of the contributory section matrices.

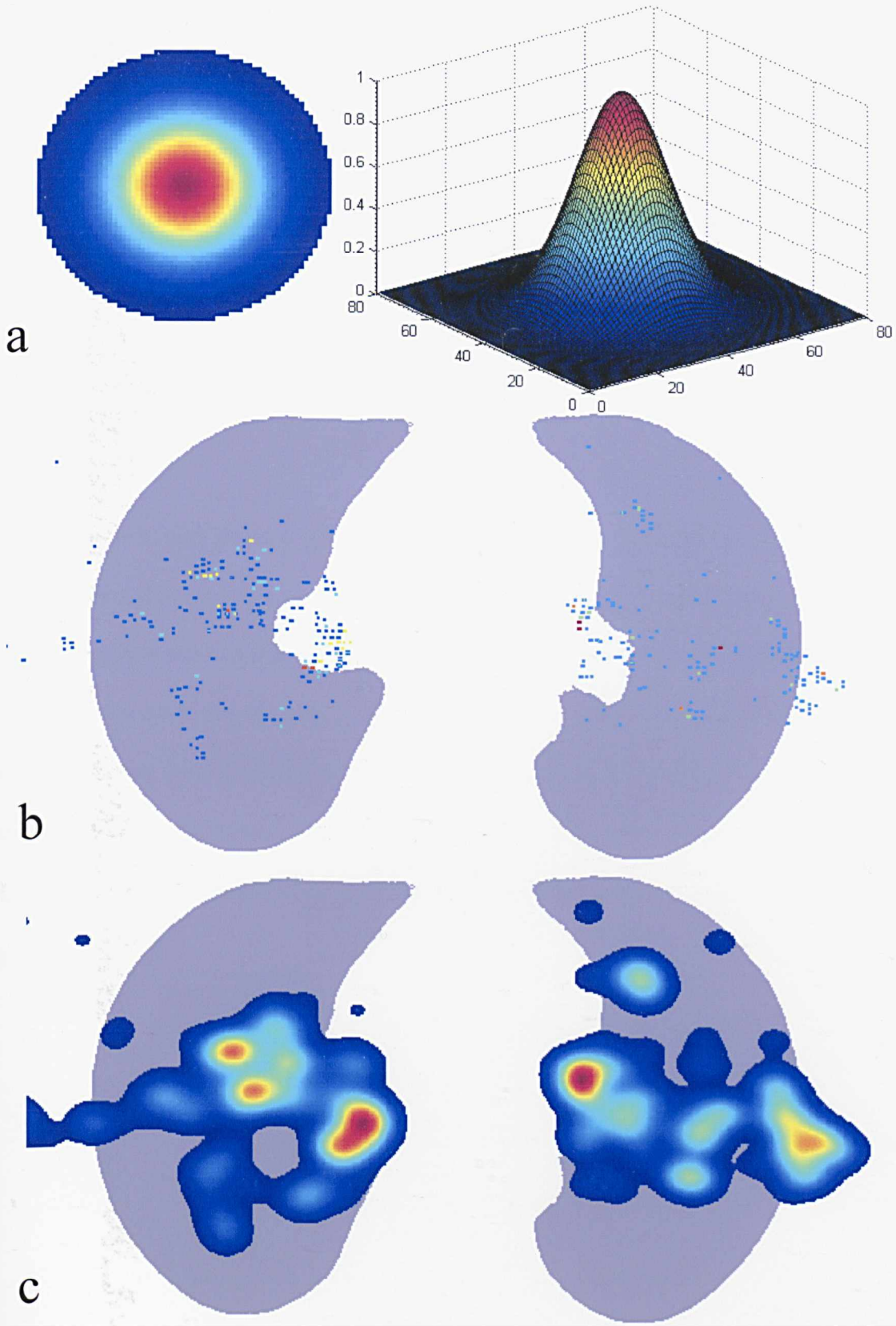


Figure 4.2: Description of the mapping of eye-tracking data points by defining each as a circular area projected by a 2° cone with density weighting determined by a Gaussian distribution. **(a)** demonstrates an isolated circular area and the variation of weighting depicted by colour and then demonstrated on a 3-dimensional plot with weighting on the y axis. **(b)** shows eye-tracking data mapped as individual points and **(c)** demonstrates the appearances when each point is plotted as a circular area following a Gaussian distribution for weighting

Raw eye-tracking data were used for the data mapping, no attempt was made to determine whether any specific data point formed part of a fixation point. The utilisation of raw data ensured that all the eye-tracking information available was used and the data point density provided some element of fixation point data; regions in which fixation occurred would record a higher data point density and therefore be apparent in this analysis.

No attempt was made to alter the area of foveal vision-based on expected distortion when viewing away from a central eye to monitor axis. The flat nature of the monitor screen would result in an oval-shaped projection of the cone of foveal vision when directed off-centre. The error inherent in this simplification was calculated as a maximum of an 18% error in projected foveal error for the geometry of the experimental set-up (Fig. 4.3), which was felt to be a minor source of systematic error by comparison to a potential 125% random error in calculate foveal area as a result of the $\pm 0.5^{\circ}$ accuracy of the eye-tracking camera.

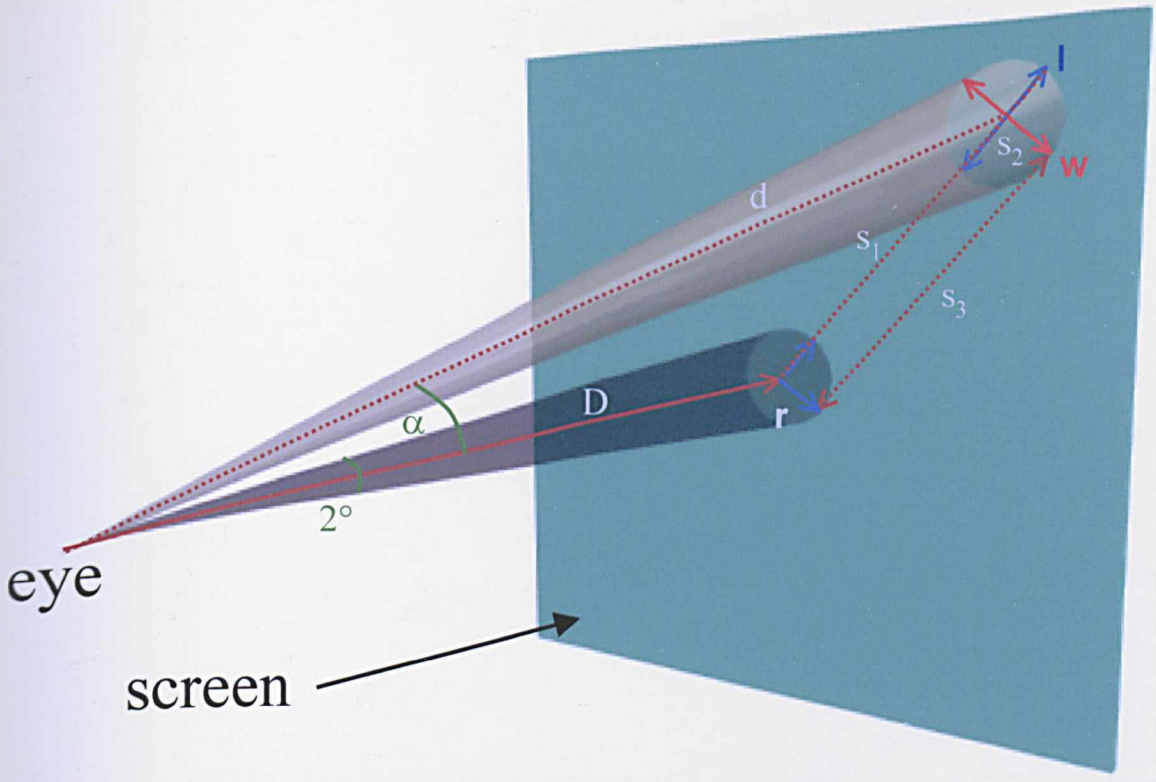


Figure 4.3: Demonstration of the distortion of the circular area projected by a cone when angled off the central axis; 'D' is the distance from the eye to the screen, 'r' is the radius of the circle of projection of the cone with the vertex at the eye and aligned along the orthogonal axis to the plane of the screen. 'w' and 'h' represent the width and length of the oval of projection of the cone when at an angle α from the orthogonal axis.

The percentage change in projected area is calculated as follows:

A_1 represents the area of the initial circle and A_2 the are of the resultant oval.

$A_1 = \pi r^2$, $A_2 = 0.8wh$

$P = (0.8wh / \pi r^2) \times 100$ (P = percentage increase in area) [1]

$r = D\tan 1$ (for a cone with a 1° vertex angle) [2]

$h = s_2 - s_1$, $s_2 = D\tan(\alpha+1)$, $s_1 = D\tan(\alpha-1)$ [3]

from [3] $h = D(\tan(\alpha+1) - \tan(\alpha-1))$

$w = 2d\tan 1$, $d = D / \cos\alpha$

$w = (2D\tan 1) / \cos\alpha$ [4]

substituting from [2], [3] and [4] into [1]

$$P = \frac{1.6\,(\tan(\alpha+1) - \tan(\alpha-1)) \times 100}{\cos\alpha\ \pi\ \tan 1} \quad [5]$$

The maximum angle α for the experimental set-up was 18 ° giving P of 118%. The reported accuracy of the eye tracking apparatus is ±0.5 °, which is 25% of the foveal visual angle and equates to potential area increase of 225%.

4.3 STATISTICAL ANALYSIS

4.3.1 Post-processing of data

The 512 x 512 arrays of data points for each side of each section created an extremely large data set with a large quantity of random noise. In order to facilitate the comparison of data sets between cases the data were simplified.

4.3.2 Masking and coarsening of data

The data file created for each case included a large amount of data points corresponding to regions outside the lung parenchyma; the area of lung parenchyma represented approximately 30-70% of the overall 512 x 512 array, as a result of the constant size of the 512 x 512 matrix with a varying area of lung parenchyma dependent on the anatomical level being represented, and most of the matrix elements not representing lung parenchyma contained zero values. Therefore, a mask derived from the "normal lung" was used to retain the matrix elements corresponding to the lung parenchyma and discard the remainder. As the same mask was used for all the data sets, the number of remaining matrix elements at each individual level, and therefore within each case, were equal for all the cases and each data element corresponded to a specific anatomical location within a section.

The selection of matrix elements corresponding to lung parenchyma reduced the size of the data set but did not reduce the influence of random noise. Therefore, the data were

coarsened according to an eight by eight grid applied to each side of each level. Finer grids (32 x 32 and 16 x 16) were considered but failed to smooth out random fluctuations (Fig. 4.4). In order to maintain a constant grid square size for all levels, the grid was defined in terms of the original 512 x 512 matrix and those grid squares occupied at lease one-eighth by lung parenchyma were included. In order to facilitate comparison between observers, who spent different lengths of time on each case, the grid square values, derived by summing the duration recorded in each of the matrix elements within the grid square, were normalised for the total time taken on each case by each specific observer such the grid square value represented the proportion of the time spent viewing that grid square, termed the time density, relative to the time spent on the case overall. Each grid square was then ranked, on a four point scale based on thresholds derived from a histogram of all the normalised data values (Figs 4.5 and. 4.6). Calculations were performed using Matlab 6.1 (Mathworks Inc., Natick MA, USA).

4.3.3 Comparison of data-sets

As a single normalised lung was used as a mask for all the cases, each case file, following mapping, masking and normalising, had the same number of data points. The resulting paired of data sets were compared using the kappa statistic (with quadratic weighting) for observer agreement. The subsequent kappa values were compared using the Wilcoxon's signed rank test and stepwise linear regression analysis to determine the significant factors affecting observer consistency in the recorded visuospatial data.

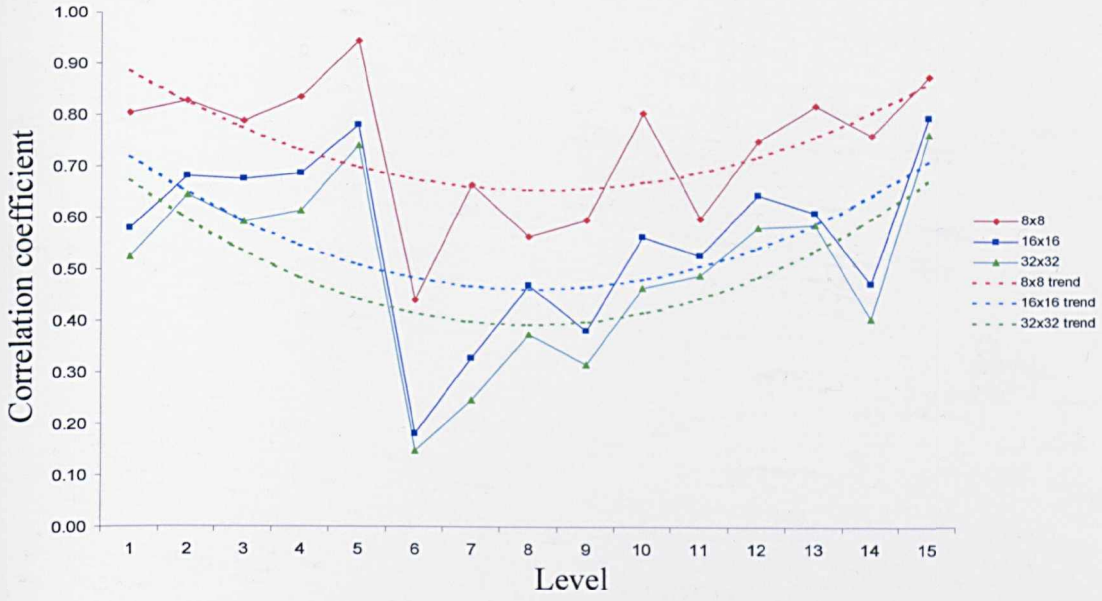


Figure 4.4: Plot of correlation coefficient against anatomical level for a comparison of the eye-tracking derived time density data for DMH and SRD for case 1. Comparisons of 32x32 grid,16x16 and 8x8 grids are plotted; reduction in data fluctuation is evident in the 8x8 grid data. The trend lines indicate higher correlation at the more cranial or caudal levels.

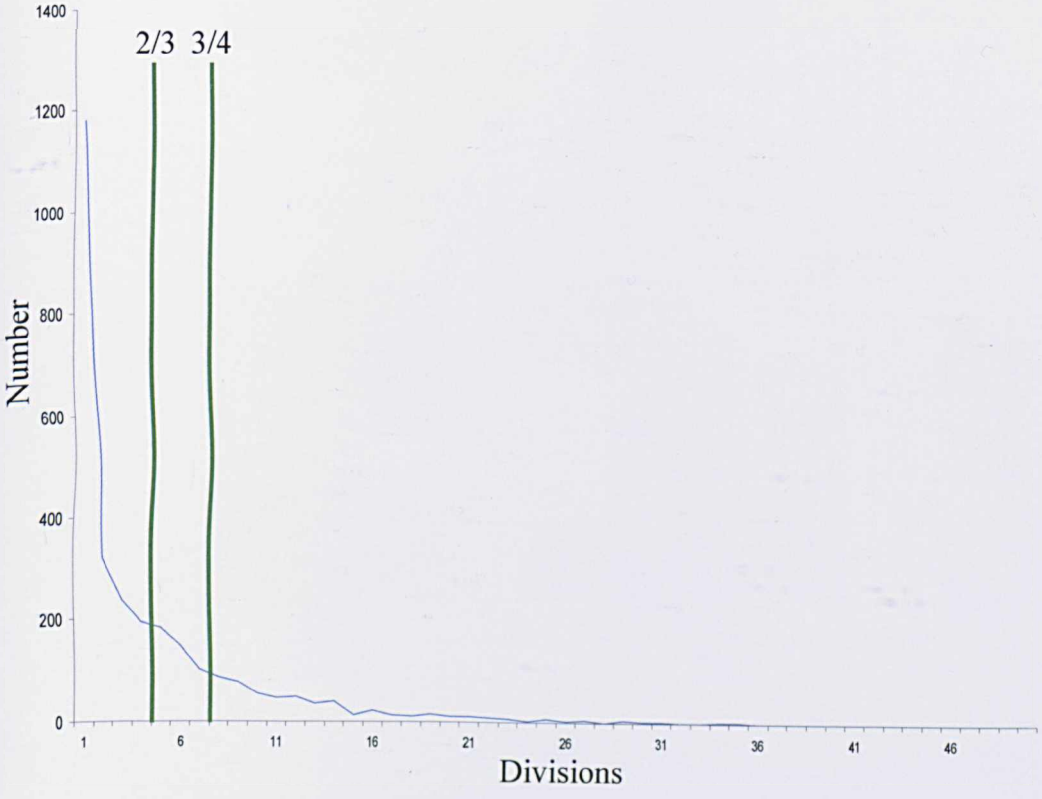


Figure 4.5: Histogram of normalised time density from the eye-tracking data: Threshold points for the calculation of ranks are shown and correspond to points at which the histogram demonstrates a departure a smooth curve, rank 1 was taken as zero, the limits for rank 2 were >0 and ≤0.00358, for rank 3 >0.00358 and ≤0.00716 and for rank 4 >0.00716.

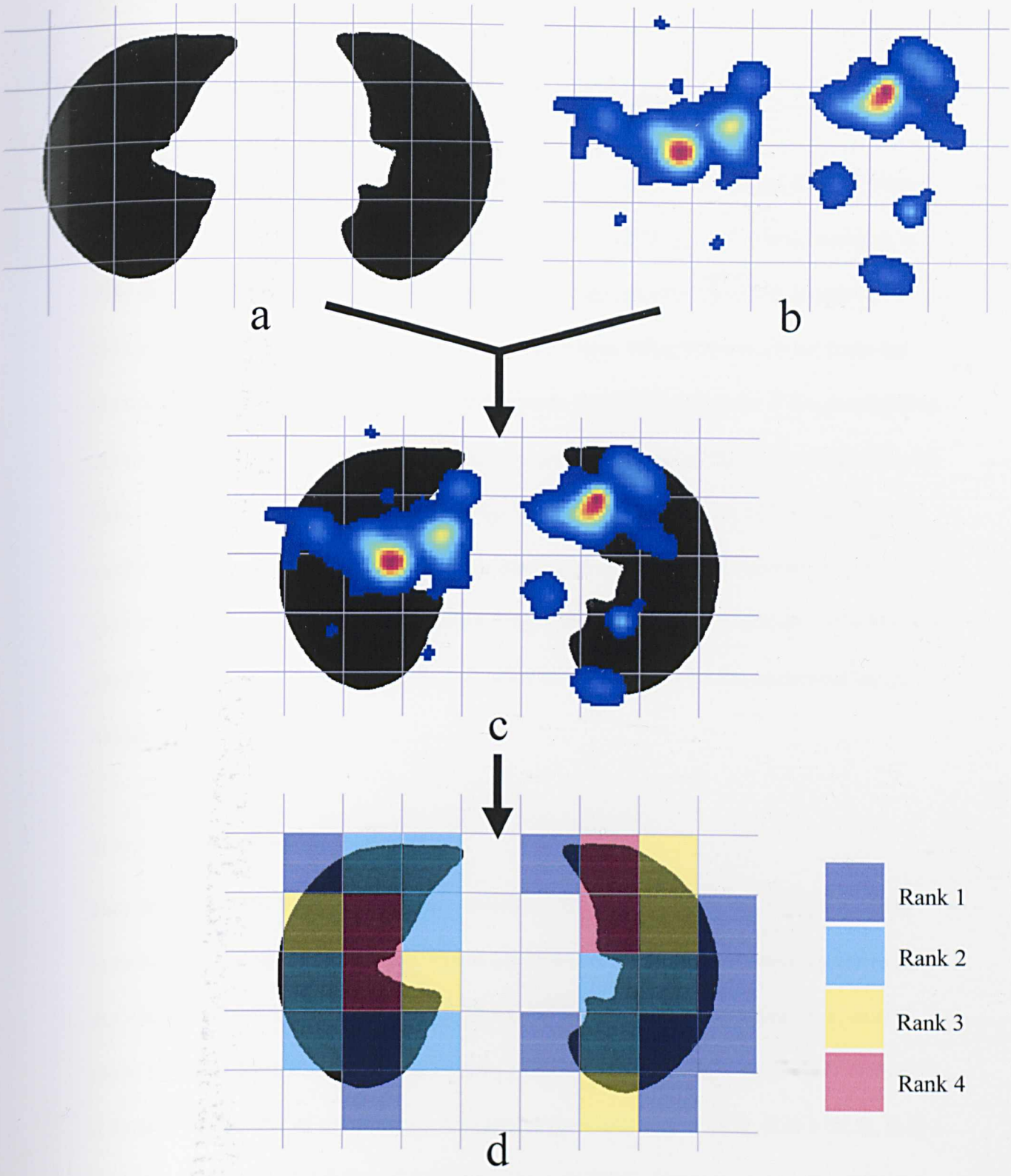


Figure 4.6: Explanation of the processing of eye-tracking data prior to analysis. **(a)** is the mask derived from the normalised lung, **(b)** is the plot of the time density distribution for the eye-tracking data at the same anatomical level following grid distortion to match the normal lung, **(c)** displays the time density distribution overlaid on the normal lung mask and **(d)** displays the ranked assignment to the grid squares dependent on the total time density found within them.

4.3.4 Inter-observer analysis

For each observer the data files for the cases presented in format A (tiled format) were concatenated, as were those for cases presented in format B (stack format), resulting in two data sets per observer. For each of six possible paired inter observer comparisons (observers 1 v 2, 1 v 3, 1 v 4, 2 v 3,2 v 4,3 v 4) a kappa value was calculated from the data file derived from cases presented in the same format to determine if the presentation format affected inter-observer visuospatial behaviour. The resulting kappa values (six for each format) were then compared using the Wilcoxon's signed rank test. A further sub-analysis was performed in which the data from cases of one of three disease types (interstitial lung disease-ILD, obstructive lung disease-OLD and no disease-NORM) was analysed separately to determine the influence of disease type on the calculated kappa values.

4.3.5 Within observer analysis.

For each observer, kappa values were calculated for every possible paired comparison between the sixteen case data sets. The result was one hundred and twenty kappa values for each observer. Each case comparison was labelled in terms of the observer, one of three potential format comparisons (format A v format A, format A v format B, format B v format B) and one of six potential disease comparisons (ILD v ILD, ILD v OLD, ILD v NORM, OLD v OLD, OLD v NORM, NORM v NORM). The three respective labels were then ranked according to their overall mean kappa values and the resulting data set

including all the within observer observations was then analysed using stepwise multiple linear regression.

4.4 RESULTS

The results for inter observer comparisons are summarised in Table 4.1. The degree of similarity in the visuospatial approach to the cases was significantly greater in the cases presented in stacked format (format B) compared to those presented in tiled format (format A) [$p<0.05$]. Only in stack format were the kappa values consistently greater than 0.4, taken as the lowest limit for acceptable observer agreement (Brennan & Silman 1992). The similarity in approach to cases of ILD cases was significantly greater than that to NORM cases but no significant difference was demonstrated between the approaches to ILD and OLD cases and OLD and NORM cases (Table 4.2).

Observers compared	Whole data set (kappa)	Tile format (kappa)	Stack format (kappa)
1 v 2	0.38	0.29	0.45
1 v 3	0.42	0.39	0.43
1 v 4	0.45	0.46	0.44
2 v 3	0.37	0.28	0.43
2 v 4	0.42	0.32	0.48
3 v 4	0.46	0.41	0.48

Table 4.1: Summary of inter-observer kappa values comparing tiled format with stack format. The kappa values for stacked format were significantly greater than those for tiled format ($p<0.05$).

Observers compared	ILD	OLD	NORM
1 v 2	0.43	0.35	0.36
1 v 3	0.45	0.40	0.40
1 v 4	0.51	0.39	0.46
2 v 3	0.34	0.42	0.35
2 v 4	0.47	0.43	0.34
3 v 4	0.49	0.42	0.48

Table 4.2: Summary of inter-observer kappa values comparing cases of different disease type. The kappa values for ILD cases were significantly greater than those for NORM cases ($p<0.05$) but no other statistically significant differences were seen.

The ranking of observers, format comparison and disease comparison, based on overall mean kappa values, is summarised in Tables 4.3, 4.4 and 4.5

Rank	Observer	Mean kappa
1	3	0.28
2	2	0.29
3	4	0.35
4	1	0.37

Table 4.3: Definition of observer ranking based on mean kappa values.

Rank	Format comparison	Mean kappa
1	Tiled v Stack	0.30
2	Tiled v Tiled	0.31
3	Stack v Stack	0.40

Table 4.4: Definition of format comparison ranking based on mean kappa values. Tiled corresponds to format A and stack to format B.

Rank	Disease comparison	Mean kappa
1	OLD v OLD	0.29
1	NORM v NORM	0.29
2	OLD v NORM	0.31
3	ILD v OLD	0.33
4	ILD v NORM	0.34
5	ILD v ILD	0.35

Table 4.5: Definition of disease comparison ranking based on mean kappa values. ILD – interstitial lung disease, OLD - obstructive lung disease, NORM - no disease.

The result of the stepwise linear regression analysis is summarised in Table 4.6. All three variables, observer, format comparison and disease comparison had a highly significant independent influence on the resulting kappa values with increasing kappa values corresponding to increasing rank.

Kappa values	Coefficient	p	95% confident interval	
Disease comparison	0.013	<0.0001	0.007	0.19
Format comparison	0.041	<0.0001	0.031	0.037
Observer	0.03	<0.0001	0.022	0.16

Table 4.6: Summary of the result of stepwise linear regression analysis on the within observer data. $R^2 = 0.25$. There is a significant positive correlation between ranking of observer, format comparison and disease comparison with the calculated kappa values.

Assuming a narrow confidence interval reflects a more powerful relationship, then disease comparison has the weakest relationship with the kappa values and format comparison and observer have an equal influence.

4.5 DISCUSSION

In Chapter 3, a difference was demonstrated in the eye-movements employed during the reading of HRCT cases dependent on which of two different formats was used to display the cases. The conclusion was that, based on the analysis of the type of eye movements alone, the observers approach to the reading of HRCT cases was more efficient in stack format compared to tiled format. From that result a more consistent visuospatial approach to cases presented in stack format compared to tiled format might be expected. The current analysis demonstrates exactly that, with greater within observer and inter-observer consistency in the visuospatial approach to cases presented in stack format as compared to cases presented in tiled format.

There are some limitations to the analytical process used in this experiment. The primary difficulty in the analysis of eye-tracking data, recorded at fifty data points per second, is the vast amount of data generated. In order to perform a meaningful analysis of the large resultant data sets, some degree of coarsening and simplification was necessary. In this experiment, although eye-tracked data points were theoretically recorded at screen pixel resolution, the actual accuracy, and therefore effective resolution, of the eye-tracking camera set-up, did not permit such high resolutions. The eye-tracking camera itself had an error of up to 0.5°, which in the experimental set-up equated to thirty to sixty pixels dependent on whether full screen or 2 x 2 display was in use. The use of distortion grids to map data points onto an anatomically normalised lung facilitated a legitimate comparison of cases that had different global anatomies. However, there is a degree of inaccuracy associated with distortion grid mapping, particularly when there is a marked disparity in cross-sectional lung parenchymal shape between the experimental case and the normalised lung; this is most marked in the lower sections where the hemi-diaphragms become significant determinants of lung parenchymal shape; the price to pay for "between case" comparisons is a reduction in data accuracy.

When mapping the data points the narrow 2° angle of foveal vision was taken to represent the area being scrutinised at any particular point. This does not take into account the input from peripheral vision, which was demonstrated in chapter 3 to contribute high-resolution information from a surprisingly large area. The foveal visual angle was employed to concentrate the analysis on the elements of visual search that are under the greatest

degree of conscious control (Rybak et al. 1998). The partitioning of mapped data into a coarse sixty four pixel sized grid (original array five hundred and twelve pixels squared) and subsequent ranking of the data for each grid square using a four point scale created manageable data files suitable for analysis using the kappa statistic. The degree of statistical significance evident in the within observer calculations suggest that such a simplification did not profoundly affect the analysis. No effort was made to identify fixation points. The type of data mapping used incorporated fixation data into a density map; a fixation point resulted in greater time recorded in a particular area. However, when comparing the data analysis from this chapter with that of the previous chapter one should be aware of the absence of fixation point data.

There was a trend toward a more consistent approach to cases of ILD by comparison to the other case types seen in the inter-observer data (reaching significance when compared to NORM cases) and in the within observer data (ILD v ILD given the highest ranking). Furthermore, comparisons involving the normal cases occupied the lowest three ranking places in the within observer analysis. The results suggests that the approach to HRCT cases that display features of disease are more consistent than that to cases displaying no disease. The implication is that the observers do not seem to have a standard anatomical-based approach to HRCT scans but display a feature targeted approach resulting in greater consistency in visuospatial data when disease features, which in interstitial lung disease follow a non-random anatomical, are present. This assertion will be examined in the following chapters where the visuospatial analysis performed here is compared to an analytical technique based on identification of disease features. Furthermore, the current

analysis has demonstrated differences in the consistency of individual observers, providing a potential measure for grading experience, which is investigated in the next experiment.

4.6 SUMMARY

1. ***The visuospatial data correlates with the raw eye movement data, presented in the last experiment, demonstrating more efficient reading of cases presented in stack format compared to tile format, as assessed by consistency of approach.***
2. ***Experienced observers confronted with normal cases do not seem to demonstrate a consistent approach.***
3. ***Greatest consistency in approach is seen in cases of interstitial lung disease.***
4. ***Observers appear to have a more feature-led approach to reading cases than an anatomy led approach.***

CHAPTER 5

ANALYSIS OF THE STRATEGY USED BY EXPERIENCED RADIOLOGISTS ASSIMILATING HRCTs.

5.1 INTRODUCTION

The experiments and analyses to date have been concerned with the general approach to HRCT images employed by experienced radiologists with particular reference to the impact of differing presentation of the images (i.e. formats). The subsequent experiments were conducted in stack format as this was read more efficiently than tiled format. In addition to comparisons of presentation format, some characteristics of visual search were identified as independent of the presentation format (see Chapter 3). The very high proportion of saccadic movements within the range of calculated useful peripheral vision and greater consistency of approach to an HRCT scan in the presence of disease compared to an absence of disease, and therefore features of disease, implies that a feature-led, as opposed to an anatomy-led, search strategy is generally used. The following experiment aimed to identify the nature of any feature-led search strategies by recording the eye movements of experienced radiologists confronted with a larger number and variety of abnormalities, more closely approximating clinical reality, than presented in the previous experiment.

5.2 MATERIALS AND METHODS

The ASL500 eye tracking camera and computer monitor were set up as described in Chapter 2

5.2.1 Case distribution.

Thirty HRCT examinations were selected and copied in DICOM format from a computer archive of extant clinical cases referred for investigation of suspected or known diffuse interstitial or obstructive lung disease, the final diagnosis was based on clinical and imaging findings with histological verification when available. For each case five images were selected that matched levels one, four, seven, nine and thirteen on the normalised lung, used for data processing in Chapter 4. The range of features found in each case is summarised in Table 5.1. The disparity in numbers of cases assigned as interstitial lung disease (ILD) compared to those assigned as obstructive lung disease (OLD) reflects the greater number of features to be found in ILD.

5.2.2 Case display

The cases were displayed sequentially in stack format on a single monitor using the same DICOM viewer employed in the experiment for Chapter 3. Reducing the number of images per case decreased the time taken to read each case and enabled the accurate reading of ten cases before re-calibration became necessary. Calibration, as described previously, was therefore performed on average three times, once at the beginning of the experiment and then repeated after every ten cases; more frequent calibration was

performed if there appeared to be an error in the eye-tracking visualised in real time on the preview monitor.

Case number	Diagnosis	Features	Disease type
1	Unspecified interstitial lung disease	GG, traction BX, Retic.	ILD
2	Constrictive obliterative bronchiolitis	mosaic	OLD
3	Sarcoidosis	nodules, Retic	ILD
4	*Final diagnosis not established*	GG, BX, mosaic	ILD
5	Bronchiectasis & emphysema	BX, mosaic, Retic	OLD
6	*Final diagnosis not established*	mosaic, nodules	ILD
7	Non-specific interstitial pneumonia	GG, Retic	ILD
8	Fibrosing alveolitis	Retic, GG, BX, fibrosis	ILD
9	*Final diagnosis not established*	GG, Retic	ILD
10	Small airways disease (Sjögren's syndrome)	mosaic	OLD
11	Non-specific interstitial pneumonia	GG, Retic	ILD
12	Lymphangioleiomyomatosis	Cysts	ILD
13	Extrinsic allergic alveolitis	nodules	ILD
14	Emphysema	mosaic	OLD
15	Respiratory-bronchiolitis interstitial lung disease	mosaic, Cysts, nodules	ILD
16	Extrinsic allergic alveolitis	GG, traction BX, Retic	ILD
17	Non-specific interstitial pneumonia	GG, traction BX, Retic	ILD
18	Alveolar proteinosis	GG, Retic	ILD
19	Small airways disease	BX, GG, mosaic, nodules	OLD
20	Pneumoconiosis	nodules, Retic	ILD
21	Wegener's granulomatosis	GG, Retic, fibrosis	ILD
22	Sarcoid, bronchiectasis & haemosiderosis	BX, mosaic, nodules, GG	ILD
23	Allergic bronchopulmonary aspergillosis	GG, BX	ILD
24	Obliterative bronchiolitis	mosaic, BX	OLD
25	Emphysema & Non-specific interstitial pneumonia	mosaic, GG, Retic, Cysts	ILD
26	Pulmonary embolism	mosaic	OLD
27	*Final diagnosis not established*	GG, BX, Retic	ILD
28	Emphysema	mosaic	OLD
29	Fibrosing alveolitis	Traction BX, Retic, GG, Cysts	ILD
30	Lymphangioleiomyomatosis	Cysts	OLD

Table 5.1: Summary of features in the cases used in the experiment; GG = ground glass, BX = bronchiectasis, mosaic = mosaic attenuation, Retic = reticulation, ILD = interstitial lung disease, OLD = obstructive lung disease.

5.2.3 Experiment sequence

Each of three experienced radiologists (DMH, MBR, and SRD) underwent the experiment once. Initially a calibration was performed, the observer then viewed ten cases, each comprising five images, in sequence, providing verbal feedback, with regard to identified features and diagnosis, for each case, without moving from the head stabilising chin rest. Their verbal feedback was recorded by hand for later review and used to resolve any difficulties in the interpretation of their eye movements. After ten cases the observers were allowed to rest for a short while away from the chin rest and then resumed reading cases following further calibration (Fig.5.1).

5.2.4 Review of the eye-tracking data.

The eye-tracked data were played back at quarter speed and overlaid on the image being viewed when that data was recorded. This review was performed for all cases and for all three observers. Any repetitive behaviour with respect to regions of the image visited or features scrutinised was recorded. Particular attention was paid to the sequence in which features or areas were viewed in order to identify apparently significant links between them. Any eye-tracked behaviour that could not be readily interpreted in terms of there being an obvious HRCT feature being scrutinised was recorded and presented to the observer as a question in a form that enabled them to view the image with markers super-imposed relevant to the question (Fig 5.2).

A second viewing of the eye-tracking data was then performed, targeted on specific behaviour patterns identified on the first viewing. The recorded eye movements were recorded in a simple binary fashion returning a yes / no answer to particular questions of behaviour.

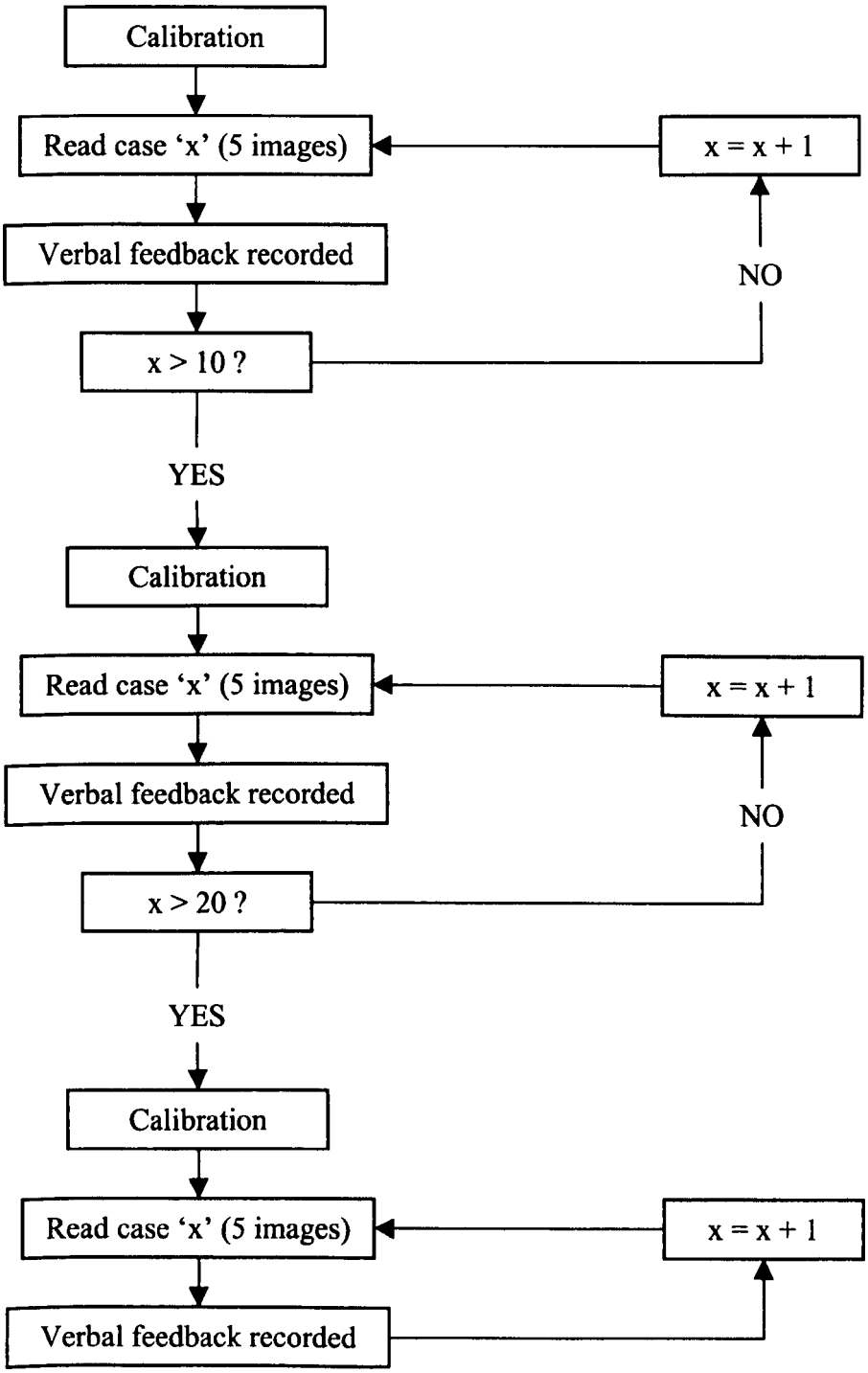


Figure 5.1: Flow diagram demonstrating the sequence of case reading and calibration followed in the experiment. x = case number.

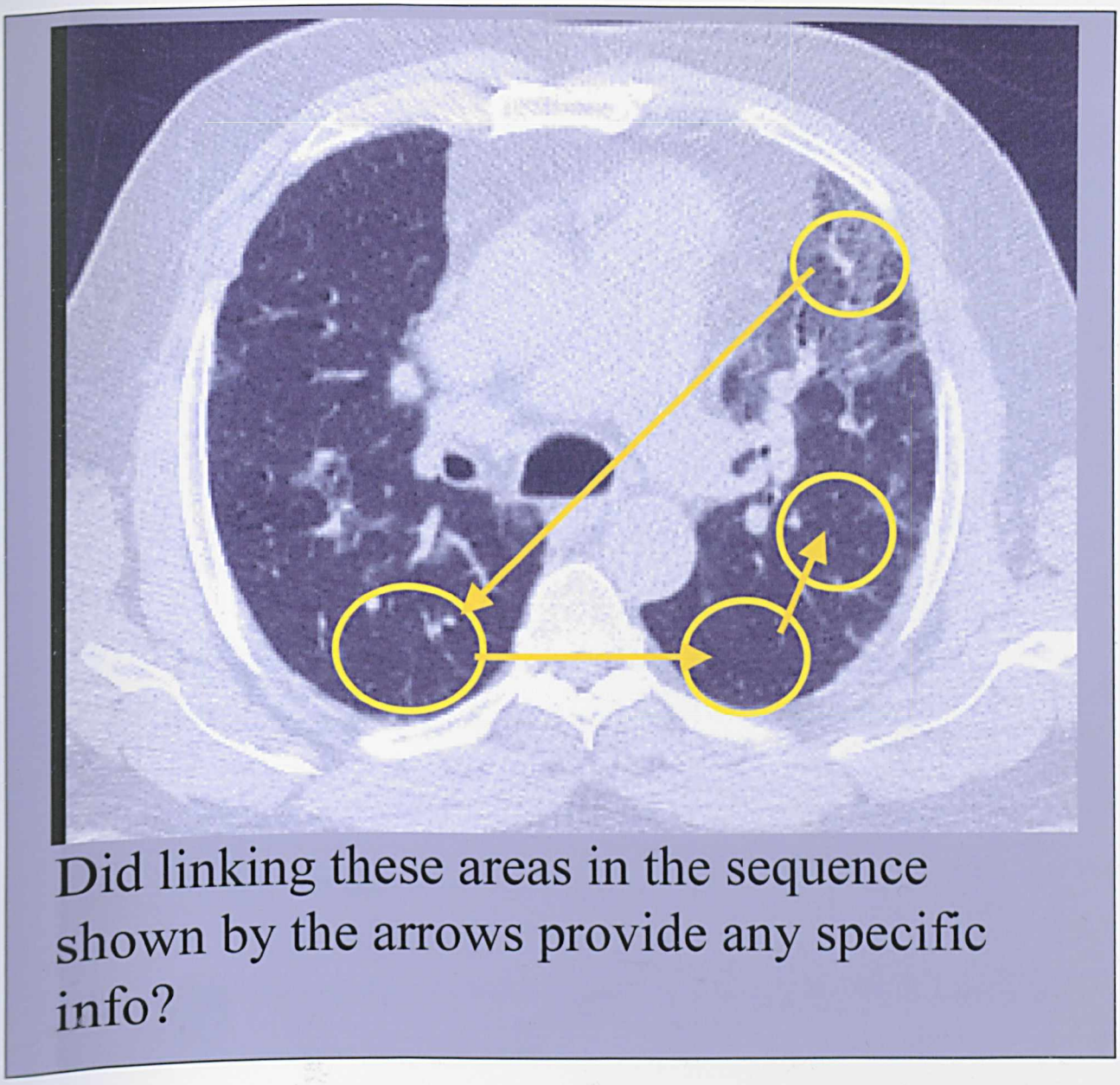


Figure 5.2: An example of the feedback questions that were given to the observers. The feedback was performed via a PowerPoint (Microsoft Corp, USA) presentation and the observers given the option to see the image in question with and without the marking circles and arrows.

5.3 RESULTS

The majority of the eye-tracked movements were readily explicable in terms of knowledge of the various presentations and patterns of common disease entities. There were very few instances of observer behaviour that was not interpretable and therefore fed back to the originating observer. In all such cases the observer was unable to suggest any rationale for the eye-movements that had not already been elucidated.

The initial review of eye-tracking data revealed the following behaviour (which was common to the approaches of all three observers).

5.3.1 Strategy with cases of interstitial lung disease:

- On most images the abnormal higher density lung (ground glass opacity) was viewed first and subsequent comparison with the normal density lung was made.
- Multiple comparisons between ground glass opacity and normal lung were made.
- The bronchovascular bundles within areas of ground glass received particular scrutiny.
- Comparison was consistently made between the normal lung and trachea or major bronchi.
- Identification of ground glass opacity in the most caudal two images led to scrutiny of the lower oesophagus.

- If nodules were identified the fissures and hila were immediately viewed if present on the same image.
- The periphery of the lungs was scrutinised only in the presence of an abnormality but did not appear to be routinely checked.

5.3.2 Strategy with cases of abnormal low attenuation lung:

- On the vast majority of images the abnormal low attenuation lung was viewed first followed by comparison with the normal attenuation lung.
- Multiple comparisons were made between the low attenuation and normal attenuation lung.
- The segmental bronchovascular bundles were subject to enhanced scrutiny by comparison with the approach to images from cases of interstitial lung disease.
- Multiple comparisons were made between the left and right segmental and sub-segmental bronchovascular bundles within the same image.
- The periphery of the lung was scrutinised only in the presence of an abnormality.
- The presence of a 'tree-in-bud' pattern prompted review of the segmental and sub-segmental bronchovascular bundles.

5.3.3 Further analysis

Based on the above findings the following behaviour patterns were analysed by reviewing the eye movements at a reduced speed; positive and negative responses to each question were recorded as one or zero respectively

- Was there abnormal high attenuation lung?
- Was there abnormal low attenuation lung?
- In the presence of abnormal higher attenuation lung (GL) was the abnormal lung scrutinised first?
- In the presence of abnormal lower attenuation lung (BL) was the abnormal lung scrutinised first?
- Was there a peripheral abnormality?
- Was the periphery scrutinised?
- Was a left – right comparison made between the bronchovascular bundles?

BL present	BL viewed first		
	Yes	No	Total
Yes	65	11	76
No	9	215	224
Total	74	226	300

Table 5.2: 2 x 2 contingency table of whether abnormal low attenuation lung was present and whether the low attenuation lung was viewed first (Chi squared test, $p<0.0001$); BL = abnormal low attenuation lung.

GL present	GL viewed first		
	Yes	No	Total
Yes	90	12	102
No	14	184	198
Total	104	196	300

Table 5.3: 2 x 2 contingency table of whether abnormal high attenuation lung was present and whether the high attenuation lung was viewed first (Chi squared test, $p<0.0001$); GL = abnormal high attenuation lung.

PA present	Periphery scrutinised		
	Yes	No	Total
Yes	110	62	172
No	20	108	128
Total	130	170	300

Table 5.4: 2 x 2 contingency table of whether a peripheral abnormality was present and whether the periphery was scrutinised (Chi squared test, $p<0.0001$). PA present = peripheral abnormality present

BL present	Left right comparison made		
	Yes	No	Total
Yes	46	30	76
No	38	186	224
Total	84	216	300

Table 5.5: 2 x 2 contingency table of whether a comparison of the bronchovascular bundles was made between the left and right lung and whether abnormal low attenuation lung was present (BL) (Chi squared test, $p<0.0001$).

5.4 INTERPRETATION OF THE RESULTS.

The striking frequency with which, in the presence of abnormal low or high lung attenuation, the abnormal attenuation lung was scrutinised first, suggests that, in the presence of differing lung attenuation, the decision as to which part of the lung parenchyma is abnormal is made near instantaneously. The subsequent comparison of adjacent areas of different lung density most likely reflected a comparison of the vessels within these areas of lung. The decision as to whether a mosaic pattern is due to abnormal higher attenuation lung or abnormal lower attenuation lung is largely dependent on the number and calibre of the vessels within the supposedly abnormal lung. Abnormal low attenuation lung ("black lung"), whether due to lung destruction, air trapping or vascular occlusion tends to have attenuated vasculature, with vessels of reduced calibre and number when compared to adjacent normal lung parenchyma. By contrast, abnormal higher attenuation lung (ground glass opacity) tends to have similar vessel numbers and calibre when compared to adjacent normal lung (Webb, Müller, & Naidich 2001c).

The failure to routinely scrutinise the lung periphery suggests that the observers screened this area of the lung with peripheral vision and only concentrated on the area if an abnormality was detected. No peripheral lesions were missed by the observers, indicating that peripheral vision was sufficient for this screening task.

The comparison of the left and right segmental and sub-segmental bronchovascular bundles occurred more frequently in the presence of abnormal low attenuation lung. The likely pathological causes for low attenuation lung can affect vessel calibre, airway size

and airway wall thickness (Webb, Müller, & Naidich 2001c). Presumably the comparison between left and right lung facilitated an assessment of whether there was an abnormality (using nature's wondrous symmetry) and enabled the assessment of whether any changes were global or focal. In general, the bronchovascular bundles were more intensively reviewed in the presence of abnormal low attenuation lung ("black lung").

The remaining eye movements, i.e. those not subjected to a quantified assessment, can also be readily explained in terms of the HRCT features found in lung disease. The eye movements within the regions of ground glass centred on the bronchovascular bundles, probably specifically the airways. Ground glass opacity may be caused by an interstitial infiltrate of fluid, cells or fibrosis; fibrosis causes retraction of the lung parenchyma and results in the pulling apart of the airway walls causing dilatation termed traction bronchiectasis (Remy-Jardin et al.1993;Westcott & Cole 1986). The identification of this feature identifies fibrosis as at least a contributor to the ground glass opacity.

The comparisons made between the major airways, particularly the trachea, and the normal lung parenchyma, in the presence of ground glass opacity, most probably reflected an attempt to gauge the level of "greyness" of the lung parenchyma and confirmed that the abnormal lung was that with a higher attenuation, rather than the lower attenuation lung. None of the observers, on questioning, were aware of making this comparison.

In the presence of ground glass opacity in the most caudal two images, all three observers consistently reviewed the oesophagus. This is probably a result of the frequency with which patients with scleroderma, a systemic disease that can cause posterobasal pulmonary fibrosis and which is characterised by oesophageal dilatation, are investigated at our institution (Webb, Müller, & Naidich 2001a). If nodules were identified (present in seven cases out of thirty) the observers paid particular attention to the fissures and then the hila. The motive behind this behaviour lies in the typical HRCT finding of sarcoidosis, namely nodules, particularly along the fissures, and hilar lymphadenopathy (Webb, Müller, & Naidich 2001b).

When confronted with abnormal low attenuation lung the observers compared normal with abnormal lung as they did when confronted with ground glass opacity; they similarly scrutinised the lung periphery only in the presence of an abnormality and, again, appeared to rely on peripheral vision. In addition, the presence of a 'tree-in-bud pattern' prompted review of the segmental and sub-segmental bronchovascular bundles. 'Tree-in-bud' appearance is due to filling and thickening of ectatic bronchioles and is usually associated with cylindrical bronchiectasis (Webb, Müller, & Naidich 2001c), hence the review of the larger airways.

5.5 DISCUSSION

The play-back of the eye-tracked data at a reduced speed enabled interpretation of the sequence of eye-movements and therefore enabled a more realistic interpretation of the search strategies, by comparison with analyses of the summed total data performed in Chapters 3 and 4.

The procedure for reviewing the eye movements reproduced the raw data played back at reduced speed and did not incorporate any attempt to determine whether dwell times were sufficient to represent fixation points. As a result, it is possible that some apparent behaviour patterns were spurious; the dwell time associated with some eye-tracked data point clusters may have been insufficient to allow high level visual analysis, assumed in the interpretation. Debate remains as to how short a dwell time is sufficient for high level perceptive processing (personal communication Dr E Krupinski, University of Arizona, USA) and the raw data technique used in this experiment ensured that behaviour patterns were not missed by the adoption of an arbitrary definition of fixation.

There were few unexpected patterns in the observers' approaches to the presented HRCT examinations. Furthermore, some general strategies dependent on the disease type presented were observed. When confronted with lung parenchyma of differing attenuation on the same image, mosaic attenuation, the observers carried out multiple comparisons between the areas of different attenuation presumably in an attempt to elucidate the reason for the difference. However, the apparent ability of the observers to

instantaneously identify the abnormal lung, suggested they were capable of making the comparison from an immediate global view. A similar "snapshot" diagnostic ability has been demonstrated in the reading of plain chest radiographs (Kundel & Nodine 1975). The differentiation of interstitial lung disease and obstructive lung disease as the cause of a mosaic attenuation lung has been shown to be readily achievable (Worthy, Müller, Hartman, Swensen, Padley, & Hansell 1997) but it is difficult to accept that a snapshot analysis could elucidate the array of ancillary HRCT signs that make this possible (Stern, Swensen, Hartman, & Frank 1995b). The cases of mosaic attenuation varied in subtlety and the most subtle were cases of abnormal low attenuation lung. It is possible from experience that the observers, faced with subtle mosaic attenuation, assumed the lower attenuation lung was abnormal until proved otherwise; in practice, when ground glass opacity is the abnormality it tends to be more obvious. An alternative explanation would be that the standardised window settings, in which the images were displayed, enabled the observers to identify the normal lung based on its greyness, and the observers therefore targeted the abnormal areas whether darker or lighter.

In all cases, the peripheral lung parenchyma (the cortical lung) was not routinely scrutinised and was presumably screened using peripheral vision The use of peripheral vision as a "screening tool" has been previously reported in the context of nodule detection on chest radiographs (Kundel 1975) acting as an adjunct to guiding foveal fixation to inconspicuous nodules (Kundel, Nodine, & Toto 1991). In the verbal feedback none of the observers failed to mention peripheral abnormalities when present, which suggested that peripheral vision was sufficiently accurate for the task of screening the

periphery of the lungs. In normal circumstances there is a high contrast boundary between the subpleural region of the lung parenchyma and the inner chest wall and the para-foveal retina is well adapted to identifying an alteration in such a boundary. The size threshold for the identification, by peripheral vision, of small sub-pleural nodules is beyond the scope of this study as there were no suitable cases viewed for such an analysis.

There were fundamental differences in the approach taken to interstitial as compared to obstructive lung disease in terms of the recorded eye movements. The bronchovascular bundles were scrutinised in both disease types; in interstitial lung disease the bronchovascular bundles in the regions of ground glass opacity were scrutinised in preference to those in the normal lung and the converse was true for cases of obstructive lung disease. Dilated airways within a region of ground glass opacity implicate fibrosis as the cause for the ground glass opacity (Remy-Jardin et al. 1993). In the presence of abnormal low attenuation lung, the segmental and sub-segmental airways may be thick walled and / or dilated indicating a primarily airway cause for the appearances and implicating obstructive lung disease as a cause. The absence of abnormalities in these airways would indicate either disease limited to the small airways, which are not readily seen on HRCT, or an alternative cause such as emphysema or chronic thromboembolic disease. The eye movements recorded were sufficiently characteristic as to identify an interpretative error with respect to SRD on a case of obstructive lung disease. The eye movements recorded were characteristic for reading a case of obstructive lung disease but the verbal feedback was of interstitial lung disease.

Overall the strategies employed for the assimilation of HRCT scans in this experiment appeared to be primarily feature-led and some simple sequential behaviour was seen, such as in alternate observation of abnormal high and low attenuation lung, the comparison of bronchovascular bundles in the left and right lung and the scrutiny of the fissures and hila following detection of nodules. This assertion will be tested in subsequent analyses where an anatomical-based comparison of eye-tracking behaviour, as performed in Chapter 4, will be compared to an analysis based on prior identification of relevant features. Furthermore the strategies employed by the observers have been incorporated into a tutorial (see Appendix) designed to improve the approach of novices to HRCT scans and their performance before and after this form of teaching will be assessed.

5.6 SUMMARY

1. ***Feature-led strategies are used by experienced radiologists when reading HRCT scans.***
2. ***There are differences in the approach to cases in which abnormal high attenuation is the dominant abnormality compared to cases characterised by abnormal low attenuation.***
3. ***Apparently subliminal behaviour was identified in reviewing the attenuation of air in the major airways in order to assess the normal density level of the lung.***
4. ***The behaviour the observers studied lends itself to an analytical technique based on feature identification.***

CHAPTER 6

VISUOSPATIAL ANALYSIS OF NOVICE'S APPROACH TO HRCT IMAGES COMPARED TO EXPERIENCED OBSERVERS.

6.1 INTRODUCTION

The extraction of information from a medical image is the least well explored aspect of the radiological diagnostic process. Eye-tracking experiments in radiological interpretation performed so far have been largely confined to plain radiographs with an emphasis on perception (Kundel, Nodine, & Carmody 1978;Kundel, Nodine, & Krupinski 1989) and search patterns (Hu et al. 1994;Krupinski 1996) in order to identify the causes of missed radiographic signs, but as yet eye-tracking techniques have not been used to identify how radiologists succeed in assimilating images.

There is increasing emphasis on the quality of care provided by health services and the advent of clinical governance in the UK requires the setting of goals in all aspects of patient care to ensure the highest standards are met (Ayers 2001). In radiology the accuracy of image interpretation, the speed with which a radiological examination is undertaken and the communication of the findings to the requesting physician all fall within the remit of clinical governance. As technological improvements in the image acquisition stage of radiology plateau and the communication with clinicians is progressively optimised through picture archiving and communication systems (PACS), so the diagnostic performance of the radiologist will increasingly become the focus of

attention; teaching methods employed in training radiologists and the mechanisms by which standards are achieved and maintained will come under scrutiny (Collins 2001). New methods for evaluating the competence of radiologists' interpretative skills are likely to be used in the future to ensure an acceptable level of performance.

In Chapter 5, the approach of experienced radiologists to a selection of HRCT scans representing obstructive and diffuse lung diseases was distilled into some simple statements summarising the experienced observers behaviour when confronted with specific features. A teaching presentation (Appendix A) was created containing general information about significant HRCT features aimed at improving feature recognition, and descriptions of the expert observers' approach when confronted with these features. In the following investigation, the visuospatial approach of novice radiologists, with little or no HRCT reading experience, was compared to that of the experienced observers. The experiment was then repeated following training with the teaching presentation created in Chapter 5; half of the novices were given only the general information and not the specific information on experienced observer behaviour. The novices' visuospatial approach to the HRCT scans before and after training was compared, and any such changes were evaluated to see whether they represented a shift towards or away from that used by the experienced observers.

6.2 MATERIALS AND METHODS

The experimental set-up and sequence of case presentation was identical to that described in Chapter 5.

The data obtained from the experiment described in Chapter 5 was used as the data for the experienced observers in the following analyses.

6.2.1 Trainee selection

Six radiology registrars, in their first year of their training, were recruited after the experiment and its aims were explained. Three were from the University College Hospital and Middlesex radiology training scheme (UCH) and three were from the King's College Hospital radiology training scheme (King's). The first year of radiology training is directed towards the identification of normal anatomy and variants, attaining knowledge of the physics relevant to the various radiological imaging modalities, acquiring an understanding of radiographic techniques and developing eye-hand co-ordination for the performance of screening procedures. Little emphasis is placed on the interpretation of axial imaging modalities and, as such, the novices recruited for this study had had very little, if any, exposure to HRCT imaging.

6.2.2 Feature recognition

For each of the thirty cases, each comprising five images, a short description of the features the novices had identified on that case was recorded and compared with a list of abnormal features identified before-hand; all of the thirty cases had abnormal features,

there were no normal cases. For each case a score was created based on the number of abnormal features identified (Swensen, Aughenbaugh, Douglas, & Myers 1992) (Table 6.1) and an overall score calculated.

Score	Description
0	No abnormal features identified
1	Less than half of features identified
2	More than half of the features identified
3	All features identified

Table 6.1: Scoring system based on the number of features identified on the HRCT images presented to the novices. The recording of features not present was marked by reducing the score for that case by one but zero remained the minimum score.

6.2.3 Case selection

The cases read by the novices were the same thirty cases as those read by the experienced observers detailed in Chapter 5. Of the thirty cases, five were singled out for detailed visuospatial analysis. Most of the significant feature types were represented in the selected cases, and these abnormalities had the least degree of spatial overlap to maximise the possibility of detecting a difference in approach to cases depicting different disease processes (Table 6.2, Figs 6.1-6.5)).

Case	Diagnosis	Features	Disease type
2	Constrictive obliterative bronchiolitis	Mosaic attenuation	OLD
3	Sarcoid	Nodules, reticulation	ILD
7	Non-specific interstitial pneumonia	Ground glass, reticulation	ILD
9	*No final diagnosis*	Ground glass, reticulation	ILD
19	Small airways disease	Ground glass, nodules, reticulation, bronchiectasis	OLD

Table 6.2: Summary of features found in the cases used for the visuospatial analysis; ILD = interstitial lung disease, OLD = obstructive lung disease

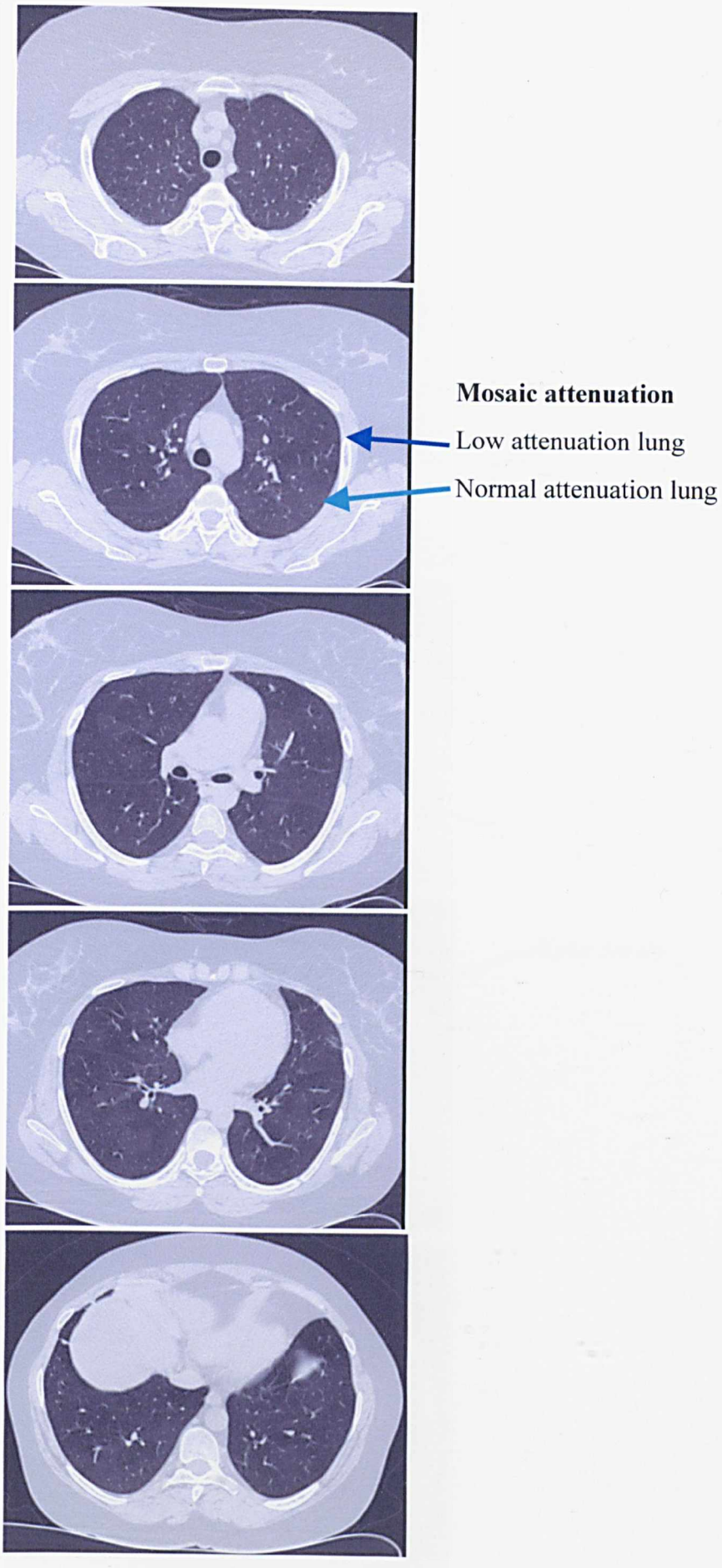


Figure 6.1: The five images comprising Case 2 with regions of low and high attenuation lung marked to demonstrate the mosaic attenuation pattern

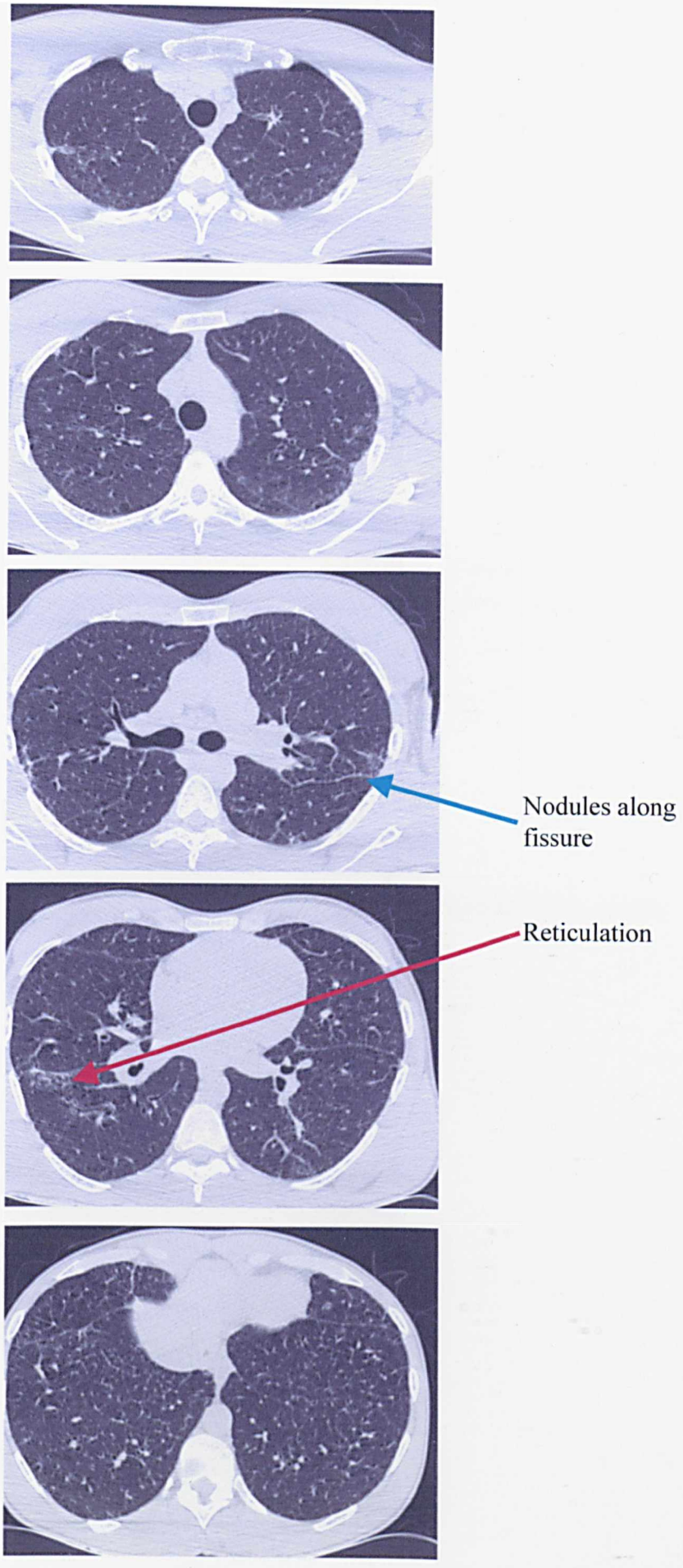


Figure 6.2: The five images comprising Case 3 with nodules and reticulation marked in this case of sarcoidosis.

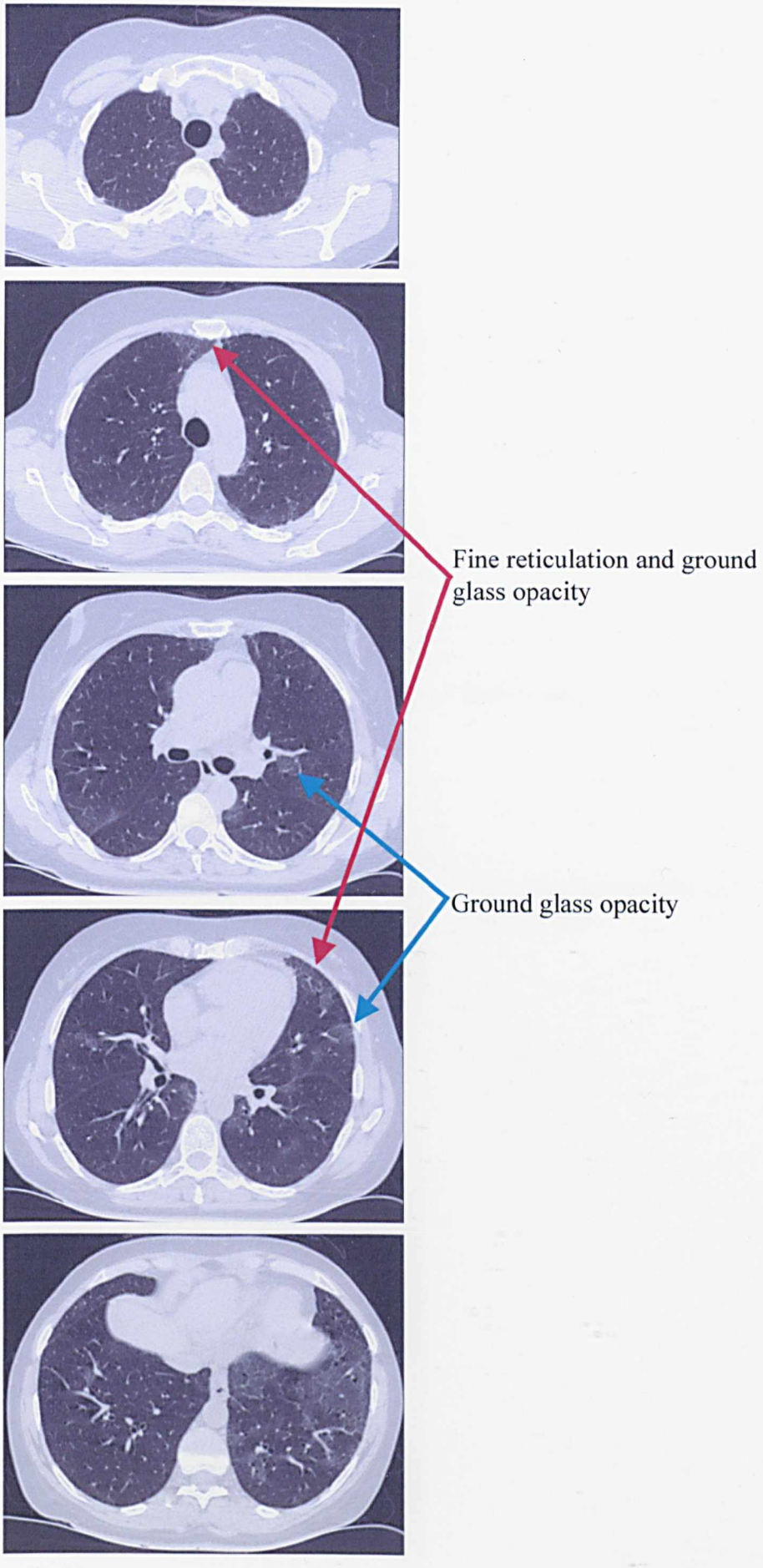


Figure 6.3: The five images comprising Case 7 with ground glass opacity and reticulation marked in this case of non-specific interstitial pneumonia.

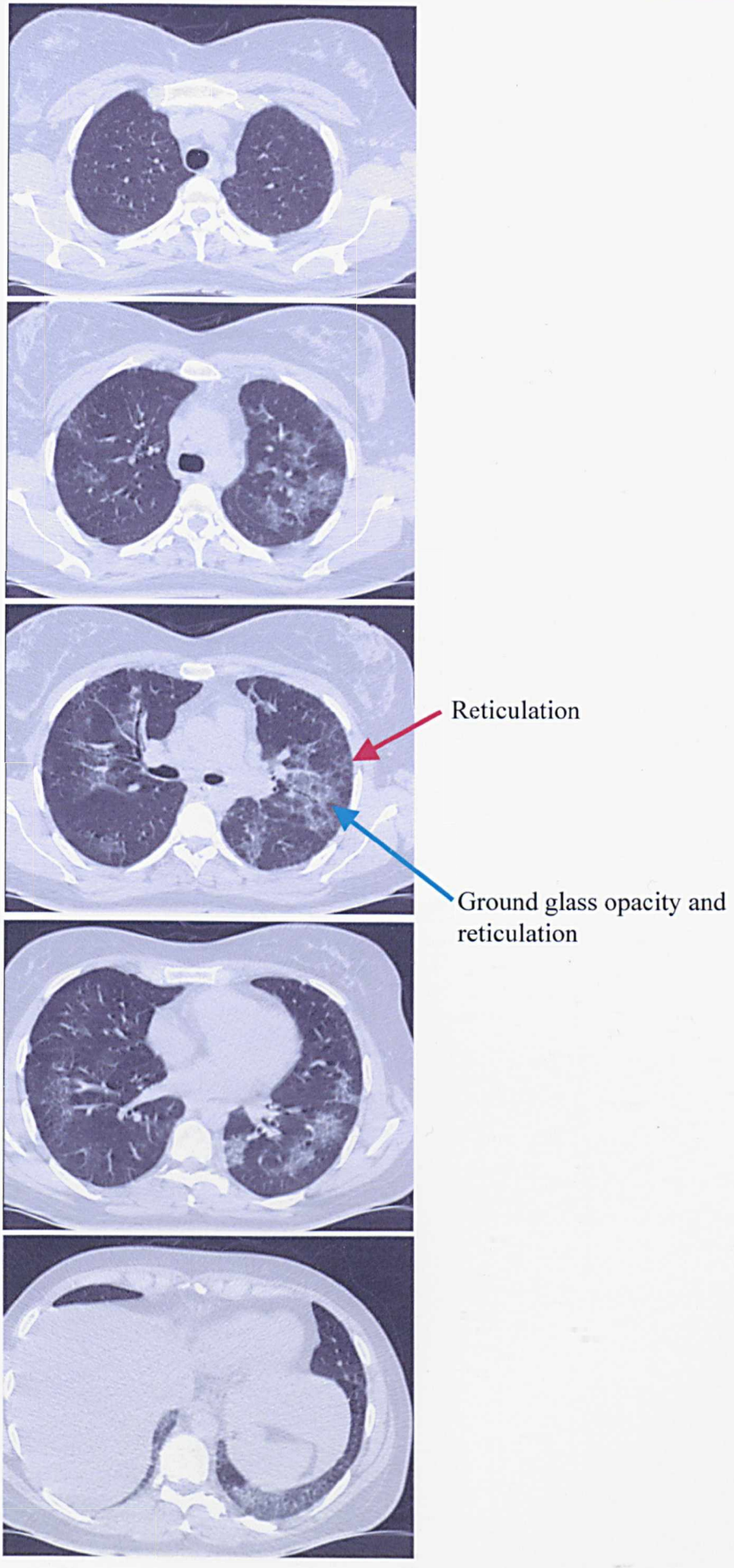


Figure 6.4: The five images comprising Case 9 with ground glass opacity and reticulation marked in this case of non-specific interstitial pneumonia.

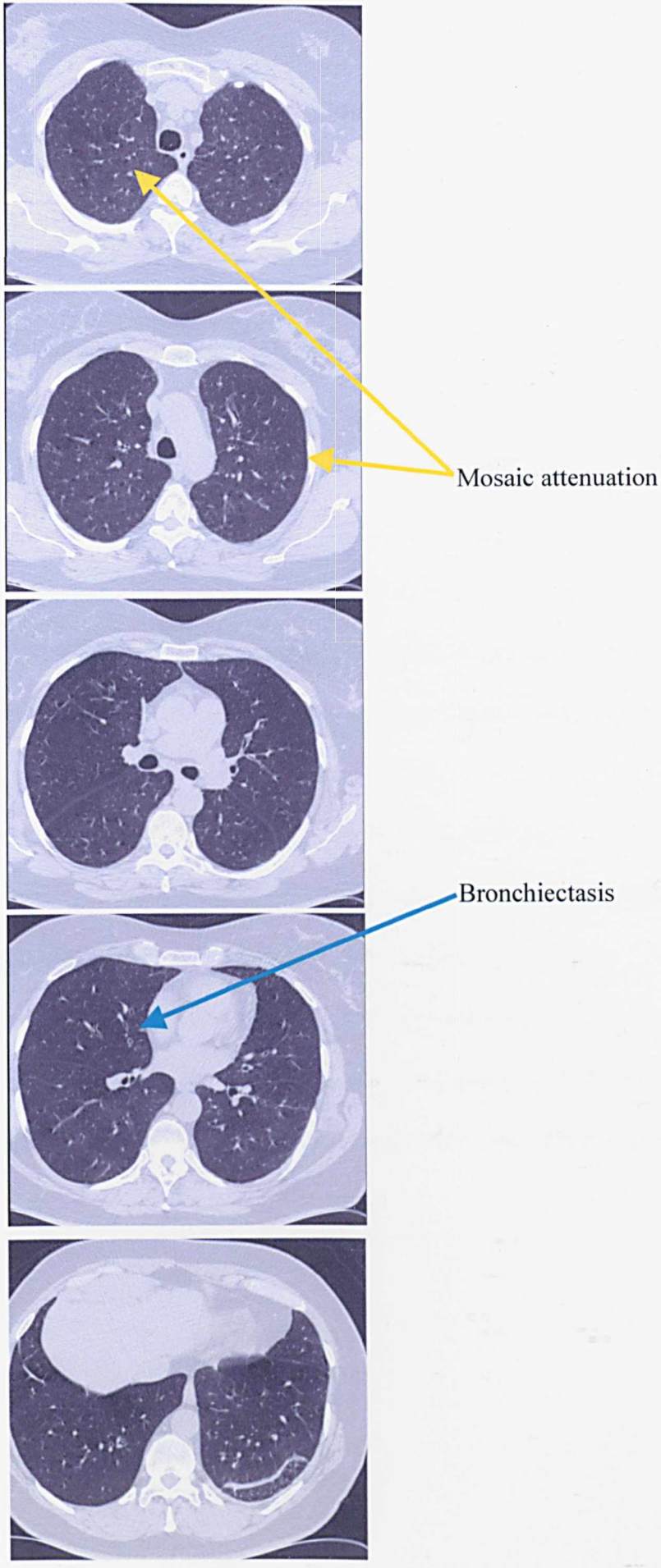


Figure 6.5: The five images comprising Case 19 with ground glass opacity, nodules, bronchiectasis and mosaic attenuation marked.

6.2.4 Training

Following the first session of viewing the HRCT images, the novices were recalled to repeat the reading of the same set of cases. The time between first and second reading ranged from three to eight months (median 4 months) and prior to re-reading the novices were provided with a PowerPoint (Microsoft Ltd, Redmond,USA) presentation (see Appendix A), which they were asked to study at least three times prior to re-reading the HRCT cases. To ensure the teaching presentation was viewed at least once, a hidden character was placed in the top right hand corner of each slide that only became visible if the presentation was viewed as a show. The novices were asked to record the hidden characters as they were revealed and present them prior to re-reading. Each novice had a personalised presentation and a unique set of hidden characters. The teaching presentations provided for the UCH novices lacked the slides containing the information pertaining to the expert approach, whereas the King's novices had this information included in their presentations. There was no communication between the novices in the different departments; both groups of novices had the general information regarding feature recognition on HRCT scans but only the King's novices were privy to the expert approach identified in Chapter 5.

6.3 STATISTICAL ANALYSIS

6.3.1 Visuospatial data

The techniques for visuospatial data analysis employed in Chapter 5 were again used. A distortion grid for mapping each image section onto the normal lung was created. Using the appropriate distortion grid the eye-tracking data points were projected onto the normal lung allowing for a two degree foveal visual angle and plotting time density as a function of a Gaussian distribution across the projected area of foveal vision (Chapter 4). The two dimensional data array of time density values corresponding to each pixel in the image was divided into an eight by eight grid and each grid square was ranked according to the proportion of time spent observing that grid region in relation to the total time spent on the case. The threshold values for determining the ranks were calculated, as in Chapter 4, from a plot of all the proportional data values against the number of grid squares recording that value (Fig. 6.6).

The resulting data files for each case for each observer contained matched-paired data, which was analysed using the kappa statistic. An analysis of variance was performed on the kappa values to determine if there was an independent influence on the resulting kappa value of the case being viewed and the type of inter-observer comparison performed. Three forms of inter-observer comparison were considered (Table 6.3).

Comparison 1	Comparison 2	Comparison 3
Experienced v Experienced	Experienced v Experienced	DMH v NoviceA
Experienced v Novice	Experienced v King's Novice	DMH v NoviceB
Novice v Novice	Experienced v UCH Novice	DMH v NoviceC
	King's Novice v King's Novice	DMH v NoviceD
	UCH Novice v UCH Novice	DMH v NoviceE
	King's Novice v UCH Novice	DMH v NoviceF
		MBR v NoviceA
		MBR v NoviceB
		MBR v NoviceC
		MBR v NoviceD
		MBR v NoviceE
		MBR v NoviceF
		SRD v NoviceA
		SRD v NoviceB
		SRD v NoviceC
		SRD v NoviceD
		SRD v NoviceE
		SRD v NoviceF

Table 6.3: Summary of the three ranges of inter-observer comparisons performed in the analysis of results. Experienced = experienced observers (DMH, MBR and SRD), King's novices were labelled A to C and UCH novices D to F.

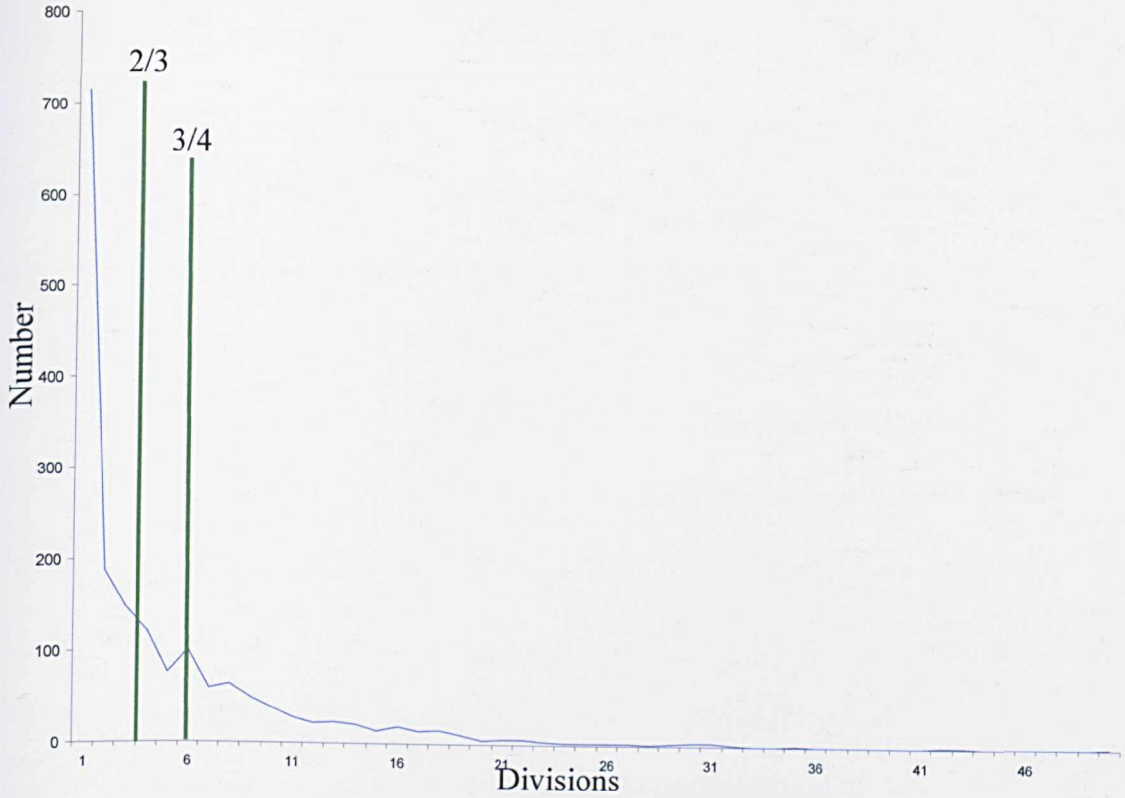


Figure 6.6: Histogram of normalised time density from the eye-tracking data: Threshold points for the calculation of ranks are shown and correspond to points at which the histogram demonstrates a departure a smooth curve, rank 1 was taken as zero, the limits for rank 2 were >0 and ≤0.053, for rank 3 >0.053 and ≤0.105 and for rank 4 >0.105.

The first inter-observer comparison was aimed at broadly determining whether the approach of the experts to the CT scans differed to that of the novices. Inter-observer comparison 2 was designed to identify differences between the two novice groups as their first year of training would have differed slightly and the training presentations made available to the novices differed as described above. Comparison 3 enabled more specific but, due to small numbers, anecdotal comparisons between individual novices and individual experts; comparison 3 was undertaken because the pooling of data may have masked a definite trend with respect to specific novices.

The following four variables, case number, comparison 1, comparison 2 and comparison 3 were all ranked according to the mean of the kappa values related to the specified variable (e.g. for 'case number' the mean of all the thirty-six kappa values for all possible inter-observer comparisons for any single case was calculated). An analysis of variance was performed on the original kappa values with respect to the ranked data.

Observer consistency was evaluated from kappa values generated from the comparison of data from the five cases viewed by that observer (resulting in ten kappa values for each of ten potential comparisons between five cases); observers were assigned as either experienced or novice and the kappa values for the two groups compared by the Wilcoxon's ranksum test to identify any significant difference between observer type in consistency of approach to the cases before or after training. Further analysis compared the data for the UCH novices and that for the King's novices before and after exposure to

the teaching presentation. The data sets for the novice groups had equal numbers of values and a Wilcoxon's signed rank test was used as the comparative test.

6.3.2 Feature identification scoring

For each novice observer the verbal report for each case was scored based on the system outlined in table 6.1. The data derived from such an analysis from the pre and post training reports consisted of paired data (identical cases), which were analysed using the Wilcoxon's signed rank test.

6.4 RESULTS

The data from the pre-training and post-training sessions are tabulated below. As the experts only performed the experiment once, their data is repeated alongside the novice post-training data and the ranking altered as appropriate.

6.4.1 Observer consistency

For each observer ten kappa values were calculated corresponding to each of the ten possible case comparisons (2v3, 2v7, 2v9, 2v19, 3v7, 3v9, 3v19, 7v9, 7v19, 9v19). The pooled data for the experienced observers (thirty kappa values) was not significantly different to that of the novice observers before or after training. However, comparison of the UCH novices and the King's novices demonstrated no significant difference in consistency before training but a significant difference after training with the UCH novices having higher values ($p<0.04$).

6.4.2 Inter-observer comparisons.

The mean kappa values from the five values (corresponding to the five cases) for each inter-observer comparison outlined in Table 6.3 were used to rank those comparisons. Each case was ranked according to the mean of the inter-observer kappa values derived from data related to that case. The resulting mean kappa values and associated ranks for the three inter-observer comparisons outlined in Table 6.2 are summarised in Tables 6.4 – 6.6.

	Pre-training		Post-training	
Inter observer comparison	**Mean kappa**	**Rank**	**Mean kappa**	**Rank**
Experienced v Experienced	.63	1	.63	1
Experienced v Novice	.57	2	.58	2
Novice v Novice	.53	3	.57	3

Table 6.4: Ranks and mean kappa values for the comparisons between the experienced observers and novice groups as a whole.

	Pre-training		Post-training	
Inter-observer comparisons	**Mean kappa**	**Rank**	**Mean kappa**	**Rank**
Experienced v Experienced	.63	1	.63	1
Experienced v King's Novice	.58	2	.56	5
King's Novice v King's Novice	.57	3	.57	4
Experienced v UCH Novice	.56	4	.61	2
UCH Novice v UCH Novice	.54	5	.6	3
King's Novice v UCH Novice	.51	6	.56	5

Table 6.5: Ranks and mean kappa values for the comparisons between the expert and novice groups separated by training centre.

	Pre-training		Post-training	
Inter-observer comparisons	**Mean kappa**	**Rank**	**Mean kappa**	**Rank**
DMH v NoviceC	.67	1	.62	4
DMH v NoviceF	.62	2	.67	1
MBR v NoviceC	.62	2	.63	3
DMH v NoviceA	.62	2	.54	8
MBR v NoviceF	.6	3	.66	2
SRD v NoviceE	.59	4	.54	9
DMH v NoviceB	.59	4	.57	7
SRD v NoviceD	.58	5	.63	3
MBR v NoviceE	.58	5	.57	7
MBR v NoviceB	.58	5	.52	10
SRD v NoviceC	.57	6	.59	6
MBR v NoviceD	.56	7	.61	5
SRD v NoviceF	.55	8	.59	6
SRD v NoviceB	.55	8	.61	5
MBR v NoviceA	.54	9	.45	12
DMH v NoviceE	.52	10	.59	6
SRD v NoviceA	.50	11	.5	11
DMH v NoviceD	.45	12	.59	6

Table 6.6: Ranks and mean kappa values for the comparisons between individual experts and individual novices.

An analysis of variance of the pre-training kappa values with respect to the ranked values of the case number and three inter-observer comparisons revealed a significant independent influence on the kappa value with respect to inter-observer comparison 1 ($p<0.03$), inter-observer comparison 3 ($p<0.05$) and case number ($p<0.0001$) but not inter-observer comparison 2. The equivalent analysis on the post-training data

demonstrated significant results not only with respect to inter-observer comparison 1 ($p<0.02$), inter-observer comparison 3 ($p<0.001$) and case number ($p<0.0001$) but also inter-observer comparison 2 ($p<0.04$). The post-training comparison 1 ranking is less clear cut than that from the pre-training data because of the increase in the mean kappa value for novice/novice comparisons. Using a Wilcoxon's ranksum test on the pre-training and post training kappa values for the comparison 1 data set, the pre-training difference between the kappa values for the experienced versus novice comparisons differed significantly from those for the novice versus novice comparisons ($p<0.05$) but after training no significant difference was found.

Prior to training, no significant difference was found between the two novice groups with respect to comparisons with the experts or in the relative observer consistency (Wilcoxon's signed-rank test). However, following training, significant differences between the UCH novices and King's novices emerged, in terms of consistency ($p<0.04$) and in the closeness of approach to that of the experts ($p<0.005$). Following training, despite the extra information made available to the King's novices, it was the UCH novices that showed both the more consistent approach and a shift in spatial behaviour closer to that of the experts ($p<0.007$).

Case 9, demonstrating ground glass opacity and reticulation, was associated with the highest inter-observer comparison kappa values, regardless of whether the observer was experienced or a novice.

6.4.3 Feature recognition score

The overall scores for each novice before and after training are summarised in table 6.7.

	Novice	Pre-training	Post-training	P value
K I N G S	A	39	45	0.32
	B	33	44	0.16
	C	22	48	0.0015
U	D	43	64	0.0019
C	E	34	58	0.0025
H	F	41	66	0.001

Table 6.7: Summary of the total scores for feature recognition from the novices for each of the two experiments.

There was no significant difference between the two groups of novices with regard to their pre-training scores. Following training there was an overall significant improvement in feature recognition scores ($p<0.0001$). This was reflected in a significant improvement in scores for King's and UCH novices considered as separate groups ($p<0.006$ and $p<0.0001$). However, two out of the six novices, when considered individually, failed to show a significant difference and both of these were King's novices resulting in a significantly greater post-training score for the UCH novices when compared to the King's novices ($p<0.002$).

6.5 INTERPRETATION OF THE RESULTS

Calculations were performed to analyse the consistency in the approach of experienced observers and novices to HRCT scans demonstrating a variety of feature characteristics and to compare the approach taken by specific observers whether experienced or novice.

Measurements of consistency based on the calculation of a kappa value for the two sets of ranked data derived from two different cases read by a single observer indicate that there is some variability in the approach to different cases, which is to some extent dependent on the observer but does not appear to relate to observer experience. Mean kappa values for consistency of approach by each observer ranged from 0.42 to 0.58 (median 0.48) prior to training and 0.42 to 0.53 (median 0.47) after training. The experienced observers were not singled out by this calculation with kappa values of 0.46, 0.48 and 0.53.

Multiple comparisons between observers and between sittings for the novices were made. The kappa values calculated from data for the same case viewed by two different observers indicates the difference in the approach of those two observers to that case. As expected, the experienced observers' approach to HRCT scans was in general in better agreement compared with that of the novices (Figure 6.7, 6.8 and 6.9), indicating some value in the analysis of approach through visuospatial data. The only global change in the inter-observer comparison following training was an increase in the kappa values for the inter-novice comparisons. Overall, there was no evidence of an alteration in novice approach toward that of the experienced observers. However, dividing the novices into two groups, dependent on the content of the training presentation provided, identified a

significant shift in approach of the UCH novices toward that of the experienced observers (figure 6.10) and no such shift in the King's novices' approach.

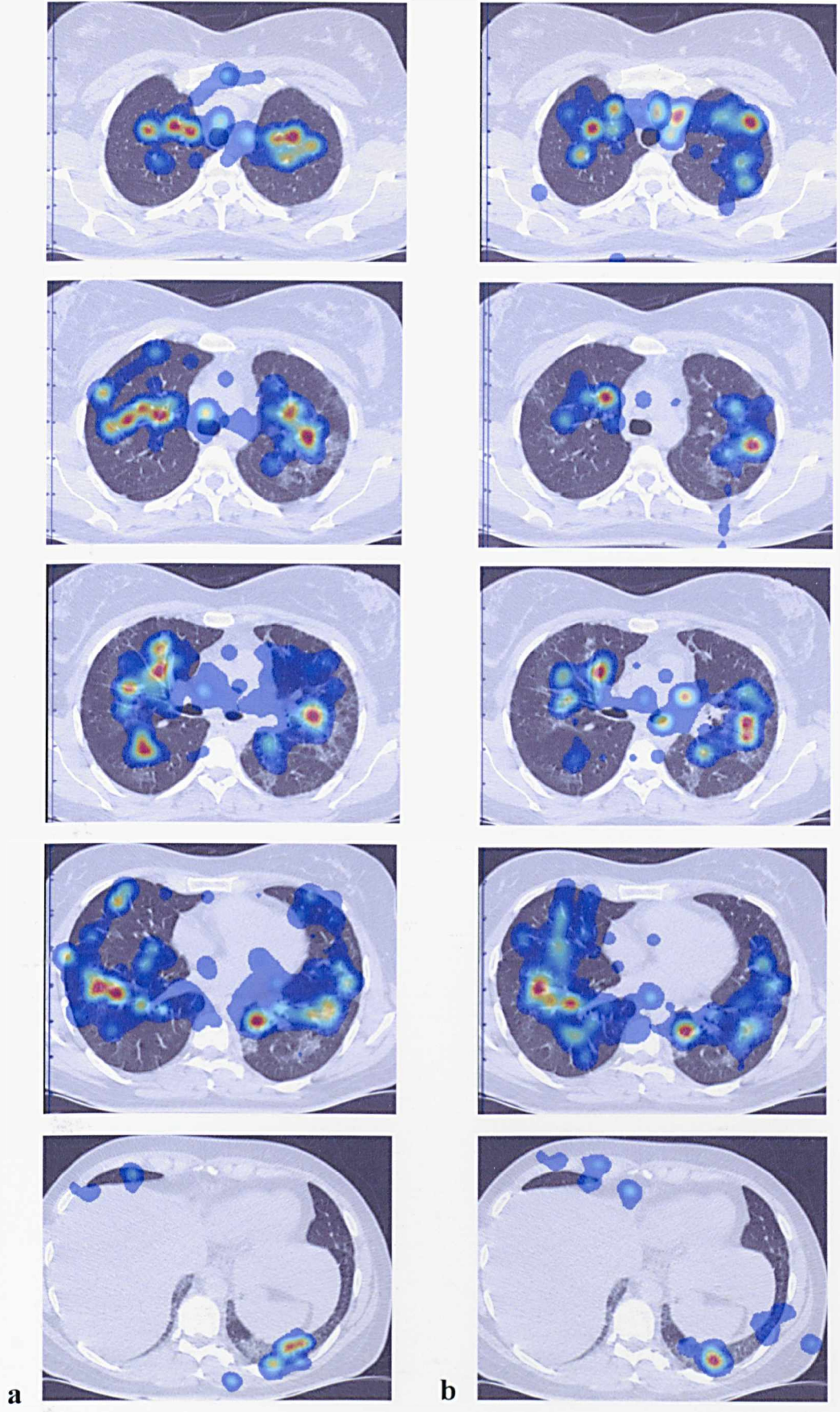


Figure 6.7: Visual projection of the native eye-tracking data for case 7 from the two experts with the greatest level of agreement (kappa = 0.79). **a)** the original images with eye-tracking density data for DMH overlaid and **b)** the original image with eye-tracking density data for MBR overlaid Density ranges from blue (least) through green and yellow to red (greatest).

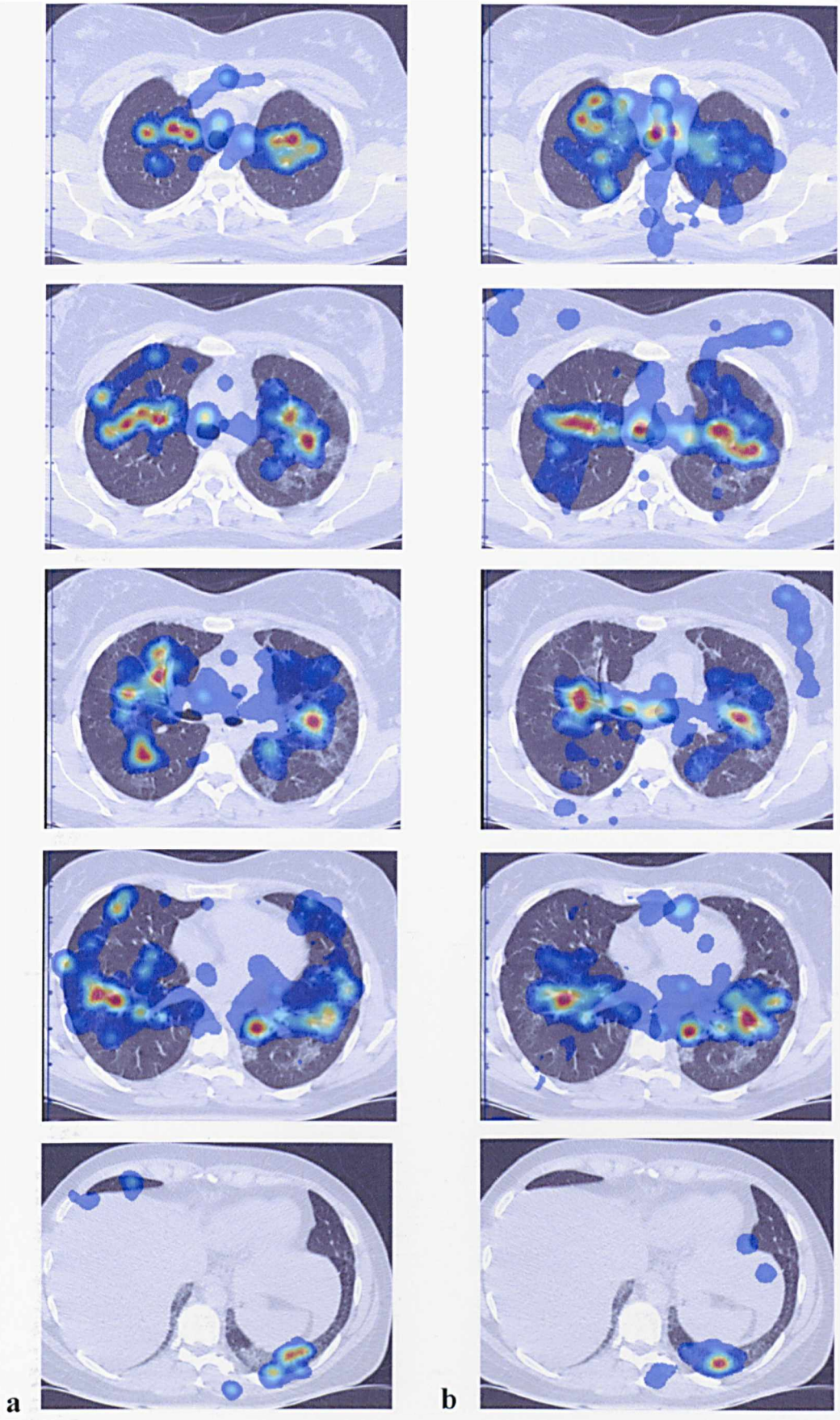


Figure 6.8: Visual projection of the native eye-tracking data for case 7 from the expert and novice with the greatest level of agreement (kappa = 0.80). **a)** the original images with eye-tracking density data for DMH overlaid and **b)** the original image with eye-tracking density data for Novice B overlaid Density ranges from blue (least) through green and yellow to red (greatest).

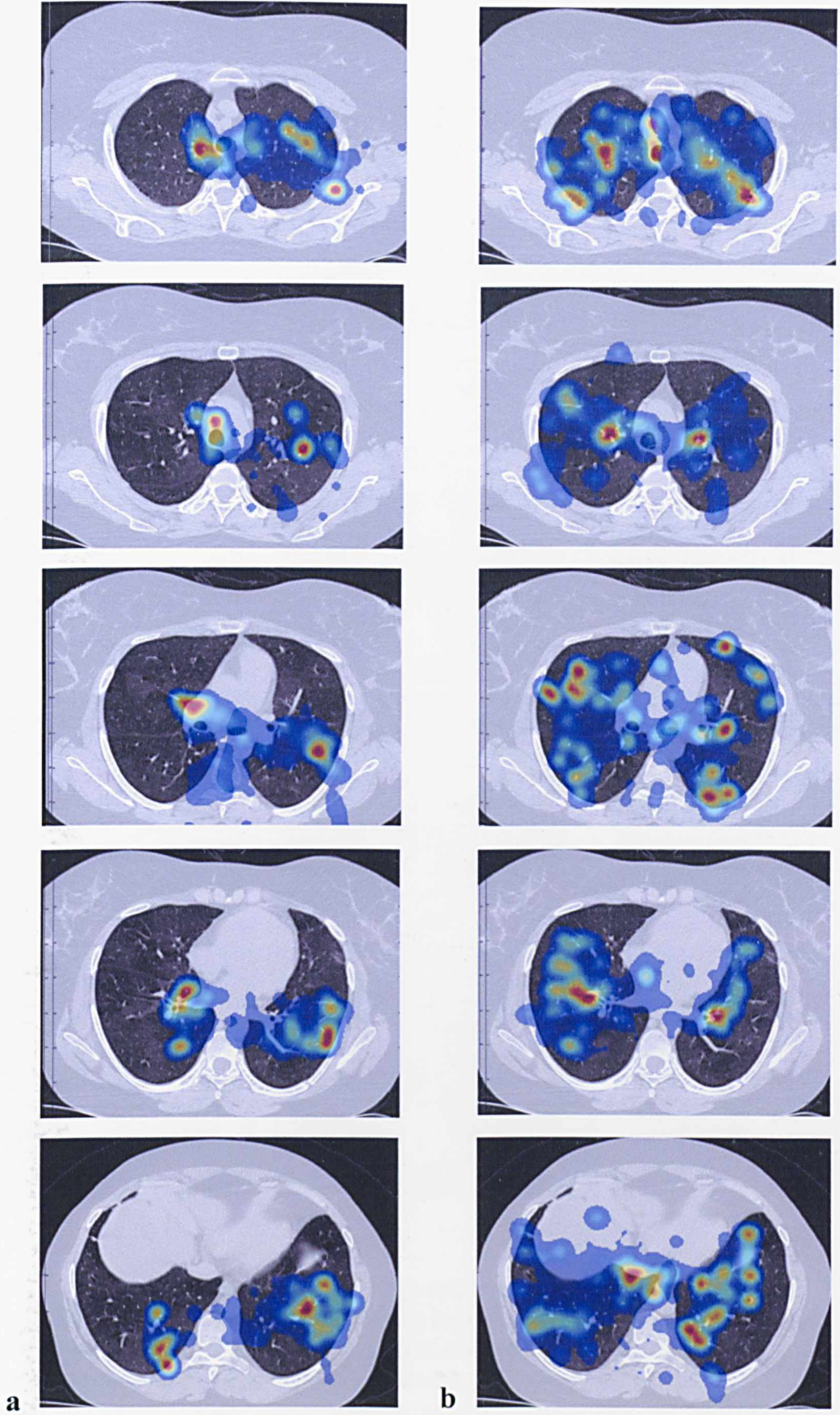


Figure 6.9: Visual projection of the native eye-tracking data for case 2 from the two novices with the least level of agreement (kappa = 0.12). **a)** the original images with eye-tracking density data for Novice A overlaid and **b)** the original image with eye-tracking density data for Novice D overlaid Density ranges from blue (least) through green and yellow to red (greatest).

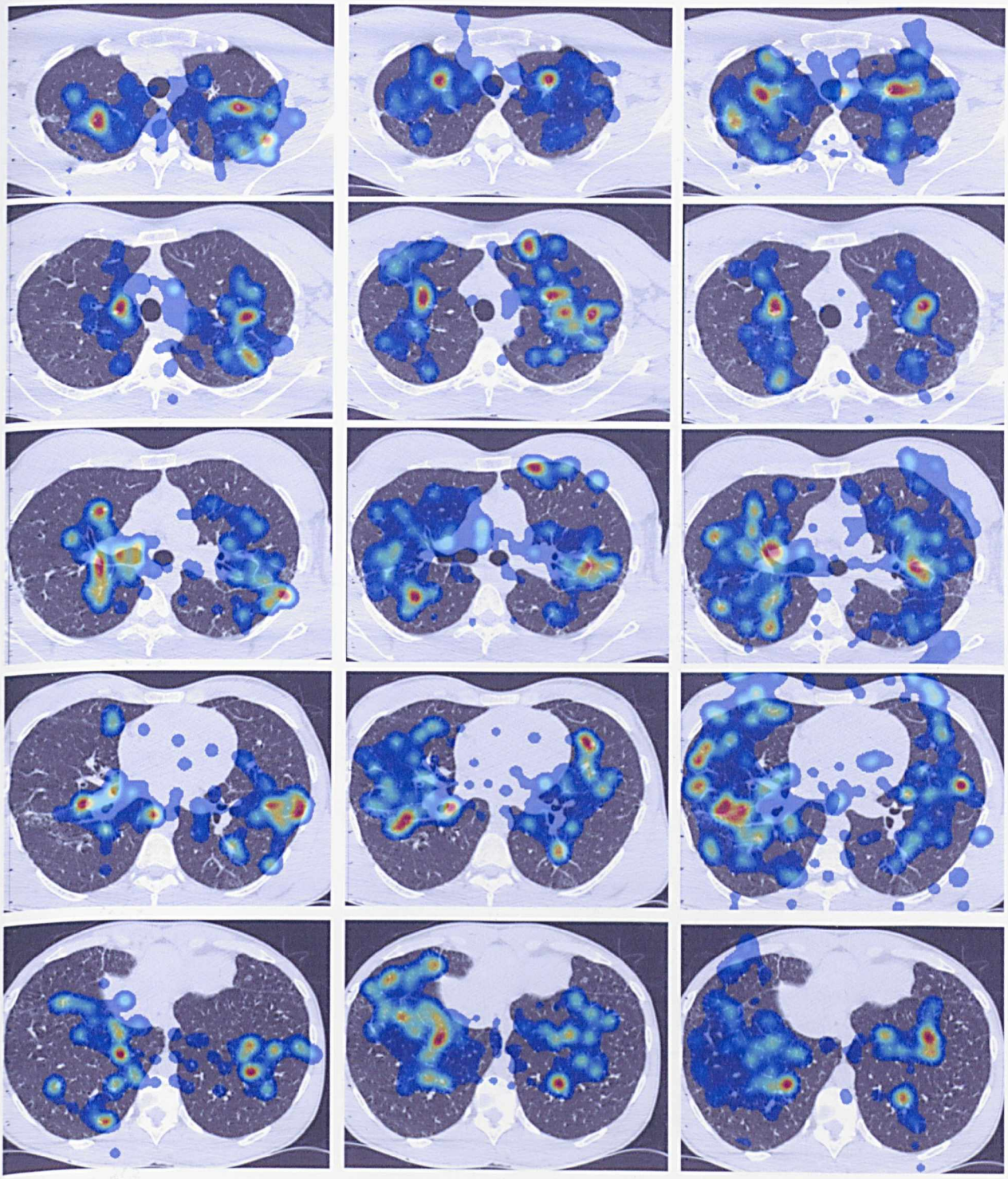


Figure 6.10: Visual projection of the native eye-tracking data for case 2 from the novice who showed the greatest elevation in kappa value between readings (kappa from 0.43 to 0.67) and from the experienced observer with whom this change occurred. **a)** the original images with eye-tracking density data for Novice D overlaid and **b)** the original image with eye-tracking density data for DMH overlaid and **c)** the original images with eye-tracking density data for Novice D post-training overlaid. Density ranges from blue (least) through green and yellow to red (greatest).

When considering comparisons between specific individuals the statistical significance of observed similarities is diminished by the small number figures involved. However, some pertinent observations are possible. The kappa values for similarity of approach between the experienced observers and the novices, both before and after training, tended to be greater when DMH was the experienced observer followed by MBR. Novice C and novice F had the greatest kappa values both before and after training but the overall shift in ranking order, based on kappa values from comparisons with the experienced observers, reflected a tendency for the UCH novices to move up the ranking compared to the King's novices, commensurate with the significant increase in kappa values recorded for the UCH novices following training.

The kappa values from this analysis ranged from 0.12 to 0.82, but the majority of the figures were clustered between 0.4 and 0.65. In some instances, ranking order was determined by as little as a 0.01 difference in kappa value. The ANOVA calculation assumes that adjacent ranks are separated by uniform differences in value and any non-uniformity could generate spurious results. It is therefore possible to over-interpret the results as presented, nevertheless the observed trends are likely to be real. The higher kappa values associated with case 9 (Fig.6.4) also identifies a potential flaw in the analysis. The most caudal image from the case includes little lung parenchyma due to bilaterally raised hemidiaphragms. Therefore, observers viewing this image, whether experienced or novice, are likely to concentrate their attention on the small section of visible lung parenchyma (in the costophrenic recesses), which as a consequence of the masking process would be the only part of the image represented in the ranked data set;

this would result in a close match of data for that image, potentially affecting the overall match for the case. In addition, this most caudal image was unlikely to be viewed for any length of time and most of the reading would concentrate on the remaining four images. An almost twenty percent reduction in the area over which the eyes are being tracked would increase the likelihood of matching time density even if randomly distributed.

The feature recognition results demonstrated an overall significant improvement in score largely due to the improvement in the scores for the UCH novices. Prior to training there was no difference between the scores for the two groups of novices but after training the UCH novices had significantly higher scores.

6.6 DISCUSSION

The overall aim of this experiment was to asses the approach of observers to HRCT scans and to determine if there was a recordable approach for experienced observers that could be used to shift the novices approach, following specific training, towards the behaviour of the experienced observers. The analytical process assumed that the approach of an observer to an HRCT scan could be recorded in term of visuospatial data, identifying - in two-dimensional space - where the observer was concentrating their gaze whilst reading an HRCT examination. Initial results demonstrated an expected agreement of approach to the presented cases between the experienced observers, which was greater than that between the novices or that between the novices and the experienced observers; this was to be expected as a novices approach with little exposure to CT imaging was more likely

to be disorganised. However, there was no stratification as a result of experience levels when the approach to different cases by the same observer were considered. Therefore, it is not a consistency of approach to different HRCT scans in general that sets the experienced observers apart from the novices. A significant improvement in consistency after training was found among the UCH but not the King's novices. As consistency of approach does not reflect experienced observer behaviour, the improvement, rather than indicating behaviour closer to that of the experienced observers, indicated a more systematic approach, perhaps relate to increased knowledge levels post-training or further experience of axial imaging outside the confines of the experiment.

The results from the second reading showed significant increases in the kappa values with respect to comparisons made with and among the UCH novices but no improvement in the results for the King's novices. The core of the training presentation was the same for both groups but more information was presented to the King's novices than to the UCH novices. The extra information did not translate into a greater improvement and the overall response of the King's novices to training was poorer than that found among the UCH novices. If the results were entirely ascribed to the influence of the training presentation, one would conclude that providing specific information with regard to how the experienced observers approached the cases unexpectedly *reduced* the efficacy of the training presentation, and thus the novices approach to the HRCT scans; it could be postulated that the information summarising the experienced observers eye movements confused the overall message with regard to what features should be assimilated when reading an HRCT scan. It is possible that there was a fundamental difference in the

responsiveness to training between the groups. The significantly greater improvement in feature recognition scores for the UCH novices when compared to the King's novices does suggest a greater benefit derived by the UCH novices from the feature description part of the training presentation; the feature recognition element of the training presentation was identical for both sets of novices. A further explanation for the findings may relate to differences in the novices' normal radiology training, which continued between the readings of the experiment cases. The UCH novices all had attachments to CT departments in this time (not foreseen at the time of selecting novice subjects) providing them with a greater level of exposure to CT scanning than the King's novices. Furthermore, the UCH novices had a longer time interval between readings. The minimum elapsed time was three months eliminating the possibility of observers recalling the cases on second viewing, but the King's novices were all re-tested within four and a half months, whereas the UCH novices were re-tested after four, five and nine months giving them greater opportunity to improve through there general training. It is possible that four to eight months training was more beneficial than two to three sessions with a training presentation. All the novices, except novice A, viewed the training presentation at least twice, novice A viewed it only once and only novice C viewed the training presentation more than three times and this was reflected in the improvement in the scores with respect to feature identification.

The above analysis was based on visuospatial data and therefore relied on consistency in the area of an image from an HRCT scan that is viewed by the observer. In diffuse lung disease there may be multiple regions in which a relevant feature could be scrutinised and

a visuospatial analysis would rely on a specific one catching the attention of the observer or the systematic scrutiny of all the instances of that feature. Agreement in visuospatial data would be enhanced by the presence of a prominent single feature or the focusing of attention on anatomical structures, which occupy a consistent spatial point. Conversely, the possibly spurious high kappa values associated with case 9 reflected a further limitation in this type of spatial data analysis, and highlights the effect of a reduction in area being viewed on the likelihood of viewing the same areas. Nevertheless, the ranking process discriminates the relative importance attributed to any specific area such that comparison of observers approach to case 9 was still of value in the analysis of observer behaviour.

The use of visuospatial data has enabled the distinction between the approaches of experienced and novice observers and has revealed alterations in novice behaviour between the two readings. The King's novices showed little improvement in either eye-tracked data, when compared to the experienced observers, or feature recognition. The UCH novices showed improvement in both, suggesting they derived greater benefit from the training presentation they were given. As the training presentation supplied to the King's novices was more complete than that given to the UCH novices, the disparity between their relative performance requires explanation and has been discussed. The study has not demonstrated a shift of novice approach toward that of the experienced observers as a result of informing the novices about the experienced observers approach. It is possible that the King's novices were less tractable or "trainable" than their UCH counterparts and as a result were unable to benefit from the extra information that was

provided. It is more likely that the differences in CT exposure between the two groups was the overriding factor, but the poor improvement in the King's novice performance suggests that the training presentation provided little benefit. It is possible that the type of analysis of the visuospatial data used in this experiment failed to identify feature-led eye-movements (indeed the extra information given to the King's novices was in the form of feature-related behaviour). An analytical technique capable of teasing out feature-led behaviour is required to determine whether any benefit was derived from the inclusion of information regarding the behaviour of the experienced observers. Furthermore, a lack of confidence may cause the novice observers to re-check areas that they initially considered normal, areas that the experts would be confident enough not to revisit; it is possible that for a portion of the time taken to view the case the novices' approach was closer to that of the experienced observers than at the end of the viewing when the unnecessary re-visiting of areas would have a confounding influence. As a result, the final eye-tracked density map may obscure an underlying pattern that closer reflects the appropriate importance of particular areas. The only way of teasing out this time-related data would be with serial density maps but this was not possible with the data sets acquired for this experiment.

Although there are limitations to using a visuospatial analysis for interpreting observer behaviour when reading HRCTs, the technique has some value and has demonstrated some significant differences in observer approaches and, significantly, a greater overall agreement between the experienced observers than the novices. This indicates that there is a definable experienced approach, which is, to some degree, recorded by this form of analysis.

6.7 SUMMARY

1. ***Experienced observers approach the same HRCT case in similar manner***
2. ***Observers, experienced or novice, vary in the consistency with which they view HRCT scans and increased consistency does not necessarily reflect more experienced behaviour.***
3. ***One group of novices demonstrated improvement in feature recognition and developed behaviour that was closer to that of the experienced observers when compared to the other group of novices.***
4. ***The other group of novices appear to have derived little benefit from the training presentation (i.e. from the more specific information regarding the experienced observers' approach).***
5. ***Differences in training experience between the two groups of novices may explain a large part of the observed differences.***
6. ***Consideration of spatial data alone may fail to identify more subtle feature-led behavioural changes.***

CHAPTER 7

ANALYSIS OF EXPERIENCED AND NOVICE OBSERVERS' FEATURE-LED ASSIMILATION OF HRCT.

7.1 INTRODUCTION

In the previous chapter a visuospatial analysis of eye-tracking data was performed using a relatively coarse anatomical grid. Some consistency in approach to HRCT scans was found between experienced and novice observers but, as discussed in chapter 6, there were two potentially confounding factors: the feature-led component of the observers' approach to the cases and the effect of confidence on the final data for a case where a more expert approach may have been masked by the re-visiting of areas that was not strictly necessary, but which would affect the overall grid square ranking. These issues are addressed in the following study that uses a technique designed to detect feature-led reading behaviour at multiple time intervals during the reading of a case. Previous eye-tracking experiments based on the recording of eye-movements to predefined regions of an image have largely been used to demonstrate deficiencies in search techniques, i.e. failure to scrutinise specific regions of a radiograph (Berbaum et al. 1996). However, the recording of eye-movements to regions on a CT image that correspond to specific anatomic and pathologic features enables the relative attention paid to those features to be quantified; such data expresses how a specific observer approaches a CT and would be affected by the relative significance, and ease of assimilation, of the features marked.

7.2 MATERIALS AND METHODS

The eye-tracking data used for the analysis in Chapter 6 was used again for this analysis. The eye movements of three experienced observers and six novices (three from each of two radiology training schemes) viewing thirty HRCT examinations, comprising five images each, were recorded.

7.2.1 Feature mapping

The five cases singled out in Chapter 6 for detailed analysis were again used. For each image of each case, the regions containing specified features were outlined by hand by a radiologist (SME), not one of the observers, in a custom designed computer program and each region labelled with a unique identifier, which was cross-referenced to a feature type. Both normal anatomical features and abnormal features were outlined and these are listed in Table 7.1. The selection of normal anatomical features was based upon the visual analysis performed in Chapter 5, i.e. those that appeared to be systematically scrutinised by the experienced observers.

Feature number	Feature type	Cases containing feature
1	Nodules	3,19
2	Abnormal low attenuation lung	2,19
3	Normal lung	2,7,9,19
4	Ground glass opacity	2,7,9
5	Reticulation	3,7,9,19
6	Bronchovascular bundles	2,3,7,9,19
7	Trachea and main bronchi	2,3,7,9,19
8	Major fissures	2,3,7,9,19
9	Atelectasis	19
10	Mediastinum	2,3,7,9,19
11	Other areas encompassed by chest wall	2,3,7,9,19

Table 7.1: Summary of the features marked for analysis and the cases in which the respective features were found.

7.2.2 Data analysis

7.2.2.1 Time density calculation

As in Chapters 4 and 6, a circular area represented the projection of each eye-tracked data point and this corresponded to a foveal visual angle of 2°. Each eye-tracked data point represented a 20 millisecond duration. The assignment of time density to the pixels found within that area was based on a Gaussian distribution with the maximum value at the centre (see Chapter 4). For each eye-tracked data point the visual duration value of all

pixels falling within the foveal visual area was incremented dependent on the distance from the centre of the projected foveal area. The overall visual duration for each feature was calculated by summing the individual visual duration values for each pixel within the region assigned to each feature and can be expressed as:

$$vd(f) = \sum_{i=1}^{i=n} vd(f_i)$$

Where $vd(f)$ is the overall visual duration of feature f, $vd(f_i)$ is the visual duration value of a pixel assigned to feature f and n is the total number of pixels assigned to feature f.

In some instances a given pixel may have been assigned to more than one feature (i.e. overlap of defined feature regions) or the foveal projection area may have covered more than one feature; in such cases, the plotting of a single eye-tracking data point increased the visual duration value of more than one feature. Any eye-tracked data point that, when projected onto the image, failed to overlap any mapped features, was assigned to normal lung, a catch-all anatomical feature definition (Fig. 7.1).

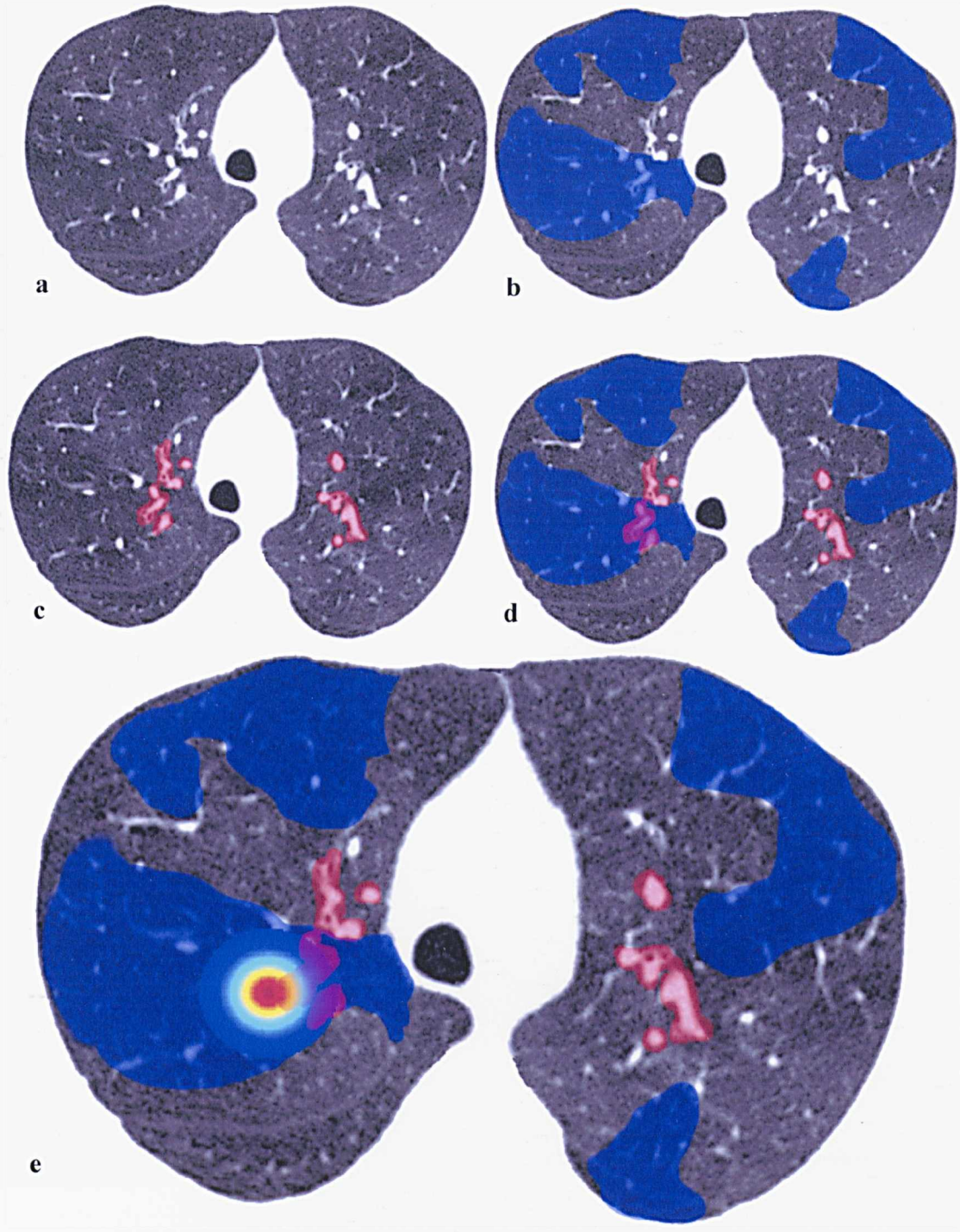


Figure 7.1: Sample case of subtle mosaic attenuation in a subject with small airways disease. **a)** the total region within which features were defined is displayed, note the mediastinum is included but the chest wall is excluded. **b)** regions of abnormal low attenuation lung are overlaid with blue, **c)** the bronchovascular bundles are marked in red and **d)** both features are combined on the one image.

In **e)** an eye-tracked data point is super-imposed on the images and feature maps to demonstrate the the recording of visual duration on two features from only one data point and the impact of overall feature area on the resulting total feature duration.

7.2.2.2 Normalisation

The total duration over which a case was read varied between observers, thus affecting the visual duration recorded for the features in a case. For comparison between observers to be feasible, the visual duration values for each feature were normalised with respect to the total visual duration for all the features.

The visual duration for a specific feature *f* is defined in equation 1.

$$td(f) = \frac{\sum_{i=1}^{i=n} vd(f_i)}{T} \quad \text{.....(2)}$$

The time density (*td*) for feature *f* is therefore defined as

Substituting in equation 2 from equation 1:

$$td(f) = \frac{vd(f)}{T} \quad \text{.....(3)}$$

Where *td(f)* is overall time density for feature *f* and *T* is the total time taken to view the case.

The visual duration recorded for each feature was a function of how close to the feature region the eye was tracked and to the overall area of overlapping between the feature region and the area of foveal vision. In the example shown in Figure 7.1 the eye-tracked data point could be directed at abnormal low attenuation lung, the bronchovascular

bundles or both. The entire area of the projected data point overlaps with the region defined for abnormal low attenuation lung and this feature will therefore record a maximum visual duration for that data point. The bronchovascular bundle feature is of a smaller area and therefore has less overlap with the projected data point and thus records a smaller visual duration from that point compared to the abnormal low attenuation lung, which does not accurately reflect the significance of the potential target of that data point. As a result, the visual duration, and ultimately the time density, calculated for a feature that occupies a large area of the CT image would be erroneously greater than that for a feature occupying only a small area but actually subject to greater visual scrutiny. The time density data for each feature was therefore normalised with respect to the area assigned to that feature such that the normalised time density (*Ntd*) can be expressed as:

$$Ntd(f) = \frac{td(f)}{n}$$

Where n as the number of pixels assigned to feature f represents the total image area of feature f and *td(f)* is defined in equation 3.

7.2.2.3 Sequential time density calculation

The calculation of normalised time density values for each feature was expressed as a histogram of normalised time density values against feature. A further data subset was calculated whereby time density/feature histograms were generated at 5 equitemporal time intervals during the reading of a case. The histograms represented cumulative data, recording all the time density values from the beginning of the reading up to the relevant point.

7.3 STATISTICAL ANALYSIS

The Kullback-Leibler distance in the form of the Kullback-Leibler information criterion (KLIC) performs a comparison of two data sets, that cannot be assumed to follow a parametric distribution, which includes the degree of difference between paired normalised values in the calculation. The result is a single value that summarises the overall distances between the values in the two data sets, a lower number corresponds to a closer match. The KLIC is explained in Chapter 2 and defined by the equation

$$KLIC(g:f) = \sum_{y=0}^{n} g(y) \times \frac{Ln[g(y)]}{Ln[f(y)]} \quad \text{.....(5)}$$

where *KLIC(g:f)* is the Kullback-Leibler information criterion for a comparison between the two distributions *g(y)* and *f(y)*.

Comparison of the KLIC values from paired observer data from each case, provides an indication of the degree of similarity between the distributions of time density of the features present, recorded from the observers eye movements.

A Spearman ranked correlation test was also performed on the paired time density/feature data sets derived from each observer for each case. This is a non-parametric test of correlation that is less powerful than the KLIC but assigns a probability to the likelihood that two data sets are independent. If the approach to a case by two observers is the same, their feature mapped time density histograms should not be independent of each other.

However, the data is ranked in this analysis such that, although the relative differences are retained, the absolute values are lost; when data values are very similar their relative ranking is somewhat arbitrary and can affect the overall result.

7.4 RESULTS

7.4.1 Expert agreement

With respect to the KLIC values there was very good agreement between the experienced observers data for case 9 and reasonable agreement in cases 2 and 3 . However, the Spearman rank correlation was only significant for all three possible comparisons between the experienced observers in case 9 (Table 7.2). The significant Spearman rank p values correspond to the lower KLIC values, as one would expect but the comparison between DMH and SRD in case 3 with a relatively low KLIC value has a p value associated with the Spearman rank correlation that fails to reach significance and probably reflects the poorer power of a rank-based correlation in these circumstances.

Case	Comparison	Spearman correlation	KLIC
2	**DMH / MBR**	p = 0.21	26
	DMH / SRD	p = 0.29	33
	MBR / SRD	p = **0.004**	**16**
3	**DMH / MBR**	p = 0.43	35
	DMH / SRD	p = 0.22	12
	MBR / SRD	p = 0.94	37
7	**DMH / MBR**	p = **0.0003**	**5**
	DMH / SRD	p = 0.35	72
	MBR / SRD	p = 0.29	50
9	**DMH / MBR**	p = **0.002**	**3**
	DMH / SRD	p = **0.001**	**13**
	MBR / SRD	p = **0.0003**	**8**
19	**DMH / MBR**	p = 0.35	107
	DMH / SRD	p = 0.31	85
	MBR / SRD	p = **0.019**	**28**

Table 7.2: Summary of comparisons between the experienced observer's results for each case with significant correlation identified by bold type.

7.4.2 Novice performance

In the absence of reasonable agreement between the data sets for the experienced observers, identification of definitive improvement or deterioration in the novice's performance was difficult (because the reference point - the experienced observer approach - was unstable). Nevertheless, the novice performance was assessed on a case by case basis with respect to each of the three experienced observer's results.

7.4.2.1 Case 2

There was no significant correlation between the novices and experienced observers either before or after training. The KLIC values showed no change in the approaches of novices A, C and D and a trend away from the experienced observers' approach in novices E and F. KLIC values for novice B remained comparatively very large indicating no similarity to the experienced observer approach.

7.4.2.2 Case 3

Before training the only significant correlation was between DMH and novice E ($p = 0.007$). After training a significant correlation was found between DMH and novices A, B, D and E ($p = 0.05$, 0.007, 0.05 and 0.04 respectively) and between SRD and novices D and F ($p = 0.007$ and $p < 0.001$ respectively). The KLIC values indicated that novices C, E and F did not alter in their approach, novice A tended towards DMH but away from SRD and MBR and novice B tended towards DMH and MBR.

7.4.2.3 Case 7

Before training there was a strong correlation between novice C and DMH and MBR (p <0.0001) and between novice B and SRD ($p < 0.001$), SRD did not correlate with DMH or MBR (Table 7.2). After training the only significant correlation was between novice A and DMH and between novices A and F and MBR, the strong correlation of novice C was no longer present.

7.4.2.4 Case 9

Before training there was a significant correlation between all three experienced observers and novices A, C, E and F (Table 7.3)

	Experienced observers					
	Before training			**Post-training**		
Novices	**DMH**	**MBR**	**SRD**	**DMH**	**MBR**	**SRD**
A	**P < 0.05**	**P < 0.007**	**P < 0.01**	**P < 0.007**	P = 0.10	**P = 0.058**
B	P = 0.09	P = 0.36	P = 0.26	**P < 0.02**	**P < 0.02**	**P < 0.007**
C	**P < 0.02**	**P < 0.007**	**P < 0.04**	**P < 0.001**	**P < 0.002**	**P < 0.001**
E	**P < 0.002**	**P < 0.03**	**P < 0.01**	**P < 0.002**	**P < 0.02**	**P < 0.004**
F	**P < 0.001**	**P < 0.0001**	**P < 0.0001**	**P < 0.01**	**P < 0.001**	**P < 0.007**

Table 7.3: Case 9 Spearman rank correlation test p values for the pre-training results from novices A, B, C, E and F compared with the results from the three experienced observers (no significant correlation was found with novice D).

Post-training novice A results ceased to correlate significantly with MBR and SRD but novice B correlation with all the experienced observers became significant. The KLIC values indicated that only the results from novices A and B changed post-training, as reflected in the correlation results. Novice A, based on KLIC, tended towards DMH and MBR but away from SRD. Novice B results tended towards DMH and MBR.

7.4.2.5 Case 19

There was no significant correlation found before or after training between any of the novices and DMH, note there was no significant correlation between DMH and the other two experienced observers. Prior to training there was a significant correlation in results between MBR and novices A, C, D, E and F and between the results for SRD and novices A, C and F (table 7.4).

	Experienced observers					
	Before training			Post-training		
Novices	DMH	MBR	SRD	DMH	MBR	SRD
A	P = 0.19	**P < 0.001**	**P < 0.02**	P = 0.10	**P < 0.02**	**P < 0.05**
C	P = 0.70	**P < 0.04**	**P < 0.01**	P = 0.75	P = 0.35	P = 0.07
D	P = 0.88	**P < 0.02**	P = 0.23	P = 0.93	**P < 0.02**	P = 0.11
E	P = 0.88	**P < 0.005**	P = 0.12	P = 0.31	**P < 0.02**	P = 0.40
F	P = 0.40	**P < 0.03**	**P < 0.01**	P = 0.80	P = 0.53	P = 0.53

Table 7.4: Spearman rank correlation test p values for the case 19 pre and post-training results from novices A, C, D, E and F compared with the results from the three experienced observers (no significant correlation was found with novice B).

After training only four comparisons had significant correlation indicating an overall trend away from experienced observer behaviour. The KLIC values indicated no change following training in the results for novices A and D, a trend away from the experienced observers in the results for novices B, C and F and a trend towards the experienced observers for novice E (not reflected clearly in the correlation significance)

7.4.3 Average normalised time density by feature

The average normalised time density for each feature from the experienced observer results for each case and the equivalent for the novices pre and post-training are plotted in Figures 7.2-7.4. The graphs demonstrate the relative importance assigned to each feature,

based on normalised time density values, in each case by the experienced observers and the novices.

7.4.4 Time sequence results

The trend in KLIC values calculated from regular time intervals during the eye-tracking recording for the novices demonstrated a tendency to reach a steady state KLIC value after 40% to 80% of the total reading time for a case with no suggestion of a tendency for the KLIC to increase toward the end of the case. An example is given in Figure 7.5.

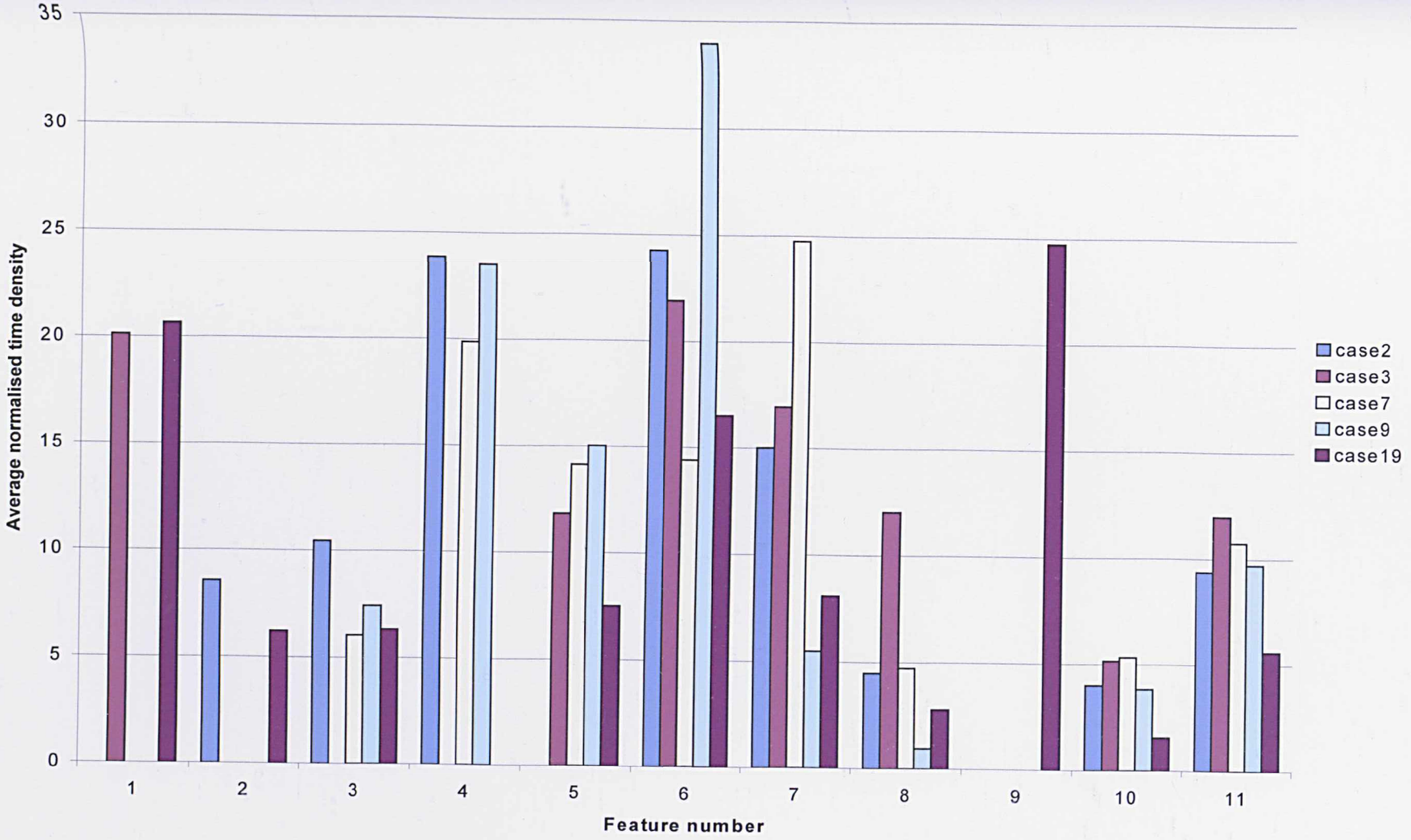


Figure 7.2: Graph of average normalised time density for each feature averaged from the results from the experienced observers for each case (different colours).

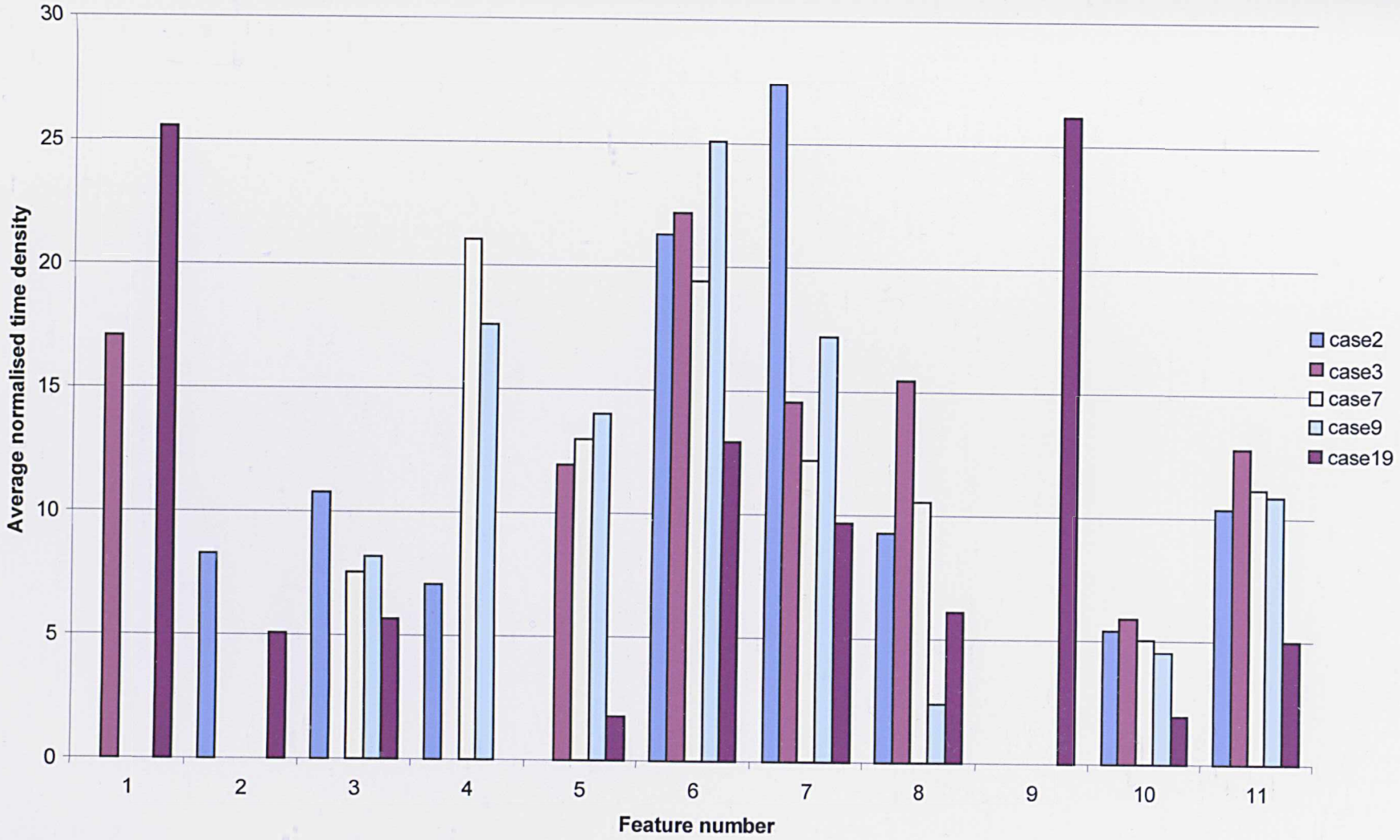


Figure 7.3: Graph of average normalised time density for each feature averaged from the results from the novice observers prior to training for each case (different colours).

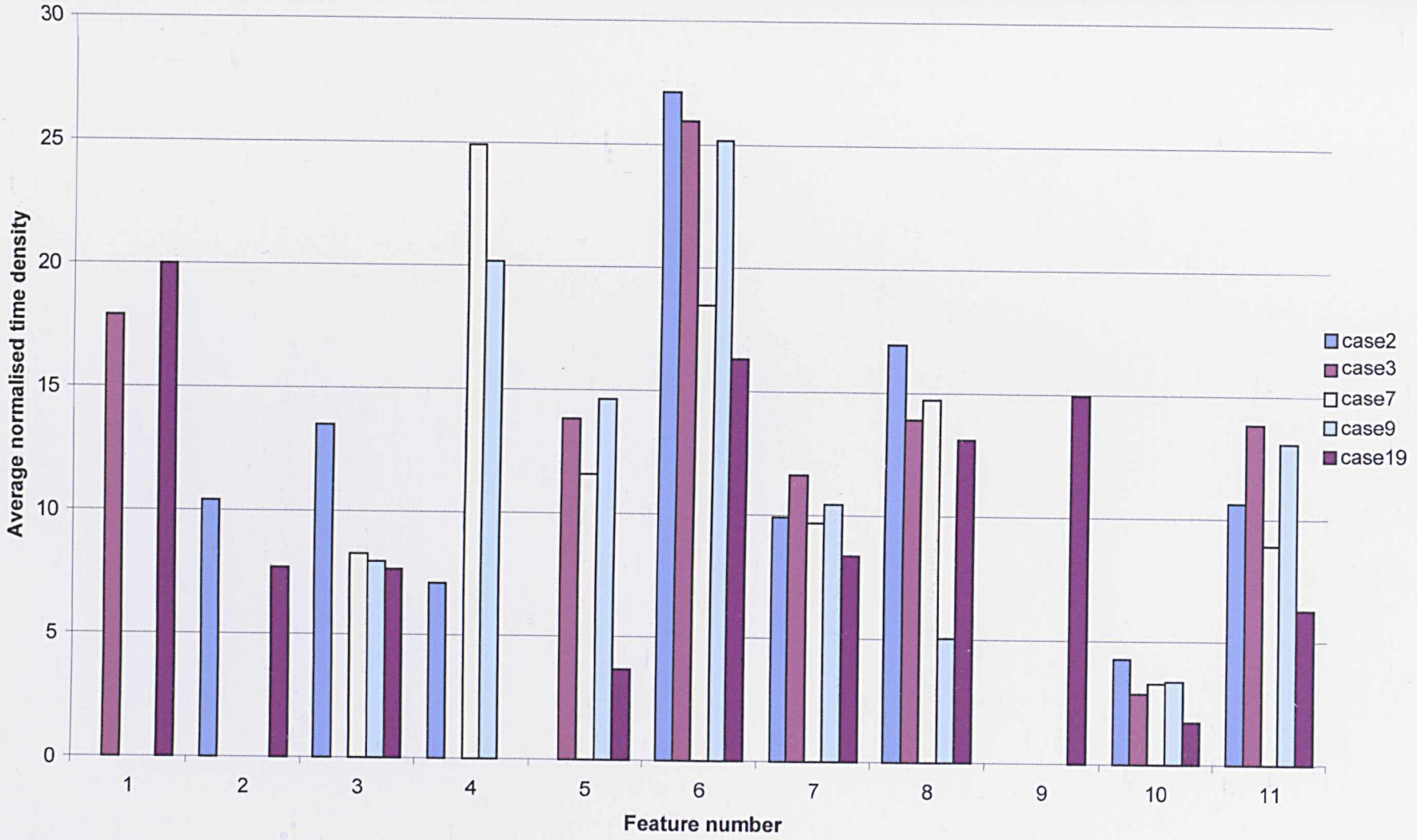


Figure 7.4: Graph of average normalised time density for each feature averaged from the results from the novice observers, after training, for each case (different colours).

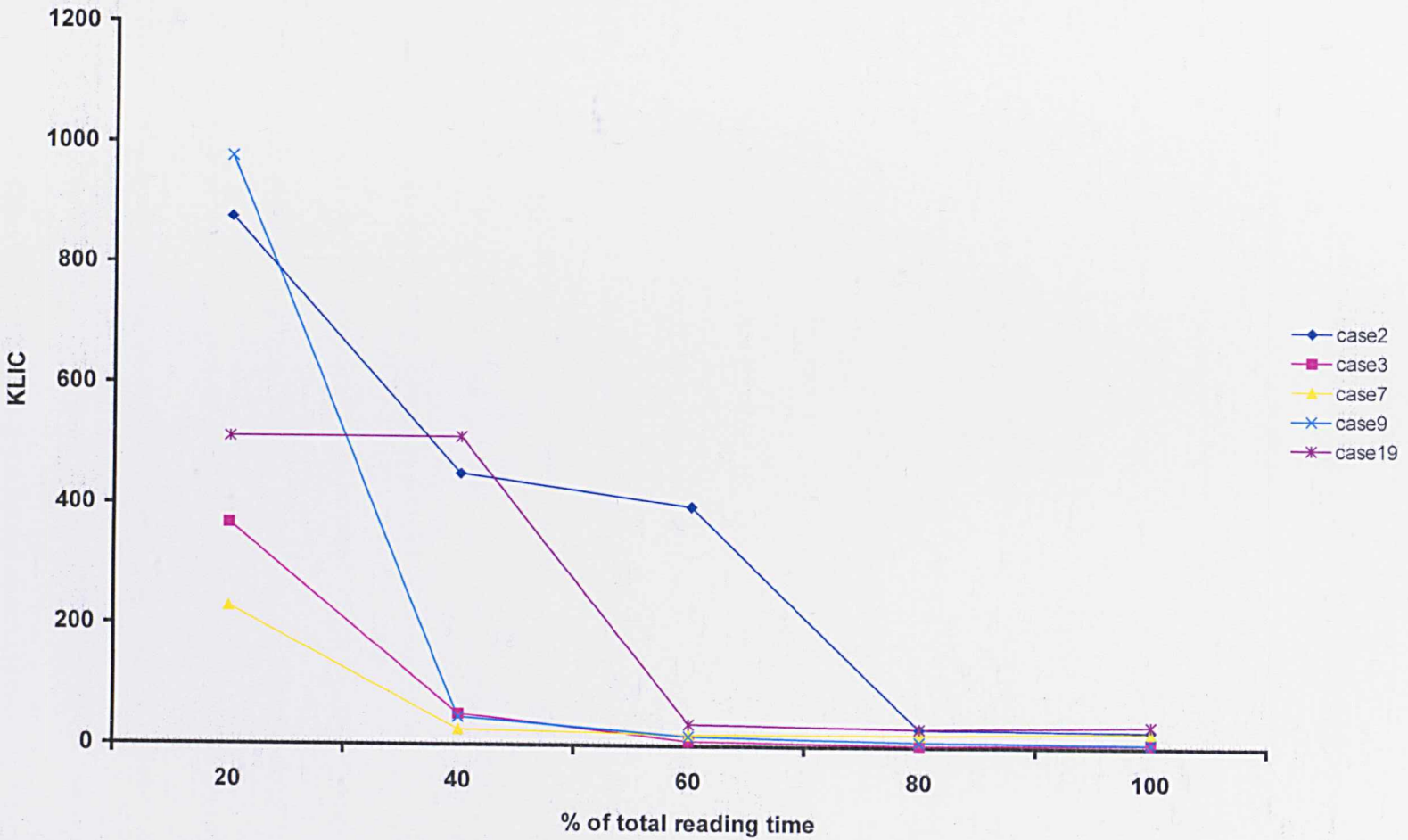


Figure 7.5: Graph of KLIC against the percentage of the total time taken for reading a case, for each case read by novice C after training.

7.5 INTERPRETATION OF THE RESULTS

There was reasonable agreement between the Spearman rank correlation and the KLIC, although some results were clearly spurious. The poor correlation between the experienced observers for some cases complicated the assessment of novice performance. In each case where there was a poor correlation between the experienced observers, two of the three observers had similar results, while the other was discrepant. In case 2 the fundamental difference between observer DMH and the other two experienced observers was a much lower normalised time density recorded for the trachea for DMH. Conversely, a high normalised time density for the trachea was the reason for the difference between SRD and the other two experienced observers in case 7. The correlation between the experienced observers in case 3 was not significant but the KLIC values suggested a reasonable level of agreement between DMH and SRD with no obvious reason for the poorer level of agreement in the comparisons with MBR. A highly significant level of agreement was seen between all the experienced observers in case 9 (a patient with lung fibrosis demonstrating ground glass opacity and reticulation in which the bronchovascular bundle were subject to particular scrutiny by all the observers) and overall KLIC values for all comparisons on the results from this case were lower than for the other cases. Case 9 had the most obvious abnormal features with considerable overlap between the regions marked as ground glass opacity and those marked as reticulation, which increased the likelihood that the time density for these two feature would increase in parallel reducing the probability spread of time density among the features mapped out. The strong agreement between the experienced observers enabled a critical but

limited analysis of any change in a novice's approach to the case. Only novices B and D had the potential to improve, by comparison with the results from the experienced observers, as the results pre-training for novices A, C, E and F significantly correlated with those from all three experienced observers. Following training novice A results remained significantly correlated only with DMH, the results for novices C, E and F were unchanged. Of the two novices with room for improvement, novice B demonstrated a shift in results that correlated significantly with all three experienced observer but novice D demonstrated no change.

In general there was no discernible influence upon the results that could be attributed to training, with or without specific information regarding the experienced observers behaviour. The lack of consistency of correlation between the experienced observers results leaves the analysis without a robust "gold standard". However, where the experienced observers differed there was typically a close correlation between two of the observers, not always the same two, with the third approaching the case in a different way. When the novice results significantly correlated with the experienced observers, whether before or after training, it was with the two experienced observers that had approached the case in the same way. This suggests that there is a routine behaviour that may be teased out of an observer's approach, but that differentiation between experienced and inexperienced observers may require a more sophisticated approach.

The averaged normalised time density results demonstrate a broad agreement between experienced observers and novices with respect to the importance attributed to specific

features. The bronchovascular bundles are scrutinised in all cases in keeping with the known multitude of pathological processes that may affect this specific anatomical element. The overall results imply that the significant features as defined in this experiment are identified by both experienced and novice observers with similar frequencies, which may explain the failure to demonstrate clear cut behavioural changes in the later analyses.

7.6 DISCUSSION

The primary aim of the feature-based analysis was to identify an alteration in a novice's approach to HRCT cases as a result of being trained in the interpretation of HRCT features and, more specifically, in the way those features are linked in terms of the probable underlying lung disease. Gaze duration, scan paths and detection times have been shown to characterise observer experience in the search strategy for reading mammograms (Krupinski 1996) but in this experiment the observer's task is one of interpretation and not detection.

The features chosen for analysis were based on those that appeared to be of most significance to experienced observers reading HRCT cases. The presence of any specific anatomical or pathological feature may vary dependent on which level, image slice, is being viewed. To accommodate the three-dimensional spatial variation in feature position, the eye-tracking data were recorded from a sequence, limited to five images, of levels allowing the analysis of the overall importance of a specific feature with respect to

an HRCT case as a whole rather than within the artificial confines of a single image. The assimilation of a case on HRCT involves the consideration of the entire imaged lung volume and thus can only be adequately assessed if the multiple images that make up a case are considered as a whole.

The relative importance of any specific feature in a given case was determined by the normalised time density value recorded for that feature from the eye-tracking data and is based on the empirical assumption that a feature's significance determines the amount of time that an experienced observer will spend scrutinising it. This is potentially a simplification, the time density data will also be a function of feature conspicuity, a feature that is readily appreciated requires less time to assimilate adequately. Furthermore, some features may be satisfactorily assessed with peripheral vision and therefore their importance may be underestimated by an approach based solely on foveal vision. A review of the normalised time density recorded for each feature averaged from the experienced observers' data and the pre and post-training novices' data separately showed consistent trends that implied that time density is a useful measure of feature significance, but not a good discriminator of experience. The failure to see any obvious change in the results following training implies that either an increase in knowledge levels does not alter the perception of feature significance or that the training failed to increase knowledge levels sufficiently to demonstrate a difference.

There were some predictable difficulties in deriving a satisfactory "gold standard" in this experiment, with one of the experienced observers recording a pattern of normalised time

density that differed to the other two in most cases. In two of the cases this discrepancy was as a result of a marked difference in the attention paid to the trachea, and it could be argued that assessment is of limited relevance to the overall interpretation of parenchymal disease. Nevertheless, the agreement between the remaining two experienced observers in such cases supported the use of those observers, data as the yardstick for assessment of the novices' data.

In only one case was the interobserver agreement between the experienced observers universal, in the remaining cases one of the experienced observers had results that did not correlate well with the other two experienced observers; there was typically a normalised time density value for one feature that skewed the results for the remaining features causing a loss of correlation.

The only attempt to include temporal data in the analysis was in the calculation of cumulative normalised time density values for the novices as they read each case. The rationale was to identify any fluctuation in the novices' results with respect to that of the experienced observers to determine whether the novices' results were closer to that of the experienced observers at some point during the reading of the case (i.e. before the end of the reading of the case had been reached). However, there was no evidence to suggest this, and the KLIC values indicated a steady state of comparison was obtained prior to ending the reading of the case. Eye-tracking experiments concerned with nodule detection on chest radiographs have demonstrated an increase in dwell time corresponds to a likely false negative response suggesting that once the reading process has been

completed there is no improvement with further viewing (Kundel, Nodine, & Krupinski 1989). The use of summated normalised time density values for an entire case may have lacked sensitivity in identifying subtle changes in the order in which features were assimilated. In particular, the temporal relationship between the accumulation of time density for different features within a case was not assessed. Such an analysis could indicate which features were temporally linked by the observers implying a synergism in their contribution to the overall assimilation of the case. It would be expected that the efficient assimilation of a case by an experienced observer would contain such temporal links between features, and the detection of such behaviour in a novice would therefore indicate a more "experienced" approach to a case.

The application of eye-tracking data to define a radiologist's approach to CT, as opposed to projectional radiographic, examinations has not been attempted before and it may be that there is no single best evaluative approach, especially as experienced observers probably do not behave in a consistent manner. The selection of normal anatomical features included in the analysis, although based on prior observation, may have confused the overall results pertaining to the pathological features. Identification of features that are not utilised by the experienced observers in the assimilation of a CT examination would not be viewed in a consistent way and would tend to add noise to the data set reducing the signal to noise ratio. The consideration of a smaller feature set may improve the degree of correlation but potentially at the expense of sensitivity, which for this analysis was maximised as any alteration in novice behaviour as a result of specific training was shown in Chapter 6 to be very subtle, if present at all.

The feature-based analysis has demonstrated the ability to record a radiologist's technique for visually assimilating HRCT examinations but has failed to identify a consistent improvement in novice approach to HRCT, as assessed by this technique, following specific training.

7.7 SUMMARY

1. ***A radiologist's visual approach to an HRCT examination can be successfully assessed in a feature-based fashion.***
2. ***There was variation in the way that experienced observers assimilate HRCT examinations (but usually marked consistency between two of three of the observers)***
3. ***Similarity in approach of novices to HRCT examinations when compared to that of the experienced observers was more determined by the case itself than through training.***
4. ***A feature-based analysis of novice approach to HRCT has failed to demonstrate a consistent improvement as a result of specific training.***

CHAPTER 8

CONCLUSION

Central to this thesis is an examination of what differentiates radiologic experts from novices. In contrast to novices, experts have completed more advanced training, developed a greater, highly specialised, domain knowledge and have years of experience. However, expertise is not an elementary matter and to cultivate expertise in trainees a greater understanding of the nature of radiologic expertise must be developed (Gunderman et al. 2001).

That experts have a greater knowledge than novices is obvious but their expertise cannot be simply defined in those terms. What sets the expert apart from the novice is the ability to integrate that knowledge with the image being assimilated, guiding the eye to key areas and linking those areas or features that have a synergistic influence on the final interpretation; expert radiologists do not necessarily have superior visual search skills (Nodine & Krupinski 1998). Because a novice tends to look more indiscriminately at an image (Hu et al. 1994), the recording of eye movements should enable the differentiation of expert from novice behaviour (Manning, Leach, & Bunting 2000) and this has been amply demonstrated in this thesis in both the anatomical and feature-based analyses.

However, the experiments did not demonstrate the adoption by the novices of more expert behaviour through the use of a training presentation, even though specific expert behaviour was highlighted. This is perhaps not surprising: studies of world-class chess

players have suggested that no short cuts to the development of such expertise exist suggesting time investments in excess of 50,000 hours are necessary, which raises questions as to what level of improvement in HRCT visual assimilation could be achieved by 1-2 hours training (Chesters & Simon 1973). Despite this, the educational implications of understanding the conceptual nature of expertise are considerable. Teaching novices the higher level thinking and processing that an expert performs should increase their ability to tackle diagnostic difficulties that arise when they are confronted by a plethora of "significant" features. Sometimes the novice may literally not know where to start but, by considering a constellation of features as a defined entity, they will be able to come to an interpretation more accurately and rapidly, and are less likely to miss a feature through the distraction of a more obvious abnormality (i.e. the well known problem of satisfaction of search (Berbaum et al.. 1990)). If this linking of significant features can be imparted to novices it should be evident in the temporal order in which their eye-movements visit various regions of an HRCT examination. This form of temporal analysis of the massive data sets acquired from 50Hz sampling of eye movements (in the order of 9000 data points per case) was beyond the scope of this thesis. In a cruder analysis, one would expect the appropriate visual linking of significant features to be reflected in the relative proportion of time spent on those features and this was apparent in the feature-based analysis. That complex tasks can be learnt by the acquisition of small, less complex, sub-elements of the overall process which are subsequently collated is likely (Lee & Andriole 2001). The failure to demonstrate a change in the novices behaviour towards that of the experts following training, may be a result of the crudeness of the analytical technique used but most likely reflects the failure

to impart a more expert approach through the training method employed. However, the use of computer-based training is likely to increase in radiology training and has been shown in these investigations, and elsewhere (Lieberman et al. 2002), to have improved the novices interpretative skills, as assessed by their verbal feedback; four out of the six novices', following training, demonstrated improved identification of significant HRCT features. Although the presented investigations were not designed to assess improvements in diagnostic accuracy, the identification of HRCT features is one step in the diagnostic process.

A feature of the earlier experiments into useful peripheral vision and the relative efficiency of viewing different formats, highlighted the significance of peripheral vision in the assimilation of HRCT images. The surprisingly effective snap-shot interpretation of chest radiographs demonstrated by Kundel et al in the 1970's (Kundel & Nodine 1975) must be a function of the use of peripheral vision, and subsequent experiments identified the use of peripheral vision as an adjunct to guiding foveal vision in the search for inconspicuous nodules on chest radiographs (Kundel, Nodine, & Toto 1991). The format related experiments demonstrated a surprising range of useful peripheral vision; the vast proportion of saccadic eye movements fell within the range of useful peripheral vision, which supports the hypothesis that peripheral vision is an adjunct to visual search strategy. The improved efficiency with which the HRCT examinations were assimilated in stacked mode as opposed to a two by two tiled mode, may in part be attributable to the more efficient use of peripheral vision in the interpretative process on the larger stacked mode images.

By its very nature, eye-tracking analysis is primarily concerned with foveal fixation and the specificity of the information derived from eye-movements with respect to the associated cognitive processes in the interpretation of what is being viewed is a complex issue. Extending the field of view to the surrounding area of useful peripheral vision, dramatically increases the amount of information that may be assimilated from an individual fixation and therefore incurs an inaccuracy in terms of the assumption of what is being viewed. The analysis of the experienced observers' approach to the abnormal HRCT examinations in Chapter 5 demonstrated the use of peripheral vision as a screening tool for viewing the lung periphery; even if eye-movements were not tracked to the lung periphery, to assume that it was not scrutinised would be erroneous. However, the detection of an abnormality does not equate to its successful interpretation and, in the presence of a peripheral abnormality, foveal vision was used to improve interpretation. Although visual input is derived from peripheral vision there is no evidence to suggest that there are shifts in visuospatial attention away from the fovea (Maioli et al. 2001).

A confounding factor in the attempt to teach more expert behaviour and subsequently test the result through eye-tracking data is the assumption that the recorded eye-movements are associated with the assimilation of the region being viewed. This is not necessarily the case as the eyes will continue to scan around an image while the observer is formulating hypotheses based on previously identified features. It is difficult to distinguish eye-movements occurring during these "thinking phases". Furthermore, a large proportion of eye-movements are beyond conscious control. Informing a novice as to where or what

they should look at will result in some direction of visual gaze but many eye-movements are stimulated by both top-down, i.e. the cognitive centres of the brain, and bottom-up, i.e. retinal visual sub-system, processes thus bypassing conscious control (Patel & Sathian 2000). Influencing eye-movements stimulated at retinal level would be a function of the way in which the image is presented, e.g. stack versus tiled mode presentation formats, in order to influence the subconscious higher level stimulated eye-movements the trainee has to be taught to think differently. Analysing the eye-movements of the experienced observers provides some insight into their interpretative process and can identify subconscious behaviour. However, successfully transferring such subconscious behaviour to the knowledge base from which the novice stimulates conscious eye-movements will fail to replicate expert behaviour as the stimulus is from the wrong cognitive source. Simply put, the human brain does not learn in this fashion and any attempt to influence subconscious behaviour by conscious control is bound to fail. A computer does not have a subconscious and interpretative behaviour of a diagnostic computer system is entirely dictated by its programmed algorithms. It may prove easier to educate a computer, than a novice radiologist, to interpret images in a more expert way through the generation of appropriate algorithms based on expert eye-tracking results. The difficulty arises in the visual cognitive processes that are performed by the human brain in perceptual and cognitive tasks that are extremely difficult to reproduce in computers. Nevertheless, there has been some success in computerised classification of nodules on chest radiographs (Vittitoe, Baker, & Floyd, Jr. 1997) and CT imaging (Reeves & Kostis 2000a)

APPENDIX

Studies suggest that computer assisted instruction, although inferior to personal interactive tutorials, can be effective (Lieberman, Abramson, Volkan, & McArdle 2002). Following the analysis of the eye-tracking data recorded from the experienced observers in Chapter 5 a teaching file (tutorial) was created as a PowerPoint (Microsoft, USA) presentation. The same cases were used as examples but care was taken to ensure the images from each case used were not one of the five used in the experiment. The tutorial concentrated on the general identification of the commonest HRCT features found in disease and indicated the likely pathological cause; the design was to facilitate feature recognition (rather than establishing a specific diagnosis). It was decided that incorporating numerous differential diagnoses would confuse the basic aim, i.e. an alteration in approach to HRCT examination reading; an improvement in diagnostic accuracy should flow from improved feature detection. At appropriate places in the tutorial, slides were included that described the approach used by the experts. The aim was to help the trainee develop an expert-like reading strategy, based on those developed by the observers through experience.

HRCT Tutorial

1

Introduction

- This presentation will demonstrate various features found on HRCT scans of diffuse lung disease.
- Go through it as often as you like but please do so at least once and let me know when I see you how many times you went through it.
- There are only 35 slides and it shouldn't take more than a 1½ hour.
- To ensure all participants have viewed all of the presentation a letter or number will appear in the top right hand corner on each slide after a click of the left mouse button. Make a record of these as you see them so that you can tell me what they were when you come for re-testing.
- This isn't lack of trust, just security against a potentially uncomfortable question that could arise in my MD viva.
- A second click will advance you to the next slide.
- Don't worry if you forget to write the letter or number down before moving on to the next slide, if you go backwards by pressing the left cursor key the letter will be visible.

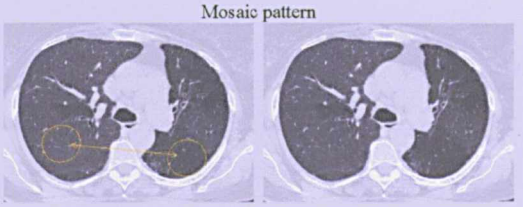
2

Mosaic attenuation

- This is usually the feature that readers find the most difficult representing patchy areas of different overall attenuation.
- The critical decision is whether the darker or lighter lung is abnormal.
- An infiltrate such as interstitial inflammation or fibrosis can give abnormally lighter lung.
- Abnormal darker lung can be due to:
  - an excess of air due to trapping not enabling normal flow into and out of alveoli (small airways disease) or due to an increase in air spaces due to destruction of lung tissue (emphysema),
  - a reduction in blood flow due to shunting away from poorly functioning air trapping or destroyed lung or obstruction to vascular flow (chronic thrombo-embolic disease).
- The best indicator is vascularity (vessel number and calibre). If the darker lung is abnormal there will be diminished vascularity in that lung compared to the lighter lung. If the lighter lung is abnormal (ground glass opacity) there should be no significant difference in the vascularity.

3

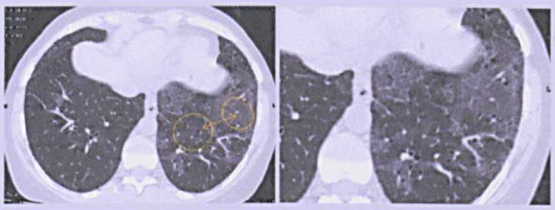
Mosaic pattern



- The above image demonstrates mosaic attenuation due to abnormal black lung.
- Note the relative lack of vessels in the black lung compared to the normal lung

4

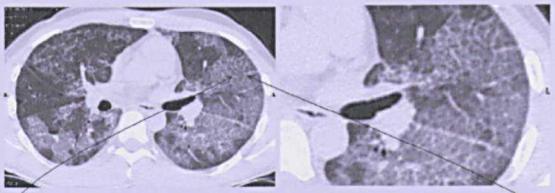
Mosaic



- The above image demonstrates mosaic attenuation due to abnormal grey lung.
- Note there is no difference in vascularity between grey lung and normal lung

5

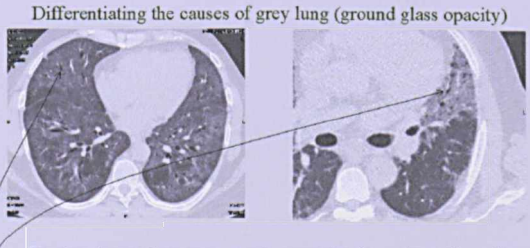
Differentiating the causes of grey lung (ground glass opacity)



- The above image demonstrates mosaic attenuation due to abnormal grey lung (ground glass opacification).
- Ground glass opacification is defined as an area of increased parenchymal density without obscuration of the vessels as occurs in consolidation.
- Note there is no dilatation of the airways to suggest traction bronchiectasis and the fissures are not being "pulled" or distorted by fibrosis. There is interlobular and intralobular septal thickening indicating an interstitial infiltrate, in this case fluid as a result of alveolar proteinosis. This is the classic "crazy paving" appearance.

6

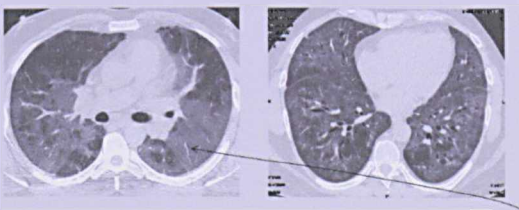
Differentiating the causes of grey lung (ground glass opacity)



- The above images demonstrate mosaic attenuation due to abnormal grey lung.
- Note there is traction bronchiectasis. The ground glass opacity on the left hand image appears "softer" than that on the right but the traction bronchiectasis indicates fibrosis as the cause in both. On the left note how dark the air in the bronchi seems compared to the adjacent lung parenchyma. This is the black bronchus sign and helps identify the grey lung as the abnormal lung.

7

Comparison of non-fibrotic and fibrotic ground glass opacity



- The above images demonstrate mosaic attenuation due to abnormal grey lung.
- The image on the right is due to interstitial fibrosis and that on the left is a cellular infiltrate. Note the difference in the airways. A normal calibre airway is seen accompanying the vessel in the image on the left. The airways in the image on the right are clearly dilated.

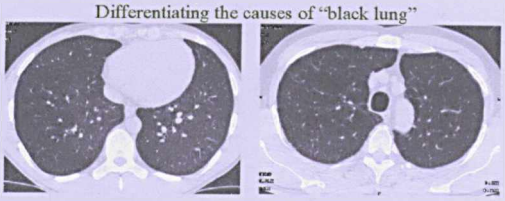
8

What the experts do.

- The experts appeared to identify abnormal ground glass rapidly and scrutinised the ground glass first.
- Comparison was made with the normal lung looking at the vessels.
- In subtle cases the experts tended to make a quick comparison with the trachea to gauge their determination of what normal parenchymal density should be (a bit greater than air).
- The airways within the ground glass were scrutinised - ?traction bronchiectasis
- Identification of ground glass led to an assessment of the oesophagus, probably a Brompton foible due to the number of scleroderma associated interstitial pneumonia cases we see here.
- The sub-pleural region was only scrutinised when an abnormality was identified and not routinely.

9

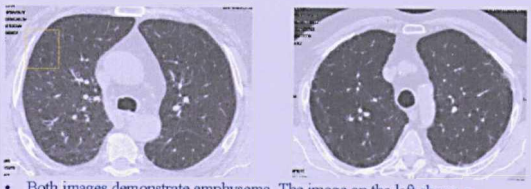
Differentiating the causes of "black lung"



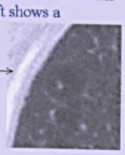
- Both images demonstrate a mosaic attenuation pattern due to abnormal black lung. The image on the left shows small airways disease, in this case constrictive obliterative bronchiolitis (OB) and the black lung is due to a combination of air trapping and physiological shunting of blood to more functional areas of lung. The image on the right is due to centrilobular emphysema.
- Similarities- reduced vascularity but due to different processes (OB-vessel shut down, emphysema- parenchymal destruction).
- Differences- The vessels in emphysema are distorted and the black lung has a permeative appearance (moth-eaten)

10

Black lung - emphysema



- Both images demonstrate emphysema. The image on the left shows a centrilobular distribution (moth-eaten lung).
- A characteristic finding is a centrilobular vessel surrounded by a lobule of destroyed lung ⟶
- The image on the right shows more extensive, panacinar, emphysema. As this becomes more severe and the number of visible vessels reduces it can be very difficult to differentiate from severe constrictive OB.



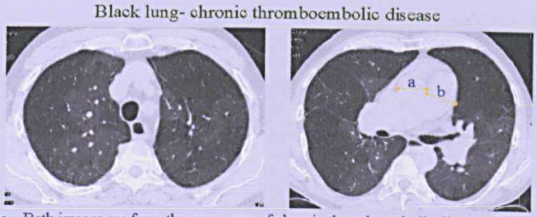
11

What the experts do.

- The experts tended to identify whether the black or grey lung was abnormal very quickly by eyeballing the vessels and scrutinised the black lung first comparing this with the grey lung.
- They then scrutinised the segmental and sub-segmental bronchovascular bundles to assess vessel and airway calibre ?bronchiectasis, PE.
- They used a comparison of left and right lung to enable comparison of parenchyma and vessels at equivalent regions of the lung.
- They tended not to scrutinise the sub-pleural region routinely, focusing attention there only if there was an abnormality.

12

Black lung- chronic thromboembolic disease



- Both images are from the same case of chronic thromboembolic disease. Note the mosaic attenuation and the reduced vascularity of the black lung. A, sometimes marked, increase in vessel calibre in the "normal" lung may help distinguish this from small airways disease (SAD). Further clues are the ancillary signs of pulmonary hypertension, likely to accompany chronic PE, such as a pulmonary outflow tract of greater calibre to the adjacent ascending thoracic aorta (b > a).

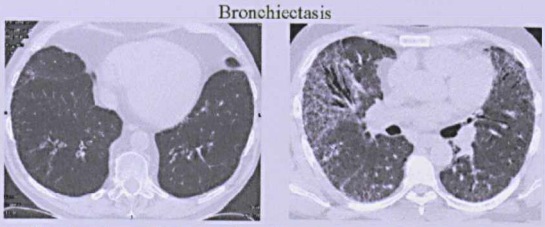
13

What the experts do.

- The presence of black lung in the absence of obvious bronchiectasis led to an assessment of the pulmonary artery to ascertain the possibility of chronic PE as a cause.
- The experts tended to do this when they alighted on an appropriate slice rather than cycle through the slices specifically for that process.

14

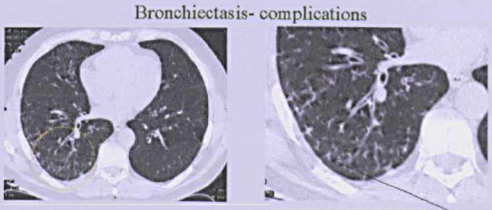
Bronchiectasis



- The above images demonstrate bronchiectasis, abnormal dilatation of the airways typically assessed by comparison of airway calibre with the adjacent pulmonary artery- the inner diameter of the airway lumen should be less than or equal to the diameter of the vessel.
- The image on the left shows cylindrical bronchiectasis as a disease entity in itself, other flavours include cystic bronchiectasis where the bronchial dilatation is not uniform. That on the right shows traction bronchiectasis secondary to parenchymal fibrosis, here the airways are literally pulled apart by the contracting lung parenchyma.

15

Bronchiectasis- complications



- The above images demonstrate bronchiectasis in an exudative state. There is an increase in secretions, typically stimulated by infection, which causes plugging of the small peripheral ectactic airways. The result is peripheral branching opacities on the HRCT with blobs at the end "tree in bud" pattern.
- When caught in cross-section the tree in bud can appear nodular. The key to diagnosing an exudative bronchiectasis is to identify the presence of bronchiectasis, if not present an alternative cause should be found such as endobronchial spread of TB or infectious bronchiolitis

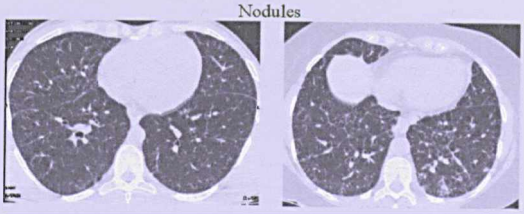
16

What the experts do.

- When a tree in bud pattern was identified a systematic search of the bronchovascular bundles was made to identify bronchiectasis.

17

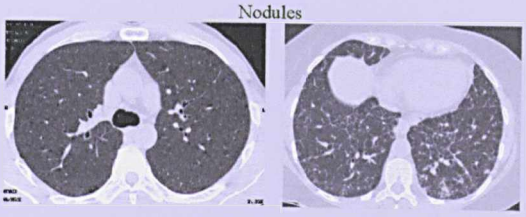
Nodules



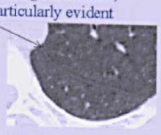
- The characterisation of nodules as centrilobular, perilymphatic or random is of value in determining the cause.
- If, as in the above cases, there are subpleural nodules then a patchy distribution and /or a predominance noted relative to the peribronchovascular interstitium, interlobular septa or subpleural regions suggests a perilymphatic distribution.
- If there are no subpleural nodules then the the pattern is most likely centrilobular.

18

Nodules



- The image on the right is unchanged from the previous slide. The image on the left demonstrates a centrilobular distribution of nodules in this case in clusters, with a "soft / fluffy quality" found in extrinsic allergic alveolitis; note the absence of nodules in the subpleural regions particularly evident along the fissure



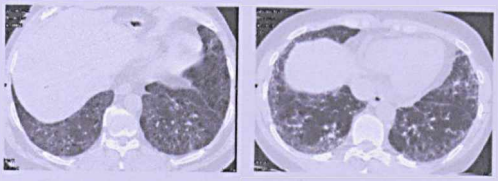
19

What the experts do.

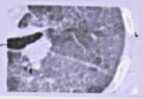
- The identification of nodules led to scrutiny of the fissures and then hila- ?sarcoid
- The subpleural region, if abnormal was scrutinised to determine if there was subpleural nodularity.

20

Reticular pattern

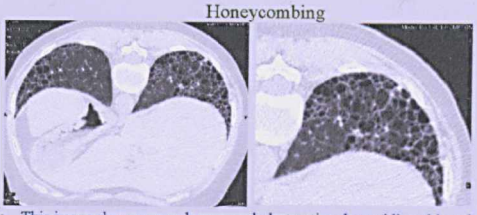


- A reticular pattern basically means multiple short interconnecting lines and represents thickening of the interstitium. Clues as to the cause of reticulation must be derived from the parenchyma overall, the presence of distortion or traction bronchiectasis would implicate, but not prove, fibrosis. In both of the above cases the cause for the reticulation is fibrosis but notice the spectrum of coarseness that can be found.
- This is a copy of the alveolar proteinosis case you have already seen to remind you of a non-fibrotic cause of a reticular pattern, heart failure could do the same.



21

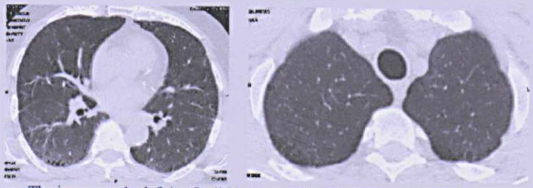
Honeycombing



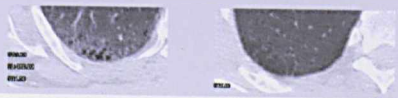
- This image shows severe honeycomb destruction due to idiopathic pulmonary fibrosis (or CFA). Multiple cystic airspaces proliferating from the subpleural region. Note the cysts abut one another and are demarcated by fibrous tissue.
- Honeycomb destruction classically begins posteriorly at the bases but as it extends cranially, maintaining a subpleural distribution, lateral and then anterior regions become involved with relatively less posterior involvement.

22

Honeycombing

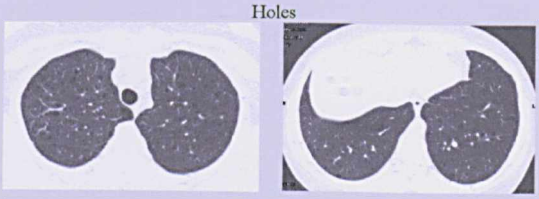


- The image on the left is of mild honeycomb destruction and not to be confused with paraseptal emphysema shown on the right.
- Note the tissue surrounding honeycomb destruction is fibrotic and tends to be of high density. The tissue adjacent to paraseptal emphysema may show the permeative appearance shown above or be demarcated by a thin wall like a bulla.

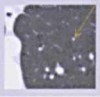


23

Holes

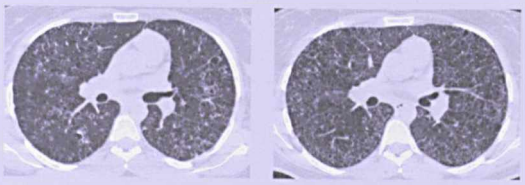


- The main causes of diffuse holes in the lung parenchyma are emphysema, lymphangioleiomyomatosis (LAM) and Langerhans cell histiocytosis (LCH). The above images are from a case of LAM (only found in women) characterised by well demarcated thin walled regular shaped cysts surrounded by normal lung parenchyma; the distribution of the cysts is uniform with no lung region spared.



24

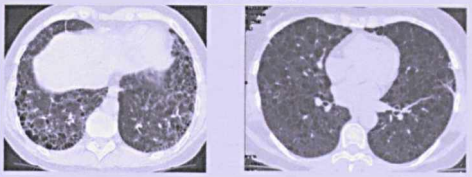
Holes



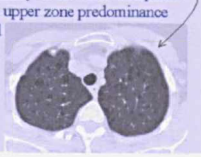
- Emphysema may be difficult to differentiate from thin walled cystic LCH, thicker walls and the presence of nodules would favour LCH (LCH cysts begin life as cavitating nodules). The above 2 images are from the same patient and separated by 3 years. Note the nodularity in the left hand image that has developed into further cysts in the right hand image. The cysts in LCH tend to be irregular in shape, have thick walls and the distribution of disease spares the extreme bases and anterior extremities of the middle lobe and lingula.

25

More severe holes



- The image on the left is of honeycomb destruction, that on the right is LAM (the upper zones show this more clearly). The cysts in honeycombing lie with a subpleural distribution in fibrotic lung predominantly in the bases compared to the diffuse distribution of cysts in LAM and an upper zone predominance in LCH where the cysts are surrounded by normal



26

What the experts do.

- Once identifying multiple cysts the experts scrutinise the surrounding parenchyma and quickly established the distribution of the cysts

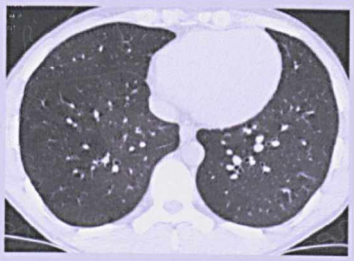
27

Self testing

- I have given you some examples of findings you may come across on HRCT and clues to differentiating them.
- There now follows a short quiz to enable you to put into practice what you have seen.
- This is just for fun, not assessed in any way but a good opportunity to enhance your recollection of what you have learned.
- There are no more hidden letters or numbers.

28

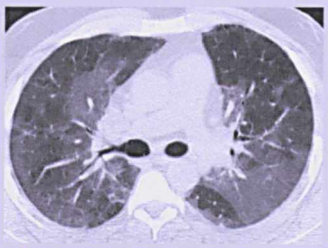
Decide for yourself the salient features on these images then press a key to reveal the answers.



Mosaic attenuation - black lung abnormal (SAD)

29

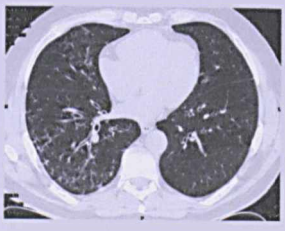
Decide for yourself the salient features on these images then press a key to reveal the answers.



Mosaic attenuation - there is patchy ground glass opacity due to an interstitial infiltrate with no evidence to suggest fibrosis.

30

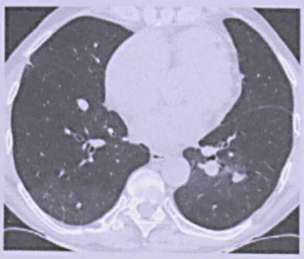
Decide for yourself the salient features on these images then press a key to reveal the answers.



Bronchiectasis and tree in bud pattern.

31

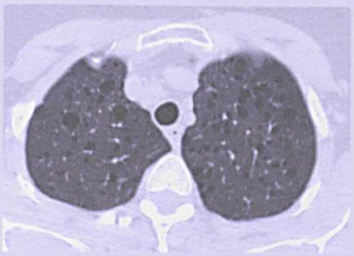
Decide for yourself the salient features on these images then press a key to reveal the answers.



Mosaic attenuation- black lung abnormal due to chronic PEs - note the calibre of the pulmonary vessels supplying the grey lung.

32

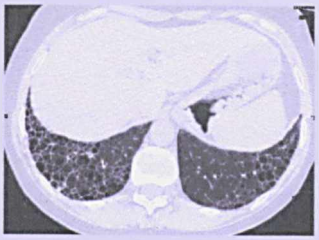
Decide for yourself the salient features on these images then press a key to reveal the answers.



Cysts diffusely involving the lung (LAM), the overall distribution of cysts would facilitate the distinction from emphysema, but note, in addition these cysts do not have central centrilobular vessels

33

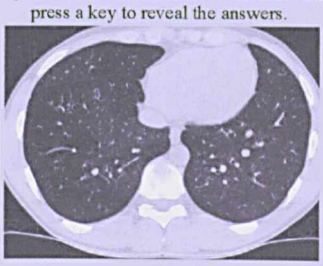
Decide for yourself the salient features on these images then press a key to reveal the answers.



Honeycomb destruction

34

Decide for yourself the salient features on these images then press a key to reveal the answers.



Mosaic attenuation due to SAD with minimal airways dilatation

35

Decide for yourself the salient features on these images then press a key to reveal the answers.



I saved this tester till last. There are areas of ground glass, black lung and normal lung - note 3 shades of grey are apparent. The cause is a process that causes small airways disease and interstitial infiltrates - extrinsic allergic alveolitis.

36

REFERENCES

Anon. 2001, "Bayesian statistics", International Society for Bayesian Analysis, www.bayesian.org/bayesian/bayes.htm.

Anthony, D. 1999, "Cluster analysis and discriminant analysis," in *Understanding advanced statistics*, Churchill Livingstone, London, UK, pp. 157-172.

Arakawa, H., Kurihara, Y., Sasaka, K., Nakajima, Y., & Webb, W. R. 2002, "Air trapping on CT of patients with pulmonary embolism", *AJR Am.J.Roentgenol.*, vol. 178, no. 5, pp. 1201-1207.

Arakawa, H., Niimi, H., Kurihara, Y., Nakajima, Y., & Webb, W. R. 2000, "Expiratory high-resolution CT: diagnostic value in diffuse lung diseases", *AJR Am.J.Roentgenol.*, vol. 175, no. 6, pp. 1537-1543.

Arakawa, H. & Webb, W. R. 1998, "Air trapping on expiratory high-resolution CT scans in the absence of inspiratory scan abnormalities: correlation with pulmonary function tests and differential diagnosis", *AJR Am.J.Roentgenol.*, vol. 170, no. 5, pp. 1349-1353.

Arenson, R. L., Andriole, K. P., Avrin, D. E., & Gould, R. G. 2000, "Computers in imaging and health care: now and in the future", *Journal of Digital Imaging*, vol. 13, no. 4, pp. 145-156.

Ayers, A. B. 2001, "Clinical governance: is it coming of age?", *British Journal of Radiology*, vol. 74, no. 888, pp. 1083-1087.

Beard, D. V., Bream, P., Pisano, E. D., Conroy, P., Johnston, R. E., Braeuning, P., McLelland, R., & Clark, R. 1997, "A pilot study of eye movement during mammography interpretation: eyetracker results and workstation design implications", *Journal of Digital Imaging*, vol. 10, no. 1, pp. 14-20.

Beard, D. V., Hemminger, B. M., Denelsbeck, K. M., & Johnston, R. E. 1994a, "How many screens does a CT workstation need?", *Journal of Digital Imaging*, vol. 7, no. 2, pp. 69-76.

Beard, D. V., Hemminger, B. M., Perry, J. R., Mauro, M. A., Müller, K. E., Warshauer, D. M., Smith, M. A., & Zito, A. J. 1993, "Interpretation of CT studies: single-screen workstation versus film alternator", *Radiology*, vol. 187, no. 2, pp. 565-569.

Beard, D. V., Johnston, R. E., Toki, O., & Wilcox, C. 1990, "A study of radiologists viewing multiple computed tomography examinations using an eyetracking device", *Journal of Digital Imaging*, vol. 3, no. 4, pp. 230-237.

Beard, D. V., Molina, P. L., Müller, K. E., Denelsbeck, K. M., Hemminger, B. M., Perry, J. R., Braeuning, M. P., Glueck, D. H., Bidgood, W. D., Mauro, M., & . 1995, "Interpretation time of serial chest CT examinations with stacked-metaphor workstation versus film alternator", *Radiology*, vol. 197, no. 3, pp. 753-758.

Beard, D. V., Pisano, E. D., Denelsbeck, K. M., & Johnston, R. E. 1994b, "Eye movement during computed tomography interpretation: eyetracker results and image display-time implications", *Journal of Digital Imaging*, vol. 7, no. 4, pp. 189-192.

Berbaum, K. S. 1995, "Difficulty of judging retrospectively whether a diagnosis has been "missed"", *Radiology*, vol. 194, no. 2, pp. 582-583.

Berbaum, K. S., Brandser, E. A., Franken, E. A., Dorfman, D. D., Caldwell, R. T., & Krupinski, E. A. 2001, "Gaze dwell times on acute trauma injuries missed because of satisfaction of search", *Academic Radiology*, vol. 8, no. 4, pp. 304-314.

Berbaum, K. S., Franken, E. A., Dorfman, D. D., Miller, E. M., Krupinski, E. A., Kreinbring, K., Caldwell, R. T., & Lu, C. H. 1996, "Cause of satisfaction of search effects in contrast studies of the abdomen", *Academic Radiology*, vol. 3, no. 10, pp. 815-826.

Berbaum, K. S., Franken, E. A., Jr., Dorfman, D. D., Rooholamini, S. A., Kathol, M. H., Barloon, T. J., Behlke, F. M., Sato, Y., Lu, C. H., el Khoury, G. Y., & . 1990, "Satisfaction of search in diagnostic radiology", *Invest Radiol.*, vol. 25, no. 2, pp. 133-140.

Bergin, C. J., Rios, G., King, M. A., Belezzuoli, E., Luna, J., & Auger, W. R. 1996, "Accuracy of high-resolution CT in identifying chronic pulmonary thromboembolic disease", *AJR Am.J.Roentgenol.*, vol. 166, no. 6, pp. 1371-1377.

Berlin, L. 1996a, "Errors in judgment", *AJR Am.J.Roentgenol.*, vol. 166, no. 6, pp. 1259-1261.

Berlin, L. 1996b, "Malpractice issues in radiology. Perceptual errors", *AJR Am.J.Roentgenol.*, vol. 167, no. 3, pp. 587-590.

Berry, M., Bannister, L. H., Standing, S. M., & Ruskell, G. 1995, "Neurology," in *Gray's Anatomy*, P. L. Williams et al., eds., Churchill Livingstone, London, pp. 1322-1361.

Birdwell, R. L., Ikeda, D. M., O'Shaughnessy, K. F., & Sickles, E. A. 2001, "Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection", *Radiology*, vol. 219, no. 1, pp. 192-202.

Brem, R. F. & Schoonjans, J. M. 2001, "Radiologist detection of microcalcifications with and without computer-aided detection: a comparative study", *Clinical Radiology*, vol. 56, no. 2, pp. 150-154.

Brennan, P. & Silman, A. 1992, "Statistical methods for assessing observer variability in clinical measures", *BMJ*, vol. 304, no. 6840, pp. 1491-1494.

Brogdon, B. G., Kelsey, C. A., & Moseley, R. D., Jr. 1983, "Factors affecting perception of pulmonary lesions", *Radiologic Clinics of North America*, vol. 21, no. 4, pp. 633-654.

Cascade, P. N., Kazerooni, E. A., Gross, B. H., Quint, L. E., Silver, T. M., Bowerman, R. A., Pernicano, P. G., & Gebremariam, A. 2001, "Evaluation of competence in the interpretation of chest radiographs", *Academic Radiology*, vol. 8, no. 4, pp. 315-321.

Chesters, M. S. & Simon, H. A. 1973, "Perception in chess", *Cognitive Psychology*, vol. 1, pp. 33-81.

Clarke, A. H. & Baartz, F.-J. Eye-tracking background and history. http://www.chronos-vision.de/products/content_products.html . 1998.
Ref Type: Electronic Citation

Colby, T. V. & Swensen, S. J. 1996, "Anatomic distribution and histopathologic patterns in diffuse lung disease: correlation with HRCT", *Journal of Thoracic Imaging*, vol. 11, no. 1, pp. 1-26.

Collins, J. 2001, "Radiology training: a program director's perspective", *AJR Am.J.Roentgenol.*, vol. 177, no. 5, pp. 1009-1010.

Doi, K. 1999, "Computer-aided diagnosis and its potential impact on diagnostic radiology," in *Computer-aided diagnosis in medical imaging. Proceedings of the first international workshop on computer-aided diagnosis, Chicago, USA*, K. Doi et al., eds., Elsevier Scienc B.V., Chicago, Ill, pp. 11-20.

Draper, S. W. A Brief Organized List of vision science numbers [www document]. Steve Draper . 10-11-1997.

Eber, C. D., Stark, P., & Bertozzi, P. 1993, "Bronchiolitis obliterans on high-resolution CT: a pattern of mosaic oligemia", *Journal of Computer Assisted Tomography*, vol. 17, no. 6, pp. 853-856.

Fitzgerald, R. 2001, "Error in radiology", *Clinical Radiology*, vol. 56, no. 12, pp. 938-946.

Foley, W. D., Jacobson, D. R., Taylor, A. J., Goodman, L. R., Stewart, E. T., Gurney, J. W., & Stroka, D. 1990, "Display of CT studies on a two-screen electronic workstation versus a film panel alternator: sensitivity and efficiency among radiologists", *Radiology*, vol. 174, no. 3 Pt 1, pp. 769-773.

Fromson, B. H. & Denison, D. M. 1988, "Quantitative features in the computed tomography of healthy lungs", *Thorax*, vol. 43, no. 2, pp. 120-126.

Ginsburg, A. P. 1988, "Current understanding of the human visual system and its application to perception of radiologic images", *Invest Radiol.*, vol. 23, no. 11, pp. 875-878.

Glenstrup, A. J. & Engell-Nielson, T. 1995, *Eye Controlled Media: Present and Future State,* BSc [www document], , Institute of computer science, University of Copenhagen.

Glenstrup, A. J. & Engell-Nielson, T. Present-day eye-gaze tracking techniques. Thesis [www document], Laboratory of Psychology, University of Copenhagen . 2002.

Goldin, J. G. 2002, "Quantitative CT of the lung", *Radiologic Clinics of North America*, vol. 40, no. 1, pp. 145-162.

Greene, R. E. 1992, "Missed lung nodules: lost opportunities for cancer cure", *Radiology*, vol. 182, no. 1, pp. 8-9.

Gregory, R. L. 1998, *Eye and brain: The psychology of seeing.*, 5th edn, Oxford University Press, Oxford.

Griffin, C. B. & Primack, S. L. 2001, "High-resolution CT: normal anatomy, techniques, and pitfalls", *Radiologic Clinics of North America*, vol. 39, no. 6, pp. 1073-90, v.

Gückel, C., Wells, A. U., Taylor, D. A., Chabat, F., & Hansell, D. M. 1999, "Mechanism of mosaic attenuation of the lungs on computed tomography in induced bronchospasm", *J.Appl.Physiol*, vol. 86, no. 2, pp. 701-708.

Gunderman, R., Williamson, K., Fraley, R., & Steele, J. 2001, "Expertise: implications for radiological education", *Academic Radiology*, vol. 8, no. 12, pp. 1252-1256.

Gurney, J. W. 1991, "Cross-sectional physiology of the lung", *Radiology*, vol. 178, no. 1, pp. 1-10.

Gurney, J. W. 1998, "Pathophysiology of obstructive airways disease", *Radiologic Clinics of North America*, vol. 36, no. 1, pp. 15-27.

Guyton, A. C. & Hall, J. E. 1996a, "The eye: I. Optics of vision," in *Textbook of medical physiology*, 9th edn, A. C. Guyton & J. E. Hall, eds., W.B. Saunders Company, Philadelphia, USA, pp. 623-634.

Guyton, A. C. & Hall, J. E. 1996b, "The eye: II Receptor and neural function of the retina," in *Textbook of medical physiology*, 9th edn, A. C. Guyton & J. E. Hall, eds., W.B. Saunders Company, Philadelphia, USA, pp. 637-648.

Guyton, A. C. & Hall, J. E. 1996c, "The eye: III. Central neurophysiology of vision," in *Textbook of medical physiology*, 9th edn, A. C. Guyton & J. E. Hall, eds., W.B. Saunders Company, Philadelphia, USA, pp. 651-672.

Hall, R. W. 1983, "Image processing algorithms for eye movement monitoring", *Computers and Biomedical Research*, vol. 16, no. 6, pp. 563-579.

Hansell, D. M. 1999, "HRCT-pathologic correlation in small airways disease", *European Radiology*, vol. 9, no. (Suppl.2), p. S89-S105.

Hansen, J. E., Ampaya, E. P., Bryant, G. H., & Navin, J. J. 1975, "Branching pattern of airways and air spaces of a single human terminal bronchiole", *J.Appl.Physiol*, vol. 38, no. 6, pp. 983-989.

He, D. M., Zhang, Y., & Zhang, Z. K. 2000, "Computer-assisted quantitative measurements by three-dimensional images on zygomatic fracture deformities", *Chin J.Dent.Res.*, vol. 3, no. 4, pp. 26-34.

Hoffman, E. A., Sinak, L. J., Robb, R. A., & Ritman, E. L. 1983, "Noninvasive quantitative imaging of shape and volume of lungs", *J.Appl.Physiol*, vol. 54, no. 5, pp. 1414-1421.

Hogg, J. C., Macklem, P. T., & Thurlbeck, W. M. 1968, "Site and nature of airway obstruction in chronic obstructive lung disease", *New England Journal of Medicine*, vol. 278, no. 25, pp. 1355-1360.

Hu, C. H., Kundel, H. L., Nodine, C. F., Krupinski, E. A., & Toto, L. C. 1994, "Searching for bone fractures: a comparison with pulmonary nodule search", *Academic Radiology*, vol. 1, no. 1, pp. 25-32.

Ikeda, M., Shigaki, T., & Yamauchi, K. 2000, "A signal-detection experiment measuring the effect of computer-aided detection on radiologists' performance", *Medical Decision Making*, vol. 20, no. 3, pp. 343-351.

Ishida, T., Katsuragawa, S., Kobayashi, T., MacMahon, H., & Doi, K. 1997, "Computerized analysis of interstitial disease in chest radiographs: improvement of geometric-pattern feature analysis", *Medical Physics*, vol. 24, no. 6, pp. 915-924.

Ker, M., Seeley, G. W., Stempski, M. O., & Patton, D. 1988, "A protocol for verifying truth of diagnosis. Preliminary report", *Invest Radiol.*, vol. 23, no. 6, pp. 485-487.

Kido, S., Kuriyama, K., Hosomi, N., Inoue, E., Kuroda, C., & Horai, T. 2000, "Low-cost soft-copy display accuracy in the detection of pulmonary nodules by single-exposure dual-energy subtraction: comparison with hard-copy viewing", *Journal of Digital Imaging*, vol. 13, no. 1, pp. 33-37.

Kido, S., Tamura, S., Nakamura, H., & Kuroda, C. 1999, "Interstitial lung diseases: evaluation of the performance of a computerized analysis system versus observers", *Computerized Medical Imaging and Graphics*, vol. 23, no. 2, pp. 103-110.

Kirkwood, B. R. 1988a, "Comparison of several means - Analysis of variance," in *Essentials of medical statistics*, Blackwell Scientific Publications, Oxford, pp. 46-56.

Kirkwood, B. R. 1988b, "Multiple regression," in *Essentials of medical statistics*, Blackwell Scientific Publications, Oxford, pp. 65-72.

Kirkwood, B. R. 1988c, "Non-parametric methods," in *Essentials of medical statistics*, Blackwell Scientific Publications, Oxford, pp. 147-152.

Ko, J. P. & Betke, M. 2001, "Chest CT: automated nodule detection and assessment of change over time- -preliminary experience", *Radiology*, vol. 218, no. 1, pp. 267-273.

Kobayashi, T., Xu, X. W., MacMahon, H., Metz, C. E., & Doi, K. 1996, "Effect of a computer-aided diagnosis scheme on radiologists' performance in detection of lung nodules on radiographs", *Radiology*, vol. 199, no. 3, pp. 843-848.

Krupinski, E. A. 1996, "Visual scanning patterns of radiologists searching mammograms", *Academic Radiology*, vol. 3, no. 2, pp. 137-144.

Krupinski, E. A., Maloney, K., Bessen, S. C., Capp, M. P., Graham, K., Hunt, R., Lund, P., Ovitt, T., & Standen, J. R. 1994, "Receiver operating characteristic evaluation of computer display of adult portable chest radiographs", *Invest Radiol.*, vol. 29, no. 2, pp. 141-146.

Krupinski, E. A. & Nishikawa, R. M. 1997, "Comparison of eye position versus computer identified microcalcification clusters on mammograms", *Medical Physics*, vol. 24, no. 1, pp. 17-23.

Krupinski, E. A., Nodine, C. F., & Kundel, H. L. 1993a, "A perceptually based method for enhancing pulmonary nodule recognition", *Invest Radiol.*, vol. 28, no. 4, pp. 289-294.

Krupinski, E. A., Nodine, C. F., & Kundel, H. L. 1993b, "Perceptual enhancement of tumor targets in chest x-ray images", *Perception and Psychophysics*, vol. 53, no. 5, pp. 519-526.

Krupinski, E. A. & Roehrig, H. 2000, "The influence of a perceptually linearized display on observer performance and visual search", *Academic Radiology*, vol. 7, no. 1, pp. 8-13.

Kullback, S. & Leibler, R. A. 1951, "On information and sufficiency. 22 (1951), 76--86.", *Annals of Mathematical Statistics* , vol. 22, pp. 76-86.

Kundel, H. L. 1975, "Peripheral vision, structured noise and film reader error", *Radiology*, vol. 114, no. 2, pp. 269-273.

Kundel, H. L. & Nodine, C. F. 1975, "Interpreting chest radiographs without visual search", *Radiology*, vol. 116, no. 3, pp. 527-532.

Kundel, H. L., Nodine, C. F., & Carmody, D. 1978, "Visual scanning, pattern recognition and decision-making in pulmonary nodule detection", *Invest Radiol.*, vol. 13, no. 3, pp. 175-181.

Kundel, H. L., Nodine, C. F., & Krupinski, E. A. 1989, "Searching for lung nodules. Visual dwell indicates locations of false-positive and false-negative decisions", *Invest Radiol.*, vol. 24, no. 6, pp. 472-478.

Kundel, H. L., Nodine, C. F., & Krupinski, E. A. 1990, "Computer-displayed eye position as a visual aid to pulmonary nodule interpretation", *Invest Radiol.*, vol. 25, no. 8, pp. 890-896.

Kundel, H. L., Nodine, C. F., & Toto, L. 1991, "Searching for lung nodules. The guidance of visual scanning", *Invest Radiol.* , vol. 26, no. 9, pp. 777-781.

Lee, F. J. & Andriole, K. P. 2001, "Does learning a complex task have to be complex?: A study in learning decomposition", *Cognitive Psychology*, vol. 42, no. 3, pp. 267-316.

Leung, A. N., Staples, C. A., & Müller, N. L. 1991, "Chronic diffuse infiltrative lung disease: comparison of diagnostic accuracy of high-resolution and conventional CT", *AJR Am.J.Roentgenol.*, vol. 157, no. 4, pp. 693-696.

Lieberman, G., Abramson, R., Volkan, K., & McArdle, P. J. 2002, "Tutor versus computer: a prospective comparison of interactive tutorial and computer-assisted instruction in radiology education", *Academic Radiology*, vol. 9, no. 1, pp. 40-49.

Lynch, D. A., Newell, J. D., Tschomper, B. A., Cink, T. M., Newman, L. S., & Bethel, R. 1993, "Uncomplicated asthma in adults: comparison of CT appearance of the lungs in asthmatic and healthy subjects", *Radiology*, vol. 188, no. 3, pp. 829-833.

MacMahon, H. 2000, "Improvement in detection of pulmonary nodules: digital image processing and computer-aided diagnosis", *Radiographics*, vol. 20, no. 4, pp. 1169-1177.

MacMahon, H., Engelmann, R., Behlen, F. M., Hoffmann, K. R., Ishida, T., Roe, C., Metz, C. E., & Doi, K. 1999, "Computer-aided diagnosis of pulmonary nodules: results of a large-scale observer test", *Radiology*, vol. 213, no. 3, pp. 723-726.

Maioli, C., Benaglio, I., Siri, S., Sosta, K., & Cappa, S. 2001, "The integration of parallel and serial processing mechanisms in visual search: evidence from eye movement recording", *European Journal of Neuroscience*, vol. 13, no. 2, pp. 364-372.

Manning, D., Leach, J., & Bunting, S. 2000, "A comparison of expert and novice performance in the detection of simulated pulmonary nodules", *Radiography*, vol. 6, no. 2, pp. 111-116.

Mathie, A. G. & Strickland, N. H. 1997, "Interpretation of CT scans with PACS image display in stack mode", *Radiology*, vol. 203, no. 1, pp. 207-209.

Mayo, J. R., Jackson, S. A., & Müller, N. L. 1993, "High-resolution CT of the chest: radiation dose", *AJR Am.J.Roentgenol.*, vol. 160, no. 3, pp. 479-481.

Milburn, A. An organisation with a memory. Department of Health . 2002.

Nakamura, K., Yoshida, H., Engelmann, R., MacMahon, H., Katsuragawa, S., Ishida, T., Ashizawa, K., & Doi, K. 2000, "Computerized analysis of the likelihood of malignancy in solitary pulmonary nodules with use of artificial neural networks", *Radiology*, vol. 214, no. 3, pp. 823-830.

Niimi, R., Shimamoto, K., Sawaki, A., Ishigaki, T., Takahashi, Y., Sugiyama, N., & Nishihara, E. 1997, "Eye-tracking device comparisons of three methods of magnetic resonance image series displays", *Journal of Digital Imaging*, vol. 10, no. 4, pp. 147-151.

Nishimura, K., Izumi, T., Kitaichi, M., Nagai, S., & Itoh, H. 1993, "The diagnostic accuracy of high-resolution computed tomography in diffuse infiltrative lung diseases", *Chest*, vol. 104, no. 4, pp. 1149-1155.

Nodine, C. F. & Krupinski, E. A. 1998, "Perceptual skill, radiology expertise, and visual test performance with NINA and WALDO", *Academic Radiology*, vol. 5, no. 9, pp. 603-612.

Nodine, C. F., Krupinski, E. A., Kundel, H. L., Toto, L., & Herman, G. T. 1992, "Satisfaction of search (SOS)", *Invest Radiol.*, vol. 27, no. 7, pp. 571-573.

Osterberg, G. 1935, "Topography of the layer of rods and cones in the human retina. Acta Ophthal., suppl. **6**, 1-103.", *Acta Ophthalmologica*, vol. suppl 6, pp. 1-103.

Padley, S. P., Hansell, D. M., Flower, C. D., & Jennings, P. 1991, "Comparative accuracy of high resolution computed tomography and chest radiography in the diagnosis of chronic diffuse infiltrative lung disease", *Clinical Radiology*, vol. 44, no. 4, pp. 222-226.

Panchangam, A., Sastry, K. V., Rao, D. V., DeCristofano, B. S., Kimball, B. R., & Nakashima, M. 2001, "Processing of medical images using real-time optical Fourier processing", *Medical Physics*, vol. 28, no. 1, pp. 22-27.

Patel, G. A. & Sathian, K. 2000, "Visual search: bottom-up or top-down?", *Front Biosci.*, vol. 5, p. D169-D193.

Penedo, M. G., Carreira, M. J., Mosquera, A., & Cabello, D. 1998, "Computer-aided diagnosis: a neural-network-based approach to lung nodule detection", *IEEE Transactions on Medical Imaging*, vol. 17, no. 6, pp. 872-880.

Piraino, D. W., Richmond, B. J., Uetani, M., Luetkehaus, T., Rockey, D., Belhobek, G., Armistead, J., & Jones, F. 1989, "Problems in applying expert system technology to radiographic image interpretation", *Journal of Digital Imaging*, vol. 2, no. 1, pp. 21-26.

Ponsoda, V., Scott, D., & Findlay, J. M. 1995, "A probability vector and transition matrix analysis of eye movements during visual search", *Acta Psychol.(Amst)*, vol. 88, no. 2, pp. 167-185.

Rajnics, P., Pomero, V., Templier, A., Lavaste, F., & Illes, T. 2001, "Computer-assisted assessment of spinal sagittal plane radiographs", *Journal of Spinal Disorders*, vol. 14, no. 2, pp. 135-142.

Reeves, A. P. & Kostis, W. J. 2000a, "Computer-aided diagnosis for lung cancer", *Radiologic Clinics of North America*, vol. 38, no. 3, pp. 497-509.

Reeves, A. P. & Kostis, W. J. 2000b, "Computer-aided diagnosis of small pulmonary nodules", *Semin.Ultrasound CT MR*, vol. 21, no. 2, pp. 116-128.

Reiner, B. I., Siegel, E. L., Hooper, F. J., Pomerantz, S., Dahlke, A., & Rallis, D. 2001, "Radiologists' productivity in the interpretation of CT scans: a comparison of PACS with conventional film", *AJR Am.J.Roentgenol.*, vol. 176, no. 4, pp. 861-864.

Remy-Jardin, M., Giraud, F., Remy, J., Copin, M. C., Gosselin, B., & Duhamel, A. 1993, "Importance of ground-glass attenuation in chronic diffuse infiltrative lung disease: pathologic-CT correlation", *Radiology*, vol. 189, no. 3, pp. 693-698.

Richmond, A. 2002. The Nature of Science. Kosmoi: worlds of science, nature and technology [www document].

Robinson, P. J. 1997, "Radiology's Achilles' heel: error and variation in the interpretation of the Rontgen image", *British Journal of Radiology*, vol. 70, no. 839, pp. 1085-1098.

Rybak, I. A., Gusakova, V. I., Golovan, A. V., Podladchikova, L. N., & Shevtsova, N. A. 1998, "A model of attention-guided visual perception and recognition", *Vision Research*, vol. 38, no. 15-16, pp. 2387-2400.

Shott, S. 1990, "Nonparametric statistics," in *Statistics for health professionals*, W.B.Saunders Company, London, pp. 229-268.

Stern, E. J. & Frank, M. S. 1994, "Small-airway diseases of the lungs: findings at expiratory CT", *AJR Am.J.Roentgenol.*, vol. 163, no. 1, pp. 37-41.

Stern, E. J., Müller, N. L., Swensen, S. J., & Hartman, T. E. 1995a, "CT mosaic pattern of lung attenuation: etiologies and terminology", *Journal of Thoracic Imaging*, vol. 10, no. 4, pp. 294-297.

Stern, E. J., Swensen, S. J., Hartman, T. E., & Frank, M. S. 1995b, "CT mosaic pattern of lung attenuation: distinguishing different causes", *AJR Am.J.Roentgenol.*, vol. 165, no. 4, pp. 813-816.

Stern, E. J., Webb, W. R., & Gamsu, G. 1994, "Dynamic quantitative computed tomography. A predictor of pulmonary function in obstructive lung diseases", *Invest Radiol.*, vol. 29, no. 5, pp. 564-569.

Strickland, N. H. & Allison, D. J. 1995, "Default display arrangements of images on PACS monitors", *British Journal of Radiology*, vol. 68, no. 807, pp. 252-260.

Stulbarg, M. S. & Frank, J. A. 1998, "Obstructive pulmonary disease. The clinician's perspective", *Radiologic Clinics of North America*, vol. 36, no. 1, pp. 1-13.

Swensen, S. J., Aughenbaugh, G. L., Douglas, W. W., & Myers, J. L. 1992, "High-resolution CT of the lungs: findings in various pulmonary diseases", *AJR Am.J.Roentgenol.*, vol. 158, no. 5, pp. 971-979.

Takasugi, J. E. & Godwin, J. D. 1998, "Radiology of chronic obstructive pulmonary disease", *Radiologic Clinics of North America*, vol. 36, no. 1, pp. 29-55.

Thiran, J. P. & Macq, B. 1996, "Morphological feature extraction for the classification of digital images of cancerous tissues", *IEEE Trans.Biomed.Eng*, vol. 43, no. 10, pp. 1011-1020.

Tuddenham, W. J. 1969, "Roentgen image perception--a personal survey of the problem", *Radiologic Clinics of North America*, vol. 7, no. 3, pp. 499-501.

van der Heyden, J. E., Inkpen, K. M., Atkins, M. S., & Carpendale, M. S. 2001, "Exploring presentation methods for tomographic medical image viewing", *Artificial Intelligence in Medicine*, vol. 22, no. 2, pp. 89-109.

Vittitoe, N. F., Baker, J. A., & Floyd, C. E., Jr. 1997, "Fractal texture analysis in computer-aided diagnosis of solitary pulmonary nodules", *Academic Radiology*, vol. 4, no. 2, pp. 96-101.

Warren Burhenne, L. J., Wood, S. A., D'Orsi, C. J., Feig, S. A., Kopans, D. B., O'Shaughnessy, K. F., Sickles, E. A., Tabar, L., Vyborny, C. J., & Castellino, R. A. 2000, "Potential contribution of computer-aided detection to the sensitivity of screening mammography", *Radiology*, vol. 215, no. 2, pp. 554-562.

Webb, W. R., Müller, N. L., & Naidich, D. P. 2001a, "Diseases characterised primarily by linear and reticular opacities," in *High-resolution CT of the lung*, 3rd edn, Lippincott Williams & Wilkins, London, pp. 193-258.

Webb, W. R., Müller, N. L., & Naidich, D. P. 2001b, "Diseases characterised primarily by nodular and reticulonodular opacities.," in *High-resolution CT of the lung*, 3rd edn, Lippincott Williams & Wilkins, London, pp. 259-354.

Webb, W. R., Müller, N. L., & Naidich, D. P. 2001c, "High resolution computed tomography findings of lung disease," in *High-resolution CT of the lung*, 3rd edn, Lippincott Williams & Wilkins, London, pp. 71-192.

Webb, W. R., Müller, N. L., & Naidich, D. P. 2001d, "Technical aspectsof high-resolution computed tomography," in *High-resolution CT of the lung*, 3rd edn, Lippincott Williams & Wilkins, London, pp. 1-49.

Webb, W. R., Stern, E. J., Kanth, N., & Gamsu, G. 1993, "Dynamic pulmonary CT: findings in healthy adult men", *Radiology*, vol. 186, no. 1, pp. 117-124.

Weibel, E. R. 1979, "Fleischner Lecture. Looking into the lung: what can it tell us?", *AJR Am.J.Roentgenol.*, vol. 133, no. 6, pp. 1021-1031.

Westcott, J. L. & Cole, S. R. 1986, "Traction bronchiectasis in end-stage pulmonary fibrosis", *Radiology*, vol. 161, no. 3, pp. 665-669.

White, K., Berbaum, K., & Smith, W. L. 1994, "The role of previous radiographs and reports in the interpretation of current radiographs", *Invest Radiol.*, vol. 29, no. 3, pp. 263-265.

Wigderowitz, C. A., Sadaiyappan, V., McGurty, D. M., & Rowley, D. I. 2001, "Computerized analysis of radiographs in Colles' fractures", *Clinical Orthopaedics and Related Research* no. 387, pp. 217-224.

Wildberger, J. E., Niethammer, M. U., Klotz, E., Schaller, S., Wein, B. B., & Gunther, R. W. 2001, "Multi-slice CT for visualization of pulmonary embolism using perfusion weighted color maps", *Rofo Fortschr.Geb.Rontgenstr.Neuen Bildgeb.Verfahr.*, vol. 173, no. 4, pp. 289-294.

Worthy, S. A. & Müller, N. L. 1998, "Small airway diseases", *Radiologic Clinics of North America*, vol. 36, no. 1, pp. 163-173.

Worthy, S. A., Müller, N. L., Hartman, T. E., Swensen, S. J., Padley, S. P., & Hansell, D. M. 1997, "Mosaic attenuation pattern on thin-section CT scans of the lung: differentiation among infiltrative lung, airway, and vascular diseases as a cause", *Radiology*, vol. 205, no. 2, pp. 465-470.

Wright, J. L., Cagle, P., Churg, A., Colby, T. V., & Myers, J. 1992, "Diseases of the small airways", *American Review of Respiratory Disease*, vol. 146, no. 1, pp. 240-262.

Wu, Y. C., Doi, K., & Giger, M. L. 1995, "Detection of lung nodules in digital chest radiographs using artificial neural networks: a pilot study", *Journal of Digital Imaging*, vol. 8, no. 2, pp. 88-94.

Yoshida, H., Doi, K., Nishikawa, R. M., Giger, M. L., & Schmidt, R. A. 1996, "An improved computer-assisted diagnostic scheme using wavelet transform for detecting clustered microcalcifications in digital mammograms", *Academic Radiology*, vol. 3, no. 8, pp. 621-627.