Reverse Hierarchy Theory and Medical Image Perception.

T. Donovan*, D.J. Manning
School of Medical Imaging Sciences, University of Cumbria, Lancaster, LA1 3JD, UK.

ABSTRACT

We are unsure about what information is extracted from an image to allow a decision about pathology to be made. Our knowledge of the interplay between top down processing or bottom up, local or global perception, perceptual or cognitive processes is uncertain. However recent research has emphasised the importance of the global or holistic look in medical image perception in which recognition of abnormalities precedes search. Reverse Hierarchy Theory [1] is a useful general theory that helps to explain this. It also enables us to understand what information is extracted from an image and how this relates to expertise. Essentially the theory states that perceptual learning begins at high levels areas and progresses down to lower level areas when better signal to noise is needed. So perceptual learning, defined as an improvement in sensory abilities after training, stems from a gradual top down guided increase in usability of first high then lower level task relevant information. Evaluation of the scan paths of groups of observers with different levels of expertise when undertaking a lung nodule perception task seems to be consistent with the theory. Experts' perception is generally immediate and holistic suggesting high level representations whereas those with an intermediate level of expertise tend to be more variable in their scan paths. Interestingly naïve observers have eye tracking metrics that are more similar to experts suggesting they take a common sense approach using perceptual skills we all have as they lack experience in being able to access the low level information from the chest radiograph.
Keywords: Reverse Hierarchy Theory, expertise, eye-tracking, image perception

1. INTRODUCTION

The importance of the initial global look or gist in understanding an image or scene is well recognised and has recently been emphasised by Kundel et al. [2, 3], who has used the term 'holistic component of image perception' in research looking at the global recognition process in the perception and identification of cancers in mammography. This process precedes the search to find mode and is more developed in experienced and proficient radiologists. This leads to the question as to what information is extracted from an image in this initial global look. The literature on the possible perceptual and neural mechanisms in recognising the gist of a visual scene indicates that scenes are processed very distributively through the visual pathways, likely using the many feedback loops in the visual system, even before the first saccade [4]. The visual hierarchy is a hierarchy of cortical areas in which lower level cells have a small receptive field and process information from the retinal field such as intensity, spatial frequency and orientation whereas at higher levels cells have a larger receptive field and representations generalise over the lower level information and respond to objects and shapes. Ahissar and Hochstein [1] propose that visual information initially travels rapidly in a fast feedforward direction and that perception begins at higher levels progressing down to lower levels via feedback connections when better signal to noise is needed. So perceptual learning, defined as an improvement in sensory abilities after training, such as the ability to detect radiological abnormalities, occurs first in higher level areas where there is greater generalisation and transfer across different tasks, whereas low level learning would be specific to the features of the stimulus. The global look or holistic perception is evidence to support this theory, and has been demonstrated by Juan & Walsh [5] in experiments using transcranial magnetic stimulation (TMS) looking at the role of V1 in visual detection. The results were consistent with reverse hierarchy theory in that visual perception follows a global to local path with neurons with a large receptive field carrying out an initial broad analysis followed by the acquisition of more detailed information from the earlier visual areas.

Although the theory can neither be proved nor disproved by data collected in an applied task such as medical image perception where the numbers of variables are so large, interpretation of eye tracking data does seem to be consistent with the theory. This paper only addresses the perceptual component of medical image interpretation and is an attempt to explain the similarity between the eye tracking metrics and scan path data of the expert and naïve observer compared to those learning the task.

*tim.donovan@cumbria.ac.uk; phone: 0044 1524384667

Medical Imaging 2009: Image Perception, Observer Performance, and Technology Assessment,
edited by Berkman Sahiner, David J. Manning, Proc. of SPIE Vol. 7263, 72631M
 · doi: 10.1117/12.812956

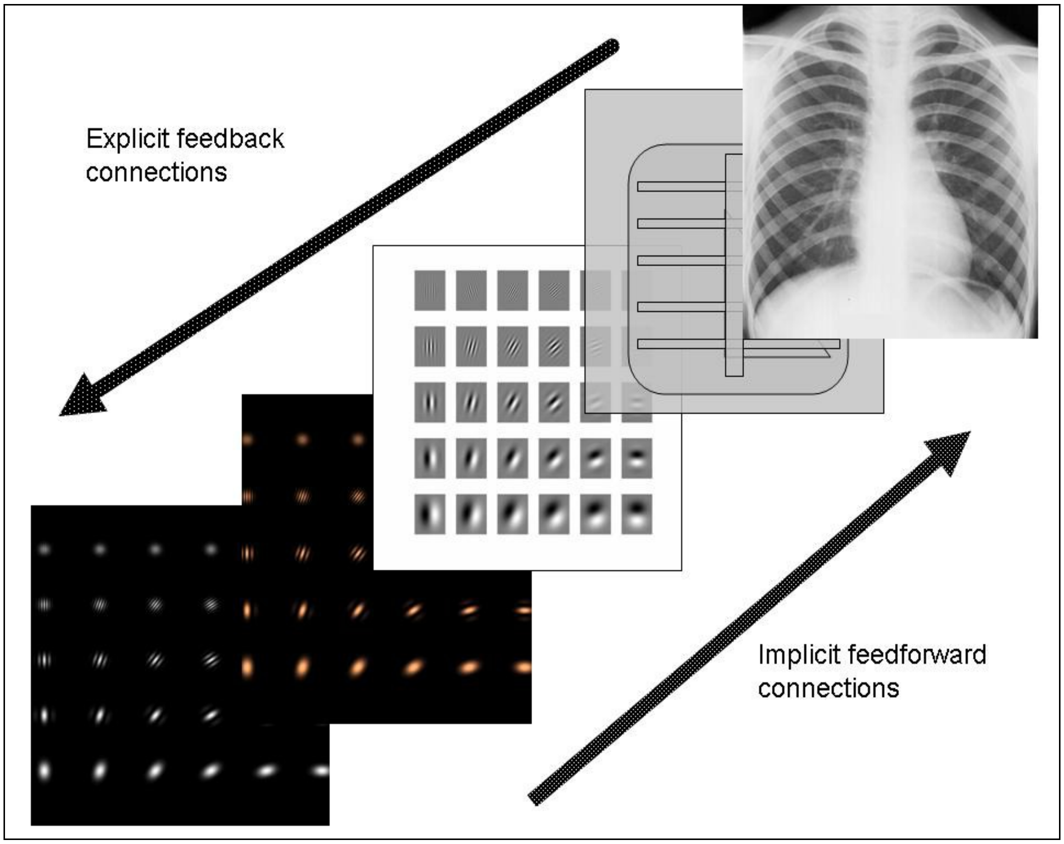

Figure 1. The initial gist of a chest radiograph for example, will depend on high level object and category representations built by implicit hierarchical processing. Later vision with scrutiny is a return to simple feature details available at low levels such as spatial frequency, orientation, contrast etc. Low level learning will be specific to these features being a low level modification by guided return down the reverse hierarchy.

2. METHODS

2.1 Observer Groups

Four groups of observer (n = 10) in each group with different levels of expertise took part in an eye tracking experiment in which they were required to search for lung nodules from a test bank of 30 chest images.

The expertise groups were naïve observers, undergraduate radiography students with 12 weeks clinical experience (Level 1), undergraduate radiography students with 28 weeks clinical experience (Level 2), and experts (8 radiologists, 2 reporting radiographers).

2.2 Test Bank

30 images were selected from a previously compiled test bank consisting of natural nodules which were histopathologically proven and simulated nodules. 15 of the images contained nodules, some with multiple nodules. Nodules were defined as discrete opacities in the lung field or mediastinum measuring between 5-30mm in diameter. Of the 28 nodules in total, 24 were histopathologically proven and the remaining 4 were simulated. Some of the images without nodules had other pathologies such as effusions.

2.3 Eye-tracking

Eye-tracking data was acquired with a Tobii x50 system (Tobii Eye Tracker and ClearView analysis software, Tobii Technology AB, 2005). This is a 50Hz scanner, i.e. it samples every 20ms.

To obtain the data to determine the eye tracking metrics rectangular regions of interest were defined around each nodule, the size of which was defined to be 1.5 times the diameter of the lesion. A fixation for the purposes of this experiment was defined as 100ms and 50 pixels, i.e. if a fixation point was more than 50 pixels away from the previous one it was classified as a separate fixation.

Eye tracking metrics calculated were as follows:

- Time to first hit. This is often used as a metric for expertise. In this experiment it is the time taken from when the image appears and a fixation is made in the ROI containing the nodule.
- Average fixation duration. This is the average length of all fixations during all recordings on the respective ROIs.
- Fixation count. This is often used to indicate efficiency of search [6] with the number of fixations overall negatively correlated with search efficiency. In this experiment it relates to the mean number of fixations in the nodule ROI on each image.
- Dwell time. This refers to the cumulative dwell time of fixations in each ROI. As a metric it is useful as it reflects the amount of visual processing occurring.

Scan paths were also qualitatively assessed.

2.4 Procedure

Ethical approval for the study was obtained. Observers were instructed to look for lung nodules and disregard any other radiological findings. They were told images may have no nodules, one nodule or multiple nodules up to a maximum of five on any one image. All observers were shown two chest images as examples, a normal one and one with multiple nodules prior to the start of the study. A calibration was carried out. The first image was presented, free search was allowed, observers were instructed to click on each nodule and indicate their confidence in their decision (data not reported in this paper). The observer terminated a search by pressing the space bar.

3. RESULTS

ANOVA was used to compares the means of the different expertise groups. Time to first hit demonstrated no significant difference between the groups [$F(3,220) = 1.883$, p=0.133]. The fixation count showed a significant difference between groups [$F(3,220) = 5.408$, $p = 0.001$], experts and naïve observers were not significantly different (p=0.93). Naïve observers had fewer fixations than Level 2 (p=0.017) and Level 1(p=0.007). Experts also had fewer fixations than Level 2 (p= 0.08) and Level 1 (p=0.04). Dwell time also showed a significant difference between groups [$F(3,220) = 5.914$, p=0.001], with experts and naïve not significantly different (p=0.65), with experts having a shorter dwell time than Level 1 (p=0.025) and Level 2 (p=0.001) and the naïve group having a shorter dwell time than Level 2 (p=0.037) and Level 1 (p=0.338).

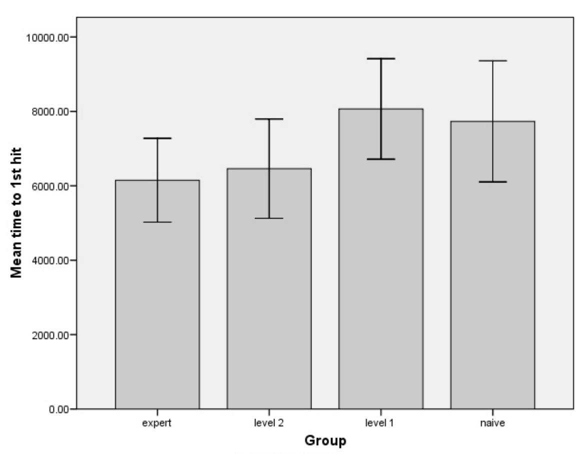

Figure 2. Mean time to 1st hit (msec) of the nodule in the ROI. There were no significant differences between the groups.

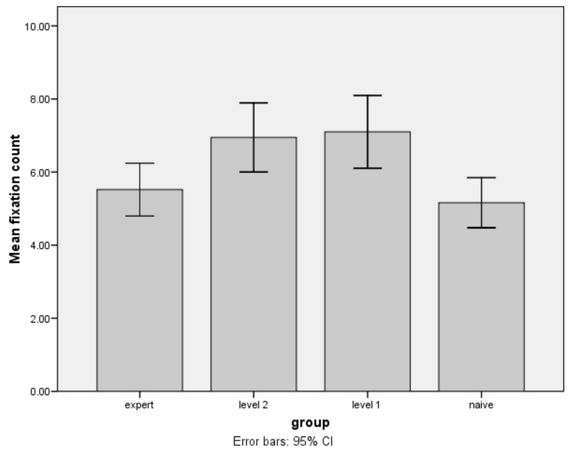

Figure 3. Mean number of fixations in the ROI. Experts and naïve observers were not significantly different from each other but they were different from Level 2 and Level 1.

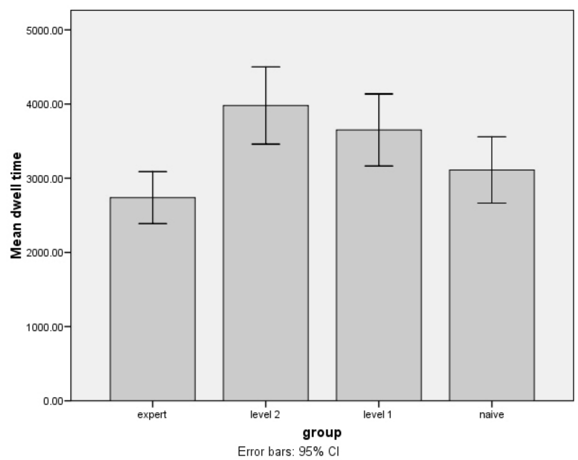

Figure 4. Dwell time (gaze time) in the ROI. Expert and naïve observers were not significantly different from each other, but they were different from Level 1 and Level 2

Qualitative assessment of scan path data.

Examples of typical scan paths, with an explanation are below: - The scan paths are representative of the way different expertise groups look at the same chest radiograph. The nodule in this example is in the left lung in the upper lobe and is at the centre of the box which represents the ROI (not visible to the observers). The cross indicates where the observer has used the mouse to click on the image where they have perceived a nodule. The different size of the fixation circles relates to the length of the fixation, with the larger circles being a longer fixation.

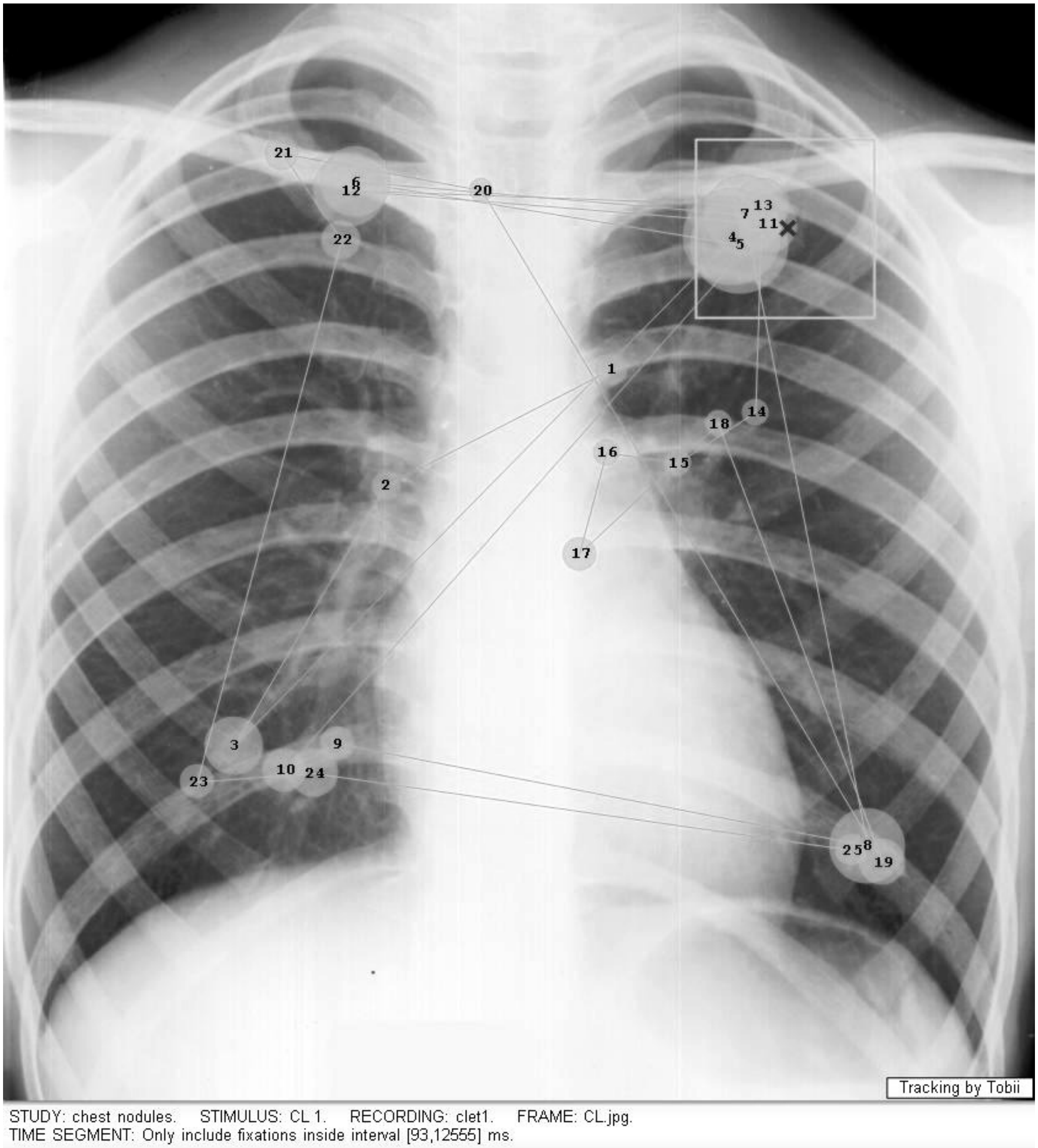

Figure 5. This is the eye movement data from an expert observer. The nodule is fixated on the 4th fixation. Search is efficient with large saccades and relatively large parts of the image are not foveated.

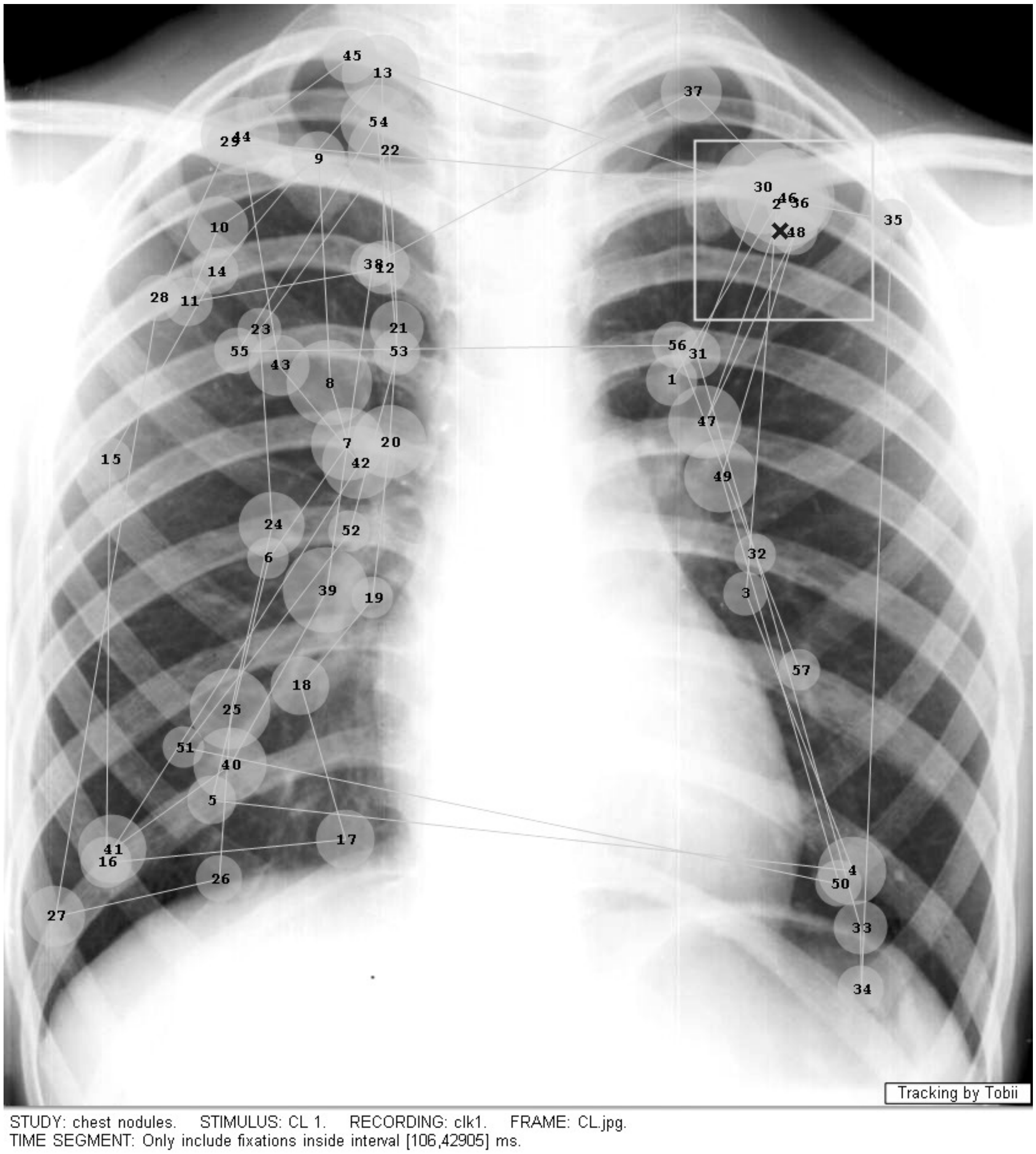

Figure 6. This is the eye movement data from a Level 2 observer also looking at the same image. Compared to the naïve and level 1 observer there are many more fixations and a less efficient scan path, although more effective then level 1 and the naïve as the one nodule in the image is correctly identified as such. The strategy used as demonstrated by the fixations indicates that they are probably accessing the low level representations such as contrast as a way of improving signal-to-noise consistent with reverse hierarchy theory.

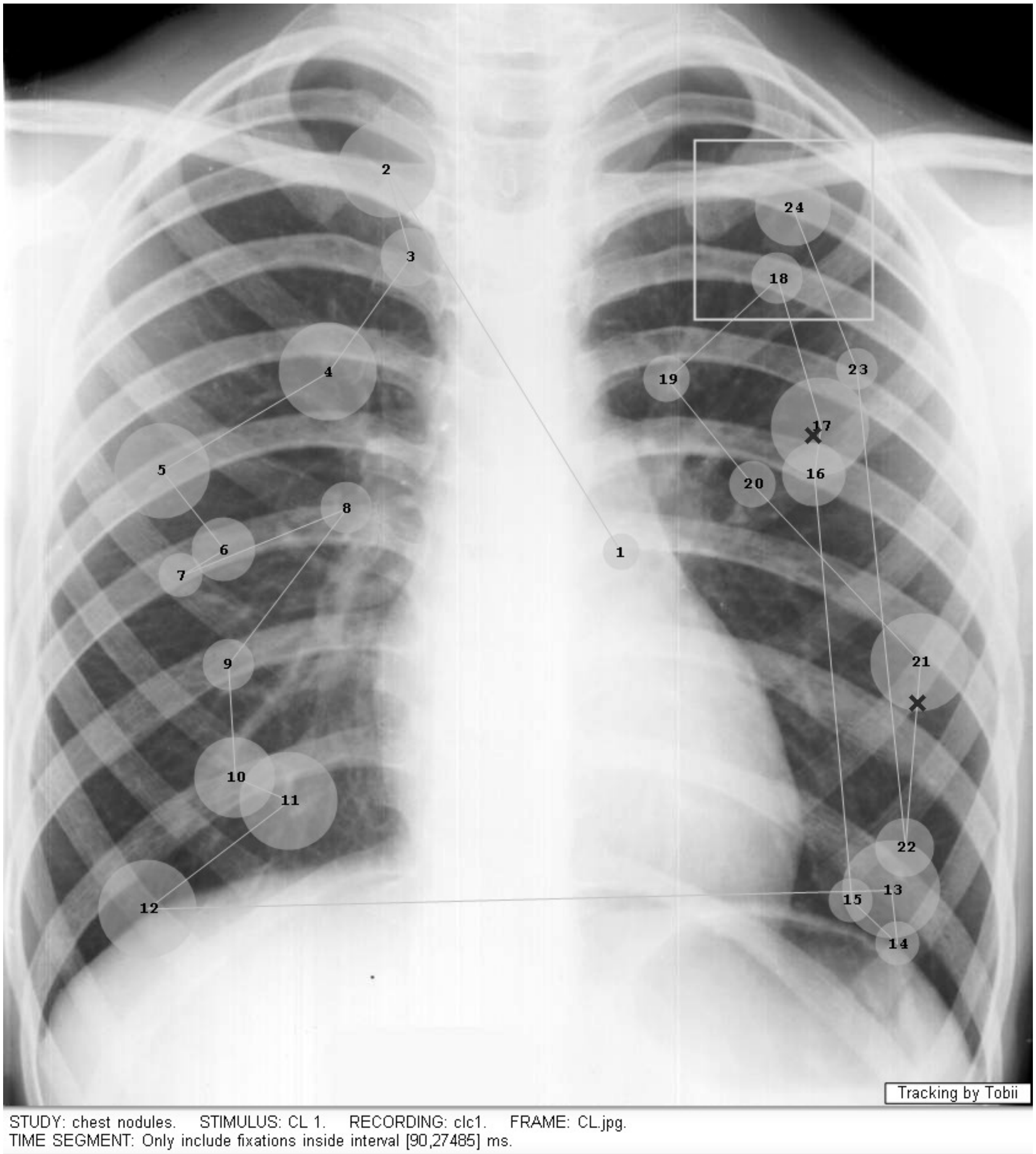

Figure 7. This is the eye movement data from a Level 1 observer. After a relatively limited clinical experience the knowledge gained about the information that can be extracted from an image has resulted in a scan path that more closely resembles that of an expert, and also probably indicates top-down guidance. Although the nodule is fixated (fixation 24) it is not identified as a nodule and two false positive sites are indicated (fixation 17 and 21) where the pulmonary shadows look circular in appearance.

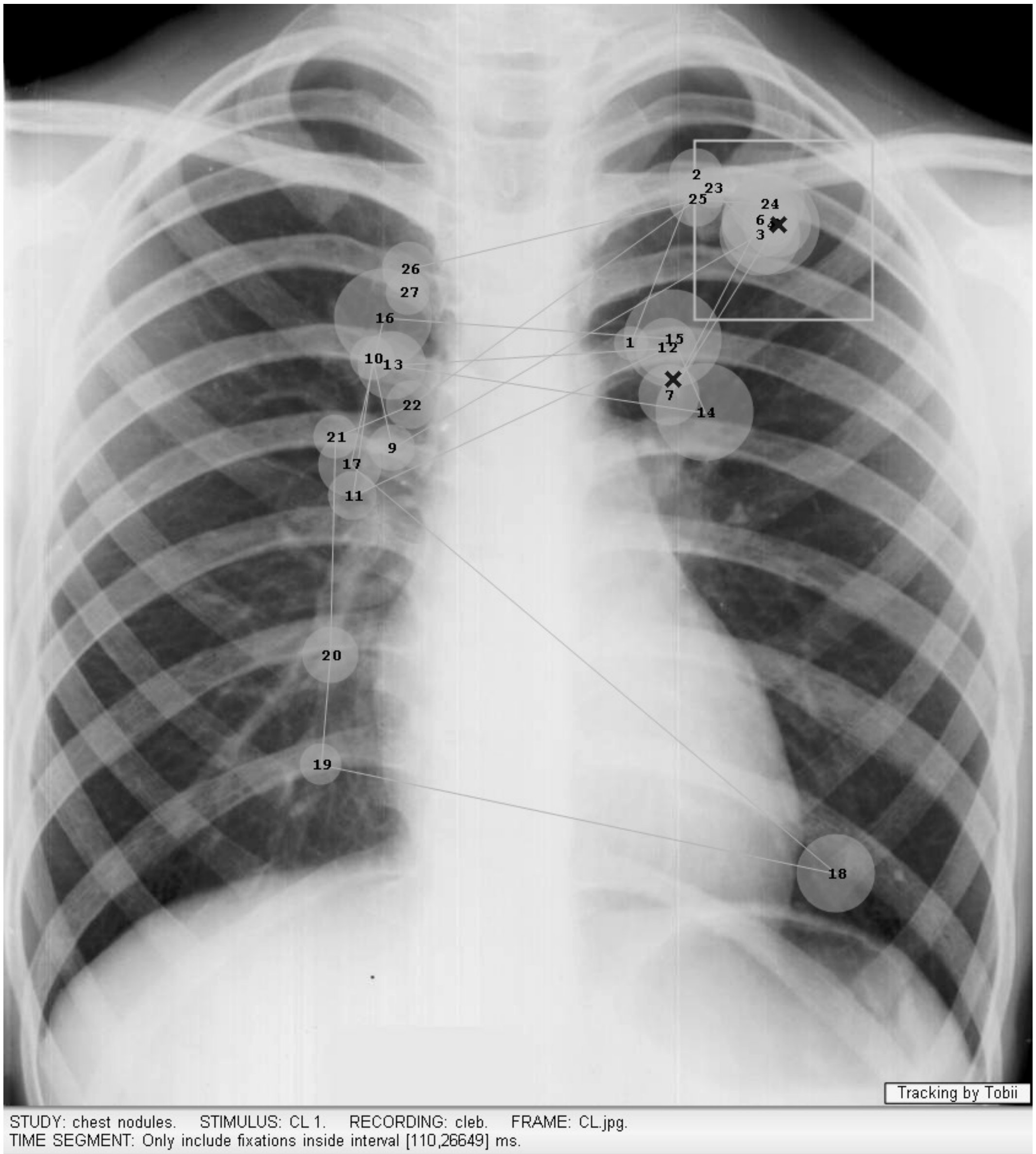

Figure 8. This is the eye movement data from a naïve observer. The nodule has been quickly identified by the 3rd fixation and the location clicked on. The search of the rest of the image is restricted to the hila, which typically contain blood vessels and the transverse processes of the thoracic vertebrae that can look similar to pulmonary nodules, i.e. well defined and circular in appearance, and one has been identified as a nodule (cross). This is consistent with RHT in that the only knowledge the naïve observers had of pulmonary nodules was the example shown prior to the study and their own understanding of the word 'nodule'. Selective attention has then been given to the stimulus attribute relevant for the task performance, which is to identify lung nodules, and probably indicates top-down guidance. The naïve observer has no knowledge of the likely location of pulmonary nodules or the many subtle appearances that could catch the radiologist out.

4. DISCUSSION

It is important to determine how radiological expertise develops so that we can improve the training of radiologists. The data from the eye tracking study indicates that in terms of eye tracking metrics experts and naïve observers have more in common with each other than with the Level 2 and Level 1 observer. Reverse Hierarchy Theory provides an explanation in that the experts and naïve observers are governed by representations at the 'top' of the visual hierarchy but for different reasons. The naïve observer is probably limited by the accessibility of task relevant information such as intensity, spatial frequency etc., whereas for the expert the perception has become automatic suggesting high level representations. The Level 1 and Level 2 observers are however possibly developing 'schema' in chest image interpretation and are learning how to make sense of the visual information.

5. CONCLUSION

Reverse Hierarchy Theory is a useful general theory for understanding how expertise relates to the information extracted from an image. This has implications for the way radiologists are trained. Experts seem to favour top-down visual information processing as demonstrated by the importance of the 'global' look or initial holistic perception, this can only be achieved by a familiarity with normal appearances so the perception of pathology becomes automatic.

REFERENCES

[1] Ahissar, M. and Hochstein, S., The reverse hierarchy theory of visual perceptual learning. *Trends in cognitive sciences,* **8**(10), pp. 457-464 (2004).

[2] Kundel, H., Nodine, C., Conant, E., Weinstein S., Holistic component of image perception in mammogram interpretation: Gaze-tracking study. *Radiology*, **242**(2), pp. 396-402 (2007).

[3] Kundel, H., Nodine, C., Krupinski, E., Mello-Thoms, C., Using gaze-tracking data and mixture distribution analysis to support a holistic model for the detection of cancers on mammograms. *Acad Radiol*, **15**, pp. 881-886 (2008).

[4] Rasche, C., Koch, C., Recognising the gist of a visual scene: possible perceptual and neural mechanisms. *Neurocomputing*, 44-46, pp. 979-984 (2002).

[5] Juan, C.-H., Walsh, V., Feedback to V1: a reverse hierarchy in vision. *Exp Brain Res*, 150. pp. 259-263 (2003).

[6] Jacob, R., Karn, K., Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In: Hyona, Radach and Deubels, eds, *The Mind's Eye: Cognitive and applied aspects of eye movement research.* Oxford, England: Elsevier Science BV, (2003).