Radiologists' eye movements reflect expertise in interpreting CT studies – a potential tool to measure resident development

Raymond Bertram, M.Sc., Ph.D.[1]; Johanna Kaakinen, M.Sc., Ph.D.[1];
Frank Bensch, M.D., Ph.D.[2]; Laura Helle, M.Sc., Ph.D.[3]; Eila Lantto, M.D., Ph.D.[2];
Pekka Niemi, M.D., Ph.D.[4]; Nina Lundbom, M.D., Ph.D.[2]

[1] Department of Psychology, University of Turku, Turku, Finland

[2] Department of Diagnostic Radiology, HUS Medical Imaging Center, Helsinki University Central Hospital, University of Helsinki, Helsinki, Finland

[3] Department of Teacher Education & Centre for Learning Research, University of Turku, Turku, Finland

[4] Department of Radiology, Turku University Central Hospital, University of Turku, Turku, Finland

Corresponding author: Raymond Bertram

Tel: +358-3335063

Email: rayber@utu.fi

Radiologists' eye movements reflect expertise in interpreting CT studies – a potential tool to measure resident development

Type of manuscript: **Original research**

Advances in knowledge:

1. During viewing of abdominal CT studies, eye fixation duration of specialists and advanced residents increased with presentation speed (b = 0.01 [95% CI: 0.004, 0.026]; *P* = .008; b = 0.04 [95% CI: 0.03, 0.05]; *P* < .0001, respectively), whereas that of early residents did not (b = -.001 [95% CI: -.01, .009]; *P* = .830).

2. In the presence of lesions in the CT image, there was a greater reduction in specialists' saccade length than in residents' (b = 0.02 [95% CI: 0.007, 0.04]; *P* = .003).

3. Unlike for specialists and advanced residents, lesion detection rate of early residents was lower in the afternoon than in the morning (OR, 0.81 [95% CI: 0.73, 0.91]; *P* = .001).

4. Detection rate performance did not differ for high contrast lesions, but early residents detected less of the low contrast lesions (45%: 13/29) than specialists (62%: 18/29; OR, 0.39, [95% CI: 0.25, 0.61]; *P* < .0001) or advanced residents (56%: 16/29; OR, 0.55 [95% CI: 0.33, 0.93]; *P*= .024).

5. Irrespective of the level of expertise, high detection rate was characterized by greater reduction of saccade length in the presence of a lesion (b = -0.10 [95% CI: -0.16, -0.04]; *P* = .002) and a greater increase in saccade length with faster presentation speed (b = 0.11 [95% CI: 0.04, 0.17]; *P* = .001).

Implications for patient care: Eye movement behavior can measure visual expertise in radiology and may offer ways to improve resident training.

Summary statement: Adaptivity to presentation speed and reactivity to lesions is reflected in eye movement behavior of expert radiologists, and could serve as indicators of achieved competence during radiology resident training.

Abstract

Purpose. To establish potential markers of visual expertise in eye movement patterns of early and advanced residents and specialists interpreting abdominal computed tomography (CT) studies.

Materials and Methods. The institutional review board approved to use anonymized CT studies as research materials and obtain anonymized eye-tracking data from volunteers. Participants gave written informed consent. Early residents (n=15), advanced residents (n=14), and specialists (n=12) viewed 26 abdominal CT studies as a sequence of images at either 3 or 5 frames per second (fps), while eye movements were recorded. Data was analyzed using linear mixed effects models.

Results. Early residents' detection rate decreased with working hours (OR, 0.81 [95%CI: 0.73, 0.91]; *P*=.001). They detected less low (but not high) contrast lesions (45%: 13/29) than specialists (62%: 18/29; OR, 0.39, [95%CI: 0.25, 0.61]; *P*<.0001) or advanced residents (56%: 16/29; OR, 0.55 [95%CI: 0.33, 0.93]; *P*=.024). Specialists and advanced residents had longer fixation durations at 5 than 3 fps (Specialists: b=0.01 [95%CI: 0.004, 0.026]; *P* =.008; Advanced residents: b=0.04 [95%CI: 0.03, 0.05]; *P*<.0001). In the presence of lesions, specialists' saccade lengths shortened more than advanced (b=0.02 [95%CI: 0.007, 0.04]; *P*=.003) and early residents' (b=0.02 [95% CI: 0.008, 0.04]; *P*=.003). Irrespective of expertise, high detection rate correlated with greater reduction of saccade length in the presence of lesions (b=-0.10 [95%CI: -0.16, -0.04]; *P*=.002) and greater increase at higher presentation speed (b=0.11 [95%CI: 0.04, 0.17]; *P*=.001).

Conclusion. Expertise in CT reading is characterized by greater adaptivity in eye movement patterns in response to the demands of the task and environment.

Abbreviations

CT	Computed tomography
EM	Eye movement
FD	Fixation duration
ms	Milliseconds
fps	Frames per second
MDCT	Multidetector computed tomography
DR	Detection rate
OR	Odds ratio
CI	Confidence Interval
b	Regression beta coefficient
HU	Hounsfield unit
PACS	Picture archiving and communications system
SD	Standard deviation
SL	Saccade length

Introduction

Resident training in radiology lasts 4 years in the US and 5 years in the EU (1). During training, residents develop visual pattern recognition skills towards an expert level. This becomes increasingly challenging with the introduction of modern complex imaging techniques (2); a radiologist typically views thousands of cross-sectional images per day. Developing necessary skills for efficient image interpretation requires therefore constant development of training and assessment methods.

Eye movements (EM) in CT reading comprise fixations and saccades. Fixation is the maintaining of gaze on a single location. Saccades are rapid movements of the eyes in between two fixations. Visual information is acquired during fixations; no information can be extracted during saccades. Saccade length and fixation duration are parameters that reflect the moment-to-moment cognitive processing of visual information (3). Measuring visual skills by analyzing EM behavior is a potential tool for the assessment of professional development in radiology (4).

Subjects direct attention to objects in a scene using both bottom-up, image-based saliency and top-down, task-dependent cues (5). Scene understanding and object recognition constrain the selection of attended locations (5) - novice and expert radiologists can be expected to differ in these skills. Expert radiologists initially perform a global analysis of the entire visual scene, before turning to focused search and detailed inspection of the image (6-8). Experts may identify lesions when images are shown for as little as 200 ms (9, 10). However, such a quick global scan increases the likelihood of undetected lesions (9-12). Less salient abnormalities require detailed visual inspection, especially when located far from the fixation point, since spatial resolution is highest at the fovea and visual acuity declines sharply outside this area (3). Searching for abnormalities in cross-sectional imaging may fundamentally change the search process and require strategic choices, for instance restricting eye movements to a small region of an organ while quickly scrolling through depth (13). EM patterns during inspection of medical images change with growing expertise (14-16). Experienced radiologists cover less image area in pulmonary nodule detection than residents (15). Similarly, pathology residents need fewer fixations and spend less time examining non-diagnostic areas with increasing training progress. Their search pattern approximates that of expert pathologists after about one year of training (16).

Our study investigated the development of visual expertise in radiology during abdominal CT reading by comparing performance and EM behavior of early residents, advanced residents, and specialists. Several factors may increase the difficulty to detect abnormalities, possibly amplifying differences in diagnostic performance and EM behavior between the expertise groups. **The following three hypotheses were tested**: number of hours worked, presentation speed, presence/absence and saliency of the lesion have different effects on residents' and specialists' detection performance (1) and EM behavior (2); detection of lesions is reflected in EM behavior independent of level of expertise (3) (see13). The **purpose** of our study was to establish potential markers of visual expertise of radiologists in eye movement patterns of early and advanced residents, and specialists interpreting abdominal computed tomography (CT) studies.

Materials and methods

Ethics. The institutional review board approved to use anonymized CT studies as research materials and obtain anonymized eye-tracking data from volunteers. Participants gave written informed consent.

Participants. Twenty-nine radiology residents from two University Hospitals volunteered to participate in this study and were classified as either early (< 2 years of experience; n = 15; mean age 32 years, range 28-41; 10 female, mean age 32 years, range 28-37; 5 male, mean age 34 years, range 29-41), or advanced (>= 2 years of experience, n = 14; mean age 35 years, range 29-41; 8 female, mean age 35 years, range 29-38; 6 male, mean age 36 years, range 30-41). In addition, twelve specialists (mean age 43 years, range 33-52; 6 female, mean age 43 years, range 35-50; 6 male, mean age 43, range 33-52) with a minimum of 2 years of experience in abdominal radiology from one University Hospital volunteered to participate. The early residents had up to 1.5 years and the advanced 1.5 to 3.5 years of experience in CT. The specialists had 2 to 22 years of experience in abdominal radiology with practice either exclusively in abdominal radiology (four) or shared between abdominal and uroradiology (three), ENT-radiology (four) or vascular radiology (one). Time spent at work on the same day before experimentation was registered for each participant.

CT studies. Twenty-six axial clinical multidetector computed tomography (MDCT) studies of the abdomen obtained on three CT scanners (GE Lightspeed VCT 64 XT, GE Lightspeed VCT 64, GE Healthcare, Little Chalfont, Buckinghamshire, United Kingdom, Siemens Definition AS+128, Siemens Healthcare, Erlangen, Germany) reformatted to 2.5 mm or 3 mm section thickness were selected and anonymized. Three specialists with 9, 20 and 22 years of experience, who were not observers (authors FB, NL, EL), independently reviewed the CT scans and came to a consensus about the lesions. Only full consensus cases were included. Three studies were without abnormalities and categorized as “normal”. Of the 23 studies with lesions, two were used as practice material before data recording. Of the remaining 21 CT studies, 18 were contrast-enhanced and 3 non-enhanced, and each contained 131-160 images and 1-6 lesions. The 21 CTs included in total 70 lesions. The saliency of lesions was assessed by size (the number of images the lesion was visible on) and by contrast between the intensity of the lesions and their background (Table 1). Parenchymal lesions were contrasted to their surrounding parenchyma (e.g. liver cyst, liver) or abdominal cavity and subcutaneous lesions to the surrounding fat. The prevalence and the saliency of the lesions were assessed by FB and NL.

Image presentation and diagnostic evaluation. The CT studies were presented once each as a continuous presentation at a fixed frame rate. Studies were allocated to two even sets (A and B). Participants viewed all studies from each set in random order with either set A or B being presented first at 3 fps and the respective other subsequently at 5 fps in a cross-over fashion. These speeds were chosen as in an earlier study (4) specialists indicated in a feedback questionnaire that with sections of 3 mm, a framerate of 2 fps or smaller is too slow and a framerate of 9 fps clearly too fast for comprehensive CT reading, whereas a framerate around 4 fps is satisfactory. Taking this feedback into consideration, while similarly being able to investigate the role of presentation speed, we chose the current framerates of 3 and 5 fps. At these presentation speeds, one image is visible for 333 ms and 200 ms, respectively. Viewing each CT study took less than 30 seconds and was followed by ticking the appropriate findings on a paper checklist (Fig 1, Appendix 1). Prior to data recording, participants were informed of technical aspects of the procedure, and that image sets could also be without findings. Clinical information and demographics were withheld. Before the experiment proper, all participants familiarized themselves with the procedure by viewing two previously determined “practice” CT studies and reporting the findings in them. The procedure took 35 minutes at most for each participant to view all CT studies from both sets.

Lighting conditions were uniform at 238 lux in quiet surroundings. Viewing distance was 60 cm to a DICOM calibrated 21.3-inch EIZO RadiForce MX210 monitor with a refresh rate of 60 Hz (EIZO Corp., Hakusan-Shi, IKW, Japan). At this distance, one degree of visual angle corresponds to 1 cm on the screen, and 2-3 cm in the patient.

[Insert Figure 1 about here]

Eye movement recording. Eye movements were recorded using a desktop model of the EyeLink 1000 eyetracker (SR Research Ltd, Mississauga, Canada). The tracker is an infrared video-based tracking system with hyperacuity image processing and spatial resolution of 0.4 degrees. Recording was monocular under immobilization by a chin and forehead rest. Prior to experimentation, the eyetracker was calibrated using a nine-point calibration grid extening over the entire computer screen.

Data analysis and statistical methods. Detection accuracy was a dichotomous variable (1 = correctly identified lesion, 0 = missed lesion). Seven lesions that were either missed by all or only detected by a single participant were excluded from further analyses (Table 1). Detection rate for each participant was calculated by dividing the number of correctly identified lesions by the total number of lesions. False positives were also calculated.

The image presentation with 3 and 5 fps ensured that changes on the screen occurred in a gradual manner, eliciting fixation-saccade cycles in EM behavior. Fixation durations (in ms) and saccade lengths (in degree of visual angle) during CT scan viewing were extracted from the EM recordings (Fig 2). We use the post-recording filtering system of the Eyelink software analysis (DataViewer) resulting from a preset saccade algorithm. The filter takes minimal fixation duration (here 50 ms) and minimal saccadic amplitude (here 0.15°) as settings. Subsequent fixations that are shorter and closer than the threshold settings stipulate are merged into one fixation (17). Given that a section was presented for 200 (5 fps) or 333 ms (3 fps), and average fixation duration was around 340 ms, both fixation duration and saccade length were typically extracted across 2 sections. The EM measures were log transformed in order to normalize the data, and observations 3 standard deviations (SD) away from the grand mean were excluded (1.16 % in fixation data, 0 % in saccade data).

Data were analyzed by generalized linear mixed effects models using lme4 package (18) for R statistical software (19). Linear mixed effects models are used to describe relationships between a response variable (e.g., detection rate) and explanatory factors/variables (e.g., presentation speed) for repeated measures data in which the response variable is measured more than once for each subject across levels (e.g., 3 fps vs. 5 fps) of one or more factors. Detection Rate, Fixation Duration and Saccade Length were analyzed separately. Correlations of Detection Rate and eye movement measures were also examined by pooling the data of all three groups to investigate whether the effects of presence of a lesion and frame rate depended on an individual observer´s detection performance. Model fitting details, coding of the variables and model estimates of these analyses are presented in Appendix Table 1. Finally, the effect of contrast on all dependent measures was explored.. Contrast was calculated by dividing the absolute difference between lesion intensity and background by 2000, the range of the Hounsfield Unit scale. This generates values between 0 and 1, with 0 indicating no contrast and 1 maximal contrast (-1000 air, 1000 bone, absolute difference 2000, contrast 2000/2000=1). The contrast effect could be assessed for a subset of the data only. Contrasts could not be calculated from 3 cases (minor fluid collection and missing spleen, 2 times); for the EM measures, sections with no lesion or more than one lesion present at the same time were excluded.

Results

Detection rate. Detection data was missing for 0.2% of the observations. Altogether, detection accuracy data consisted of 2,577 observations (Table 1). The DR of specialists (60%: on average 38 lesions out of 63 were detected, 38/63) was higher than that of residents (52%: 33/63; OR, 1.38 [95% CI: 1.08, 1.76]; *P* = .0107). Advanced residents detected more lesions (55%: 35/63) than early residents (49%: 31/63) (OR, 1.49 [95% CI: 1.02, 2.18]; *P* = .038). For all groups, DR was higher at 3 than at 5 fps (specialists: 3 fps 62%: 39/63, 5 fps 58%: 37/63; advanced residents: 3 fps 59%: 37/63, 5 fps 52%: 33/63; early residents: 3 fps 52%: 33/63, 5 fps 45%: 28/63 (OR, 0.68 [95% CI: 0.56, 0.82]; *P* < .0001), and the size of a lesion increased the likelihood of it being detected (OR, 1,03 [95% CI: 1.01, 1.04]; *P* = 0.001). Time spent at work prior to viewing these CT data sets decreased lesion detection rate for early residents (OR, 0.81 [95% CI: 0.73, 0.91]; *P* = .001), but not for advanced residents (OR, 1.07 [95% CI: 0.95, 1.19]; *P* = .260) or specialists (OR, 1.05 [95% CI: 0.83, 1.33]; *P* = .70) (Fig 3). The number of false positives did not differ significantly across groups (on average 8 for each group, p > .4).

In the analysis of the effect of lesion contrast on detection rates, there were a total of 2,391 observations. Early residents detected less of the low contrast abnormalities (45%: 13/29) than specialists (62%: 18/29; OR, 0.39, [95% CI: 0.25, 0.61]; *P* < .0001) or advanced residents (56%: 16/29; OR, 0.55 [95% CI: 0.33, 0.93]; *P*= .024), whereas there were no differences between groups in the detection of high contrast abnormalities (early residents (52%: 16/30) vs. specialists (56%: 17/30): OR, 0 .76 [95% CI: 0.44, 1.31]; *P* = .322; advanced residents (51%: 15/30) vs. specialists (56%: 17/30), OR, 0.75 [95% CI: 0.44, 1.27]; *P* =.285).

Fixation durations. The final EM data set consisted of 97,813 fixations.Mean FD ranged from 327 ± 34 ms (specialists, 3 fps, no lesion) to 374 ± 77 ms (early residents, 5 fps, lesion present) (Table 2). FDs were on average 10 ms longer with a lesion present in the image than without (b = 0.03 [95% CI: 0.02, 0.04]; *P* < .0001). FDs were on average 4 ms longer at 5 fps than at 3 fps for specialists (b = 0.01, [95% CI: 0.004, 0.026]; *P* = .008) and 11 ms for advanced residents (b = 0.04 [95% CI: 0.03, 0.05]; *P* < .0001), but not for early residents (b = -.001 [95% CI: -0.01, 0.009]; *P* = .830) (Fig 4).

The analysis of the effect of lesion contrast on FD included 11,820 fixations. High contrast lesions elicited longer FD than low contrast lesions across groups (344 ms vs 313 ms; b = 0.04 [95% CI: 0.02, 0.06], *P* =.0002).

Saccade lengths. The final EM data set consisted of 98,110 saccades.SL ranged from 2.41 ± .23 º (specialists, 3 fps, lesion present) to 2.68 ± .46 º (advanced residents, 5 fps, lesion present). There were significant interactions between expertise and presence of lesion, (see Appendix Table 1) indicating that the impact of a lesion on saccade lengths depended on the level of expertise: average saccade length was shorter with a lesion present than absent, and the effect of a lesion was greater for the specialist group (.17º, b = -0.05 [95% CI: -0.07,-0.04]; *P* < .0001) than for the advanced (.10º, b = -0.03 [95% CI: -0.04,-0.02]; *P* = .001) or early residents (.10º, b = -0.03 [95% CI: -0.04,-0.02]; P < .0001) (Fig 5). Saccade lengths were overall .12º longer at 5fps than at 3fps (b = 0.04 [95% CI: 0.03, 0.05], *P* < .0001).

In the analysis of the effect of lesion contrast on SL 11,909 saccades were included. SL was shorter for high contrast than low contrast lesions across groups (1.59° vs 1.91°, b = -0.18 [95% CI -0.27, -0.09], *P* <.0001).

Summary. Table 3 presents the key findings: the effects of work time, lesion size, contrast, presentation speed, and the presence of lesions on detection rate and EM measures for each group.

Eye movement measures and detection performance. When pooling the data of all groups, we found that the presence/absence of a lesion caused different adjustments in SL as a function of DR (b = -0.10 [95% CI: -0.16, -0.04]; *P* = .002). In order to illustrate this effect, the estimated SLs were computed for individuals with DR of 1 SD below and above the mean. For individuals with low DR, saccades were estimated to be .09 degrees shorter in the presence of a lesion in comparison to the absence of a lesion, whereas for individuals with high DR, this was .16º

(Fig 6). DR was also related to the effect of frame rate (b = 0.11 [95% CI: 0.04, 0.17]; *P* = .001). Participants with low DR showed a relatively small effect of frame rate with saccades being .09º longer in 5fps than in 3fps. Participants with high DR generated saccades that were .18º longer in 5fps than in 3fps (Fig 6). No effects were found for FD.

Discussion

In our study, specialists showed more flexibility in eye movement behavior than residents; specialists reacted to the presence of lesions in CT images by *longer* fixation durations and *shorter* saccades and adapted to faster presentation speeds by *longer* fixation durations and *longer* saccades. The performance of early residents differed from advanced residents' and specialists' in several aspects: they had lower detection rates, which further decreased towards the afternoon and were especially present for low contrast lesions. This entails that growing visual expertise goes hand in hand with a growing ability to detect visually less salient abnormalities. EM behavior of early residents was the least flexible; although increased presentation speed presented an obvious challenge for this group as evidenced by decreased detection performance, it did not lead to increased fixation durations. In the presence of lesions, they did not adapt the saccade length to the same extent as specialists. In contrast, the advanced residents' EM adaptation to increased presentation speeds was similar to that of specialists: although their diagnostic performance also declined with increased presentation speed, they adapted by increasing fixation durations. However, - as the early residents - the advanced residents reacted to the presence of abnormalities by only a small decrease in their saccade length.

Increased presentation speed led to lower detection rates and longer saccades for all groups and to increased fixation durations for advanced residents and specialists. An earlier study (4) showed that increased presentation speed leads to shorter rather than longer saccades for naïve participants viewing abdominal CT studies. The longer saccades in all our groups under fast presentation thus suggest acquired behavior during radiology training. Presumably, this strategy allows covering the visual scene more efficiently under time pressure. Longer fixations at high frame rates is a sensible strategy considering the continuous changing of image composition in CT. A single lesion appears on several images, and longer fixation durations allow for better characterization. Our data suggests that the ability to adapt fixation duration to increased speed is acquired during resident training, be it at a later stage than saccadic adaptation. Similar findings can be observed in studies on driving expertise, where adaptation of EM patterns to road type correlate with driving experience (20, 21), suggesting a similar pattern of visual development.

Our finding that specialists’ saccades became markedly shorter in the presence of a lesion concurs with previous studies (4). This decrease in saccade length probably reflects a switch from searching to closer inspection and characterization of a lesion. The smaller decrease in saccade length for the resident groups indicates that residents are still learning to use this strategy.

Our findings showed that – unlike for specialists and advanced residents - detection performance of early residents participating in the afternoon was lower than that of early residents participating in the morning (Fig 3). This finding is in agreement with that of Krupinski et al. (22), who showed that residents’ detection rates of pulmonary nodules were higher early in the morning before work than in the late afternoon after working time, independent of other factors. The afternoon drop in detection performance of early residents can be linked to them still being in an active learning phase, where complex semantic links must be reasoned and validated one step at a time, which takes a lot of mental effort. Further in the learning process, semantic connections are firmly established around overarching concepts and knowledge becomes more directly accessible (23), allowing cognitive processing to become less strenuous. The development of visual expertise is assumed to include structural and functional changes in brain areas related to perception and cognition (24), allowing to perform at a high level throughout the day.

Our study had a number of limitations. The participants viewed CT images as a standardized presentation without being able to change presentation speed, direction, or zoom, to allow measuring their EM patterns under standardized conditions, making the conditions of the experiment inevitably deviate from a typical clinical setting. The study was designed to make perfect diagnostic performance hard or impossible to achieve, in order to emphasize differences between the groups. The experiment was conducted under partly attenuated lightning to allow for the viewing of the relatively high contrast CT studies while still being able to fill in the checklist. We chose a large variety of lesions to stimulate the participants to search and detect as well as they could. Our main focus was to study the participants´ reactions to the lesions, and their diagnostic performance was of secondary interest. Also, instead of performing an areas-of-interest analysis of the eye movements (25, 26), we chose to analyze more general and easily extractable EM measurements as indices of expert performance. Finally, in our experiment the participants did not immediately indicate the presence of a lesion. However, we coded the presence/absence of a lesion for each section and assessed the EM behaviour as a function of this variable in our regression analyses. That specialists detected more lesions than residents and at the same time resorted to smaller saccades in response to the presence of a lesion implies a direct link between EM behaviour and lesion detection.

Practical implications of our findings include that early residency should be viewed as a “hatching” period, where residents are more likely to be affected by exhaustion than advanced residents and specialists, which should be taken into consideration when assigning their daily schedule. Furthermore, our results support the present inclination to abandon rigid time schedule-based resident training in exchange for a more individually tailored approach. Analyzing development of expertise by EM behavior could be useful to monitor individual residents’ learning curves, and contribute to a further shift towards competency-based resident training.

In sum, adaptivity to presentation speed and reactivity to lesions is reflected in eye movement behavior of expert radiologists, and could serve as indicators of achieved competence during radiology resident training. Our results indicate acquired skills in expert radiologists such as reacting to appearing lesions by *shortening* saccades - more than residents do - and adapting to faster presentation speed by *lengthening* saccades and *increasing* fixation duration. In line with what was found for detection performance, EM reactivity and adaptivitity of advanced residents lie between early residents' and specialists'. Moreover, good detection performance irrespective of group correlated with reactivity in saccade length. This suggests that - already early in resident training - saccade length reactivity may be useful to detect a visually talented radiologist. In conclusion, our study shows that development in eye movement behavior is indicative of achieved competence during radiology resident training.

Appendix 1

<u>FINDINGS:</u>	**PARTICIPANT:**		**TRIAL NUMBER**	
NO FINDINGS	☐			
DIAPHRAGM	☐ hiatus hernia	☐ esophageal varices		
LIVER	☐ cyst	☐ metastatic disease		
GALL BLADDER	☐ inflammation	☐ concrement		
BILE DUCTS	☐ dilation			
SPLEEN	☐ abnormal	☐ removed		
PANCREAS	☐ tumor	☐ inflammation	☐ fluid collection	
PANCREATIC DUCT	☐ dilation			
KIDNEYS	☐ cyst	☐ Transplant	☐ shrunken kidney	☐ infarction
AORTA	☐ aneurysm			
SMALL BOWEL	☐ occlusion	☐ incarceration		
LARGE BOWEL	☐ tumor	☐ inflammation	☐ diverticulosis	
APPENDIX	☐ inflammation			
GYN. ORGANS	☐ tumor			
ABDOMINAL CAVITY	☐ free fluid	☐ perforation	☐ enlarged lymphatic glands	
MUSCLES	☐ tumor	☐ fluid collection		
SUBCUTIS	☐ tumor	☐ hernia		

References

1. European Society of Radiology. European Training Curriculum for radiology, 2013. Downloaded on 9.5.2014 at http://www.myesr.org/html/img/pool/20_08_2014_ESR_2013_ESR-EuropeanTrainingCurriculum_web.pdf

2. Grimm LJ, Kuzmiak CM, Ghate SV, Yoon SC, Mazurowski MA. Radiology resident mammography training: Interpretation difficulty and error-making patterns. Acad Radiol 2014; 21:888-892.

3. Rayner K. Eye Movements in Reading and Information Processing: 20 Years of Research. Psy Bul 1998; 124(3): 372-422.

4. Bertram R, Helle L, Kaakinen JK, Svedström E. The Effect of Expertise on Eye Movement Behaviour in Medical Image Perception. PLoS ONE 2013; 8(6): doi:10.1371/journal.pone.0066169

5. Itti L, Koch C. Computational modeling of visual attention. Nat Rev Neurosci; 2: 194-203: doi: 10.1038/35058500

6. Krupinski EA. Human factors and human-computer considerations in teleradiology and telepathology. Healtcare 2014; 2: 94-114.

7. Carmody DP, Nodine CF, Kundel HL. Finding lung nodules with and without comparative visual scanning. Percept Psychophys 1981; 29(6): 594–598.

8. Kundel HL, Nodine CF, Krupinski EA, 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 2008; 15 (7): 881–886.

9. Kundel HI, Nodine CF. Interpreting chest radiographs without visual search. Radiology 1975; 116: 527–532.

10. Mugglestone MD, Gale AG, Cowley HC, Wilson ARM. Diagnostic performance on briefly presented mammographic images. Proc SPIE 1995; 2436: 106–115.

11. Carmody DP, Nodine CF, Kundel HI. An analysis of perceptual and cognitive factors in radiographic interpretation. Perception 1980; 9: 339–344.

12. Kundel HL, Nodine CF, Conant EF, Weinstein SP. Holistic component of image perception in mammogram interpretation: gaze-tracking study. Radiology 2007; 242 (2): 396 – 402.

13. Drew T, Vo ML, Olwal A, Jacobson F, Seltzer SE, Wolfe JM. Scanners and drillers: Characterizing expert visual search through volumetric images. J Vis. 2013; 13 (10), 1–13.

14. Reingold EM, Sheridan H (2011) Eye movements and visual expertise in chess and medicine. In Liversedge SP, Gilchrist ID, Everling S (Eds.), Oxford handbook on eye movements, pp. 767–786. Oxford, UK: Oxford University Press.

15. Krupinski EA. Visual scanning patterns of radiologists searching mammograms. Acad Radiol 1996; 3 (2): 137–144.

16. Krupinski EA, Weinstein RS. Changes in visual search patterns of pathology residents as they gain experience Proc. SPIE 2013; 8673.

17. SR Research. EyeLink data viewer user’s manual. Document version 1.8.221, 2002-2007, http://www.sr-research.com/dv.html

18. Bates D, Maechler M, Bolker B and Walker S. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7, 2014, http://CRAN.R-project.org/package=lme4.

19. R Core Team. R: A language and environment for statistical computing, 2013. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

20. Crundall DE, Underwood, G. Effects of experience and processing demands on visual information acquisition in drivers. *Ergonomics*, 1998, *41*(4): 448-458.

21. Underwood G, Chapman P, Brocklehurst N, Underwood J, Crundall D. Visual attention while driving: sequences of eye fixations made by experienced and novice drivers. *Ergonomics*, 2003, *46*(6): 629-646.

22. Krupinski EA, Berbaum KS, Caldwell RT, Schartz KM, Madsen MT, Kramer DJ. Do long radiology workdays affect nodule detection in dynamic CT interpretation? J Am. Coll. Radiol. 2012; 9: 191-198.

23. Boshuizen HP, Schmidt HG. The development of clinical reasoning expertise. *Clinical reasoning in the health professions*, 2008: 113-121. Elsevier: Amsterdam.

24. Harel A, Kravitz D, Baker CI. Beyond perceptual expertise: Revisiting the neural substrates of expert object recognition, Frontiers in Human Neuroscience, 2013, 7: 00885 DOI: 10.3389/fnhum.2013.00885

25. Helbren E, Halligan S, Phillips P, Boone D, Fanshawe TR, Taylor SA, et al. Towards a framework for analysis of eye-tracking studies in the three dimensional environment: a study of visual search by experienced readers of endoluminal CT colonography. Br J Radiol 2014;87: 20130614.

26. Mallett S, Phillips P, Fanshawe TR, Helbren E, Boone D, Gale A, et al. Tracking Eye Gaze during Interpretation of Endoluminal Three-dimensional CT Colonography: Visual Perception of Experienced and Inexperienced Readers. Radiology 2014; 273(3):783-92. doi: 10.1148/radiol.14132896

Table 1. The 21 cases (Case) including in total 63 lesions (Lesion) used as test materials, the number of sections lesions were visible (NrSlides), the contrast between the intensity of the lesion and its background (Contrast) and the mean detection rates for specialists (DR_S), advanced residents (DR_AR) and early residents (DR_ER).

Case	Lesion	NrSlides	Contrast	DR_S	DR_AR	DR_ER
1	Large bowel, diverticulosis	52	0.41	0.67	0.93	0.8
	Large bowel, inflammation	52	0.04	0.92	0.57	0.67
2	Appendix, inflammation	24	0.08	0.58	0.5	0.53
	Gall bladder, concrement	33	0.06	1	1	1
	Gall bladder, inflammation	36	0.06	0.92	0.93	0.87
	Bile ducts, dilation	15	0.02	0	0.21	0.07
	Abdominal cavity, free fluid	6	0.06	0.83	0.43	0.33
4	Kidneys, infarction	34	0.07	1	0.79	0.93
5	Diaphragm, hiatus hernia	11	0.08	0.5	0.29	0.13
	Large bowel, diverticulosis	19	0.36	0.92	0.93	0.6
	Large bowel, inflammation	19	0.05	1	0.79	0.67
6	Pancreas, fluid collection	34	0.06	0.75	0.43	0.47
7	Liver, metastatic disease	68	0.04	1	1	1
	Large bowel, tumor	10	0.08	0.75	0.29	0.2
8	Liver, metastatic disease	13	0.01	0.92	0.86	0.67
	Large bowel, diverticulosis	12	0.11	0.33	0.29	0
	Tumor in subcutaneous fat	40	0.08	0.08	0.07	0.33
9	Pancreas, fluid collection	16	0.03	0.92	0.79	0.87
	Pancreas, inflammation	63	0.06	0.92	0.79	0.8
	Abdominal cavity, free fluid	50	0.05	0.33	0.29	0.4
10	Appendix, inflammation	12	0.08	0.64	0.79	0.2
	Liver, cyst	9	0.09	0.91	0.86	1
	Kidneys, cyst	5	0.10	0.27	0.21	0.2
	Large bowel, inflammation	20	0.07	0.64	0.36	0.67
11	Kidneys, cyst	8	0.11	1	0.79	0.67
	Spleen, removed	18	NA	0.08	0.29	0.4
	Large bowel, inflammation	54	0.34	0.5	0.5	0.27
	Gall bladder, concrement	20	0.09	0.67	0.79	0.8
12	Aorta, aneurysm	23	0.07	0.5	0.5	0.33
	Kidneys, cyst	39	0.02	0.83	0.93	0.93
	Small bowel, occlusion	98	0.50	1	0.93	0.93
	Large bowel, diverticulosis	28	0.27	0.83	0.57	0.53
13	Gyn., tumor	7	0.05	0.33	0.14	0.33
14	Gyn., tumor	7	0.08	0.33	0.57	0.21
15	Liver, cyst	2	0.03	0.25	0.14	0.13
16	Gyn., tumor	12	0.02	0.17	0.07	0.27
	Pancreas, tumor	9	0.10	0.42	0.29	0.13
	Pancreatic duct, dilation	19	0.05	0.25	0.29	0.2
	Kidneys, cyst	7	0.07	0.58	0.43	0.67
	Large bowel, diverticulosis	70	0.24	0.5	0.36	0.2
	Intrahepatic bile ducts, dilatation	24	0.06	0.92	0.79	0.8
17	Small bowel, incarceration	13	0.07	0.42	0.29	0.2
	Small bowel, occlusion	77	0.12	0.75	0.79	0.87
	Large bowel, diverticulosis	12	0.47	0.5	0.57	0.13
	Abdominal wall hernia	11	0.10	0.5	0.71	0.53
	Abdominal cavity, free fluid	72	0.05	1	1	0.87
18	Muscles, tumor	20	0.01	0	0.21	0
	Liver, cyst	2	0.03	0.42	0.36	0.4
	Kidneys, cyst	9	0.05	0.42	0.5	0.33
	Large bowel, diverticulosis	58	0.12	0.67	0.57	0.47
19	Kidneys, shrunken kidney	21	0.07	0.92	0.79	0.47
	Kidneys, cyst	7	0.01	0.33	0.36	0.27
	Kidneys, transplant	42	0.08	0.83	0.93	0.87

	Spleen, removed	59	NA	0.33	0.29	0.6
	Large bowel, diverticulosis	76	0.31	0.75	0.71	0.27
	Large bowel, tumor	13	0.06	0.17	0	0
20	Pancreas, tumor	23	0.03	0.58	0.29	0.33
	Liver, cyst	10	0.02	0.67	0.86	0.8
	Abdominal cavity, free fluid	24	0.06	1	0.71	0.67
	Abdominal cavity, perforation	76	0.53	0.83	0.79	0.4
21	Diaphragm, esophageal varices	20	0.11	0	0.14	0.07
	Spleen, abnormal	66	0.10	0.42	0.57	0.6
	Abdominal cavity, free fluid	38	NA	0.42	0.79	0.27

Table 2. Means and standard deviations of the eye movements measures for specialists, advanced and early residents as a function of presentation speed and presence of a lesion in the image.

		Presentation speed			
		3 fps		5 fps	
Measure	Expertise group	Lesion absent	Lesion present	Lesion absent	Lesion present
FD (ms)	Specialists	327±34	342±38	334±23	345±29
	Advanced residents	339±62	349±69	351±65	362±64
	Early residents	371±83	378±75	371±79	374±77
SL (º)	Specialists	2.50±.30	2.41±.23	2.61±.37	2.47±.31
	Advanced residents	2.56±.37	2.56±.39	2.68±.46	2.64±.34
	Early residents	2.50±.42	2.45±.46	2.61±.52	2.64±.51

FD: Fixation Duration; SL: Saccade Length

Table 3. Summary of main findings showing the effect of lesion saliency, lesion contrast, presence or absence of lesion, frame rate and working hours on detection rate, fixation duration and saccade length in the three groups.

Manipulated Factor	Measure	Specialists	Advanced Residents	Early Residents
Lesion saliency ↑	DR	↑	↑	↑
Lesion contrast ↑	DR	−	−	↑
	FD	↑	↑	↑
	SL	↓	↓	↓
Lesion present	FD	↑	↑	↑
	SL	↓↓	↓	↓
Frame rate ↑	DR	↓	↓	↓
	FD	↑	↑	−
	SL	↑	↑	↑
Working hours ↑	DR	−	−	↓

DR: Detection Rate; FD: Fixation Duration; SL: Saccade Length

↑Increase; ↓ Decrease; ↓↓Larger decrease than in other groups; − no effect

Appendix Table 1

1. Final model estimates for lesion detection. The model is based on 2577 responses recorded from 41 participants for 63 lesions. Participants and lesions were included in the model as crossed random factors. For the Expertise group was dummy coded, specialists served as a reference group. The presentation speed was contrast coded (3fps = -.5, 5 fps =.5), as was gender (female=-.5, male = .5). Visibility and work time were centered. Note that the fixed effect estimates for the presence of abnormality and presentation speed represent slopes in the specialist group. Significant interactions were followed by computing simple slopes for each expertise group.

Random effects	Variance	SD		
Participant (intercept)	0.08606	0.2934		
Lesion (intercept)	1.88482	1.3729		
Fixed effects	B	SE	z	P
Intercept	0.609292	0.232541	2.620	0.008789
Saliency	0.027430	0.007823	3.506	0.000455
Presentation speed	-0.383282	0.094709	-4.047	<.0001
Gender	-0.412927	0.343667	-1.202	0.229544
Work time	0.046463	0.119593	0.389	0.697637
Resident group (advanced)	-0.333586	0.195059	-1.710	0.087232
Resident (early)	-0.807994	0.192728	-4.192	<.0001
Resident group (advanced)*Gender	0.739143	0.412165	1.793	0.072922
Resident (early)*Gender	-0.045868	0.415141	-0.110	0.912023
Resident group (advanced)*Work time	0.018088	0.132726	0.136	0.891598
Resident (early)*Work time	-0.251640	0.133333	-1.887	0.059119

2. Final model estimates for fixation duration. The model is based on 97,819 fixations recorded from 41 participants during viewing of 24 cases. Participants and cases were included in the model as crossed random factors. Expertise group was dummy coded, specialists served as a reference group. The presence of abnormality during fixation was contrast coded (yes= .5, no = -.5) as was presentation speed (3fps = -.5, 5 fps =.5). Note that the fixed effect estimates for the presence of abnormality and presentation speed represent slopes in the specialist group. Significant interactions were followed by computing simple slopes for each expertise group.

Random effects	Variance	SD			
Participant (intercept)	.0186	.1365			
Case (intercept)	.0005	.0224			
Residual	.2169	.4657			
Fixed effects	B	SE	Df	t	P
Intercept	5.703	.0398	39	143.341	<.0001
Presence of abnormality	.03179	.0036	12240	8.769	<.0001
Presentation speed	.01488	.0056	97730	2.652	.00801
Expertise group (advanced)	.01971	.0538	38	.366	.71638
Expertise group (early)	.07803	.05303	38	1.472	.1494
Expertise group (advanced)*Presentation speed	.02107	.00769	97690	2.739	.00617
Expertise group (early)*Presentation speed	-.01601	.00767	97730	-2.088	.03679

3. Final model estimates for saccade lengths. The model is based on 98,110 saccades recorded from 41 participants during viewing of 24 cases. Participants and cases were included in the model as crossed random factors. Expertise group was dummy coded, specialists served as a reference group. The presence of abnormality during fixation was contrast coded (yes= .5, no = -.5) as was presentation speed (3fps = -.5, 5 fps =.5). Note that the fixed effect estimates for the presence of abnormality and presentation speed represent slopes in the specialist group. Significant interactions were followed by computing simple slopes for each expertise group.

Random effects	Variance	SD			
Participant (intercept)	.00937	.0968			
Case (intercept)	.00296	.0544			
Residual	.21741	.4663			
Fixed effects	B	SE	Df	t	P
Intercept	1.139	.03210	52	37.696	<.0001
Presence of abnormality	-.0541	.00597	95080	-9.077	<.0001
Presentation speed	.0382	.00309	98030	12.379	<.0001
Expertise group (advanced)	.02836	.03828	41	.7410	.46298
Expertise group (early)	.00026	.03769	41	.0007	.99445
Expertise group (advanced)*Abnormality	.02245	.00767	98030	2.926	.00344
Expertise group (early)*Abnormality	.02283	.00765	98030	2.982	.00286

4. Final model estimates for saccade lengths with detection rate as a continuous fixed factor. The model is based on 98,110 saccades recorded from 41 participants during viewing of 24 cases. Participants and cases were included in the model as crossed random factors. Detection rate was centered. The presence of abnormality during saccade was contrast coded (yes= .5, no = -.5) as was presentation speed (3fps = -.5, 5 fps =.5).

Random effects	Variance	SD			
Participant (intercept)	.008739	.09348			
Case (intercept)	.002961	.05442			
Residual	.217390	.46625			
Fixed effects	B	SE	Df	t	P
Intercept	1.149	.01842	63	62.393	<.0001
Presence of abnormality	-.0383	.00369	79380	-10.399	<.0001
Presentation speed	.0382	.00309	98030	12.376	<.0001
Detection rate	.2823	.1461	41	1.931	.06036
Detection rate * Presence of abnormality	-.0982	.03153	98030	-3.115	.00184
Detection rate * Presentation speed	.10595	.03158	98030	3.352	.00080

Figure 1. Two subsequent trials as presented in the experiment. After having viewed a CT study the participant filled out a checklist to indicate the observed abnormalities if any. When ready a button was pressed to make the next examination appear. The whole experiment started with 2 practice trials, after which 12 CT studies appeared with a frame rate of 3 fps followed by 12 CT studies with a frame rate of 5 fps. Of the 24 experimental trials, 21 contained 1-6 abnormalities.

Figure 2: Fixations and saccades projected on a CT image of a patient with liver metastases. Each circle represents a fixation and the size of a circle reflects fixation duration. Around fixations, fine image details can be perceived. Saccades are represented by lines with arrowheads between fixations; the amplitude of saccades varies both within and across radiologists.

Figure 3: Detection rates (y-axis) as a function of work time (x-axis). Specialists are marked with black dots, advanced residents with diamonds, and early residents with gray triangles.

Figure 4: Model estimates for the fixation durations as a function of frame rate and expertise level. Error bars represent standard error.

Figure 5: Model estimates for saccade lengths as a function of the presence of lesion and expertise level. Error bars represent standard error.

Figure 6: Model estimates for SLs computed for radiologists with DR of 1 SD below (low DR, left panel) and above the mean (high DR, right panel) in all groups combined. For radiologists with high DR, presence of lesion in the image induced greater shortening in SL (8%) than for radiologists with low DR (5%). Increasing the presentation speed induced a greater increase in the SL for radiologists with high DR (9%) than for those with low DR (5%). Error bars represent standard error.