Original contribution

Characterizing the development of visual search expertise in pathology residents viewing whole slide images

Elizabeth A. Krupinski PhD\textsuperscript{a,\,*}, Anna R. Graham MD\textsuperscript{b}, Ronald S. Weinstein MD\textsuperscript{b}

\textsuperscript{a}Department of Medical Imaging, University of Arizona, Tucson, AZ 85724
\textsuperscript{b}Department of Pathology, University of Arizona, Tucson, AZ 85724

Received 28 March 2012; revised 25 May 2012; accepted 30 May 2012

Keywords:
Expertise;
Visual search;
Decision making;
Radiology;
Whole slide images;
Digital pathology;
Telepathology

Summary
The goal of this study was to examine and characterize changes in the ways that pathology residents examine digital whole slide images as they progress through the residency training. A series of 20 digitized breast biopsy whole slide images (half benign and half malignant biopsies) were individually shown to 4 pathology residents at four points in time—at the beginning of their first, second, third, and fourth years of residency. Their task was to examine each image and select three areas that they would most want to zoom in on in order to view the diagnostic detail at higher resolution. Eye position was recorded as they scanned each whole slide image at low magnification. The data indicate that with each successive year of experience, the residents’ search patterns do change. Overall, with time, it takes significantly less time to view an individual slide and decide where to zoom, significantly fewer fixations are generated overall, and there is less examination of nondiagnostic areas. Essentially, the residents’ search becomes much more efficient. These findings are similar to those in radiology, and support the theory that an important aspect of the development of expertise is improved pattern recognition (taking in more information during the initial Gestalt or gist view) as well as improved allocation of attention and visual processing resources. Progression in improvements in visual search strategies was similar, but not identical, for the 4 residents.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Key indications of expert interpretation of medical images are consistent, accurate and efficient diagnostic performance, which require not only dedicated training and experience but some degree of talent, aptitude and motivation [1]. A major question, however, is what are the best training methods and what types of experiences do trainees (ie, pathology residents) require in order to optimally develop their diagnostic skills? [2–5] As pathology is a visual specialty, it would seem that to optimize training, an understanding of how visual perception skills develop and change as a function of experience would be beneficial [6,7]. Surprisingly there has been little investigation on the development of visual expertise in pathologists. However, in recent years techniques used to study the development of visual expertise in another image-based medical specialty, radiology, have been successfully applied to pathology [6–9].

1.1. Visual information extraction and expertise

One of the most interesting research findings about expertise is the ability of the expert to extract the “gist” or key components of a visual scene after a single brief glance. For example, Kundel and Nodine [10] flashed chest
radiographs for 200 milliseconds to experienced radiologists (to prevent visual scanning) and found that 70% of the time they rendered correct interpretations. Classic studies of master level chess players found that masters were far better at recalling game piece positions after very brief 5-second presentations than less-skilled players [11,12]. Experts are better able to search, process, and interpret larger perceptual units than those with less skill because they are better able to recognize these units more efficiently and effectively as configurations or chunks of information rather than individual pieces. The mechanism behind this phenomenon is thought to be due to perceptual and/or cognitive tuning to the visual task as the observer encounters more exemplars over time with suitable training and feedback [1].

Pathology involves clinicians examining images, extracting relevant diagnostic information, and rendering diagnostic decisions. Thus, it seems likely that the development of expertise has a number of common underlying themes and/or mechanisms. A number of studies have been conducted to understand what makes a “good” radiologist or pathologist, and more generally, how we can assess whether someone has an underlying aptitude for interpreting medical images [6,7,13–16]. In radiology the examination of visual search patterns has been used as a tool for understanding development of expertise because the images are either film displayed on a view box or digital images on a computer so eye-tracking has been relatively easy [17–23]. Conventional light microscopy involves viewing histopathology glass slides through a light microscope which excludes eye-position recording. All of that changed in recent years as whole slide images became available, and studies of expertise in pathology are now appearing [6,8].

1.2. Using eye-tracking to study visual search and expertise

In 1963, Llewellyn-Thomas and Lansdown [17] conducted the first reported eye-position study in medical imaging. It demonstrated that search patterns are somewhat unique to the individual and tend not to be uniform in image coverage. Since then, a number studies have examined such issues as why errors occur (false negatives and false positives) [18–21], how experts differ from novices [22–24], and how different display parameters affect diagnostic accuracy and visual search efficiency [25,26].

With the development of technologies to create whole slide images (WSI), the study of expertise has expanded to include examination of the development of expertise in pathology. In 2003, Tiersma et al [27] found that pathologists viewing WSI generally adopt search strategies similar to those of radiologists. Readers are attracted to some slide/specimen locations for relatively short periods of time and to others for longer periods of time. Those examined for longer periods of time often contain more relevant diagnostic information. Roa-Pena [28] in a more recent study found (as did Kundel et al [19]) that areas examined for extended periods of time tend to be relatively common between readers. There are regions of interest (ROIs) containing diagnostic information that attract visual scrutiny across all readers. That is not to say that every reader looks at the exact same areas, but there are certain areas that attract attention more than others. These are referred to as common or coincidence ROIs, while areas examined by only one reader are sporadic or individual ROIs. Although these studies revealed some important aspects of visual search, neither dealt directly with expertise and/or experience of readers.

To address this we initially assessed eye movements of medical students, pathology residents, and practicing pathologists examining WSI [6]. There were significant differences in search behaviors as a function of level of experience. Fully trained pathologists spent significantly less time overall scanning WSI compared to residents or medical students, and spent the majority of the overall search time examining and selecting ROIs they would want to zoom in on to visualize diagnostic information and render a diagnostic decision. Residents and medical students also spent time examining (ie, fixating or directing high-resolution foveal vision to) the ROIs they eventually selected for zooming, but they also spent more time looking at other locations not subsequently selected. An examination of the saccades (eye movements made between fixations that move the eyes across image locations) revealed that pathologists had longer average lengths (seconds), shorter distances (degrees of visual angle), and faster velocities (length/seconds) than residents and medical students. These results are very similar to those found in search studies with radiologists and radiology residents [1].

1.3. The present study

The present study builds upon these initial observations and addresses the question of when do residents start to become more efficient in their search behaviors? The study longitudinally examined changes in search patterns of pathology residents as they progressed through their training program. The goal was to characterize changes that took place to determine (1) how long it takes before their patterns resemble those of expert pathologists and (2) what is the nature of those changes. In radiology, a key difference between novices and experts is search efficiency. Residents take longer to find relevant targets, search the entire image, make decisions, and generally look at more areas than experts [1]. A secondary question of the present study was to determine if the same is true in pathology.

2. Methods and materials

2.1. Images

Twenty breast core biopsy surgical pathology cases (half benign and half malignant cases verified by the original
Development of visual search expertise in viewing WSI

2.2. Readers and study protocol

Four pathology residents (readers) were recruited at the beginning (within the first month) of their first year in residency before any significant training. Each reader participated in a single session lasting approximately 45 minutes. Readers were fitted with the eye-tracker (Fig. 1) and told that they would be viewing images at low magnification one at a time. They were seated approximately 60 cm from the display for initial calibration, but afterwards could adjust their position forward or side-to-side as desired. They were instructed to select the top 3 locations that they would zoom onto if they were going to view the image in greater detail to render a diagnostic decision. Each reader saw the images in a different random order. The residents returned at the beginning of their second, third, and fourth years of residency and repeated this exercise with the same images re-randomized each time.

2.3. Eye-tracking

The ASL SU4000 Eye-Tracker system (Applied Science Labs, Bedford, MA) computes line of gaze and dwell time based on pupil and corneal reflection parameters. An infrared light-emitting diode and phototransistor detector are mounted on the headband (Fig. 1). Infrared light is emitted and reflects off a reflective visor into the left eye, reflecting back off the pupil and cornea to the visor, which then reflects it back to a charge-coupled device camera (complete details can be found in [6]).

The eye-position data were analyzed using standard methods [30]. Briefly, the accuracy of the system (spatial error between true eye position and computed measurements) is less than 1 degree. The SU4000 samples eye positions every 1/60 of a second to generate raw x-, y-coordinate eye-position data. Fixations are formed by grouping x- and y-coordinates of the raw data using a running mean distance calculation having a 0.58 radius threshold. Dwell time can be calculated for each fixation, summed across fixations, then associated with a given region of interest or location in the stimulus image. Using the conventional concept of the useful visual field, we correlated fixation data with the 3 image locations manually selected by each reader. If a fixation was within a radius of 2.5° of a marked location, it was a hit, with the restriction that it had to be on the same piece of tissue the reader marked. If multiple fixations were associated with a given location (ie, the observer looked at a location, looked somewhere else, and returned), they were grouped into fixation clusters and dwell times cumulated.

3. Results

3.1. Eye-position results

Fig. 2 shows the search patterns of one resident on the same WSI for Year 1 (A) through Year 4 (D). Circles represent fixations or locations where the eye lands with high-resolution foveal vision and lines show the order in which they were generated. Circle size reflects dwell time, with larger circles indicating longer dwells. Details about the changes averaged across readers are provided below, but these figures show that as the resident got more training and experience, there were definite changes. Search became more efficient—fewer fixations were generated, fewer locations revisited multiple times, individual locations received less intense scrutiny, and jumps between fixations (ie, saccades) were longer. Throughout the presentation of the results, the data are pooled across observers as there were no significant interobserver differences.
The eye-position data were analyzed using an analysis of variance. The first analysis examined total number of fixations per slide as a function of residency year. With each subsequent year there was a significant (F = 850.076, P < .0001) drop in average fixations per slide (Fig. 3); with a high of 129.64 (SD = 21.96) in Year 1 to a low of 24.63 (SD = 8.29) in Year 4. Post hoc protected least squares difference tests revealed that each year differed significantly from the other.

The second analysis examined total time spent dwelling (fixating) on the three ROIs selected for zooming, referred to as ROIs 1, 2, and 3 indicating the order they were selected. A 2-way analysis of variance revealed a significant main effect due to year in residency (F = 83.95, P < .0001) and a main effect due to location number (F = 108.08, P < .0001). With each year in residency they spent less total time dwelling on selected ROIs; and they tended to spend less time on each ROI (Fig. 4). The trend continued throughout subsequent years, but the difference between ROI vs non-ROI decreased.

The saccade data were analyzed using the parameters in the previous study [6]. Saccade length (seconds) significant increased with residency year (F = 23.30, P < .0001) (Fig. 5).

Saccade distance (degrees of visual angle) differed (F = 38.56, P < .0001) as a function of residency year. It was shortest in the fourth, longest in the second and third, and intermediate between these extremes in the first year (see Fig. 6).
There were also significant differences across all readers in saccade velocities (distance/sec) as a function of year of residency ($F = 387.43$, $P < .0001$). Velocities were shortest in the first year, increased in the second, fell slightly in year 3, then fell significantly in year 4 (Fig. 7).

The content of the ROIs was examined to determine (1) if there were consistencies across or between readers, and (2) if the content was diagnostically relevant. A board-certified pathologist reviewed each area to determine whether it contained diagnostically relevant information. In terms of consistency of ROIs chosen by an individual reader, 20% of the time they chose the same location in all 4 years; 51% of the time they chose a location from a previous year more often than selecting a new one; 19% of the time they chose approximately the same number of prior and new locations, and 10% of the time they selected more new than prior locations. Fig. 8 shows one slide with the locations selected by one resident across the 4 years.

In terms of diagnostic relevance, even at the beginning of the first year 96% of the ROIs selected contained diagnostically relevant information and that percentage steadily increased to 100% by year four ($\chi^2 = 3.54$, $P = .017$). The results are shown in Table.

### 4. Discussion

The results support those from our previous study [6] as well as those of Crowley et al. although they did not record search patterns but rather used a think-aloud protocol to assess how readers search slides [9]. Overall, it is clear that there are major changes in the way residents search pathology WSI as they progress through residency. With more experience, search patterns progress from resembling those of medical students to those of experienced pathologists. Search times decrease, time spent on areas selected as potentially diagnostic decreases, and saccades are more efficient (longer in time but shorter in distance, suggesting they are more purposeful). Therefore, the visual search of
WSIs becomes more efficient as a function of each passing year of residency. Residents appear to take in more information during the initial global or gist view [1,10–12], allowing them to efficiently and effectively scan only the most relevant image details. These changes in efficiency resemble those observed in radiology residents as they become more expert at viewing radiographic images. There seems to be a common pathway towards expertise characterized by increased efficiency in visual search strategies as residents become more familiar with the content to expect in images and what image details and characteristics are indicators of relevant information for rendering diagnostic decisions.

The changes in search and efficiency appear to occur at two points in time. Sometime between the first and second year (summer break perhaps) there is a change from rather haphazard jumping around the image and generating lots of fixations to a more systematic and efficient search with fewer fixations and longer (perhaps more directed) saccades between locations. Before this change, the total time spent on selected locations is rather high, indicating extensive attention paid to these locations while making the decision to select them for zooming. After the first year, however, there is a progressive increase in efficiency that extends through the third year. Residents generate fewer fixations, rapidly fixating on diagnostically relevant ROIs. They have shorter dwells on these locations indicating less hesitation in making the decision that this is where diagnostic information is likely to be. Comparing the beginnings of the third and fourth year, there is a further increase in efficiency. By the fourth year, residents are not yet at the point of completely duplicating search strategies of Board Certified pathologists, but they are clearly progressing in that direction.

As noted by Nodine [1], expertise in interpreting medical images occurs as a function of time but also as a function of the number of images viewed and interpreted in that time frame. As they become more familiar with what normal and abnormal, benign and malignant, and what one disease vs another looks like (as well as possible variations), they become more efficient at extracting relevant image features required to render accurate diagnoses. In this breast core biopsy series, the residents learned what tissue characteristics are more likely associated with a malignant vs benign breast tissue diagnosis. Thus, as time and residency training progress, they more quickly fixated on ROIs likely to contain useful diagnostic information rather than scan every possible location.

Interesting questions to consider are how many cases does a pathologist need to interpret in order to be considered proficient to function independently? These are difficult question to answer, but there are relevant studies. In radiologic interpretation of mammograms, for example, the Mammography Quality Standards Act requires interpreting physicians to have had at least 3 months of formal training in interpreting mammograms, a minimum of 60 hours of medical education in mammography, and have interpreted at

### Table

<table>
<thead>
<tr>
<th>Locations chosen in every year that were diagnostically relevant or not relevant</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevant</td>
<td>96.25%</td>
<td>99.58%</td>
<td>99.17%</td>
<td>100%</td>
</tr>
<tr>
<td>Not relevant</td>
<td>3.75%</td>
<td>0.42%</td>
<td>0.83%</td>
<td>0%</td>
</tr>
</tbody>
</table>
least 240 mammographic examinations within the immediate prior 6-month period to becoming certified [31]. In the United Kingdom mammographers are required to interpret 5000 exams per year [32]. Interestingly, however, Beam et al studied the association between self-reported annual interpretation case volume and radiologist accuracy in screening mammography and found no significant association between case volume and accuracy [33]. It may be feedback, rather than simple case reading volume is a stronger influence on the development of expertise in medical image interpretation [22,34,35].

There are few if any studies in the pathology literature on the number of cases according to specific types of organs or diagnosis residents need to review in order to be considered adequately trained. The Accreditation Council on Graduate Medical Education requires demonstration of competency in specific areas, that is, surgical pathology and cytopathology (more than simply completing a prescribed schedule of coursework in terms of weeks or months), but assessing competency in pathology in a reproducible manner has been a challenge [36,37], as has the definition of what actually constitutes an error [38]. Of note, however, are recent articles promoting the use of WSI for improved and more frequent training opportunities to enhance residents’ interpretation skills [37,39–42].

The present study is one of the first to examine search patterns of residents longitudinally through training. It is limited by the fact that we only studied the residents at the beginning of each year, but not at the end of the final year. Nevertheless, the trends were clear. The similarities among the four residents were striking and allow for making generalizations about the evolution of search strategies over time. This provides us with a unique window into the impact of training on search behaviors and the specific and quantifiable changes that tend to occur. Observations at shorter time intervals and in relation to specific resident rotations (ie, surgical pathology vs. blood banking) are required to more precisely understand when these transitions occur and whether they can be reliably correlated with other variables such as number of cases/images viewed or influence of non-job related factors such as aging. At some point, further study into the nature of the development of expertise in medical image interpretation may also provide us with the means to predict at the onset of training who is more likely to develop or enhance their search and interpretation skills and thus become an expert pathologist.

Acknowledgments

We would like to extend our thanks to the four residents who participated as readers in the study: Jennifer Gardetto, MD, Charles Mosher, MD, Claire Pollard, MD, and Carlo Valentino, MD. We also note that an early non-peer-reviewed version of the paper was published: Krupinski EA. On the development of expertise in medical imaging. SPIE Proceedings Human Vision & Electronic Imaging XVII, 2012:82910R.

References


