



HHS Public Access

Author manuscript

Proc Hum Factors Ergon Soc Annu Meet. Author manuscript; available in PMC 2024 October 25.

Published in final edited form as:

Proc Hum Factors Ergon Soc Annu Meet. 2023 September ; 67(1): 953–958.

doi:10.1177/21695067231192929

Shifting Perspectives: A proposed framework for analyzing head-mounted eye-tracking data with dynamic areas of interest and dynamic scenes

Haroula M. Tzamaras¹, Hang-Ling Wu², Jason Z. Moore², Scarlett R. Miller¹

¹Pennsylvania State University Industrial Engineering

²Pennsylvania State University Mechanical Engineering

Abstract

Eye-tracking is a valuable research method for understanding human cognition and is readily employed in human factors research, including human factors in healthcare. While wearable mobile eye trackers have become more readily available, there are no existing analysis methods for accurately and efficiently mapping dynamic gaze data on dynamic areas of interest (AOIs), which limits their utility in human factors research. The purpose of this paper was to outline a proposed framework for automating the analysis of dynamic areas of interest by integrating computer vision and machine learning (CVML). The framework is then tested using a use-case of a Central Venous Catheterization trainer with six dynamic AOIs. While the results of the validity trial indicate there is room for improvement in the CVML method proposed, the framework provides direction and guidance for human factors researchers using dynamic AOIs.

INTRODUCTION

While researchers have been trying to decode people's thoughts for years, actual mind reading is still decades away (Kioustelidis, 2011). For now, we have to rely on other methods to understand human cognition. One such method is eye-tracking which is based off of the belief that movement of the eye is correlated to mental processing (Pfu yczka, 2018). The field of eye-tracking started with the observation of eye movement using mirrors. After this, light was incorporated. The original light-based eye-trackers included a chin rest and required the head to be completely still to be used accurately. It wasn't until the 90s that immobilizing the heads of participants became unnecessary (Pfu yczka, 2018).

Modern eye-tracking devices use cameras that can be wearable or remote that use infrared light to track both the movement of the eyes and when they are still, often referred to as eye fixations (Holmqvist & Andersson, 2011; Morimoto & Mimica, 2005). There are two main types of eye trackers that are commonly used today, head-mounted eye trackers that a participant wears, and computer-mounted or remote eye trackers that are fixed on a surface, usually a screen (Morimoto & Mimica, 2005). The scenes in eye-tracking studies can be static or dynamic. Static scenes refer to stationary images whereas dynamic scenes refer to videos or situations with interactivity and movement.

Fixations are useful because measuring their duration can help assess levels of comprehension or interest for a task, and tracking the movement of the eyes can help understand relationships between fixations and how the gaze pattern moves throughout time, referred to as gaze mapping or fixation sequences (Plu yczka, 2018; Rayner et al., 2006). Areas of interest (AOIs), or specific locations where the eyes may become fixated, are commonly used as markers of importance when researchers are studying a specific procedure or process (Ashraf et al., 2018). The main metric of interest could be AOI hits, or number of times the gaze coincided with an AOI (Holmqvist & Andersson, 2011). When these metrics are collected in dynamic scenes, it adds complexity to the analysis (Kok & Jarodzka, 2017).

Because of its utility, eye-tracking has been used in a variety of fields, from aviation to social interaction; eye-tracking has been used to understand attention and learning (Rosch & Vogel-Walcutt, 2013; Schilbach, 2015; van de Merwe et al., 2012). An increasingly important application of eye-tracking is its use in medical training (Kok & Jarodzka, 2017). Eye-tracking in medical education has been used to understand student learning in cadaver dissection (Sánchez-Ferrer et al., 2017), operator perception during diagnostic interpretation (Brunyé et al., 2019), and expertise and skill level (Tien et al., 2014). For some medical procedures, eye-tracking has been used to determine when a trainee has reached an expert level of performance (Chen et al., 2019). The most used eye-tracking metrics in medical training are fixations on areas of interest (Ashraf et al., 2018; Fichtel et al., 2019), length of fixation (Kok & Jarodzka, 2017), and fixation sequences (Kok & Jarodzka, 2017).

Tobii Pro Lab, a robust analysis software, is commonly used for eye-tracking data, but each added area of interest adds extra time and subjectivity into the analysis (Jongerius et al., 2021). In Tobii Pro Lab, an eye-tracking video can be manually mapped, meaning the person doing the mapping draws the AOIs on the first frame and then checks frame by frame that the outlines they drew in the first frame still line up with the AOIs stopping to modify size, location, and orientation of the AOI labels. Tobii Pro also has the option of automapping AOIs, meaning the person draws the AOIs on the first frame and then Tobii automatically tracks them throughout the recording. Additionally, several researchers have developed their own analysis tools for complex eye-tracking studies, however, in these frameworks the eye-tracker was fixed, the AOIs were located on a screen, or the application was so specific that the analysis method could not be modified for other purposes, like medical training (Fichtel et al., 2019; Jongerius et al., 2021; Kumari et al., 2021).

The drawbacks with current analysis methods lead us to one major question which is: can we automate the process for analyzing remote eye-tracking data based on AOI hits when the axis and AOIs are moving from frame to frame and the overall scene is dynamic? For the remainder of this paper, we propose and test a framework for automating this type of eye-tracking analysis.

COMPUTER VISION + MACHINE LEARNING (CVML) EYE-TRACKING FRAMEWORK

To overcome the shortcomings of traditional eye-tracking analysis for more complex problems, we developed a framework to automatically locate and track moving AOIs. The goal of this framework was its specific application to medical training, and as such, eye-tracking videos from a medical simulator were used for all model training. The steps of the framework are summarized in Figure 1.

Step 1

The preliminary step in our framework is to clean and prepare the gaze data that came from the eye-tracking glasses. For recording, we used a Tobii Pro Glasses 3 with 50 Hz gaze collection and a scene camera of 1920×1080 pixels and 25 frames per second (fps). The gaze data output from Tobii includes a .txt file with timestamp and 2D gaze data, 3D gaze data, and left eye and right eye gaze origin, direction and pupil diameter information for each point. For mapping AOIs, the data of interest is the 2D gaze coordinates output in XY format. The file is imported into excel and the data is cleaned so the only remaining columns are the timestamp and the X and Y coordinates.

Step 2

The second step in our framework is to utilize Matlab (R2022a) to plot the gaze coordinates on the gaze video from the glasses as a red circle to match what the Tobii software does. This is done in the code in four distinct steps:

1. *Break into frames* - The frame rate of the video is approximately 25 fps. The video needs to be broken down into frames at 25 fps so that each frame can be plotted with the appropriate coordinates.
2. *Find true X and Y for each gaze coordinate* - The coordinates are recorded by the glasses at 50 Hz, but the true rate when accounting for error is 49.84935. Because the sampling rate is not exactly twice that of the video frame rate, interpolation is used to find the true coordinate values for each frame to be able to plot the circle.
3. *Plot true X and Y for each frame on image as circle* - On each frame, a circle is drawn around the true coordinates with a radius of 30, chosen arbitrarily, and a new image with the circle printed on it is saved.
4. *Recombine annotated images back into a video* - To see the gaze moving in real time and to analyze the areas of interest with computer vision more efficiently, all red circle images were recompiled into a video at 25 fps.

Step 3

The third step in our framework is to build a computer vision machine learning (CVML) model in Roboflow. For this application, two separate computer vision models were used. The first computer vision model was previously developed for tracking tools during the use

of a medical simulator to give live feedback during training (Brown & Wu, 2022). This model was expanded to also include the computer used in the trainer pictured in Figure 2.

The second computer vision model was developed solely for this framework and its purpose is to identify the red circle, which represents the gaze point, in the image. Output from these respective models can be seen in Figure 2. When the CVML model detects an object, it creates a bounding box around that object and provides a percentage of confidence that it has labeled that object correctly. For each prediction that it makes, it provides the coordinates and dimensions of the bounding box in addition to the confidence. For our models, we used the Roboflow default confidence value of 50, meaning that the bounding box was only created if the model was at least 50% certain that it was labeling correctly.

Step 4

Finally, to employ the CVML models, python was used to compare the two models to each other and see if the input matches. To do this, both models were imported into Pycharm. The python code pulls in the video file output by Matlab, and provides a detection output frame by frame. The Python code compares the coordinates of each bounding box from the computer vision models on the image, and if there is any overlap between the red circle bounding box and that of any of the tools (**refer to** Figure 2), the code outputs what was detected and otherwise reports no detection.

CASE STUDY: Medical simulation

In central venous catheterization (CVC), a complex medical procedure with a high complication rate, training is important because the more experienced the operator of the procedure is, the less likely they are to have adverse outcomes (McGee & Gould, 2003). With CVC, the most difficult skills for trainees to master are mechanical, indicating the importance of understanding when and how to use each medical tool (McGee & Gould, 2003). Figure 3 shows the tools of importance during the CVC procedure as exhibited by a CVC simulator and labeled in the order of use.

The application of eye-tracking to CVC training is complex. If eye-tracking were to be applied to analyze the acquisition of specific mechanical skills, each tool labeled in Figure 3 would need to be considered a separate AOI. This adds complexity to the situation because eye-tracking studies generally rely on larger AOIs to avoid excess noise during analysis (Hessels et al., 2016). Additionally, the medical tools each move at least once during the procedure as they are being used by the doctor. Because the tools are moving, the doctor needs to be able to move their body around freely; therefore, the eye-tracker needs to be head-mounted which means that the axis is shifting throughout the recording.

Considering the complications with analyzing eye-tracking data taken from CVC training, we wanted to see if the CVML analysis framework could be applied. In this case study, we considered three separate analysis methods. The first method was AOI tracking conducted in Tobii Pro by a manual rater. The second method was AOI tracking conducted in Tobii Pro by its automatic detection software (auto method) and not adjusted by a manual rater. The third method was AOI tracking conducted by the CVML framework. The manual method

was used as the control method; it was checked by a rater frame by frame and is therefore the most accurate.

Metrics

The AOIS used in this case study were the computer, the needle, the guidewire, the scalpel, the dilator, and the catheter (**refer to** Figure 3). These tools were chosen because they are all significant aspects of CVC and the process of learning the mechanical skills of the procedure (Graham et al., 2007). The two metrics of interest used in this case study were total number of AOI hits and level of agreement between methods. An AOI hit is anytime the gaze is located on an AOI in a single frame. Total number of hits per AOI is a useful metric because it shows how many times the gaze was located on a certain AOI and allows direct comparison between the output of the three models. Agreement between methods, based on Cohen's Kappa of interrater reliability, is a useful metric because it allows us to understand if the automatic and CVML methods are different from the control method.

Model Analysis

The same two-minute eye-tracking video was run through the three methods of interest. For the Tobii manual method, one independent rater mapped AOIs. This was done by going through the video frame by frame for 2994 frames and correcting the AOI outlines each frame to make sure they were lined up with each tool. For the auto mapped case conducted in Tobii Pro, AOIs were drawn in the first frame by the same independent rater and the rest were auto calculated by Tobii based on where the AOIs were in the first frame. For the CVML framework, the video was input to the python code. All output was in the form of AOI hits per tool for each frame where a 1 indicated alignment of the gaze and the AOI and a 0 indicated no alignment.

Next, the metrics of interest were calculated from the AOI hits. For total AOI hits in each model, the sum of all 1 values of each tool were added together. Cohen's kappa for interrater reliability was found by comparing the total output for each tool frame by frame in the CVML method and the automatic Tobii method to the control.

Results

The total number of AOI hits found by each model for each tool can be seen in Figure 4. The control method and the CVML framework found a similar number of hits for the computer and the catheter whereas the auto method was not close for any of the tools. The auto method also did not detect any AOI hits for the needle, scalpel, dilator, or catheter.

Cohen's Kappa was run in SPSS to measure agreement through interrater reliability between the various methods for each tool compared to the control. The data met all required assumptions. All results from the Cohen's Kappa and the strength of each agreement (Landis & Koch, 1977) can be found in Table 1. The auto method had fair agreement for the computer ($\kappa = .375$, $p < .05$) and poor agreement for the guidewire ($\kappa = -.037$, $p < .05$). The CVML method had very good agreement for the computer ($\kappa = .831$, $p < .05$) and poor agreement for the guidewire ($\kappa = .190$, $p < .05$). Agreement with all other tools was poor ($p > .05$).

Discussion

The results of the CVML framework indicate that it was more precise for AOI detection compared to existing automapping tools. Specifically, the CVML framework identified all six of the AOIs while the automapping feature was unable to detect four of the six AOIs (needle, scalpel, dilator, and catheter). This may be due to the fact that the automapping software relies on the automatic AOI tracking, which is not always able to account for smaller shapes and parts that move often (Hessels et al., 2016). Additionally, the automapping method detected the computer over 400 times, likely because of its large area compared to some of the other tools (refer to Figure 2).

In addition, the CVML framework had a higher level of agreement with the control condition compared to the automapping condition. While the automapping had significant agreement with the control condition for two AOI's, one was only "fair" agreement (computer) while the second was not only "poor" agreement, but the negative kappa represented less than chance agreement. On the other hand, for CVML framework the level of agreement was very good for "computer". In addition, for the guidewire, the level of agreement was approaching "fair" agreement (0.20). However, agreement was not significant for four of the six AOIs.

There are several reasons why the AOI hits for the CVML model may not have reached significant agreement with the control. First, the threshold for confidence in the CVML framework was set to 50%. In other words, if the framework was detecting the gaze point (represented by a red circle) or a tool (AOI) and was less than 50% confident in its selection, it was not listed as being seen by the model and thus no bounding box was created. In addition, there may have been inaccuracies in the integration of the Computer Vision and Machine Learning components. For example, there were times where the CVML model would detect the gaze point but not the AOI, or vice versa, and thus would not return a hit.

The CVML framework introduced here provides direction and guidance for human factors researchers interested in analyzing gaze data for dynamic scenes. The results of this study identify three promising areas that deserve further investigation. First, there were some strength in agreement between the CVML framework and the control method, indicating that additional work on the CVML framework may lead to increases in inter-rater agreement and thus accuracy. Second, this framework provides a method for automatic analysis in experimental setups where the eye tracker is head-mounted and the person is doing an active task. Finally, the CVML framework is an efficient means of analyzing these complex datasets – the case study presented here took less than an hour to analyze AOI hits, which is significantly more efficient than manual mapping which took the rater 3 hours and 53 minutes. The manual method took substantially more time because of AOI drift that occurred from frame to frame. For example, Figure 5 shows AOI drifting over the span of 1 second, with the images being taken every 5 frames. Because of this drift, the manual rater needed to make modifications to the AOIs almost every frame. This type of manual coding is infeasible when analyzing gaze data of larger participant pools.

While the CVML framework shows promise, there are several limitations of this study. First, the manual eye tracking analysis method that was used as the control condition has been shown to be the most accurate for analyzing eye gaze data, but it is not without criticism (Jongerius et al., 2021). However, there is currently no better option to ensure complete accuracy. Additionally, only one video was tested on the proposed framework, and a separate video was used to train the model. More trials need to be conducted with more videos to determine the transferability of these results.

Future work should focus on improving the accuracy of the CVML model and the attributes. There were several cases where the CVML model did not detect the red circle. This could be due to the size and appearance of the red circle, which was not varied or tested during this case study. More shapes, colors, and sizes of the circle will be tested to determine if this impacts the detection.

CONCLUSION

This paper outlines the CVML framework and an initial validation through a use-case. This framework uniquely applies to eye-tracking with dynamic areas of interest. This CVML framework is necessary because the existing analysis methods either do not account for an axis that shifts every frame, or take an infeasible amount of time to apply. The CVML framework was shown to have higher detection and accuracy compared to other automatic eye-tracking analysis methods. However, there are areas for improvement in this methodology. Future work should focus on increased methods of identifying gaze points and CVML model robustness.

ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health (NIH) under Award Number RO1HL127316. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Coauthors Jason Moore and Scarlett Miller own equity in Medulate, which may have a future interest in this project. Company ownership has been reviewed by the University's Individual Conflict of Interest Committee.

REFERENCES

- Ashraf H, Sodergren MH, Merali N, Mylonas G, Singh H, & Darzi A (2018). Eye-tracking technology in medical education: A systematic review. *Medical Teacher*, 40(1), 62–69. 10.1080/0142159X.2017.1391373 [PubMed: 29172823]
- Brown D, & Wu H (2022). DMD2022–1020 COMPUTER VISION ENABLED SMART TRAY FOR CENTRAL VENOUS. 3–6.
- Brunyé TT, Drew T, Weaver DL, & Elmore JG (2019). A review of eye tracking for understanding and improving diagnostic interpretation. *Cognitive Research: Principles and Implications*, 4(1). 10.1186/s41235-019-0159-2
- Chen HE, Sonntag CC, Pepley DF, Prabhu RS, Han DC, Moore JZ, & Miller SR (2019). Looks can be deceiving: Gaze pattern differences between novices and experts during placement of central lines. *American Journal of Surgery*, 217(2), 362–367. 10.1016/j.amjsurg.2018.11.007 [PubMed: 30514436]
- Fichtel E, Lau N, Park J, Henrickson Parker S, Ponnala S, Fitzgibbons S, & Safford SD (2019). Eye tracking in surgical education: gaze-based dynamic area of interest can discriminate adverse events and expertise. *Surgical Endoscopy*, 33(7), 2249–2256. 10.1007/s00464-018-6513-5 [PubMed: 30341656]

- Graham A, Ozment C, Tegtmeyer K, Lai S, & Braner D (2007). Central Venous Catheterization. 10.1056/NEJMvcm055053
- Hessels RS, Kemner C, van den Boomen C, & Hooge ITC (2016). The area-of-interest problem in eyetracking research: A noise-robust solution for face and sparse stimuli. *Behavior Research Methods*, 48(4), 1694–1712. 10.3758/s13428-015-0676-y [PubMed: 26563395]
- Holmqvist K, & Andersson R (2011). *Eyetracking: A comprehensive guide to methods, paradigms and measures* (Issue March). Oxford University Press.
- Jongerius C, Callemeyn T, Goedemé T, Van Beeck K, Romijn JA, Smets EMA, & Hillen MA (2021). Eye-tracking glasses in face-to-face interactions: Manual versus automated assessment of areas-of-interest. *Behavior Research Methods*, 53(5), 2037–2048. 10.3758/s13428-021-01544-2 [PubMed: 33742418]
- Kioustelidis J (2011). Reading minds. *New Scientist*, 210(2818), 32. 10.1016/S0262-4079(11)61504-2
- Kok EM, & Jarodzka H (2017). Before your very eyes: The value and limitations of eye tracking in medical education. *Medical Education*, 51(1), 114–122. 10.1111/medu.13066 [PubMed: 27580633]
- Kumari N, Ruf V, Mukhametov S, Schmidt A, Kuhn J, & Küchemann S (2021). Mobile eye-tracking data analysis using object detection via YOLO v4. *Sensors*, 21(22), 1–17. 10.3390/s21227668
- Landis R, & Koch G (1977). An Application of Hierarchical Kappa-type Statistics in the Assessment of Majority Agreement among Multiple Observers Author (s): Landis J. Richard and Koch Gary G . Published by: International Biometric Society Stable URL: <https://www.jstor.org/stab>. *Biometrics*, 33(2), 363–374. [PubMed: 884196]
- McGee DC, & Gould MK (2003). Preventing complications of Central Venous Catheterization. *The New England Journal of Medicine*, 348(12), 1123–1133. 10.1016/j.jamcollsurg.2007.01.039 [PubMed: 12646670]
- Morimoto CH, & Mimica MRM (2005). Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding*, 98(1), 4–24. 10.1016/j.cviu.2004.07.010
- Plu yczka M (2018). The First Hundred Years: a History of Eye Tracking as a Research Method. *Applied Linguistics Papers*, 4/2018(25), 101–116. 10.32612/uw.25449354.2018.4.pp.101-116
- Rayner K, Chace KH, Slattery TJ, & Ashby J (2006). Scientific Studies of Reading Eye Movements as Reflections of Comprehension Processes in Reading. *Scientific Studies of Reading*, 10, 241–255. 10.1207/s1532799xssr1003
- Rosch JL, & Vogel-Walcutt JJ (2013). A review of eye-tracking applications as tools for training. *Cognition, Technology and Work*, 15(3), 313–327. 10.1007/s10111-012-0234-7
- Sánchez-Ferrer ML, Grima-Murcia MD, Sánchez-Ferrer F, Hernández-Peñalver AI, Fernández-Jover E, & Sánchez del Campo F (2017). Use of Eye Tracking as an Innovative Instructional Method in Surgical Human Anatomy. *Journal of Surgical Education*, 74(4), 668–673. 10.1016/j.jsurg.2016.12.012 [PubMed: 28126379]
- Schilbach L (2015). Eye to eye, face to face and brain to brain: Novel approaches to study the behavioral dynamics and neural mechanisms of social interactions. *Current Opinion in Behavioral Sciences*, 3, 130–135. 10.1016/j.cobeha.2015.03.006
- Tien T, Pucher PH, Sodergren MH, Srisakandarajah K, Yang GZ, & Darzi A (2014). Eye tracking for skills assessment and training: A systematic review. *Journal of Surgical Research*, 191(1), 169–178. 10.1016/j.jss.2014.04.032 [PubMed: 24881471]
- van de Merwe K, van Dijk H, & Zon R (2012). Eye Movements as an Indicator of Situation Awareness in a Flight Simulator Experiment. *International Journal of Aviation Psychology*, 22(1), 78–95. 10.1080/10508414.2012.635129

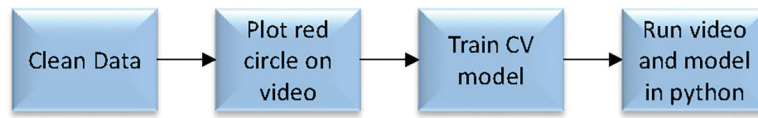


Figure 1:
Summarized steps of proposed framework

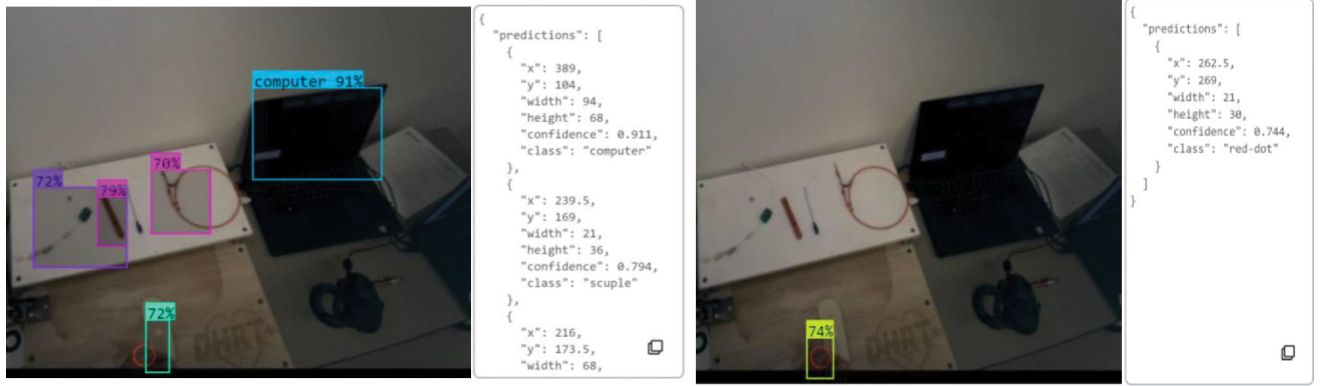


Figure 2:
(left) output from the tool tracking CVML model; (right) output from the red circle tracking CVML model

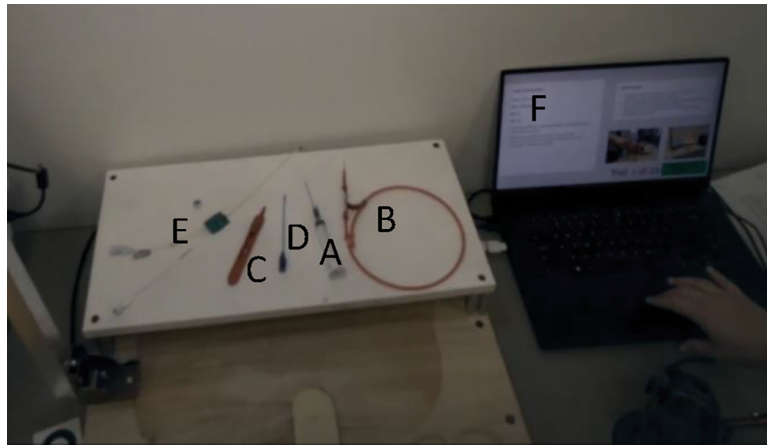


Figure 3:
The simulator of interest with the tools labeled; A. needle, B. guidewire, C. scalpel, D. dilator, E. catheter, F. computer

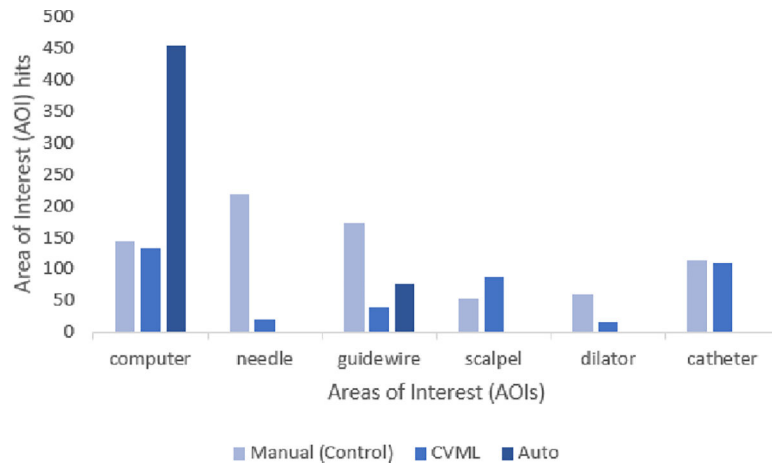


Figure 4: AOI hits for each method. The lighter bar for the manual method represents the control condition for AOI hits – *proximity to this bar indicates increase similarity in AOI detection but not necessarily the accuracy of when it was detected*



Figure 5:
AOI drifting over time during the manual mapping method in Tobii Pro

Cohen's Kappa indicates higher level of agreement (accuracy) for CVML when compared to the manual model

Table 1:

Manual to Auto	Kappa	Strength	Significance	Manual to CVML	Kappa	Strength	Significance
<i>Computer</i>	.375	Fair	<.001	<i>Computer</i>	.831	Very Good	<.001
<i>Needle</i>	.000	Poor	-	<i>Needle</i>	-.013	Poor	.197
<i>Guidewire</i>	-.037	Poor	.027	<i>Guidewire</i>	.190	Poor	<.001
<i>Scalpel</i>	.000	Poor	-	<i>Scalpel</i>	-.023	Poor	.194
<i>Dilator</i>	.000	Poor	-	<i>Dilator</i>	.019	Poor	.190
<i>Catheter</i>	.000	Poor	-	<i>Catheter</i>	-.020	Poor	.267