Modeling Gaze Behavior for Real-Time Estimation of Visual Attention and Expertise Level in Augmented Reality

Dong Woo Yoo*
Northeastern University

Hamid Tarashiyoun†
Northeastern University

Mohsen Moghaddam‡
Northeastern University

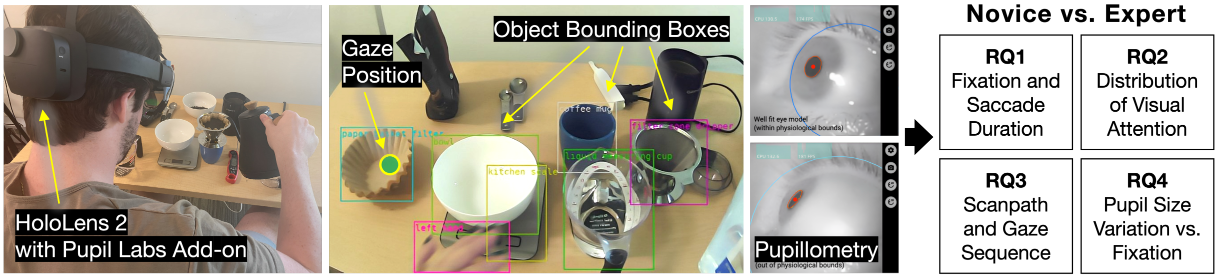

Figure 1: Overview of the study apparatus (left), gaze tracking and pupillometry data (middle), and research questions (right).

Abstract

Augmented reality (AR) technologies have recently gained substantial attention within the industry due to their potential applications in on-the-job training and assistance across diverse industrial settings. However, personalizing AR instructions and feedback interventions that cater to individual user needs and skill levels remains a relatively less explored area of research. This paper aims to bridge this gap by utilizing eye tracking data coupled with computer vision to examine the gaze and pupil behaviors of individuals with various levels of expertise performing AR-guided procedural tasks. The main goal is to investigate the relationship between eye tracking data, visual attention, and expertise by exploring four research questions associated with (1) differences in fixation and saccade duration between novices and experts, (2) variations in visual attention allocation to action-relevant areas of interest (AOI) between novices and experts, (3) the influence of expertise on scanpath and transitions between AOIs, and (4) the correlation between pupil size variations and fixation/saccade behaviors. The findings of a study on humans that focused on two procedural tasks are reported. The study uses synchronized gaze, pupillometry, and egocentric videos to analyze gaze interactions with AOIs and background stimuli based on object detection models. This research advances our understanding of the relationship between gaze behaviors, visual attention, and expertise, thus offering new insights into enabling adaptive and personalized interventions in AR. These insights are specifically relevant to AR use cases centered on training or on-the-job assistance.

Index Terms: Eye tracking—Pupillometry—Visual attention—Expertise

1 Introduction

Augmented reality (AR) technologies have garnered considerable attention in recent years due to their potentially transformative capabilities for training and assistance in diverse industrial environments [1]. However, a relatively less explored avenue is how AR can effectively customize instructions and feedback interventions to accommodate the specific needs and skill levels of individual users during training or task execution. Personalized scaffolding is crucial [2] as it aligns with users' attention and expertise levels to help them construct knowledge [3]. However, current methods largely offer generic "one-size-fits-all" instructions, disregarding the unique needs and knowledge of the individual. The potential of AR to provide personalized and adaptive learning experiences is evident, especially when visual gaze overlays and instructional cues are used in real time to improve learning and task performance. To dynamically adapt interventions according to users' unique needs and preferences in AR, it is essential to ascertain their level of task-specific expertise and attention patterns in real-time. This paper delves into the gaze and pupil behaviors of users equipped with AR headsets to differentiate between novices and experts during AR-guided execution of procedural tasks. The primary objective is to link eye-tracking metrics with task performance, paving the way for algorithms that adaptively guide user attention in AR based on expertise. For example, in surgery, distinct visual attention patterns emerge between novices and experts as the complexity of the task increases [4]. Another surgical study revealed how the level of expertise of a surgeon dictates the distribution of attention during procedures [5]. Similarly, in aviation, the correlation between a pilot's gaze and performance has been studied, spotlighting the role of experience in flight attention [6,7]. Furthermore, innovative research in various fields has underscored the value of gaze-centric displays in the construction of intuitive, attention-aware systems [8,9]. This study focuses on using these insights to enhance AR experiences tailored to individual skill levels.

Despite significant recent progress in AR eye tracking research, there is still a lack of systematic modeling of gaze and pupil behavior during AR-guided tasks. Drawing on data from advanced eye-tracking technology in AR, the findings presented in this paper will serve as the basis for intelligent and adaptive AR systems for training and task assistance. These systems, designed with the capability to dynamically adjust in real-time based on the user's expertise and attention, promise to offer robust and reliable results that can further augment the existing body of knowledge.

1.1 Research Questions

Previous research in this field has made some advances in understanding and evaluating visual attention and expertise using various eye-tracking technologies [5, 7, 10]. Yet, current knowledge lacks a

*e-mail: yoo.d@northeastern.edu
†e-mail: tarashiyoun.s@northeastern.edu
‡e-mail: mohsen@northeastern.edu

comprehensive understanding of the intricate details regarding gaze interactions with areas of interest (AOIs), as well as pupil dilation or contraction at the action level, and their correlation with the user's attention and skill level during task execution. Driven by this motivation, our objective is to advance the understanding of the following research questions.

> **RQ1. Fixation and Saccade Duration** Do novices and experts demonstrate significant differences in the average duration of fixations and saccades while performing AR-guided procedural tasks?

The eye-mind hypothesis posits a direct connection between the visual fixation of the eyes and the cognitive processing of the mind [10]. That is, fixation is known to be correlated with cognitive workload and attention. On the other hand, saccades are commonly attributed to the degree of uncertainty or hesitation exhibited by an individual while tackling complex tasks, where a longer duration of saccades indicates a greater degree of indecisiveness [11]. However, there is a lack of agreement on the relationship between expertise and the durations of fixation and saccade. According to the hypothesis of long-term working memory, it is anticipated that more skilled individuals will exhibit shorter fixation durations compared to novices [12]. This is attributed to the experts' ability to efficiently and quickly retrieve the necessary information from their long-term working memory, thanks to their domain-specific knowledge and extensive experience [13]. However, there is an opposing viewpoint that suggests longer fixation durations provide more stability for information processing, potentially leading to improved outcomes. That is, the relationship between gaze behaviors and expertise is influenced by the learners' familiarity with the task and its complexity. This provides a strong impetus to delve deeper into this research question to gain a more comprehensive understanding of how fixation and saccade behaviors are related to the level of expertise.

> **RQ2. Distribution of Visual Attention** Is there a significant difference in the level of visual attention experts allocate to action-relevant AOIs compared to novices?

The distribution of visual attention refers to how the user's gaze points are spread across various sources of information, indicating the general allocation of visual focus [14]. There is evidence that experts tend to focus more on relevant information, while novices are frequently distracted by irrelevant information [7]. Specifically, according to the information reduction hypothesis, experts aim to concentrate their processing efforts on task-relevant AOIs while minimizing the allocation of visual attention to redundant information [15]. An AOI refers to a specific portion of a stimulus that captures the user's visual attention and plays a crucial role in investigating the disparities between experts and novices using eye tracking techniques [7]. The information reduction hypothesis also explains the inconsistent findings regarding differences in fixation duration between experts and novices. Experts may deliberately extend their fixation duration on AOIs relevant to the task, while reducing their processing time on less important areas. However, in the context of AR-guided task execution, it is crucial to verify this relationship, as AOIs are not predetermined and must be identified and examined in tandem with action-level gaze patterns. This allows for an understanding of the duration of user fixation on AOIs that are relevant to the task, as well as those that are irrelevant to the task and background stimuli.

> **RQ3. Scanpath and Gaze Sequence.** Does the level of expertise influence the frequency and temporal sequence of transitions between AOIs?

A scanpath refers to the sequential pattern of fixations and saccades, represented by the temporal sequence of gaze coordinates [16]. Assessing the expertise level of an individual based on their gaze behavior during a specific action is heavily based on this crucial metric. Experts are expected not only to be familiar with the relevant AOIs but also to demonstrate a clear understanding of the sequential order in which these AOIs are utilized during the action. According to the holistic image perception model, experts exhibit superior efficiency compared to novices in processing stimuli at both the global and local levels [17]. They possess the ability to quickly analyze the overall scene at a glance, enabling them to distinguish important information from irrelevant details. Experts possess advanced visual recognition skills, which enable them to rapidly and accurately recognize meaningful information and their relationship compared to novices. But a greater understanding of this hypothesis is still required, particularly within the context of AR-guided task execution in unstructured and dynamically evolving environments.

> **RQ4. Pupil Size Variation.** Is there a correlation between changes in pupil size and fixation or saccade behaviors?

The pupil diameter, as a physical attribute, can serve as a valuable component in user modeling for AR. It is influenced by various factors, from lighting conditions to emotions, cognitive workload, and attention [18]. By monitoring the changes in pupil diameter, an AR system can gather information about the user's level of attention and potentially even their emotional state. Also, it can be utilized to detect the user's focal point on specific AOIs, enabling the AR system to deliver more pertinent information or engage with virtual objects more seamlessly and intuitively. However, there is insufficient evidence on the potential correlation between variations in pupil size and the durations and patterns of fixations or saccades. This lack of evidence hinders our understanding of whether longer fixations during a specific action truly indicate cognitive processing and attention.

1.2 Proposed Work

This paper presents the findings of exploratory research conducted to address the aforementioned research questions. The study examined the gaze behavior of a group of participants while performing two procedural tasks: cooking and making coffee. These tasks were specifically designed by the MIT Lincoln Laboratory to evaluate DARPA's Percentually Enabled Task Guidance (PTG) program [19]. Specifically, we collected synchronized gaze, pupillometry, and egocentric videos from the participants, and implemented an object detection model to analyze frame-wise interactions between gaze and AOIs as well as background stimuli. A pre-study questionnaire was administered to collect participants' self-reported expertise in performing the two tasks. In this context, expertise was characterized as the level of experience and proficiency attained in the tasks, considering both the quality and the quantity of experience [20]. All participants received initial training on the two tasks, the AR app and the HoloLens 2 headset. Data were collected and then used to quantify several metrics required to answer RQs 1-4.

The key contributions of this paper are summarized as follows:

- We propose a new framework for data-driven modeling of user attention and expertise level in AR based on gaze and pupil tracking data captured from commercial AR headsets, as well as a computer vision method for the study of fixation on task-related and action-related stimuli. We validate the framework through a study of human subjects.
- We contribute to the understanding of fixation and the duration of saccades. Experts exhibit longer fixation durations and

more accurate saccades compared to novices, indicating differences in attentional concentration and information processing efficiency.

- Our study contributes to the field of visual attention by revealing that experts allocate their attention more effectively to action-relevant AOIs compared to novices. This highlights the role of expertise in guiding attentional skills and has implications for training programs targeting visual attentional skills.
- We provide insights into scanpath and gaze sequences, showing that experts display more efficient and deliberate scanpaths with fewer revisits and smoother transitions between tasks. This emphasizes the importance of expertise in organizing gaze behavior and suggests the need for improved attention guidance strategies for novices.
- Our study addresses the correlation between pupil size variation and gaze behaviors, revealing significant associations between changes in pupil size and fixation or saccade behaviors. This highlights the role of cognitive processes, attentional demands, and expertise in shaping pupil responses during task performance.

2 Materials and Methods

In this section, a detailed description of the study materials and methods is presented. This includes a detailed description of the user study and the procedures implemented for data collection, as well as the computational techniques developed for data analysis and synthesis. Additionally, the metrics established to tackle the research questions specified in Section 1 are described.

2.1 User Study

Participants. 15 students were recruited to participate in the study; 5 participants were experts (1 woman, 4 men) and 10 novices (5 women, 5 men), all graduate students. Prior to the experiments, a pre-study questionnaire was conducted to collect their perceived level of expertise in performing the two tasks. In this context, "expertise" refers to individuals who have the necessary knowledge and significantly more extensive and superior experience to perform the two tasks [20]. Also, all participants received preliminary training on the tasks, the AR app, and the HoloLens 2 headset through the HoloLens Tips App.

Apparatus. The study involved the execution of two procedural tasks of making pour-over coffee and pinwheels with the help of an AR app (see Figure 1). Participants manually interacted with the AR app instructions, using hand gestures to navigate through the training content. The app offered a variety of instructions for each step, including guides, checklists, safety instructions, and videos (Figure 2.1). To collect data, we use the HoloLens 2 headset equipped with the Pupil Labs add-on, enabling the capture of gaze and pupillometry data (Figure 2.2). In addition, egocentric videos were recorded to identify AOIs within each frame. Participants' interactions with AR content were recorded and gaze data was collected using the built-in eye tracker of the HoloLens 2 headset to investigate their visual engagement with AR content. The eye trackers of the Pupil Labs add-on operate at a sampling rate of 120 Hz, while the world camera records at 200 Hz and were synchronized using Unix timestamps.

Tasks. The sequence of actions for the two tasks is described as follows: [19]. *Pour-over coffee task*: (1) Measure 12 ounces of cold water. (2) Transfer the water to a kettle. (3) Place the dripper in the mug. (4) Take the paper filter and fold it in half, creating a semicircle. (5) Fold it in half again to form a quarter-circle shape. (6) Position the paper filter inside the dripper and shape it into a cone. (7) Empty the coffee beans from the bag. (8) Weigh 25 grams of coffee beans. (9) Add the coffee beans to the coffee grinder. (10) Grind the coffee

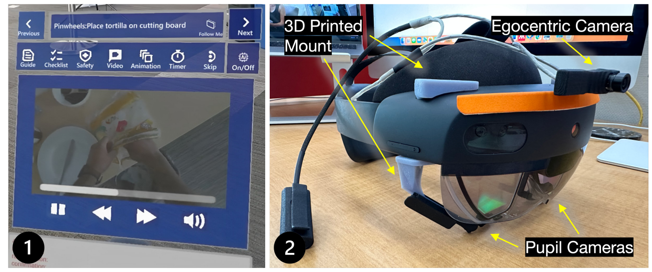

Figure 2: Data collection setup. (1) The AR interface provides guidance, a checklist, safety instructions, videos, animations, and a timer per step. (2) HoloLens 2 equipped with a Pupil Labs add-on to collect, gaze, and pupil as well as wide-POV egocentric videos.

for a duration of 20 seconds. (11) Transfer the ground coffee to the filter cone. (12) Verify the temperature of the water. (13) Pour a small amount of water over the coffee grounds. (14) Slowly pour the remaining water in circular motion. (15) Dispose of the paper filter and the coffee grounds used. (16) Hold the cup of coffee in front of you. *Pinwheels task*: (1) Place the tortilla on a cutting board. (2) Use a knife to gather the peanut butter. (3) Spread the butter on the tortilla. (4) Clean the knife. (5) Use the knife to gather jelly. (6) Spread the jelly on the nut butter. (7) Roll the tortilla. (8) Insert a toothpick. (9) Trim the end of the roll. (10) Discard both ends. (11) Place the floss underneath the tortilla and cut using the floss. (12) Place the pinwheels on a plate.

Procedure. After completing the prestudy questionnaire, the participants received a brief introduction to the task and the AR app from the researchers. To ensure their familiarity with head mounted AR, particularly HoloLens 2, participants underwent a 10-15-minute training session using the HoloLens Tips app. They will also perform a standard calibration process for the Pupil Labs add-on eye trackers to ensure accurate measurements. Following this initial preparation, each participant performed each task once, following the instructions provided by the AR app. We considered unchanged laboratory lighting throughout the experiments. Additionally, the AR app did not include any large or shiny holograms that could affect pupil size. We recorded their interactions with the user interface and captured gaze data using the HoloLens, as well as the gathered gaze, pupillometry, and egocentric videos using the Pupil Labs add-on. The logged data from the HoloLens were used to label and segment gaze, pupillometry, and egocentric video data according to task steps (i.e., fine-grained actions), enabling us to calculate the completion time for each step. At the end of the session, participants completed a post-study questionnaire that included reporting their cognitive load (NASA-TLX) (Table 2), self-efficacy, experience with the HoloLens/AR app, and providing general feedback through structured and open-ended questions.

Figure 3: Materials for the pour-over coffee task (left) and the pinwheels task (right). Courtesy of DARPA PTG [19].

2.2 Modeling and Analysis

Data Preprocessing and Synchronization. Data from eye tracking and egocentric videos were synchronized to study visual behavior in AR tasks. Eye-tracking data were filtered for precision and then calibrated to the AR coordinates. The pupil data, including variations and timestamps, was aligned with the eye-tracking data. Egocentric videos helped to pinpoint AOIs and background stimuli. We utilized an object detection model to identify objects in each video frame and aligned this with gaze and pupil data. The resulting comprehensive dataset contained gaze positions, fixations, saccade durations, pupil size variations, and related AOIs and stimuli. The quality-aligned data facilitated a deep exploration of gaze behavior during AR tasks. AOIs were selected based on their relevance to each step of the task, guided by a previously established task model with detailed instructions.

Object Detection Model. Using the egocentric camera on the AR headset, we used a top-tier computer vision model [21]. This model, Faster R-CNN, efficiently detects objects. Comprising two parts: the region proposal network (RPN) and a region-based convolutional network, it identifies potential objects and classifies them with precision. Trained on the extensive COCO dataset, the model can recognize a wide array of object categories. During live AR operations, the model assesses each video frame and predicts object presence and locations, seamlessly merging with gaze data to observe user interactions with AOIs and stimuli.

AFD and ASD Estimation. AFD captures the average time that participants' gaze stays fixated on a specific area during AR tasks. In our study, the data were classified by subject, video, and step. Using Pupil Lab software, we confirmed fixations and determined the average fixation duration for each step. We then averaged these durations to obtain the AFD for each video. The AFD was computed separately for both experts and novices.

ASD measures the average duration between rapid eye movements at different focal points. This metric reflects the participants' visual scanning speed and efficiency during task transitions. Using Pupil Lab data, we identified each fixation's duration and calculated the time between consecutive fixations to determine the saccade length. The mean of these lengths gave the ASD for each video. The study also derived separate ASD averages for experts and novices.

Gaze Distributions and Scanpaths Inference. To understand participants' visual attention and cognitive approaches during AR-guided tasks, we analyzed gaze distributions and scanpaths using eye tracking data. This revealed the areas that most captured the participants' focus and how they distributed their attention. We sourced gaze positions from a Pupil Lab file, which offered timestamps, gaze coordinates (x, y), and a confidence value for accuracy. The timestamp with the highest confidence was used per frame.

Using an object detection model, we identified collisions between gaze positions and objects in each frame. This enabled analysis of gaze interactions with both step-related and nonstep-related objects, facilitating comparisons between novices and experts.

For scanpath creation, we observed changes in gaze direction between object pairs. A 2D list documented transitions of gaze between object IDs. By analyzing this data, we generate a scan path showing the sequence of gaze movements across objects.

Pupillometry. We analyze pupil diameter changes to gauge participants' cognitive processing and arousal. Pupil dilation indicates cognitive load, while constriction suggests focused attention or familiarity with the task. We used pupillometry data to better understand participants' visual and cognitive reactions, enhancing our evaluation of their task expertise. The data was sourced from Pupil Lab's exported file, which contains timestamps, confidence percentages, and pupil dilation durations. Only data with a confidence level above 60 % (as per Pupil Lab's recommendation) was considered. Post-filtering, we computed the average and standard deviation for pupil sizes, then incorporated these findings into our main analysis.

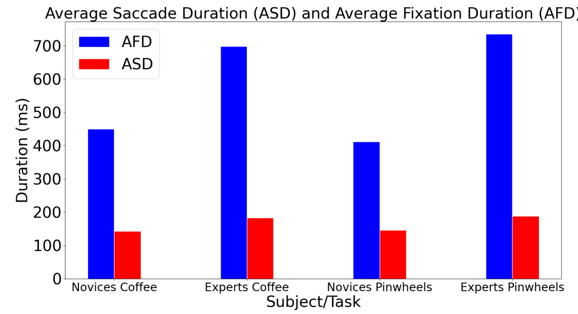

Figure 4: AFD and ASD results for the two tasks (RQ1).

2.3 Metrics

Average Fixation Duration (AFD). Guided by the eye-mind hypothesis [10], we studied the average fixation duration, considering the fatigue of the eye from the AR headset interactions.

Average Saccade Duration (ASD). We assessed learner indecisiveness through saccade durations, factoring in potential eye fatigue from AR headset use [11].

Visual Distribution. We analyzed gaze distribution among various stimuli [14], measuring the AFD on action-relevant/irrelevant/background sources using object detection and gaze positioning.

Scanpath Transitions. Our scanpath analysis, derived from fixation and saccade sequences, delves into transitions between stimuli types and discerns cognitive strategies from gaze patterns [16]. This paper [4] highlighted its relevance in distinguishing between expert and novice surgeons.

Pupil Size Variation. Through eye-tracking, we examined variations in pupil size, shedding light on cognitive load, arousal, and attentional engagement levels. Changes in pupil size are indicative of cognitive demands and emotional responses during the execution of the task.

3 Results

This section presents the findings obtained from the analysis of the collected data, focusing on addressing the research questions outlined in the study.

Data quality control measures are designed to maintain the integrity and reliability of the dataset and minimizing potential confounding factors.

3.1 RQ1: Fixation and Saccade Duration

Figure 4 details the measurements for the coffee and pinwheel tasks, differentiating between novices and experts. For each task, we note the AFD and ASD metrics and summarize the data for the two levels of expertise. Our t-test analysis uncovers significant differences in fixation durations (AFD: t =-10.68, p = 0.0087) and saccade durations (ASD: t = -12.28, p = 0.0066) between the groups.

The analysis highlights that experts tend to have longer and more focused fixations, paired with efficient gaze transitions. On the contrary, novices show briefer fixations and saccades, indicative of less precise attention and gaze patterns. These distinctions underline the impact of expertise on attention efficiency and offer insights for crafting specialized training programs. Such findings can boost novice performance and accelerate their transition to expert levels, enriching overall skill growth.

Table 1: Frequency of visits between AOIs.

Task	Novices	Experts
Pour-over coffee task	2822	520
Pinwheels task	2797	2266.6

3.2 RQ2: Distribution of Visual Attention

Figure 5 illustrates that experts and novices differ in their focus during tasks. The t-test showed that experts and novices allocate visual attention differently to action-relevant AOI (p = 0.0121). However, for fixed frames without step-related steps, both groups were similar (p = 0.8136). Similarly, the attention given to fixated frames related to steps was comparable between both groups (p = 0.4226). In essence, while experts and novices differ in overall attention to action-relevant AOIs, they show no difference for non-step or step-related frames. In the coffee task, the results of the t-test reveal negligible differences between the groups in the distribution of attention to background stimuli, non-step-related AOIs and step-related AOIs, suggesting that both groups allocate attention similarly in this task.

3.3 RQ3: Scanpath

The scanpath analysis brings insights into the sequential gaze behavior exhibited by the participants. The findings reveal that experts exhibit a lower number of gaze transitions in both the coffee and pinwheel tasks (Table 1). This suggests that novices engage in active scanning and exploration of objects on average, potentially indicating a more thorough allocation of attention. Conversely, the reduced frequency of gaze transitions among experts implies a more focused or selective attention strategy. The greater frequency and rapidity of gaze transitions between objects in novices may signify a heightened cognitive load, as users process and integrate information from multiple sources, like AR instructions. On the contrary, the reduced number of gaze transitions may indicate a lower cognitive load, suggesting a task that is comparatively simpler or less demanding.

3.4 RQ4: Pupil Size Variation

The accuracy and precision of video-based gaze estimation are known to be influenced by the fluctuations in pupil size observed during fixation [22]. Fixation of pupil dilation introduces errors in estimating the gaze direction, reducing accuracy and precision. This study [23] found that pupil dilation was more pronounced for familiar stimuli, indicating its potential as an objective measure of cognitive familiarity. The data presented in Figures 6 and 7 illustrate variations in pupil size between novices and experts during coffee and pinwheel tasks. Notably, experts exhibited a significant peak in pupil size when searching for the garbage disposal location (Steps 9-10 of the Pinwheels task), which involved nonstep related AOIs. This observation suggests that cognitive familiarity influenced the variation in pupil size among experts. The ANOVA in Figure 8 highlights the link between pupil size during tasks and cognitive familiarity, with dilation suggesting increased cognitive engagement.

Table 2: Average NASA TLX responses per group.

Question (0: Very Low - 10: Very High)	Novices	Experts
How mentally demanding were the tasks?	0.912	1.622
How physically demanding were the tasks?	2.316	1.277
How hurried or rushed was the pace of the tasks?	0.824	1.255
How successful were you in accomplishing what you were asked to do?	8.698	7.012
How hard did you have to work to accomplish your level of performance?	1.774	1.535
How insecure, discouraged, irritated, stressed, and annoyed were you?	0.502	0.4

4 Limitations and Future Work

Despite the promising findings reported, this research has certain limitations. It focuses on two procedural tasks, questioning its applicability to various AR settings. The precision of eye tracking, especially during fast eye movements, could affect the reliability of the data. Classifying participants as 'novices' and 'experts' could overlook nuanced gaze behaviors of intermediate skill levels. The limited sample size and task-specific focus mean findings may not be universally applicable. Future studies should include diverse samples, diverse tasks, and deeper measures to better understand gaze behavior and cognitive correlations. The proposed data-driven expertise modeling approach will lay the foundation for future AR-based industrial training systems to allow adaptive and personalized interventions tailored to the level of expertise of individual users. Novices may need more support, while experts need fewer interventions to prevent distraction or overreliance on AR. Future AR-based industrial training systems can facilitate personalized, proficiency-based, dynamic training and on-the-job assistance.

5 Conclusions

The exploratory study presented in this paper shows that the gaze and pupillometry behaviors of users during AR-guided task performance can be used effectively to estimate their level of expertise. Specifically, experts demonstrated more efficient gaze strategies, marked by longer fixation duration associated with task/step-relevant areas of interest (AOIs) as well as more intentional saccades between the AOIs. That is, experts demonstrated fewer transitions between AOIs compared to novices. Experts also showed higher variations in pupil size during less familiar steps, which may correspond to variations in cognitive load. This is a preliminary effort that aims to automate the process of estimating the level of expertise based on

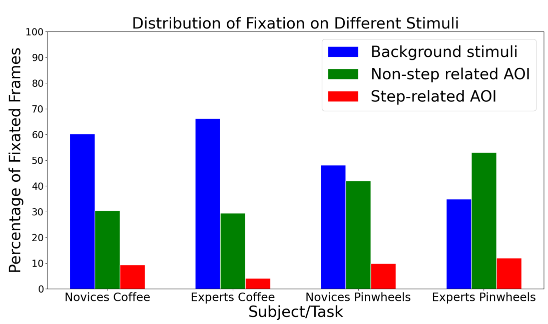

Figure 5: Distribution of visual attention between different stimuli: step-related, non-step-related, and background (RQ2).

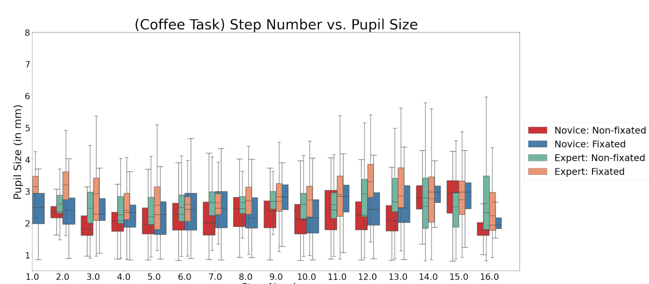

Figure 6: Variations in pupil diameter throughout different steps of the Pour-Over Coffee task for novices versus experts (RQ4).

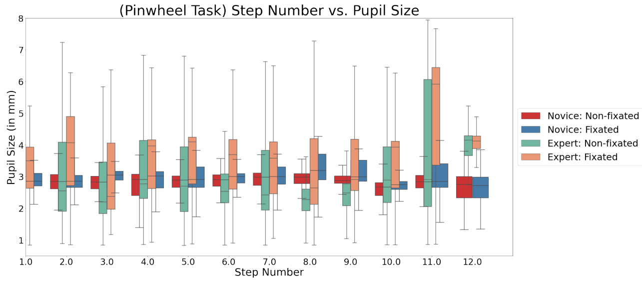

Figure 7: Variations in pupil diameter throughout different steps of the Pinwheels task for novices versus experts (RQ4).

eye tracking data in AR to effectively adapt the level of guidance and instructions to the individual needs of the user and their familiarity with different steps of the task. Future research will build on the findings reported in this paper to develop new objective metrics and data-driven models to estimate expertise and integrate them into AR systems designed for training and assistance.

Acknowledgments

This work is supported by the DARPA Perceptually Enabled Task Guidance (PTG) Grant No. HR00112220001 and the NSF Grant No. FW-HTF-2128743. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the NSF or DARPA.

References

[1] Mohsen Moghaddam, Nicholas C Wilson, Alicia Sasser Modestino, Kemi Jona, and Stacy C Marsella. Exploring augmented reality for worker assistance versus training. *Advanced Engineering Informatics*, 50:101410, 2021.

[2] Tzu-Chiang Lin, Ying-Shao Hsu, Shu-Sheng Lin, Maio-Li Changlai, Kun-Yuan Yang, and Ting-Ling Lai. A review of empirical evidence on scaffolding for science education. *International Journal of Science and Mathematics Education*, 10:437–455, 2012.

[3] María-Blanca Ibáñez and Carlos Delgado-Kloos. Augmented reality for stem learning: A systematic review. *Computers & Education*, 123:109–123, 2018.

[4] Shan Li, Melissa C Duffy, Susanne P Lajoie, Juan Zheng, and Kevin Lachapelle. Using eye tracking to examine expert-novice differences during simulated surgical training: A case study. *Computers in Human Behavior*, 144:107720, 2023.

[5] Tony Tien, Philip H Pucher, Mikael H Sodergren, Kumuthan Sriskandarajah, Guang-Zhong Yang, and Ara Darzi. Differences in gaze behaviour of expert and junior surgeons performing open inguinal hernia repair. *Surgical endoscopy*, 29:405–413, 2015.

[6] Gal Ziv. Gaze behavior and visual attention: A review of eye tracking studies in aviation. *The International Journal of Aviation Psychology*, 26(3-4):75–104, 2016.

[7] Stephanie Brams, Gal Ziv, Oron Levin, Jochim Spitz, Johan Wagemans, A Mark Williams, and Werner F Helsen. The relationship between gaze behavior, expertise, and performance: A systematic review. *Psychological bulletin*, 145(10):980, 2019.

[8] Andrew P Bayliss, Emily Murphy, Claire K Naughtin, Ada Kritikos, Leonhard Schilbach, and Stefanie I Becker. "gaze leading": Initiating simulated joint attention influences eye movements and choice behavior. *Journal of Experimental Psychology: General*, 142(1):76, 2013.

[9] Alexander Toet. Gaze directed displays as an enabling technology for attention aware systems. *Computers in Human Behavior*, 22(4):615–647, 2006.

[10] Marcel A Just and Patricia A Carpenter. A theory of reading: from eye fixations to comprehension. *Psychological review*, 87(4):329, 1980.

[11] Eric Castet and Guillaume S Masson. Motion perception during saccadic eye movements. *Nature neuroscience*, 3(2):177–183, 2000.

[12] KA Ericsson and W Kintsch. Long-term working memory. *Psychological review*, 102(2):211–245, 1995.

[13] Gonca Gokce Menekse Dalveren and Nergiz Ercil Cagiltay. Insights from surgeons' eye-movement data in a virtual simulation surgical training environment: effect of experience level and hand conditions. *Behaviour & Information Technology*, 37(5):517–537, 2018.

[14] Jan-Louis Kruger, Esté Hefer, and Gordon Matthew. Attention distribution and cognitive load in a subtitled academic lecture: L1 vs. l2. *Journal of Eye Movement Research*, 7(5), 2014.

[15] Andreas Gegenfurtner, Erno Lehtinen, and Roger Säljö. Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational psychology review*, 23:523–552, 2011.

[16] Shivsevak Negi and Ritayan Mitra. Fixation duration and the learning process: An eye tracking study with subtitled videos. *Journal of Eye Movement Research*, 13(6), 2020.

[17] Harold L Kundel, Calvin F Nodine, Emily F Conant, and Susan P Weinstein. Holistic component of image perception in mammogram interpretation: gaze-tracking study. *Radiology*, 242(2):396–402, 2007.

[18] Siyuan Chen and Julien Epps. Using task-induced pupil diameter and blink rate to infer cognitive load. *Human–Computer Interaction*, 29(4):390–413, 2014.

[19] Bruce Draper. Perceptually-enabled Task Guidance. `https://www.darpa.mil/program/perceptually-enabled-task-guidance`, 2021. [Online; accessed 1-June-2023].

[20] Norbert M Seel. *Encyclopedia of the Sciences of Learning*. Springer Science & Business Media, 2011.

[21] Zekun Zhang and Minh Hoai. Object detection with self-supervised scene adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21589–21599, June 2023.

[22] Kyoung Whan Choe, Randolph Blake, and Sang-Hun Lee. Pupil size dynamics during fixation impact the accuracy and precision of video-based gaze estimation. *Vision research*, 118:48–59, 2016.

[23] Léon Franzen, Amanda Cabugao, Bianca Grohmann, Karine Elalouf, and Aaron P Johnson. Individual pupil size changes as a robust indicator of cognitive familiarity differences. *PloS one*, 17(1):e0262753, 2022.

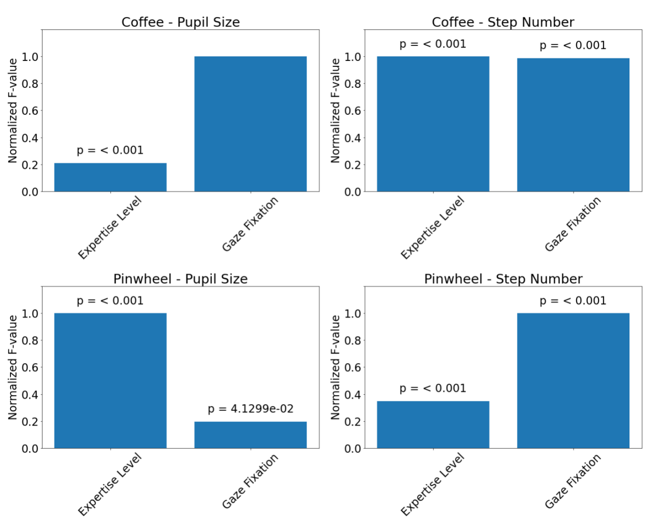

Figure 8: ANOVA results for variations in pupil size (RQ4).