Using Eye Tracking to Evaluate and Develop Innovative Teaching Strategies for Fostering Image Reading Skills of Novices in Medical Training

Nora Castner¹, Shahram Eivazi¹, Katharina Scheiter², and Enkelejda Kasneci¹

¹ Perception Engineering, University of Tübingen, Tübingen, Germany
² Leibniz-Institut für Wissensmedien, Tübingen, Germany

Abstract. In the medical domain, developing skills, such as performing surgery, involves copious training and time. The eye movement behavior of experts during domain specific tasks shows measurable differences to those from a student. With this study, we analyze eye movement data from expert and novice microsurgeons while viewing surgical videos. Our results show that the level of expertise had no effect on fixations and saccades during the task. However, there was a likely effect of expertise on the relationship of saccade length and fixation duration. A comparative analysis of areas of interest with respects to the level of expertise was also performed and differences were found. Understanding the visual search strategies of experts in specific domains of the medical field can influence the training of students, such as models for gaze-based adaptive learning employed during training of medical students.

Keywords: Eye Tracking, Visual Search, Expertise, Microsurgery, Medical training, Gaze-based training

1 Introduction

Expertise remains an appealing topic; one of its most interesting research areas being devoted to measure the effectiveness of training programs on progression from novice to expert. Expertise, for any specific skill set, not only develops over time, but also requires intense practice. In the medical field, expertise has been shown to be related to faster and more accurate identification of anomalies, such as in radiograph interpretation [2, 14]. It has also been found that expertise relates to the efficiency of performance, where expert laparoscopic surgeons perform surgery more quickly than their novice counterparts [24] and were also more likely to have more fluid hand motions and more “elegant” surgical maneuvering [23].
Expertise in any domain is heavily reliant on visual processing, and notable differences are apparent regarding the eye movement behavior of both experts and novices. Experts hone their attention more on task-relevant information and reduce any noise from non-important information [4, 17]. Gegenfurtner and colleagues [7] assumed that these eye movement behaviors are indicative of rapid information processing (shorter fixation duration [8, 17]), selective attention allocation (fixation on relevant areas), and thorough global image analysis (longer saccade length [8, 17]). Differences between experts and novices could also be due to more systematic approaches employed by experts and novices’ lack of knowledge and practice [7, 8, 16].

Typically, eye movement behavior patterns, such as the aforementioned, are similar in experts across domains [1, 7]: Although task appropriate eye movements are also visible [3, 5, 8, 11]. For instance, expert laparoscopic surgeons focus their attention more to target locations while performing surgeries rather than to the tools employed whereas novice surgeons shift their gaze more often between tools and target locations [24]. On the other hand, Kundel and colleagues [13] found that visual search patterns of medical students resembled a localized central search method, whereas the search patterns of staff radiologists resembled a wider circular pattern. Consequently, the task-related eye movements during surgery are likely to be more procedural [21] in comparison to the analyzing and interpreting of radiographs, where the task-related eye movements are more reflective of employing visual search strategies [8, 17].

Recently, to develop systematic viewing strategy in students, low-cost interventions, such as viewing expert scanpaths [10], have been proposed to augment the conventional training methods. These gaze-based learning interventions have been found to improve systematic viewing strategies for both experts and novices, though task performance (e.g. anomaly perception) only improved for the experts [6, 10]. To improve student perceptual performance, more individualized feedback catered to each of the levels of understanding can be employed. Meaning a more high-cost gaze-based intervention that employs eye movement in order to adapt instructional feedback to guide the student based on learning models. Therefore, systematic viewing strategies and perception can progress in a more parallel fashion and create a more effective learning environment.

In this study, we explore the differences in eye movement behavior of experts and novices in the medical domain. Specifically, we looked at the eye movement data of microneurosurgeons with different levels of expertise while watching video segments of surgery. We expect that their eye movements may be more similar to those of a visual search task since they are watching a video rather than performing a surgical task. These eye movement patterns can offer insight into what we can expect for future studies: For instance, one concentrating on visual search and radiography. Since we know that eye movement behavior is indicative of expertise, we intend to support the previous literature and further supplement them with our data from microneurosurgeons. We also aim to explore not only expert novice differences but also differences in the intermediary level as well. Furthermore, we will look at expert areas of interest in contrast to non
expert areas of interest. We intend to gain insight from these area differences in order to set the foundation for future research regarding gaze-based learning paradigms for medical students.

2 Method

The experiment was conducted at the Neurosurgery Department of Helsinki University Central Hospital. Forty neurosurgeons ($M_{age} = 41.46$ years, $SD_{age} = 10$, 2 female) with varying degrees of expertise participated in the experiment. All participants had at least 2 years of experience in neurosurgery and performed at least 10 surgeries. The average surgeons’ years of experience was 13.17 ($SD = 1.59$) and average number of hours performing surgeries was 2213.5 hours ($SD = 2458.76$). Subjects were divided into three groups: Novice (students and residents with less than 1000 hours of surgery experience), intermediate (between 1000 and 2000 hours of surgery experience), and expert (with 2000 and over hours of surgery experience). The group divisions resulted in the intermediate group being the smallest, however it was determined based on the range of the self-reports. Data for one subject was excluded for the last video segment due to calibration errors.

2.1 Experiment design and apparatus

For recording the eye movement data, a Tobii T120 eye tracker with a 17 inch screen was used with a 60-65 cm viewing distance, 1024 by 768 pixel resolution, and 60Hz sampling rate. We recorded micro-neurosurgeons’ eye movements while they watched video segments of a brain aneurysm clipping. The video was divided into six segments, each with an average length of 20 seconds. After each segment, the video was paused and then during the frozen video image, participants answered three questions related to the state of the neurosurgery. The questions were related to a general awareness of the surgery situation, and prompted participants to orally describe 1) the objects just seen on the image, 2) the progress of the surgery, and 3) the expectations about the next step in the surgery. All participants answered these questions correctly.

Prior to the experiment, consent and medical background information were taken by the instructor. First, participants received a brief introduction to the experiment and then were asked to perform nine points calibration routine for the eye-tracker. After calibration, participants viewed an instruction slide and then proceeded to the video segments.

The study generated a large volume of data and a preliminary analysis indicated that there was great variety in the strategies exhibited. To focus our analysis in this paper, we employed only the data from the first and last video segments. We did not analyze the other segments because they exhibited a slight effect of blurring. Analysis of the gaze data from the chosen video segments was conducted with the software EyeTrace [12]. For simplicity and presentation of Area of Interest (AOIs), we chose to look at one random image from each segment rather than the whole video since the video clips were not highly dynamic.
2.2 Statistical analysis

Aligning with much of the previous literature, we measured eye movement differences in experts, intermediates, and novices. This measurement was based on two principle measures of visual attention: fixations and saccades. Specifically, we looked at the differences in number of fixations, mean fixation duration, mean saccades length, and mean saccade duration for both video segments. To evaluate whether segment presentation had an effect on eye movements, we compared eye movement variables between the first and last video segment. Then, for each video segment, we compared the eye movement behavior between expertise groups. Lastly, we measured the areas of interest (AOI) for each group over the time course of each video clip.

Using the EyeTrace software [12], which employs the I-BMM algorithm [19], fixations were calculated as having a minimum duration of 100ms and saccades were calculated from the time between two fixations. For the AOIs, we calculated gradient-based heatmaps from the fixation data for each group. Heatmaps were created for 5 second intervals for each video segment. Since the videos were not highly dynamic regarding dramatic movements, tool or otherwise, we chose a random frame from each segment for the AOI overlay visualization. The heatmap calculation based off of fixations is a two dimensional Gaussian distribution. For the calculations, we chose $\sigma = 10$ as determined empirically. We also defined fixation inclusion (a pre-threshold value in Eyetrace) to 5%, meaning at least 5% of the total fixations could be considered for an AOI candidate. Thus, for each of the 5 second intervals, this translated to an average minimum of two fixations.

3 Results

Eye movement data was compared between the first and the last video segment. A paired samples t-test revealed significantly more fixations in the first segment ($M = 30.74$, $SD = 9.25$) than in the last ($M = 26.10$, $SD = 7.79$); $t(38) = 2.5$, $p = .017$. There were also significant differences between the mean duration of fixations for the first ($M = 678.60$ ms, $SD = 274.96$ ms) and last segment ($M = 576.07$ ms, $SD = 179.90$ ms); $t(38) = 2.09$, $p = .044$.

Concerning saccades, both length and duration were shorter in the first segment than in the second: $M = 53.85$, $SD = 15.50$ and $M = 60.83$, $SD = 22.06$ for saccade length (in pixels) in both segments respectively and $M = 61.89$, $SD = 16.50$ and $M = 66.26$, $SD = 17.82$ for saccade duration (in ms) in the respective segments. However, there were no significant differences regarding both saccade measures in the first to last segment.

3.1 Expertise Differences and Eye Movement Behavior

Table 1 details the eye movement features for each expertise group. Regarding expertise level, a one-way ANOVA revealed no significant differences regarding a main effect of expertise on eye movements behavior in either video segment.
(see Table 2). Saccades and fixations during either video segment were then not indicative of level of expertise. Consequently, we looked at the correlation between fixations and saccades to see whether level of expertise had an effect on fixations and saccades.

Table 1: Mean and standard deviations of subgroups

<table>
<thead>
<tr>
<th>Variable</th>
<th>Segment 1</th>
<th>Segment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Novice</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Num. Fixations</td>
<td>31.44 (8.45)</td>
<td>27.86 (7.56)</td>
</tr>
<tr>
<td>Mean Dur. Fixations</td>
<td>607.53 (178.58)</td>
<td>521.83 (180.03)</td>
</tr>
<tr>
<td>Mean Len. Saccade</td>
<td>54.96 (18.26)</td>
<td>55.52 (11.12)</td>
</tr>
<tr>
<td>Mean Dur. Saccade</td>
<td>60.97 (9.17)</td>
<td>71.36 (25.54)</td>
</tr>
</tbody>
</table>

\[ a \] Duration is in ms, length is in pixels
\[ b \] Calibration issue for one participant resulted in data loss for one subject.

Table 2: One way ANOVA effects of expertise on eye movements

<table>
<thead>
<tr>
<th>Variable</th>
<th>Segment 1</th>
<th>Segment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Num. Fixations</td>
<td>( F(2,39) = .483, p = .621 )</td>
<td>( F(2,38) = 2.465, p = .085 )</td>
</tr>
<tr>
<td>Mean Dur. Fixation</td>
<td>( F(2,39) = .277, p = .760 )</td>
<td>( F(2,38) = .542, p = .586 )</td>
</tr>
<tr>
<td>Mean Len. Saccade</td>
<td>( F(2,39) = .142, p = .868 )</td>
<td>( F(2,38) = 1.424, p = .254 )</td>
</tr>
<tr>
<td>Mean Dur. Saccade</td>
<td>( F(2,39) = 2.138, p = .132 )</td>
<td>( F(2,38) = .122, p = .886 )</td>
</tr>
</tbody>
</table>

\[ a \] Levene’s Test for Homogeneity of Variance with \( F(2,37) \) revealed equal variances could be assumed for all variables, \( p > .05 \)
\[ b \] Levene’s Test for Homogeneity of Variance with \( F(2,36) \) revealed equal variances could be assumed for all variables, \( p > .05 \)

A partial correlation analysis revealed that in the first video segment, hours of surgery experience had an effect on the correlation of saccade length and fixation duration \( (r = -.394, p = .014) \) as well as saccade length and number of fixations \( (r = .470, p = .003) \). However, the effect was not significant in the last video segment. Years of experience also revealed an effect on the same correlations as well: \( r = -.400, p = .013 \) and \( r = .476, p = .003 \) respectively. Figure 1 depicts the relationship between saccade length and fixation duration as well as number of fixations. It is worth noting that without the effect of level of expertise, both saccade length and fixation duration for the first video segment was significant \( (r = -.405, p = .01) \) and saccade length and fixation number for the first segment had significant correlations independent on level of expertise \( (r = .479, p = .002) \).

From the correlation analysis, it is clear that longer saccade lengths correlate to both shorter fixation durations and higher numbers of fixations for the first video segment. These correlations were in turn affected by level of expertise,
Fig. 1: Correlation of eye movements for the first video segment: Mean fixation duration and mean saccade length (left) and mean number of fixation numbers and mean saccade length (right). Level of expertise is marked on the graph and shows its effect on both correlations.

though it is unclear as to how much since correlations existed independent of expertise.

3.2 AOI Patterns of Eye Movements Over Time

To measure the effect of level of expertise on areas of interest over the time course of the video segments, we calculated gradient-based heatmaps based on the fixation information for 5 second intervals in each segment. Even though the video was not highly dynamic, differences in the areas of interest with respect to level of expertise is apparent along with their changes over time. Figure 2 shows the differences in AOI shapes and clustering as a result of level of expertise for the last video segment. Here, we notice that the overall positioning of the AOIs has a high overlap, which is due to the surgical procedure remaining in a condensed area of the images (i.e. relatively static). However, we can see that the overall number of AOI clusters for the novices is higher than that of the experts, with 25 clusters for novices, 29 for the intermediates, and 17 for the experts. To an extent, the size of the main central focal areas, or largest clusters, are larger for the novices compared to both intermediates and experts. These trends in the size of the clusters were also apparent in the first segment, though there was an exception regarding number of clusters: with 39, 23, and 29 for novices, intermediates, and experts respectively.

4 Discussion

In this study, we used eye tracking technology to quantify neurosurgeon’s eye movements when watching video segments of a microscopic brain surgery. We
found that experts are more likely to employ both shorter and fewer fixations in association with longer saccades. This differs from other studies that showed experts employ shorter fixations and longer saccades [7,10,16,21]. In our dataset the gaze measures were not significantly affected by the expertise level of the surgeons, however there is an effect of expertise on the relationship between fixations and saccades.

There are important dependencies contingent on the progress of the treatment and this is reflected in the subject viewing behavior in the first stage of surgery compared to last stage. We exposed these intricate relationships by analyzing eye movements in time. We found that initially when watching the surgery segments, all participants had significantly higher and longer fixations for the first segment compared to last segment. The highest difference were in the mean number of fixations (8.77) for the novice group. This could indicate novices employed less fixations over the time course as a result of acquisition to the task. In this sense, task acquisition was also apparent for the experts, though to a lesser extent (mean difference of 4).

Number of fixations can be indicative of the information-reduction hypothesis, so that experts may reduce the number of fixations on areas of redundant information and centralize their fixations on only the relevant areas [7]. This model could also be relevant for temporal viewing behavior as we noticed by the trends in our data, such that selective attention becomes more precise over time, though possibly faster for experts.
It is possible, that in the first 20 seconds, even an expert may take some time to process the task at hand, though he or she may be inherently faster at detecting the relevant areas as opposed to the novice. In our data, the mean fixation duration data also supported the information-reduction hypothesis, but not in support of expertise effects. Fixation duration was significantly shorter for all groups in the last video segment, meaning relevant areas were determined faster by the last video segment. This overall decrease in fixation duration can suggest changes in long-term working memory regardless of group [7]. Here again, experts had a lower mean fixation duration compared to novices in the last video segment (not significant statistically), indicating they may be more likely to process information more efficiently.

We reported that expertise affected the eye-movements mainly regarding fixation areas. The size and the clustering of the AOIs differed between groups as well as over time. Although, the video images were highly localized and exhibited minimal tool use, there were time intervals where novices’ fixation regions revealed attention to the tool and sometimes more clusters outside the central focal area when compared to intermediates and experts in the same time interval. Overall, this behavior was found for both segments.

Concerning the experts and the novices, their areas of attention can be indicative of their cognitive processing of the task at hand. The experts tended to attend to areas organized as “chunks” close to each other and novices determined a large main area to selectively focus then scattered their attention to smaller areas near the outer regions of this central focal area. A more in-depth analysis can offer better insight into areas of interest as determined by level of expertise. Such an analysis could be comparing the area of AOI clusters between groups or even center extraction of clusters and then looking at the distance from the nearest expert cluster.

From the AOI data, we see that the intermediate group employed more concentrated fixations. This behavior can be attributed to two presumptions, one being that the surgeons grouped as intermediate, though less experienced in terms of total hours of surgery, still have a wealth of experience. This group may have also taken the experiment more seriously compared to the experts. If an expert assumes that the task is too simple or contextually different, he may be less likely to provide an accurate representation of his expertise. To our knowledge, many of the studies regarding expertise may not take into account the level of challenge and expert may need to illicit generalizable performance. The second presumption for the intermediate eye movement behavior could be that there were only seven participants compared to the 16 and 17 in the novice and expert groups. Recruiting more participants meeting the criteria for an intermediate microneurosurgeon sample could possibly provide a more comprehensive understanding of AOIs for this level of expertise.

The eye movement behavior of all groups could have been influenced by the context: Subjects were watching a video of microneurosurgeries and not actually performing the surgeries themselves. Actually performing the surgeries creates an environment that facilitates context based behaviors, such as hand or eye
Eye Tracking and Innovative Teaching in Medical Training

movements [18, 22]. Simply watching a video may not be enough to evoke the natural task-based behaviors. From our data, we had initially suspected that we would not see precise surgery-based eye movements, but rather general eye movements indicative of expertise. More importantly, we expected to see more eye movement behavior related to interpretation of the task. Generally, a surgeon is debriefed in advance concerning what type of surgery he needs to perform, but if the surgeon is rather watching with little context, he or she responds differently to the stimuli. For our data, we asked the subjects to watch the video and interpret the surgery.

Visual attention skills, both procedural and analytical, are essential for developing and mastering surgical techniques [9, 15, 20, 24]. In microneurosurgery, moreover, the learning curve for eye-hand coordination is long and may depend on the subjective psychomotor abilities. Surgeons have to learn to scan across the operating field while concentrating on tiny objects in the operating field. As such, they have to maintain attention to follow the procedure and avoid obstacles as they safely move the instrument through delicate structures. Our work employed gaze-based measures that are more likely to be indicative of visual search strategies for interpreting and analyzing these structures. Measuring the eye movement data in this context as perceptual and analytical rather than procedural offers a new understanding of search strategies that are employed by microneurosurgeons. Consequently, both these forms of visual attention and their interplay in both expert and novice surgeons becomes evident. In the general medical domain, the importance of perceptual aspects of visual attention is a concept of interest, since it is crucial to accurate medical image processing, such as interpreting radiographs.

4.1 Future work

Gegenfurtner and colleagues [7] promote that for any skill learning, training aids often direct novices on specific information to focus on and in which order they should focus on it. When students are trained to employ expert visual search strategies, improvements in their perceptual performance were measurable [8]. For instance, Eye Movement Modeling Examples is a training technique in which the experts’ visual behavior (e.g. scanpaths) can guide learners’ visual attention to relevant areas [10]. This model and other gaze-based adaptive training techniques are of interest because this information can facilitate learning.

We propose a more high-cost intervention that employs a gaze model to determine the deviation the current scanning and the scanning of a slightly advanced subject and visualizes that difference. Such a training technique would look similar to the one portrayed in Figure 3, where information of expert scanpath models is presented during a training session in the form of salient features when the student’s gaze deviates to a certain extent. Initially, intensity differences would be very minimal, however, if the novice appears lost or confused, the intensity could then increase. Thereby, the subject can derive that the displayed scanning is advantageous over his current approach and ultimately adapt his strategy. The idea behind this approach is that the gaze pattern is altered directly during
the fluent scanning movement which can lead to a better incorporation in the students own strategy.

![_intensity overlay of expert scanpaths during novice training for 1s interval_](image1.png) ![Adaptive feedback as intensity increase if novice appears lost](image2.png)

Fig. 3: Example of gaze based adaptive training simulation. Salient features of expert scanpaths may change as a result of adaptive gaze feedback indicating that the novices has deviated from the model.

The results from this work are likely to be indicative of eye movement behaviors we can expect to find regarding experts and novices employing visual search strategies on the medical imaging fields such as radiograph. We propose a future study with a concentration on radiograph interpretation among dentistry students, novice and advanced, and certified practitioners and will attempt to model expertise development. We aim to develop a technique to investigate and model the change in viewing behavior during expertise development so it can be used for assessment of medical practitioners and furthermore can be employed as a learning tool for students. The proposed study intends to evaluate massed practice, domain knowledge and task experience. Then, develop a high-cost intervention based on eye movement modeling. This intervention will be compared against other gaze-based interventions (e.g. EMME).

Reading radiographs, especially for dentists, is non-trivial due to the routine use and likelihood for detection errors [16], research into effective training techniques is pivotal. From this work with microneurosurgery, we can expect that expert defined AOIs during a time interval can be different to the those defined by novices. Real time evaluation of the novice differences could be performed; then, slight intensity indicators can guide the novice towards more expert strategies. In the longterm, this proposed research will build off of the greatly understood work of visual behavior and expertise, our current findings on expertise and area of interest, and finally adaptive learning environments and it’s impact on medical image analysis.
References


