Lessons Learned from Developing Gaze Tracking Tools for ATM and Flight Deck Training

Author(s):
Hildebrandt, Michael; Langstrand, Jens-Patrick; Nguyen, Hoa Thi

Publication Date:
2020-03

Permanent Link:
https://doi.org/10.3929/ethz-b-000407655 →

Lessons Learned from Developing Gaze Tracking Tools for ATM and Flight Deck Training

Michael Hildebrandt[✉*], Jens-Patrick Langstrand[*], and Hoa Thi Nguyen[*]

[*]Biometrics Lab, Institute for Energy Technology, Norway

Eye tracking (ET) is a proven and valuable research tool, but unique challenges arise when applying it in day-to-day training. This paper discusses lessons learned across five years of developing eye tracking tools for training. The paper reports on feedback from ATM instructors when presented with pure gaze point videos from ET glasses, and how this feedback was used to develop real-time gaze analysis functionality. The paper also discusses requirements for aviation training ET systems, such as the ability to model the work environment accurately. It discusses some of the challenges we encountered, such as restrictions in accessing real-time simulator data, and how such data can significantly increase the value of ET analysis. Finally, we discuss how we applied these experiences to develop SYNOPTICON [1, 2], a non-commercial ET software tool for aviation training.

1 INTRODUCTION

Simulator training is an essential corner stone of aviation safety and remains the focus of research efforts aimed at improving both its efficiency and effectiveness.

It is therefore not surprising that the training domain is open to new approaches and technologies that can improve training, while at the same time being conscious of the need for a solid business case (i.e. cost-benefit analysis) before proceeding with new approaches.

One such approach is the use of eye tracking (ET) for aviation training [3]. ET has long been used as a research tool in aviation and other domains, and it is therefore not surprising that there have been proposals to also leverage ET as a training tool. However it is only recently that there have been serious attempts by the aviation industry and ET vendors to deploy such technologies in commercial training simulators. This trend is likely related to progress made in the automotive industry, where ET starts to become a standard technology for Advanced Driver Assistance Systems (ADAS), for instance monitoring for distraction and fatigue [4]. The adoption of ET in the automotive industry has been enabled by technological progress in the automatic, real-time analysis of ET data, more specifically through the use of machine learning models. Such technologies are likely to be needed for ET to make similar breakthroughs in the aviation sector.

However, there are important differences between the automotive and aviation use case that need to be considered. The automotive use case typically involves the detection of cognitive or physiological states (distraction, fatigue, etc.), or the simple detection of gaze objects (e.g., are the eyes on the road?). The "consumer" of the ET data is an intelligent ADAS on board the vehicle. For the aviation use case (air traffic and flight deck training), the relevant gaze information is far more complex (e.g. scan patterns), more variable, and more context-dependent (e.g. depending on the lesson plans and goals, evaluation criteria, etc.). The consumer of the information is a human, i.e. the instructor and/or trainee, whose understanding, interpretation, conclusion and use of the data is far less predictable than when the information is used by, for instance, an ADAS. That means that ET for aviation training presents a unique set of challenges with only limited opportunity for applying the advances in automotive ET.

Another influence on future uses of ET in aviation training is the long tradition of ET use in aviation research. In both air traffic management (ATM) research and flight deck Human Factors, ET has

✉Contact: michael.hildebrandt@ife.no

Article №9 · Pages 78–85 · https://doi.org/10.3929/ethz-b-000407655

provided valuable insights into visual attention, workload, and other factors. However it is important to recognize and consider the differences between research and training, and deliberate if and how research measures can be used in training. Research is usually conducted under highly controlled conditions by trained investigators who are very knowledgeable about the ET measures and their meaning. The same assumption cannot be made for training instructors in a busy simulator center. Research studies usually look for statistical differences in experimental conditions across a large pool of participants. In training, measures must be valid and reliable for an individual trainee. Research data is typically analyzed offline, while for training, results need to be available in real-time or immediately after the simulator session. Research data often comes with many qualification and caveats, whereas training data should be either unambiguous, or easily (and fairly) interpreted by all stakeholders. Finally, participants in research studies are protected from negative effects of their performance, while any human performance data collected in aviation training may affect the trainee's career.

All these arguments suggest that a unique set of requirements exist around the use of ET for aviation training, and that we cannot simply transfer approaches from the research field or from other industry domains. Instead, the use of ET should be based on a bottom-up approach, rooted in an understanding of the information needs of instructors and trainees.

2 METHOD

In this paper, we present experiences from working with instructors, training centers and simulator vendors to develop solutions that may help make ET a practical, usable tool for aviation training. Over the past five years, we conducted a number of workshops with training center staff to identify requirements and identify potential functionality for eye tracking tools in aviation training. Using a participatory, iterative design approach, we then proceeded to develop and test this functionality.

3 RESULTS

3.1 Feedback from Stakeholders

About five years ago, we started working with an air traffic control (ATC) training center to explore what kind of eye tracking tools could benefit ATC training.

As an initial step, we conducted a baseline data collection in an ATC control tower. An ATC officer was wearing mobile ET glasses, while an instructor was monitoring the live gaze replay as well as the officer's general work. After a one-hour recording, both the officer and the instructor were debriefed. We were interested to understand what the participants saw as potential applications as well as limitations of this data source.

It became immediately obvious that the instructor had great difficulty following the very fast eye and head movements of a tower controller. She even reported feeling nauseous after some time of following the ET live video. She also reported that it was difficult to divide attention between the ET video and monitoring the controller's work. She thought that it is essential to be able to monitor both, as without knowing what the controller is doing and what the situation was, it would be difficult to judge if scan patterns are appropriate. She also pointed out that monitoring the overall work is the priority for an instructor, with ET as an add-on, but that the effort required for following the ET would invert this situation, with disproportionate amount of attention required for the eye tracking. While she found it useful to monitor if, for instance, the controller had scanned the runway before giving a clearance, she believed that some of this information was already available to her from monitoring head movements alone. She also criticized the lack of aggregated summary data that could provide meaningful information, and therefore considered the value of monitoring pure gaze videos as low.

From a practical standpoint, both participants believed that data must be available immediately during the debriefing session, while the ET glasses we used at the time required a transcoding period of 15-20 minutes before replay was possible. During this test we also encountered practical problems with calibration and with participants wearing glasses. Correction lenses, as available from most ET vendors, were considered cumbersome and did not work for participants with bifocal lenses. Participants also reported that they felt their vision uncomfortably restricted by the ET glasses, especially when looking at the strip board or when making quick upward and side-way glances at the runway or taxi ways. That meant that more head movement than normal was necessary to perform the required scans. Such gaze restrictions or "gaze costs" could have effects on gaze strategy, and could potentially negatively affect training. Based on these negative reactions, we believe serious testing and evaluation of such risks is necessary before ET glasses are recommended or used as everyday aviation training tools.

In a final workshop with instructors and controllers, we were interested to find specific examples of how they believed ET information could be used. As expected, participants pointed out that certain scan patterns are expected from controllers, and that ET could help assess these. However, in digging deeper on this issue, participants found it difficult to unambiguously define and operationalize these patterns. How rigidly should a controller adhere to a pattern? Does an additional fixation break the pattern? How long must a controller glance at information to consider that it was "seen"? How quickly does information age (e.g. if wind direction was checked 20 seconds ago, does it have to be checked again or is it acceptable to take the information from memory?). How context-dependent is the judgement of what a right or wrong scan pattern is? How should contextual information be taken into account (e.g. task goals)? What level of knowledge about ET does an instructor need to make use of the information? How to take into account the varying levels of accuracy in the ET data?

A clear outcome of this case study was that looking at gaze videos was not an approach that would work for ATC training. Better visualization tools and more automated data processing would be needed to make ET useful. We therefore set out to develop a non-commercial software tool called SYNOPTICON [1, 2] to specifically address the need of training centers (see video at the link). The core functionality of this system, and how it addresses training needs, will be described in the next section.

3.2 A Gaze Analysis System for Training

It quickly became clear that a system for automatic gaze object detection must have a model of the environment (simulator) to map the gaze data onto. Some eye tracking vendors use QR markers or computer vision to detect gaze objects, but these approaches were quickly rejected as (a) it is not feasible to apply QR codes in a commercial simulator, (b) the resolution and wide-angle nature of scene cameras in ET glasses severely limits the accuracy of computer vision and would not work at larger distances, and (c) the low lighting conditions in most simulators would further undermine the quality of computer vision analysis.

We therefore took a different approach, namely to merge the gaze vector data from the eye trackers with position data from a motion capture system. By attaching retro-reflective position tracking markers to the ET glasses, the position and orientation of the glasses can be measured at an accuracy of $\pm$1mm and at a high temporal resolution (e.g. 240Hz). When this data is merged with gaze vectors from the ET glasses, a gaze vector in world coordinates (i.e. relative to the simulator environment) is obtained.

3.2.1 Modeling of the Physical Environment. To identify gaze objects in real-time, this gaze vector must be projected onto a virtual model of the environment. To do that, a highly accurate model must be constructed and aligned to the position tracking system. Since conventional modeling approaches (e.g. laser distance measure or LIDAR point clouds) would not be able to generate these kind of models, we developed a custom modeling solution using the same optical motion capture system as used for the tracking of the ET glasses. This is achieved through us of a "modeling wand", which is simply a rod with position markers attached. The system is calibrated to know where the tip of the want is in space. Objects are then modelled by touching the corners of the objects with the tip of the wand and hitting a key in the modeling software to signal that this point should be saved. For instance, a display can be modeled by simply registering three of its corner. The tool provides support for polygons and curved surfaces, so that for instance the non-rectangular cockpit windows, the curved throttle quadrant, the flight yoke, or the curved projection wall of a tower simulator can be modeled.

This approach allowed us to model the highly detailed cockpit of a Boeing 737 simulator at an accuracy of ±1mm within about 45 minutes (Figure 1). Surfaces in the model can be overlaid with screen textures for a more realistic look, or, where available, live textures from the simulator can be streamed.

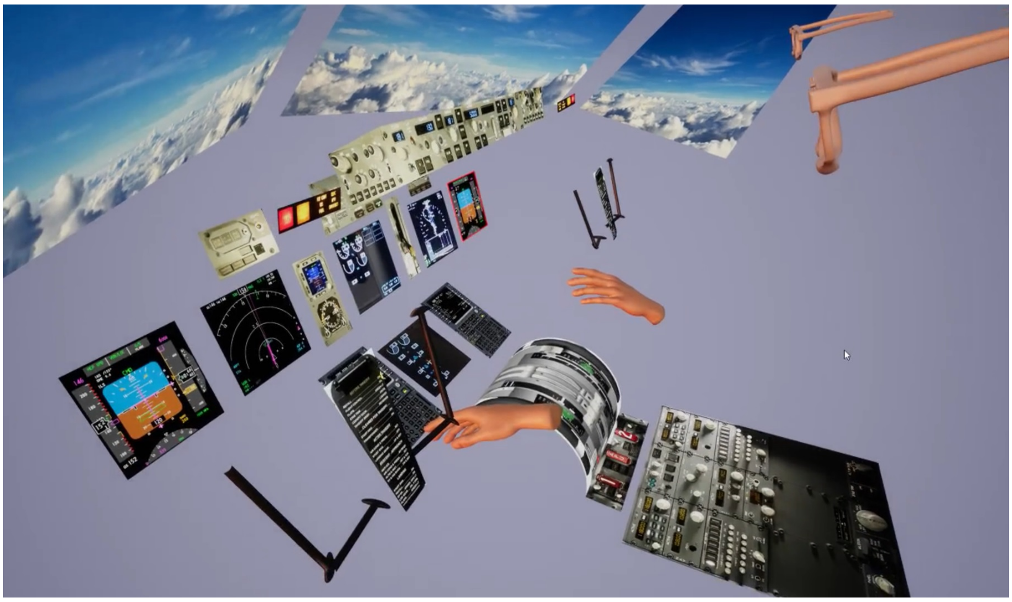

Figure 1: Model of a Boeing 737 cockpit simulator, including eye and hand tracking.

3.2.2 Gaze Mapping. The next required functionality is a collision detection algorithm that measures where the gaze vector intersects with the model. This functionality is required to later generate statistics about the gaze objects. As many simulators are used for multi-crew training, the system needs to keep track of multiple participants. This is achieved by creating an "actor" and assigning one or more sensors to it. For example, two pilots would be modeled as two separate actors, receiving data streams from an eye tracker and a position tracking marker each. Such a system provides a lot of flexibility of configuration. For instance, if one of the participants cannot be eye-tracked, then a gaze approximation can still be achieved by head orientation (motion capture) data alone. The system also supports other sensors, such as heart rate or galvanic skin response; remote (screen-mounted) ETs, and pose estimation data from webcam streams. With the actor model, all these sensors can be organized and assigned to the right participant, so that data can later be exported with the right labels for further processing.

In testing the system, it became clear that results can be misleading if only the calculated gaze point is considered. With the inherent variability in ET data, it is not possible to determine with absolute certainty if the participant is looking at instrument A, or another instrument B just a few centimeters adjacent. In subsequent testing at a flight academy, it also became clear that peripheral vision needs to be taken into account when assessing gaze objects. The system therefore contains a "gaze cone" functionality, whereby a user-defined field of vision is used to measure which instruments the participant had in view.

3.2.3 Object Tracking. When testing the system in an ATC tower training center, it quickly became clear that a static environment model would not be adequate. Monitors occasionally get moved around or nudged. Many modern simulators now have motorized desks where the height can be individually adjusted; and there are mobile objects such as tablets which must be modeled. We therefore developed functionality whereby a whole set of displays can be referenced to a specific position tracking marker. Thereby, a marker can be attached to the screen, so that the system knows when the desk height is changed, and adjusts the model in real-time. Similarly, mobile objects such as tablets can be tracked by attaching markers to them.

In Figure 1, two hands and two flight yokes are visible. These were also tracked through markers, so that yoke and hand motion could be recorded. The hands were modeled as solid objects without finger movements. In the system, touch-Areas of Interest (AOI) can be defined, and when the hand model intersects with a touch-AOI, a touch event is logged. The hand tracking is sensitive enough to differentiate between press and turn motions, so that for instance heading changes in the auto pilot can be detected.

Yoke tracking is especially useful for simulator where flight data can not be accessed, as it will show yoke movements performed by the auto pilot.

3.2.4 Dynamic Areas of Interest. When developing functionality for ATC tower simulators, it quickly became clear that a static model alone is not sufficient, as many relevant events take place on the projection screen (e.g. monitoring an aircraft as it taxis to the runway). We therefore provide functionality for dynamic AOIs based on real-time data from the simulator. The simulator can send screen coordinates for where an AOI has appeared, along with meta data (e.g. labels). The system creates an AOI and updates its position as new data is received from the simulator. The dynamic AOI is handled the same as any other AOI, and eye data is logged and visualized. When the aircraft has departed, the simulator can send a signal for the AOI to be deleted. Such functionality is essential for the efficient use of ET in ATC tower simulators.

3.2.5 Gaze Visualizations. It remains an open question what type of gaze visualizations and aggregations would most benefit aviation training stakeholders. Until this question can be resolved through detailed co-design and evaluation studies with stakeholders, we chose to provide a number of generic visualizations in our system. This includes heat maps, gaze tracks and sequence diagrams. In initial discussions with instructors, sequence diagrams (Figure 2) were considered particularly useful, as they provide an easy understanding of shifts in visual attention over time. They also allow identification of periods when an instrument or AOI were not looked at, for instance failure to look out for the runway during approach and landing, because of fixation on instrumentation. Such visualizations are particularly valuable if they can accumulate data from multiple participants, e.g. to ensure that at least one of the pilots regularly checks airspeed on approach.

3.2.6 Simulator Events. In discussions with instructors, it became clear that any gaze visualization must be configurable. Aggregation of data across a full scenario can make the data meaningless, as different flight phases and events may require different scan patterns. It is therefore important that the aggregation period for visualizations can be customized. Ideally this should be achieved automatically based on simulator events (e.g. start of engine failure until touch-down). We have therefore provided functionality to exchange messages between the simulator and the gaze analysis system. Any event sent to the analysis system is logged as a marker, and these markers can be used to define aggregation periods in the software. In addition, markers can be inserted manually by the instructor, either during the scenario, or later during debriefing. This provides maximum flexibility for defining data aggregation periods.

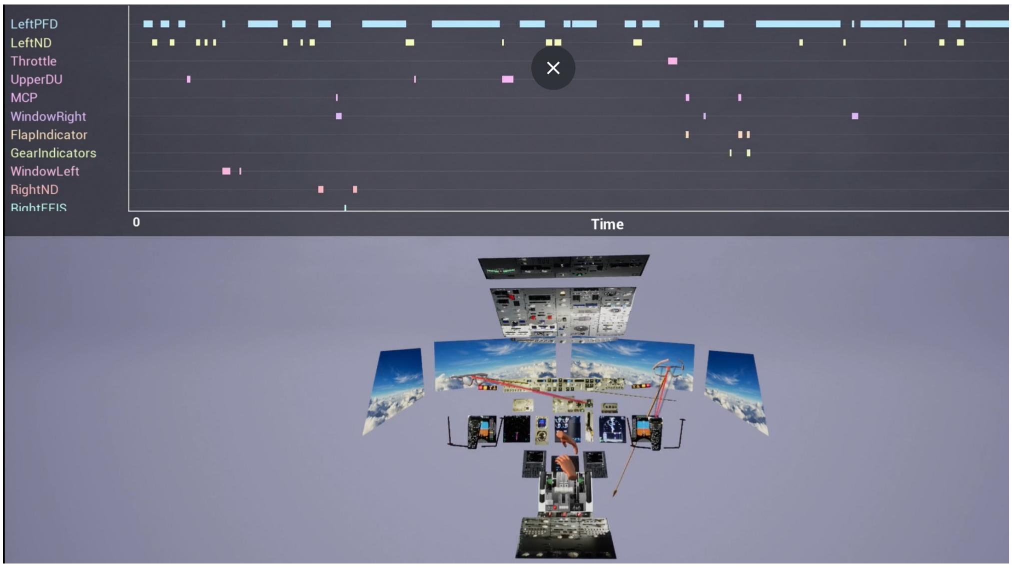

Figure 2: Example of sequence diagram.

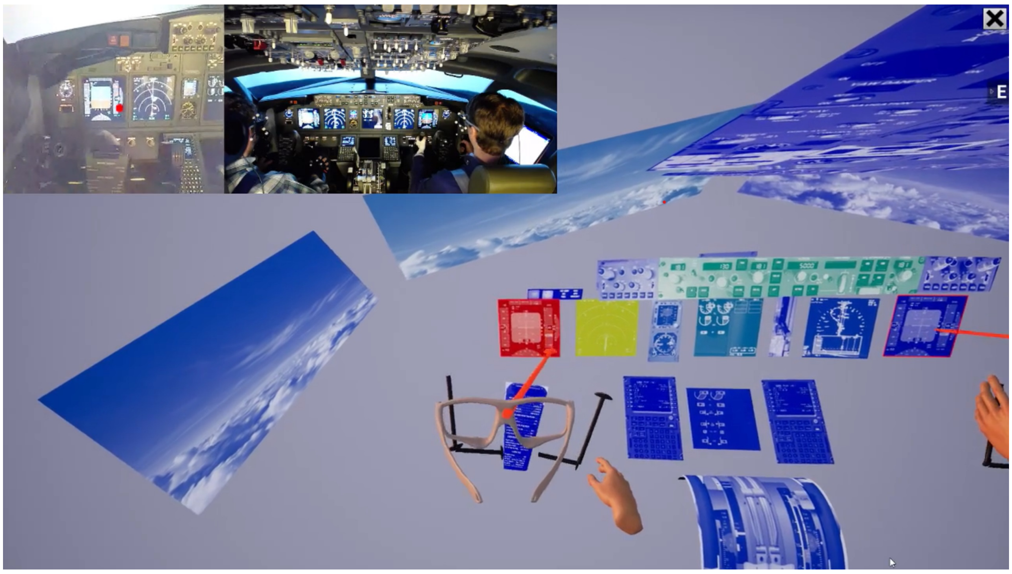

Figure 3: Heat map visualization (red on primary flight display), with gaze vectors (red line), webcam video (center-top) and original ET video (top left).

3.2.7 Flight Data. An important lesson from our stakeholder study is the contextualized nature of ET data. Ideally, gaze information should be presented alongside, or integrated with, other scenario data. We therefore included functionality to receive, plot and record simulator data such as flight parameters or other relevant data. Unfortunately it has been our experience that many commercial aviation simulator vendors do not provide easy access to such data, instead providing a "black box" system not open to third parties. This severely limits the ability to develop advanced human performance tools that could jointly analyze flight and behavioral data. It should be a matter of urgency for stakeholders to start discussions with vendors to provide open software interfaces.

3.2.8 Networking Features. Working with multi-sensor systems, including ET glasses for two pilots, a position tracking system and physiological sensors and webcams, it became clear that the volume of data cannot be handled by a single computer. It was therefore necessary to develop a networked architecture that would allow the system to be highly scalable. All data in our system is exchanged via network protocols. This has proven critical for achieving good system performance. It has also lead us to abstract the data streams from different sensor vendors to a common protocol. That means that the system can support multiple sensor vendors easily.

3.2.9 Audio-Visual Data. To understand context during debriefing, it is important to enable audio-visual data recording alongside with ET recording. In our system, multiple video streams as well as the original gaze video from the ET glasses can be synchronized and replayed immediately after the recording. Such functionality is essential if flight events cannot be added during the scenario and must be entered manually later. If the instructor tags events while running the scenario, these markers will appear in the recording, and the instructor can jump right to critical scenario phases for replay.

3.2.10 Calibration. During our stakeholder interviews, we noticed that handling ET devices can be a daunting task for instructors. We therefore took care to provide easy-to-use functionality for the calibration of the joint ET and motion tracking system. The instructor can freely choose calibration points and mark them out in the 3D model of the simulator. Those could for instance be three corners of a monitor. The participant is then asked to look at each of these points in turn, and each calibration point is acknowledged by pressing the space bar. It is easy to check calibration accuracy as the gaze vectors are shown in real-time. This approach has proved successful, and during a subsequent feasibility test with 20 pilots, we managed to equip each pair of pilots with eye trackers and hand markers, seat them and calibrate them within under 5 minutes. For a practiced simulator instructor, calibration of the ET glasses can be achieved within under 3 minutes.

3.2.11 Hardware and Software. The SYNOPTICON system supports head-mounted eye tracking glasses from *Tobii*, *SMI* and *Pupil Labs*. As a position tracking component, it supports the *Optitrack* motion capture platform. Additional hardware can easily be added, as all data exchange is handled through *publish-subscribe* data protocols such as *WAMP* or *MQTT*. The system supports physiological data streaming of *Shimmer Labs* sensors.

SYNOPTICON runs on Windows PCs or laptops. Due to its networked software architecture, software tasks can be distributed across multiple computers to improve performance. For instance, the eye tracking and motion capture servers can be run on separate machines, which will stream their data to the computer that merges and visualizes the data. Synopticon is built on the 3D game engine *Unreal Engine 4*.

4 CONCLUSIONS AND FUTUREWORK

Our experience so far suggests that it is both necessary and feasible to develop ET data analysis systems specifically for the needs of the aviation training domain. Key characteristics of such systems are the ability to aggregate data in a coherent visual display, to make it easy to identify changes in visual attention over time through appropriate visualizations, and to contextualize the data through integration with simulator events and flight data. We believe that such features should be considered in any system that employs ET for aviation training. In addition we have highlighted a number of practical and technological requirements, such as easy calibration and networking.

What remains a major challenge is working with instructors to identify specific visual scan patterns or expectations on visual attention. After several years of experience in this field, we believe that progress in this area can only be made by analyzing simulator training lesson plans, lesson goals and evaluation criteria. Once gaze patterns have been identified from these foundations, they can then be operationalized, and functionality can be built into the system for automatic identification of patterns.

There are two important areas we have not explored yet. One is the use of ET by the trainee. Our discussion have been mainly with instructors, resulting in an improved understanding of their needs. It is likely that for the moment, instructors are the primary users of ET data. However, the pilots and ATC officers themselves may also benefit from engaging with this information source. This would open unique opportunities and challenges. For instance, a trainee may want to compare visual attention performance over a period of time (say, a year).

The other unexplored field is the use of ET for generating measures of cognitive states such as workload or fatigue. This topic has been mentioned by several instructors, and there is clearly a need for such measures if they can be obtained reliably and consistently. We have so far avoided this topic as we do not believe in the technological readiness of such systems for daily training. However, advances in sensor technologies and analysis techniques may soon makes this a possibility.

User monitoring technology is evolving, and the performance of purely camera-based tracking systems is rapidly improving. We have tested such systems and added them to our system. Our conclusion is that they provide adequate estimates for head orientation (which can be sufficient for some applications), but still have a long way go on eye tracking performance.

REFERENCES

[1] Michael Hildebrandt, Jens-Patrick Langstrand, and Hoa Thi Nguyen. Synopticon: a real-time data fusion platform for behavioral research. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings*, pages 494–497, 2019.

[2] Jens-Patrick Langstrand, Hoa T Nguyen, and Michael Hildebrandt. Synopticon: Sensor fusion for real-time gaze detection and analysis. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 62, pages 311–315. SAGE Publications Sage CA: Los Angeles, CA, 2018.

[3] Julius Jakubowski and Wen-Chin Li. Enhancing aviation simulator de-briefs through the integration of student eye tracking: The instructor's perspective. In *International Conference on Human-Computer Interaction*, pages 259–270. Springer, 2019.

[4] Junli Xu, Jianliang Min, and Jianfeng Hu. Real-time eye tracking for the assessment of driver fatigue. *Healthcare technology letters*, 5(2):54–58, 2018.