Augmenting pilot training through an unobtrusive eye-tracking system

K. Wilson PhD, A. Robinson, R. Myers, M. Lenné PhD
Seeing Machines Limited
Canberra, ACT; Melbourne, VIC, Australia; San Francisco, CA, USA; Canberra, ACT, Australia
kyle.wilson@seeingmachines.com, alexander.robinson@seeingmachines.com, rama.myers@seeingmachines.com, mike.lenne@seeingmachines.com

M. Corbett
Institute of Aviation Medicine (RAAF)
Adelaide, SA, Australia
mark.corbett@defence.gov.au

ABSTRACT

The anticipated increase in demand for pilots over the coming years, combined with a push toward evidence-based training aids, call for new tools to support instructors to objectively measure pilots' scanning performance. Eye-tracking based training tools may offer a solution to the so-called 'monitoring problem' as they may inform the instructor what a pilot has looked at and when, while also highlighting what they have not looked at.

Eye-tracking systems have only recently matured where they may be practical to use in an operational training context. This paper reports on quantitative and qualitative measurements of the value and utility of an unobtrusive eye-tracking based training aid installed in a Lead-In Fighter Training Hawk 127 Full Mission Simulator.

In Study 1, seven instructors conducted operational training sessions both with and without the use of an eye-tracking based training aid. Questionnaire responses indicated that using the eye-tracking tool provided benefits such as increased awareness of trainees' scanning behavior. In Study 2, five instructors viewed video replays of flight sequences performed by both trainees and instructors. Some replays displayed pilots' scanning behavior while others did not. Pilots' identities were hidden, and instructors had to judge whether the replay featured a trainee or instructor. There was no detectable difference in correct judgments for replays with versus without eye-tracking information. This was likely due to the remote nature and low-quality of replays, and the advanced performance of the trainees, however.

Interviews with instructors provided important context with which to interpret the quantitative results, and offered insights which will aid future development and integration of eye-tracking based training aids in operational contexts. The current research demonstrates that eye-tracking tools can offer significant benefit to instructors. However, it is crucial that these tools are properly integrated into the training environment.

ABOUT THE AUTHORS

Kyle Wilson, PhD is Senior Scientist & Team Lead (Human Factors) at Seeing Machines. He received his PhD in Psychology at the University of Canterbury for his research on friendly fire incidents.

Alexander Robinson leads Seeing Machines' military and aviation training & simulation division. He is a current private pilot and former military pilot, with an invested interest in improving pilot training.

Rama Myers leads the product management & partnerships functions within Seeing Machines' aviation division. Rama collaborates closely with research institutions and industry to better understand and address challenges in crew fatigue, pilot monitoring, and operator workload.

Mark Corbett is the lead Human Factors researcher at the Royal Australian Air Force's Institute of Aviation Medicine. He has a background in psychology and military aircrew, with international experience in pilot training.

Mike Lenné, PhD is Chief Science & Innovation Officer at Seeing Machines. He delivers key advances in operator state technology and works with customers, government, and industry partners to promote the implementation of operator state monitoring technology that makes transport modes safer. Mike holds a BSc (Hons), Global Executive MBA and a PhD in Human Factors Psychology. He is an Adjunct Professor at the Monash University Accident Research Centre (MUARC) in Melbourne.

N.B. Opinions are the authors' and not those of the Department of Defence or the Department of Veterans' Affairs.

Augmenting pilot training through an unobtrusive eye-tracking system

K. Wilson PhD, A. Robinson, R. Myers, M. Lenné PhD
Seeing Machines Limited
Canberra, ACT; Melbourne, VIC, Australia; San Francisco, CA, USA; Canberra, ACT, Australia
kyle.wilson@seeingmachines.com, alexander.robinson@seeingmachines.com, rama.myers@seeingmachines.com, mike.lenne@seeingmachines.com

M. Corbett
Institute of Aviation Medicine (RAAF)
Adelaide, SA, Australia
mark.corbett@defence.gov.au

INTRODUCTION

Pilot training is one of the most important predictors of operational performance (Todd & Thomas, 2013). Training pilots is often time- and resource-intensive, and the anticipated increased demand for military (Thompson, 2018) and civilian pilots (Boeing estimates that 763,000 additional pilots will be recruited over this time; The Boeing Company, 2020) as well as the unknown effects of COVID-19-induced lack of currency (Flight Safety Foundation, 2020), is set to put unprecedented pressure on training providers and operators. There is also a move towards 'evidence-based training', with militaries, airlines, and training providers demanding new tools to support their instructors' ability to objectively measure and train expert scan and gaze patterns in their pilots.

The ability to train, measure, and monitor could be greatly enhanced by the application of an appropriate technology. One of the challenges that instructors face is the difficulty in ascertaining the exact nature of a pilot's scan, based on the instructor's vantage point and observation from behind or even beside the aviator. This is a key challenge within simulator training (Li et al., 2014) and it may undermine the instructor's ability to know when and how to provide constructive feedback on scanning strategies to the trainee (Sullivan et al., 2011). Although instructors can observe crew posture and behaviors, they are inhibited in their ability to assess the crew's scanning behavior with any degree of certainty, which is a concern when pilots acquire a significant portion – up to 90% in some cases – of their flying information visually (Draeger, 2009). As a result, the cognitive demand of instructors is likely increased, as they piece together an assessment of how the trainee is visually gathering information, formulating their decision making, and acting based on their monitoring. The high speed at which eye movements typically occur (Fuchs, 1967), and that pilots themselves are often unaware of their actual scanning techniques (Robinski & Stein, 2013), further highlight the challenges of assessing monitoring behavior in the current training environment. Eye-tracking based training tools may offer an aid to the monitoring problem as they could inform the instructor what the pilot has looked at and when, and will also highlight what they have not looked at.

Despite evidence showing the value of eye-tracking for understanding pilot performance emerging well over fifty years ago (Fitts et al., 1950), it is yet to be widely adopted in training settings. The cumbersome nature of many eye-tracking systems is likely a key factor in why this technology has not yet gained traction (Vidulich & Tsang, 2019). Only recently has technology reached a point whereby eye-tracking systems may be practical to use in training settings. For instance, it is now possible to unobtrusively track a pilot's gaze in real-time, with high accuracy. The significant value that eye-tracking can offer in aviation for teaching pilots improved scan patterns and estimating monitoring skills is discussed in a research review paper by Peysakhovich and colleagues (2018).

This paper describes the process used to integrate an unobtrusive eye-tracking system into a military operational pilot training environment with actual pilots-in-training, and reports on a study to test whether using an eye-tracking system could aid instructors. The project that encompassed the current research was broad in nature. Here, we present a sub-section of this research. Specifically,

- Study 1 was based on the hypothesis that use of the eye-tracking system would improve instructors' awareness of trainees' scanning behavior.
- Study 2 was based on the hypothesis that the system would help instructors to determine the proficiency level of pilots.

First, we describe the process used to develop and integrate the eye-tracking system into the chosen training environment.

SYSTEM DEVELOPMENT

Before the research commenced, it was acknowledged that the eye-tracking system must be developed with input from relevant military stakeholders. The location of the sensor (see Figure 1) was selected as it provided a suitable balance between 1) reliable, accurate, and high precision tracking coverage across a range of instruments and areas of interest, and 2) it did not interfere with normal cockpit operations or distract the pilot.

Figure 1. Hawk 127 Full Mission Simulator with Crew Training System installed (yellow highlight).

Flight simulator

The Royal Australian Air Force (RAAF) uses three Hawk 127 Full Mission Simulators (FMS) for Lead-in Fighter Training. All data recorded with the eye-tracking system occurred in a Hawk 127 Full Mission Simulator (Figure 1). These CAE-built simulators consist of high-fidelity replica cockpits, each in their own dome visual display system. Each FMS has its own Instructor Operating Station (IOS). The pilots were trainees from two Introductory Fighter Courses (IFC66 and IFC67, hereafter referred to as Course 1 and Course 2).

Eye-Tracking System

The eye-tracking system (hereafter 'Crew Training System'; CTS) consisted of a vision-based sensor (i.e., a camera) fitted into the simulator cockpit (Figure 1) and an associated software application integrated into the IOS (Figure 2). The sensor was focused on the pilot's face and recorded head and eye movements. The pilot's eye movements were overlaid (visually represented by a colored cross indicating the real-time fixation point, or 'gaze vector,' & a colored line indicating the 'gaze trail' or scanning path of the pilot) on top of a repeated view of the cockpit instruments and external visuals.

Instructor Information Requirements

The CTS interface was designed with input from the RAAF's 78 Wing and the Institute of Aviation Medicine. Design considerations included defining and incorporating optimal Areas of Interest for scanning locations (Table 1) and ensuring that key information was presented to instructors clearly and in real-time. This design effort included:

1) Layout and sizing of Areas of Interest (Figure 2);
2) Coloring of a visual attention time plot (termed the CandyBar® - a Seeing Machines proprietary visualization tool, displaying a variable time scale along the Y-axis which is 'colored-in' with the pilot's real-time gaze across Areas of Interest);
3) Gaze vector size, color, and trail length; and,
4) Several features that were incorporated to assist Qualified Flying Instructors (QFIs) in interpreting the real-time gaze of a trainee. The IOS was used to display pilots' gaze visualizations.

Table 1. Area of Interest definitions.

Area of Interest	Description *(Italicised - CTS Console and CandyBar legend descriptor)*	User Interface Comments
HUD / HGS	Head-Up Display (*HUD*) / Head-Up Guidance System (*HGS*)	Primary display on CTS Console – gaze vector & trail, border highlighting, and on CandyBar.
Head-Down Flight Instruments	Comprised of: 1) Heading Indicator (*HI*) 2) Calibrated Airspeed Indicator (*CSI*) 3) Head-Down Flight Director (*HDFD*) 4) Altimeter (*ALT*) 5) Vertical Speed Indicator (*VSI*) 6) Engine Display (*ED*) 7) Angle of Attack Indicator (*AOA*)	Displayed on CTS Console – border highlighting, and on CandyBar.
Navigation Display	Central Multifunction Display (*NAV MFD*)	Displayed on CTS Console – border highlighting, and on CandyBar
Secondary Multifunction Displays	Left and Right Multifunction Displays (*LHS MFD*, *RHS MFD*)	Not displayed on CTS Console in any format in the interests of space conservation and ease of interpretation. CTS Console is capable of displaying these.
Out the Window	OTW – defined as the Areas of Interest to the left and right of the HUD (*Outside LHS*, *Outside RHS*).	Displayed on CTS Console on the CandyBar.
Other	Areas of Interest not defined in the geometrical cockpit model for CTS Console (*Other*).	Displayed on CTS Console on the CandyBar.

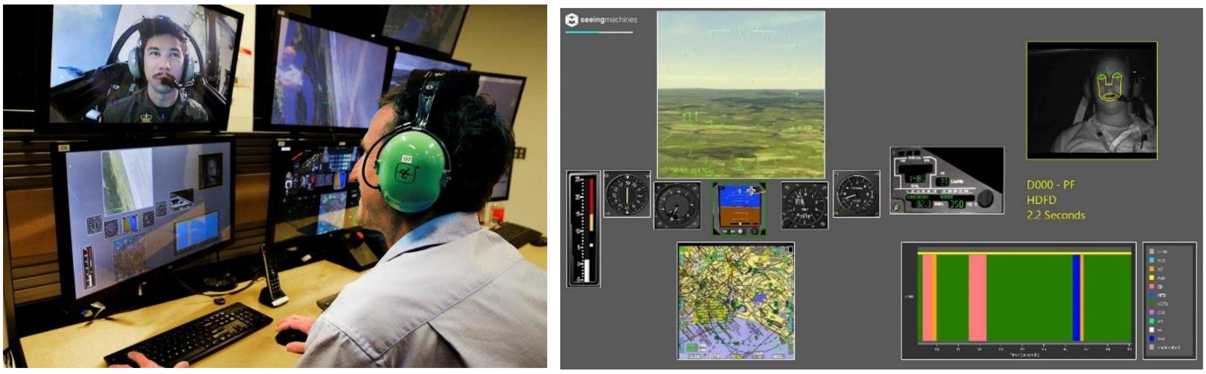

Figure 2. IOS (left) with CTS Console (right – and integrated into the bottom left IOS screen) as presented to instructors. CandyBar visual attention time plot is located on the bottom right of the CTS Console.

STUDY 1 – INSTRUCTORS' AWARENESS OF TRAINEES' SCANNING BEHAVIOR

Despite the promise of eye-tracking for pilot training, no research existed to show that providing flight instructors with objective, real-time eye movement data from trainees could benefit instructors. Study 1 tested the hypothesis that using the eye-tracking system would improve instructors' awareness of trainees' scanning behavior. A group of instructors performed training sessions both with and without the availability of real-time eye movement data. Outcomes were measured using a combination of questionnaire data from instructors following selected simulator training sessions, and interviews at the end of the courses. It was predicted that instructors would report that both their awareness of scanning behavior and their situation awareness was enhanced when eye movement data was available. In addition, instructors' workload was examined to understand whether using Crew Training System imposed any additional workload on them. This was important to determine, given simulator sessions tend to be busy and instructors already have high task demands.

STUDY 1 – METHODS

Participants

Seven instructors who were involved with two consecutive Introductory Fighter Courses at 79 Squadron at RAAF Base Pearce participated. The first course ran from September 2019 through to early 2020 and the second course ran from early 2020 through to September 2020. Ethical clearance for this project was provided by the Departments of Defence and Veterans Affairs Human Research Ethics Committee and informed consent was obtained from participants prior to commencing the study.

Research Design

A control condition was used, where the eye-tracking visualizations and information were blinded to the instructor, by turning off the CTS display. The control data consisted of 25 training sessions and included data from four different instructors. For the experimental condition, CTS collected eye movement data from trainees and presented this data to instructors via the CTS display on the IOS. The experimental condition included data from four training sessions (2 instructors) during Course 1, and nineteen training sessions (four instructors) during Course 2. The variable manipulated here was whether the Crew Training System display was ‘on’ or not. This design afforded a robust experimental comparison between current training methods (i.e., a ‘baseline’) and training with the aid of CTS.

Questionnaires

Instructors answered the questions below using a visual analog response scale. Each response was later manually measured using a ruler and converted to a numerical value (0-100) to aid analysis.

1. *Workload: What level of workload did you experience during the training session while monitoring the trainee?*
2. *Situation Awareness: What was your level of Situation Awareness of the trainee’s scanning behavior during the training session?*
3. *Ability to interpret scanning: How well could you interpret the trainee’s scanning patterns during the training session while monitoring the trainee?*

Two additional questions were asked of instructors when CTS was ‘on’, but not when it was ‘off’:

- *Usefulness during training: How useful was the eye-tracking system for you during the training session while the trainee was flying?*
- *Ease of understanding data: How easy to understand was the eye-tracking data?*

Interviews

An interview was conducted with six instructors. The purpose of this interview was to further explore whether CTS provided value as a training tool (and if so, how). While two instructors had previously used CTS during their simulation instruction, four instructors had not used CTS to the extent that was anticipated. To familiarize them with CTS, they were instead asked to each fly a base turn and an ILS intercept in the simulator, while they were recorded by CTS. Following this, their respective recordings were replayed to them individually in the adjacent debrief room and they were subsequently interviewed.

Procedure

Each training session involved a trainee pilot completing a session in the simulator as part of the IFC curriculum, and instructors performed their teaching duties normally. Trainees’ eye and head movements were recorded with CTS. For the control condition, no eye movement data was displayed to instructors. For the experimental condition, the eye movement data was presented to instructors on one of the IOS monitors. Instructors used a ‘screen switch’ button to toggle between the standard display and the CTS display. They were not instructed on when or how often to use the button to display CTS; they were only asked to use CTS at least some of the time during the training sessions. After each training session, instructors completed the questionnaire with pen and paper. The semi-structured interviews took place outside of pilot training times.

STUDY 1 – RESULTS

Questionnaires

The mean responses from instructors are shown below in Table 2. No statistical inference testing was performed on the data due to the small sample sizes.

Table 2. Mean average responses (0-100 scale) and standard deviations (± SD) for each of the questionnaire items across Course 1 and Course 2.

Measure	Course 1 (CTS off); *N* = 25	Course 1 – sim session 4 (CTS on); *N* = 4	Course 2 (CTS on); *N*= 19
Workload	**33.6** ± 10.6	**38.2** ± 15.3	**62.4** ± 18.8
Situation Awareness	**36.2** ± 19.5	**45.2** ± 21.3	**34.5** ± 30.0
Ability to Interpret scanning	**29.8** ± 18.3	**48.4** ± 37.5	**28.6** ± 28.8
Usefulness during training session	n/a	**32.8** ± 18.8	**25.0** ± 26.1
Ease of understanding data	n/a	**81.8** ± 16.4	**30.0** ± 28.6

The data indicated that CTS was beneficial when it was switched on during Course 1 for simulator session 4, as shown by higher ratings for instructors' situation awareness and their ability to interpret trainees' scanning behavior. At this point in time (sim session 4), instructors also reported that the eye-tracking data as visualized on CTS was easy to understand, and appeared to offer utility (i.e., usefulness) during the training session. The workload question indicated that there may have been a marginal increase in workload after CTS was switched on, if any at all.

The ratings provided during Course 2, however, suggest that it was not benefitting instructors at this point. During this course, several training sessions involved 'night flying,' which prevented the instructors from using the eye-tracking system due to incompatibility between the eye-tracking system's near-infrared illumination, and the Night Vision Imaging Systems the pilots used. Accuracy issues that instructors perceived to be related to non-calibration also likely contributed to the decrease in positive sentiment toward CTS during Course 2. Comments provided by the instructors (as part of the questionnaire) illustrate this (Table 3). Only themes that were mentioned twice or more in the comments are included here.

Table 3. Comments made during Course 2 and their frequency.

Theme	Number of times mentioned
Calibration or accuracy of the gaze information	14
Lack of debrief facilities and/or time available	9
Eye movements distracting or presented too quickly	4
Recording stopped or 'froze' partway through	3
Size or resolution of display	2
Sharing of the CTS display with the standard IOS display	2

Interviews

Direct quotes from instructors are presented below. Several instructors referred to using the system for checking where a trainee was looking.

> *"When I'm watching [the eye-tracking data], it has confirmed to me where the student was looking."*

> *"When I say 'check the VSI (Vertical Speed Indicator)', I want to know that they've checked the VSI."*

> *"Eye-tracking is helpful to judge the discipline of a pilot by how consistently they look at certain areas. This indicates a level of experience or lack of experience."*

Using the system for remediation purposes was also a key theme.

> *"… where I'd use it is for a student having a lot of problems . . . for remediation . . . If someone has a perfect scan, we wouldn't be using it."*

"If a guy or girl has a good training session, I won't sit them down and go through it."

"If I need to do some remedial training with someone, it will come in handy."

While one instructor felt that CTS would be highly useful for debriefing following the training, other instructors highlighted potential barriers to its use during debrief, such as time constraints. One instructor said that while the current integrated debrief system they have was 'great,' it was rarely used due to the additional time required to operate it. Another reported barrier was that many instructors did not have the required government-furnished or personal computing equipment that would have enabled them to replay recorded videos following the training session. Several instructors thought that the tool would be useful for briefing trainees, for example, if an exemplar of good scanning behavior could be shown to trainees prior to their training sessions.

"It's a great training tool – I can see the advantage of briefing this with a student [regarding] what they should look at, but it's tricky to do in real time."

Instructors were asked whether they thought CTS would be beneficial in earlier-stage flight training, such as basic and advanced military flying training. There was a strong consensus that CTS would be highly useful there, and more useful than it was at the current lead-in fighter training level.

"I would see it as a huge advantage at [Basic Flying Training]."

"I haven't seen anyone with a bad scan here. It's harder [to use Crew Training System to its full capability] when we've just got more advanced students."

"…for [basic and advanced flying] teaching . . . the basics . . . I can see it as being really valuable."

One instructor who had used CTS during training noted that he was still becoming accustomed to the system.

"I'm finding I'm having to teach myself where I can use it, where it's handy."

The temporary integration of the CTS display into the IOS meant that viewing the CTS display was not optimized for use. During the study, one of the standard IOS monitors was shared between the CTS interface and the original display and it took several seconds to switch back and forth between the displays.

"The eye tracker steals one of the screens… it takes 5 seconds for us to flick back [to the other display normally on that monitor] . . . you don't just push [the screen-switch button] & go."

Finally, consistent with comments made on the questionnaires, two instructors noted that the gaze information was not accurate or visible at certain times. As noted in the questionnaire results above, this was primarily an issue during Course 2.

STUDY 1 – DISCUSSION

Findings from the questionnaires indicated that instructors benefitted from using the eye-tracking system, however, this was dependent upon the system being 1) available to instructors to use (e.g., night flying largely prevented its use), and 2) accurate in terms of showing where trainees were looking. When the system was functioning as intended, instructors reported enhanced situation awareness and an improvement in their ability to interpret scanning patterns. This likely reflects that the system provided instructors with information they cannot usually access due to their vantage point, and that assessing what instrument or area a pilot is looking at is normally a very subjective process.

The main reported benefit of the eye-tracking information was for 'checking' scanning quality. Instructors used the eye-tracking to identify where pilots were looking in real-time, but also how often they were looking at specific areas. The latter was reportedly used to judge the discipline of a pilot, and relatedly the level of experience. Another commonly reported benefit of the system was for remediation purposes. As pointed out by instructors though, the system would likely offer greater benefits if used at an earlier stage of training, when trainees' scanning is at an earlier stage of development.

The temporary nature of the eye-tracking integration into the prime simulator resulted in some accuracy and usability issues that at times prevented instructors from gaining value from the system. When the gaze vector did not accurately reflect where the trainee was looking (i.e., Course 2), there was a reduction in instructors' situation awareness and ability to interpret trainees' scanning to the levels reported at the baseline (where gaze information was not shown to instructors). Furthermore, instructors reported increased workload which likely reflects that they had to actively ignore the gaze information during this time. Regarding usability, some instructors reported frustration with the time required to switch back and forth between the CTS display and the standard IOS monitor. These issues are readily resolved through eye-tracking system integration into the simulator environment.

Lastly, inexperience with CTS may have prevented instructors from extracting the most value from the system, particularly during early use. Instructors were not given any extensive training on how to use the tool, or any specific guidance on how they should interpret scanning behavior. This was reflected in one instructor's comment that he was having to teach himself where to use it.

STUDY 2 – THE PROFICIENCY LEVEL OF PILOTS

It was hypothesized that presenting trainees' eye movements to instructors in real-time would help instructors to determine the proficiency of the pilot. A controlled experimental design was used, whereby recordings of flight sequences (flown prior to the test) both with (CTS 'on') and without (CTS 'off') a visualization of eye movements were presented to instructors. Instructors were tasked with judging the proficiency of the pilot in each video. The set of recordings contained sequences flown by both trainee pilots (less proficient) and qualified flying instructors (more proficient). It was expected that more correct judgments (regarding whether the pilot in the video was a trainee or an instructor) would be made when CTS was 'on' versus when it was 'off'. Recordings with CTS 'on' were also expected to lead to quicker and more confident judgements, because in addition to flight performance data, instructors could use data on scanning behavior (e.g., through CTS' real-time gaze information and the CandyBar to inform their judgements).

STUDY 2 – METHODS

Participants

Five instructors from 79 Squadron participated. Ethical clearance for this project was provided by the Departments of Defence and Veterans Affairs Human Research Ethics Committee and informed consent was obtained from participants prior to commencing the study.

'Expert vs Novice' videos

The set of videos included CTS recordings of trainees' standard pattern base turns (Figure 3) and Instrument Landing System (ILS) intercepts (Table 4) performed during simulator mission training sessions and recordings of the same flight sequences performed by four instructors.

Table 4. Definitions of base turns and ILS intercepts for the purposes of the video replays.

Sequence	Parameters
Base Turn	Start: Application of bank at end of downwind leg and start of base turn (i.e., >10 degrees angle of bank (AoB). Stop: Roll-out on final with wings level (i.e., <10 degrees AoB)
ILS Intercept	Start: Application of bank at start of the turn to intercept the ILS localizer (i.e., >10 degrees AoB). Stop: Passing 400 feet on final (HUD altitude)

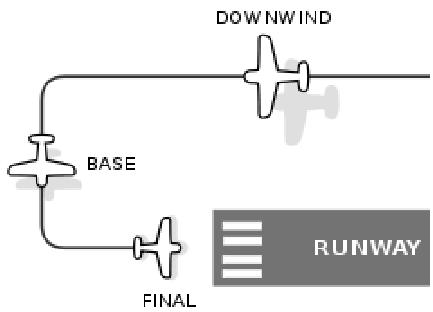

Figure 3. Standard pattern downwind-base-final legs.

For the purposes of the experiment, the recordings of instructors represented more proficient pilots (i.e., 'experts') and the trainees represented less proficient pilots (i.e., 'novices').

The recordings were duplicated to create copies of each. They were then edited so that:

1) Half of the recordings were presented with gaze tracking information ('CTS-on') and half were presented without ('CTS-off').
2) The 'face camera view' normally visible on the CTS display was hidden, preserving pilots' anonymity.

As the face camera view was turned off, the instructors had no way of identifying the experience of the pilot in the video based on recognition of the pilot alone. Instead, they had only the eye behavior data (only for CTS 'on' videos) and visible performance data (e.g., instrument readings and other aircraft characteristics) with which to decide. There were 26 videos in total, however each participant was shown between 17 and 25 videos ($M = 23$) due to time constraints. The order of the videos was randomized, and each participant was shown a roughly even number of videos with CTS 'on' versus 'off'.

Research Design

The dependent variables included 'judgement accuracy' (correct or incorrect), 'reported confidence level' (rated using a 5-pt Likert item with '1' representing no confidence and '5' representing extremely confident), and the time that it took to arrive at a decision (seconds). The independent variable was the visibility of eye movement data ('on' vs. 'off').

Procedure

The videos were presented to the five instructors (individually). A video conferencing tool and its screen-sharing function was used to display each video, which was played remotely from the researcher's laptop. The participants watched each video and indicated to the researcher the point at which they had made a judgement on the pilots' proficiency, as well as the judgement itself; together with their 'confidence level' (on a 5-point scale). The researcher also asked the participant questions concerning why they had made that judgement, and what information or data they specifically used to make the judgement.

STUDY 2 – RESULTS

Judgments of proficiency level

A total of 115 judgments were provided across the five instructors involved in the test. Correct judgments were made more frequently when CTS was off ($M = 90.0\%$; $SD = 0.14$) as opposed to when it was on ($M = 70.7\%$; $SD = 0.14$). The difference was not statistically significant however, $t(4) = 2.4$, $p > .05$. There was no apparent difference in confidence ratings between CTS 'on' ($M = 4.08$, $SD = 0.29$) and CTS 'off' ($M = 4.05$, $SD = 0.18$), $p > .05$. There was also no apparent difference in the time to arrive at a decision between CTS 'on' ($M = 36.7$ sec, $SD = 9.4$) and CTS 'off' ($M = 35.3$ sec, $SD = 12.13$), $p > .05$.

Instructors' remarks

Using screen-share to present videos over an internet connection led to quality issues with playback of the videos. As a result, the eye gaze vector and trail were often difficult to interpret and reportedly were not always relied upon. This was sometimes perceived as a calibration issue, or a general 'jumping around' of the gaze vector.

> *"Bad calibration makes it extremely difficult as I don't know if it is a calibration issue or a student looking at the wrong areas."*

A common theme was around requiring familiarity with the system and the information it provides.

> *"It's easier to tell whether it's a trainee or an instructor without the eye-tracking, as I'm not used to it. The eye-tracker jumping around can be distracting."*

> *"Useful tool, but instructors need exposure to become familiar with it."*

> *"[The eye-tracking is] pretty good. I'm getting used to the information."*

Other comments indicated that CTS provided helpful information that assisted instructors in interpreting and diagnosing proficiency. Information on aircraft performance and eye-tracking appears to be complementary.

> *"Occasionally the eye-tracking [showed that the pilot] was looking at areas he shouldn't have been . . ."*

> *"It was good with the eye-tracker, being able to see what he was looking at. I think this is an instructor – quick scan, on height, on speed. Not focusing for too long, checking analogue Vertical Speed Indicator – they keep the scan moving."*

> *"Performance and eye-tracking are equally relevant for base and ILS; the ILS shows more of the instrument scan, and a constant gaze track. That is a rapid scan, looking at the right things quickly, back to the HUD velocity vector, finding information quickly and using it, knowing where to look."*

Instructors indicated that the information on the CandyBar plots complemented the real-time gaze information, for example, when pilots looked out the side windows (at which point the gaze vector disappeared).

> *"This trainee is looking at the HUD a lot – they're fixated on the HUD – I noticed this from the CandyBar. When the scan goes out the window, I look at the CandyBar."*

> *"The eye-tracker was really good here. The CandyBar caught my attention, the chopping and changing of instruments highlighted that the scan was deliberate and not HUD fixated. The CandyBar is helpful in this case as it shows a very deliberate scan."*

> *". . . the CandyBar was useful in confirming inside/outside view."*

> *"Again, the CandyBar is really good – a delineated and deliberate scan to other instruments."*

Finally, instructors indicated that the system would be useful for replaying scanning performance to trainees.

> *"If I had this as a tool I could highlight where their scan breaks down, and refocus and prioritise where it needs to be improved. [As an instructor] you can tell [them] where it breaks down, but you can't show them."*

> *"This is a great training tool to show them what they looked at, and what they should have been looking at."*

As a follow-up to the Expert vs. Novice test, time-scale plots were produced of pilots' visualisations covering the beginning to the end of the base turns across each of the AOIs, using the quantitative eye movement data captured by CTS (Figure 4). These afforded an objective comparison of scan patterns between each of the pilots.

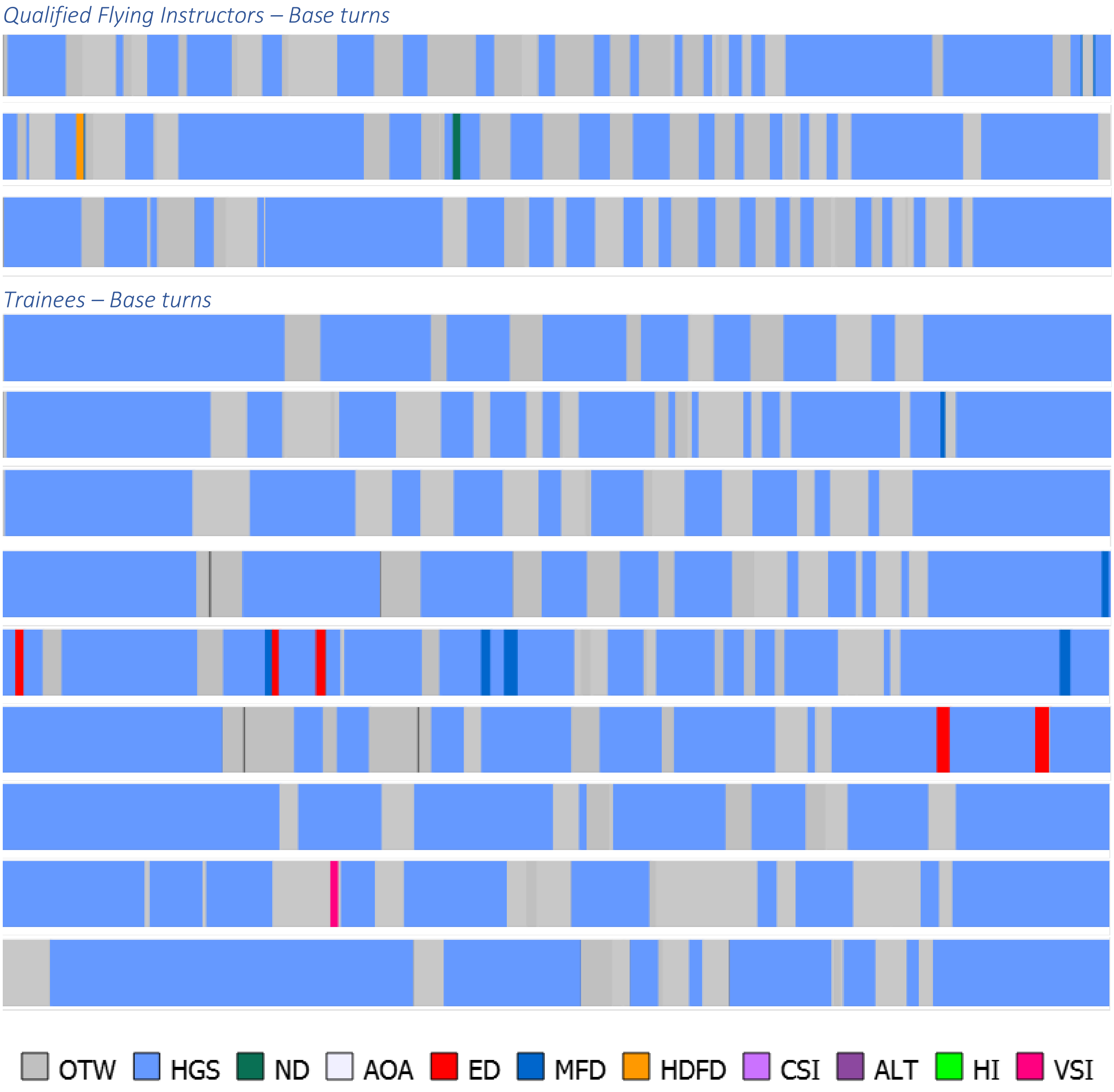

Figure 4. CandyBar plots of base turns that featured in the Expert vs Novice test. See Table 1 for acronyms.

STUDY 2 – DISCUSSION

There are several factors which may contribute to the unexpected results regarding correct judgements, including:

1) Stop-start nature of the videos (due to bandwidth) may have been disruptive to the real-time gaze data;
2) Calibration or accuracy of the eye gaze information was insufficient on some video replays and therefore may have been misleading;
3) Instructors may rely more on performance metrics than eye gaze data, particularly when they are inexperienced with CTS; and,
4) Pilots assigned to the 'novice' group (IFC trainees) had more experience than typical novices as they were qualified pilots undertaking a highly advanced phase of training.

Instructors' comments about how they made judgments when CTS was 'on' indicate two important points. First, there is a familiarization period where instructors that have had minimal or no prior exposure to the CTS console have some difficulty in becoming accustomed to the eye-tracking data presented and using it to interpret training performance. This is unsurprising, given that having access to trainees' eye-tracking data has not been possible until now. A familiarization period following the introduction of new technology is not uncommon. One implication of this is that instructors using such a tool would likely benefit with some basic training that shows what the different features on the console portray, and suggestions for how to interpret the overall scanning data and incorporate it into effective teaching methods.

Second, despite being unfamiliar with the CTS console, all five of the instructors appeared to have no difficulty in discerning ‘good’ scanning from poorer scanning behaviors. Further, the attributes that they cited as evidencing this (e.g., “looking at the right things quickly”) were consistent across the instructors and appeared to align closely with the objective data shown on the CandyBar plots.

The CandyBar plots helped instructors to quickly identify whether a pilot had spent time scanning internally at the Head-Down Flight Instruments, as well as externally (both HUD and outside HUD). This reportedly helped instructors to identify ‘HUD fixation’, which appears to be linked to the phenomenon of tunnel vision and failures to perceive critical aircraft information (Li et al., 2014). The plots also helped instructors determine whether scans were deliberate, structured, and disciplined in taking in key AOIs; as well as the relative rate and fixation of a scan. Also, when the scan gaze vector was not on-screen due to the trainee looking beyond the primary AOI (e.g., out the window), some instructors then used the CandyBar to determine where the trainee was looking.

GENERAL DISCUSSION

This research examined the efficacy of an unobtrusive eye-tracking system in an operational training environment with instructors and pilots-in-training. It was hypothesized that using the eye-tracking system would:

1) Improve instructors’ awareness of trainees’ scanning behavior, and
2) Help instructors to determine the proficiency level of pilots.

Study 1 addressed the first hypothesis. The findings indicated that instructors benefited from using the eye-tracking system. As hypothesized, instructors reported improvements in their awareness of trainees’ scanning behavior and greater situation awareness. Despite providing instructors with additional information, the system did not appear to increase their workload. These positive findings were dependent upon the system functioning as intended, however, as well as the nature of the training sessions.

Study 2 addressed the second hypothesis. With respect to instructors’ judgments concerning ‘expert’ (instructor) or ‘novice’ (trainee), there were no observable differences in instructors’ ability to determine pilot proficiency between observations supported by the eye-tracking system and observations without it. However, challenges around conducting the test and instructors’ lack of prior exposure with the system appear to have been largely responsible for this result. Interviews with instructors provided important subjective context with which to interpret the findings, as well as valuable insights on how to best integrate into operational environments.

A key learning from the current research is that systems integration—in terms of hardware installation, computing requirements, and software/user interfaces—is critical in achieving a high level of performance, reliability, and ease of operation in a high-tempo military flight training environment. Additionally, instructional workflow integration in terms of instructor training and understanding; and appropriate and configurable user interfaces for specific flight phases and sequences, as well as instructor preferences, is important to achieve the underlying objective of an eye-tracking based training enhancement – to better enable instructors to interpret and instruct.

Another important takeaway is that instructors should first be familiarized with the technology as there is a learning curve. There is also the question of whether instructors should be given guidance on how to interpret scanning behavior. However, instructors’ comments indicated that they were aware of attributes that are associated with ‘good’ scanning behavior and used these to inform their judgments. Examples include scanning that looked more ‘rapid,’ ‘disciplined,’ and ‘looking at the right things quickly’ (i.e., not dwelling for an extended period). These scanning attributes have been widely reported in prior research with ‘expert’ pilots (e.g., Ziv, 2016).

Although the system provided benefit to instructors, there was consistent agreement that it would be more useful at an earlier stage of training. The pilots involved with this study were fully qualified military pilots, selected for fast jet training on account of their higher-than-average performance during training, with several years of flight training experience; therefore, they had already adopted a relatively proficient scanning behavior. As a result, many instructors reported that they used the system for ‘checking’ scanning performance but would see it as more useful for remediation purposes (when inadequate scanning is identified) or with pilots who are still developing their scan patterns. The relatively advanced proficiency of trainees in the current research also likely contributed to the failure to observe differences between 'expert' vs 'novice' judgments in Study 2.

While the system was not used for debriefing purposes, instructors felt that being able to show trainees replays of their scanning behavior would enhance training, for example, through 'showing' trainees as opposed to just 'telling.' Instructors also indicated that the system would provide value for briefing prior to training sessions, as trainees could be shown examplars of other pilots (e.g., examples of high performance). The significant potential that eye-tracking based training aids offer during briefing and debriefing has been discussed in prior research (Peysakhovich et al., 2018). As noted above, the system must be suitably integrated with both existing computing infrastructure and instructor workflow to support its use.

There are several opportunities for future research that would build on the current studies. The first involves conducting an evaluation following a familiarization period and over a longer period. Secondly, while the qualitative data presented here suggests that an eye-tracking system would add value during briefings and debriefings, an experimental evaluation of this is required. Thirdly, research with earlier-stage pilot candidates is needed to confirm the additional benefit that eye-tracking may offer for pilots who are still developing their scanning behavior. Importantly, all the above should be tested with a system that is sufficiently integrated.

Conclusion

Despite the mixed results observed over the two studies, it is evident that eye-tracking systems can benefit instructors and have considerable potential to improve training outcomes in an operational context. This research has provided insights into the practicalities of implementing an eye-tracking based training tool into an established training context. To ensure optimal benefit to trainees and instructors, the technology must be well integrated into the overall training system to enable reliable presentation of information, and ease of use for instructors.

REFERENCES

Draeger, J. (2009). Visual ergonomy and flying safety. The importance of recognizing the instruments. *Ophthalmologe 106*(4), 370-373.

Fuchs, A. F. (1967). Saccadic and smooth pursuit eye movements in the monkey. *The Journal of Physiology*, *191*(3), 609.

Li, W. C., & Braithwaite, G. (2014). The investigation of pilots' eye scan patterns on the flight deck during an air-to-surface task. In *International Conference on Engineering Psychology and Cognitive Ergonomics* (pp. 325-334). Springer, Cham.

Peysakhovich, V., Lefrançois, O., Dehais, F., & Causse, M. (2018). The Neuroergonomics of Aircraft Cockpits: The Four Stages of Eye-Tracking Integration to Enhance Flight Safety. *Safety*, *4*(1), 8.

Robinski, M., & Stein, M. (2013). Tracking Visual Scanning Techniques in Training Simulation for Helicopter Landing. *Journal of Eye Movement Research*, *6*(2), 1–17. https://doi.org/10.16910/jemr.6.2.3

Sullivan, J., Yang, J. H., Day, M., & Kennedy, Q. (2011). Training simulation for helicopter navigation by characterizing visual scan patterns. *Aviation, Space, and Environmental Medicine*, *82*(9), 871–878.

The Boeing Company (2020). *Pilot & Technician Outlook 2020–2039.*

Thompson, N. (2018). Avoiding a pilot retention death spiral: the pressures of pilot shortage and Department of Defense retention efforts to maintain an effective fighting force. [Unpublished master's thesis]. Joint Advanced Warfighting School, Virginia. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/1051136.pdf

Todd, M. & Thomas, M. (2013). *Pilot experience and performance in an airline environment*, (AR-2012-023). Canberra. Australia: Australian Transport Safety Bureau.

Vidulich, M. A., & Tsang, P. S. (2019). Improving Aviation Performance through Applying Engineering Psychology: Advances in Aviation Psychology, Volume 3. CRC Press.

Werfelman, L. (2020). Anticipation. *Flight Safety Foundation*, 08 March 2020.

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