INSPECT: Understanding Trainee Cognitive Processes in ATC Training

Thomas Bellucci, Emmy Gabriel, Petra ten Hove, Maykel van Miltenburg, Daniela Pistone, Steven Niedenzu, Thimo Willems

Royal Netherlands Aerospace Centre (Royal NLR)

Amsterdam, The Netherlands

Thomas.Bellucci@nlr.nl, Emmy.Gabriel@nlr.nl, Petra.ten.Hove@nlr.nl, Maykel.van.Miltenburg@nlr.nl, Daniela.Pistone@nlr.nl, Steven.Niedenzu@nlr.nl, Thimo.Willems@nlr.nl

ABSTRACT

In the highly demanding field of air traffic control (ATC), where complex skills and extensive workload management are required, it is essential to have insight into the cognitive processes of trainees in order to facilitate effective training. Performance-based training (PBT) has been widely adopted in recent years to train air traffic controllers (ATCOs) on technical and procedural skills, such as adherence to communication protocols and separation procedures. However, the development of software tooling in support of PBT has primarily focused on the assessment of technical skills, and as a result, the assessment of non-technical (cognitive) skills, such as information perception and workload management, has remained challenging.

In this paper, we present INSPECT – a technology demonstrator with the aim of supporting instructors in coaching of non-technical competencies by providing objective insights into the cognitive processes of trainees. Our approach involves the use of eye tracking to relate the eye movements and pupil dilation of the trainee to the information presented on the trainee's radar screen, enabling one to derive a range of objective metrics, covering three major cognitive skills: situational assessment, workload management, and problem solving/decision-making. By leveraging data obtained from eye tracking and the flight data presented on the radar screen, INSPECT is able to derive a range of objective metrics concerning the perception of information, anticipation of inbound flights, mental workload, visual scanning cycle, and decision-making; results are then displayed in a dashboard for the purpose of debriefing. Moreover, to support the instructor during a training session, a live tool is developed to display the trainee's real-time eye movements on the radar screen, allowing the instructor to follow the scanning cycle of the trainee. Ultimately, our research demonstrates the potential of instructor support tools to transform raw data into insights, enhancing the effectiveness of ATCO training.

ABOUT THE AUTHORS

Thomas Bellucci, MSc. is an R&D engineer at the Royal Netherlands Aerospace Centre (NLR) at the department of Training & Simulation and works on research projects involving Artificial Intelligence (AI), software engineering, data science and machine learning. He studied Artificial Intelligence (AI) at the University of Amsterdam (UvA) and the Vrije Universiteit (VU), Amsterdam, completing both his BSc and MSc programs with distinction (summa cum laude).

Dr. Emmy Gabriel obtained a PhD in experimental particle physics from the University of Edinburgh in 2020, working for the Large Hadron Collider (LHCb) collaboration at CERN. She is currently employed as an R&D engineer in the Training & Simulation department of the Royal NLR. Her projects focus on data science, simulation operations and simulation technology.

Petra ten Hove is a training expert at the Royal Netherlands Aerospace Centre (NLR), working in the areas of training & simulation for civil and military aviation. She graduated from the University of Twente in 2014 with a Master's degree in Educational Science and Technology. Petra has a broad experience in training research across the aerospace sector and investigates the implementation of new training concepts and technologies such as performance-based training.

Maykel van Miltenburg earned his Master's degree in Applied Cognitive Psychology (MSc) in 2014 and is an R&D Engineer within Royal NLR's department of Safety and Human Performance. He participated in many of the recent NLR human-in-the-loop studies in the military domain. His expertise lies on measurement tools such as rating scales and a combination of

(psycho)physiological measurements such as electroencephalography, functional near-infrared spectroscopy, eye-tracking and electrocardiography.

Daniela Pistone, MSc. is a cognitive psychologist at the Royal Netherlands Aerospace Centre (NLR). She specializes in performance-based and personalized training innovations, focusing on understanding and optimizing cognitive processes to enhance training and assessment.

Steven Niedenzu, MSc. studied theoretical particle physics at the University of Amsterdam (UvA) and Vrije Universiteit Amsterdam (VU). After obtaining his degree, he started working at the Training & Simulation department of the Royal Netherlands Aerospace Centre (NLR) as an R&D Engineer. His research topics include data science, cross-domain security and simulation technology.

Thimo Willems, MSc. has completed a master in Industrial Design Engineering at the University of Twente and has since focused on the field of usability and human machine interaction. He has launched a startup that supports decision-making and programming of the regional energy transition in the form of an interactive data visualization platform. Currently, he works at the Royal Netherlands Aerospace Centre as R&D Engineer in the Human Machine Interaction Technology team.

INSPECT: Understanding Trainee Cognitive Processes in ATC Training

Thomas Bellucci, Emmy Gabriel, Petra ten Hove, Maykel van Miltenburg, Daniela Pistone, Steven Niedenzu, Thimo Willems

Royal Netherlands Aerospace Centre (Royal NLR)

Amsterdam, The Netherlands

Thomas.Bellucci@nlr.nl, Emmy.Gabriel@nlr.nl, Petra.ten.Hove@nlr.nl, Maykel.van.Miltenburg@nlr.nl, Daniela.Pistone@nlr.nl, Steven.Niedenzu@nlr.nl, Thimo.Willems@nlr.nl

INTRODUCTION

Air traffic control (ATC) is a high-stakes, safety-critical profession that demands exceptional skill, attention and problem-solving abilities. The primary task of an air traffic controller (ATCO) is to organize the flow of air traffic in a designated region of airspace, known as a sector, in order to optimize traffic throughput, coordinate landings and ensure safe separation of aircraft in the airspace. As global air traffic continues to grow, the need for well-trained ATCOs is becoming increasingly important. The safety and efficiency of air traffic operations rely heavily on the ability of a controller to make quick and accurate decisions, in often complex and dynamic situations. To achieve this, ATCOs must continually monitor aircraft movements, issue clearances and instructions to pilots, and collaborate with surrounding ATC centers to ensure seamless transition of traffic between airspaces.

While technical skills, such as adherence to communication protocols and separation procedures, are essential for ATCOs to master in order to efficiently and safely manage air traffic, it is equally important to develop strong non-technical cognitive skills, such as situation assessment and workload management, which are of utmost importance for air traffic management under high workload (Lee, Jeon, & Choi, 2012). However, coaching ATCOs on non-technical abilities poses a significant challenge as these are often difficult to quantify, observe, and assess.

In this work, we present INSPECT – a technology demonstrator with the aim of supporting instructors in coaching non-technical skills by providing comprehensive insights into the cognitive processes of the trainee. INSPECT works by integrating eye tracking and data analytics algorithms into military ATC training. In training of military ATCOs, trainees engage with a simulated ATC environment consisting of a radar display and a Touch-Input Display (TID), mimicking a real-world air traffic control scenario (Figure 1). By fusing data obtained from a display-mounted eye tracker with the flight data presented on the trainee's radar display, INSPECT can derive a range of metrics related to the cognitive processes of the trainee, ranging from metrics measuring information perception and anticipation of flights entering one's sector, to mental workload and visual scanning behavior. In doing so, INSPECT goes beyond mere analysis of technical competencies by revealing aspects of the cognitive processes of the ATCO. Moreover, by analyzing the contents of the controller's command logs and radio communication, various metrics concerning task-load, problem-solving and decision-making were implemented which can aid the instructor in guiding the training process.

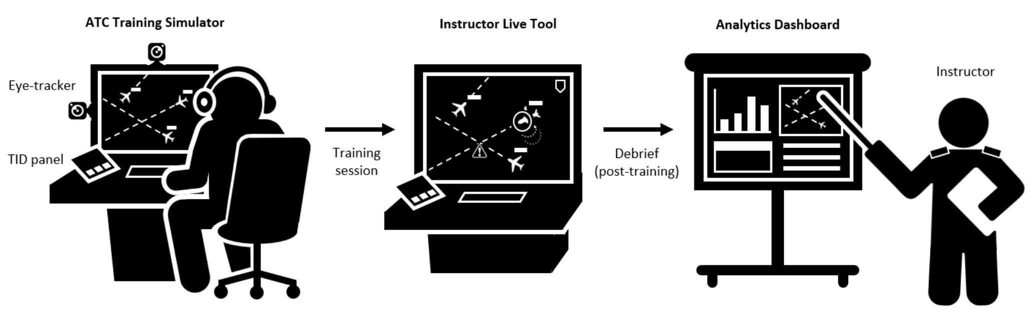

Figure 1: Conceptual Illustration of the INSPECT Demonstrator in a Military ATC Training Environment

To achieve this, two software functionalities were implemented as part of INSPECT:

- A live visualization tool to display the trainee's eye movements as an overlay on a copy of the radar display, allowing an instructor to monitor scanning patterns and identify infrequently observed aircraft while the trainee manages traffic.
- A dashboard for post-session debriefing to display a range of quantitative metrics concerning the non-technical abilities of the ATCO. This dashboard contains a data analytics pipeline to process eye tracking data and data from the ATCOs radar display, enabling the extraction of quantitative, time-varying metrics concerning the perception of information, anticipation of flights, radio communication, scanning cycles, mental workload, task-load and decision-making.

We elaborate on the INSPECT demonstrator and present some preliminary results of user tests.

BACKGROUND AND RELATED WORK

Performance-Based Training in ATC

Air traffic management involves a multitude of tasks that place significant demands on cognitive capabilities. As a result, air traffic controllers must possess a unique set of skills, which can be challenging to acquire. To assess the cognitive demands of the ATCO, a structured approach is necessary to identify the factors that contribute to the complexity of air traffic controllers' cognitive processes. To this end, the main provider of air traffic control services in the Netherlands – Luchtverkeersleiding Nederland (LVNL) – has proposed the ATC Cognitive Process & Operational Situation (ACoPOS) model (Schuver-van Blanken, Huisman, & Roerdink, 2010), which distinguishes between three key cognitive processes (Figure 2):

1. *Situational Assessment* (SA) describes an ATCO's mental representation of the current situation, involving three phases: *perception of information*, *interpretation* (creation of a complete situational picture), and *anticipation of the future situation*. SA is cognitively demanding and requires continual monitoring of the radar display to maintain.
2. *Workload Management* (WLM) involves the direction of attention to relevant information sources in order to maintain situational awareness. ATCOs use visual abstractions to represent essential characteristics of the situation, allowing them to regulate and set priorities to avoid information overload and reactive action.
3. *Problem Solving and Decision-Making* (PS&DM) involves naturalistic decision-making, where ATCOs search for patterns and relevant visual cues to establish a course of action, rather than 'rationally' comparing possible alternatives to select a certain optimal action. ATCOs continuously develop and adjust plans to safely manage traffic under constraints, creating alternative solutions and backup plans as needed.

The outcome of these processes is a set of *actions*, which may comprise radio communication with pilots to issue clearances or instructions, as well as coordination of actions with controllers of other sectors.

An alternative model of the ATC cognitive process, from the perspective of conflict detection and resolution, is the Conflict Life Cycle model proposed by (Nordman, Meyer, Klang, Lundberg, & Vrotsou, 2023). This model involves a four-stage process: first, a conflict between two aircraft is recognized, then a solution in the form of a directive call is proposed and implemented; finally, the solution is monitored to observe whether a found solution actually resolves the problem.

Learning Analytics

To generate objective data for specific competencies from the ACoPOS competency model, Learning Analytics (LA) are required. Learning Analytics concern the collection, analysis and reporting of data about trainees and the training context. LA can support instructors in monitoring trainee performance, identifying learning trends, and in visualizing progress during a trainee's learning trajectory. A learning ecosystem facilitates performance-based training by supporting the (real-time) collection, storage, and distribution of data, as well as efficiently unlocking that data for LA purposes, providing insights into the data and analyses through effective data visualizations.

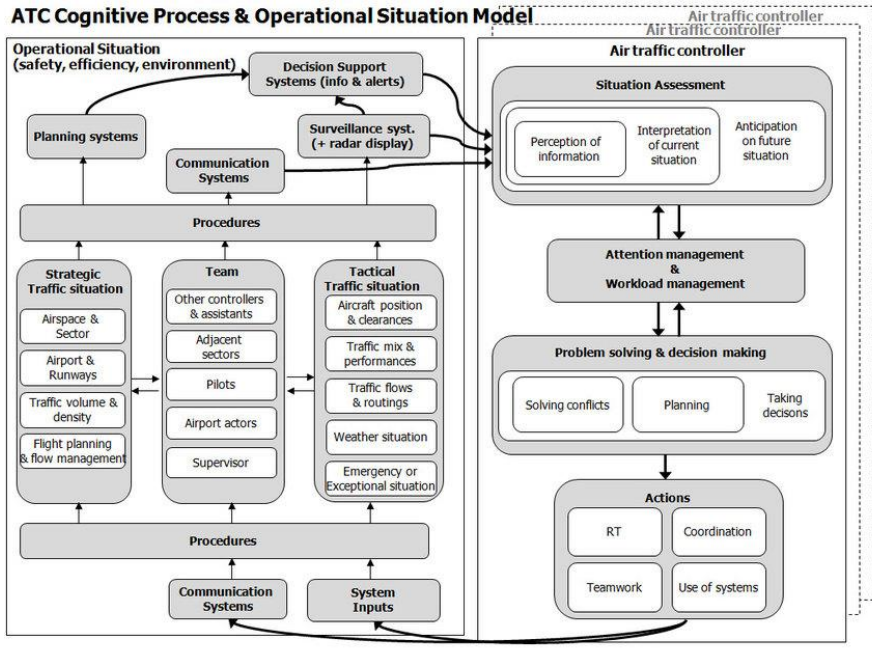

Figure 2: The ATC Cognitive Process & Operational Situation (ACoPOS) Model. Reproduced from (Schuver-van Blanken, Huisman, & Roerdink, 2010)

Eye Tracking

Eye tracking involves the use of specialized cameras and algorithms to record and analyze the movements of an individual's eyes as they perform a task. In the context of ATC, eye tracking provides a window into the cognitive processes of controllers, allowing researchers to examine the ATCO's visual scanning cycle in a non-invasive manner and determine which information on the radar display is observed by the ATCO and which information is not. While it is not possible to infer from mere behavioural data that the perceived information is also understood and interpreted correctly by the ATCO; by analyzing the eye movements of ATC trainees, researchers can discover patterns in their visual attention and scanning cycle, as well as gain insights indicative of various decision-making strategies (Joseph & Murugesh, 2020).

Derived metrics from eye trackers, such as fixation duration, saccade length, and pupil dilation have additionally been shown to be reliable indicators of cognitive load, attention, and mental workload (Joseph & Murugesh, 2020; Tole, Stephens, Harris Sr, & Ephrath, 1982; Ahern & Beatty, 1979). In ATC, these metrics have been used to identify moments of high cognitive demand or fatigue, such as during conflict resolution, which can guide training (Jonasson, 2023; Rodriguez, 2015). Using eye tracking-derived metrics to assist in coaching non-technical skills with simulator data may thus be able to augment and complement the military ATC training process used today.

SYSTEM OVERVIEW

ATC Training Environment

INSPECT was developed with Approach Control (AC) training in mind, as practiced at Schiphol Airport for the military airbases in the Netherlands. In military AC, a controller is tasked with guiding arriving aircraft from their en route phase to their final approach and landing at a specific airbase, while simultaneously taking corrective actions to prevent a loss of separation between aircraft. At Schiphol, ATC training is carried out using a whole-task ATC simulator capable of mimicking real-world air traffic scenarios. The simulator consists of a pair of controller stations, each equipped with a radar display that can be flexibly deployed as a controller or instructor station.

Data collection sessions were organized in order to develop the INSPECT software and infrastructure. These sessions consisted of complex and congested traffic scenarios in Dutch airspace focusing on the Eindhoven area. Background radio telephony was simulated with pseudo-pilots. The instructor was positioned to the left of the controller, using the display of the neighboring station for observation.

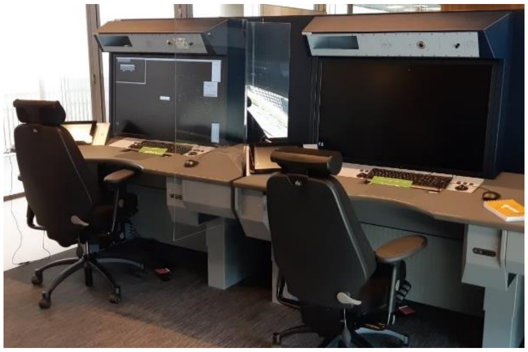

Figure 3: Instructor Station (left) and Controller Station (right)

Figure 4: Close Up of the Controller Radar Display

The controller position was equipped with a Smart Eye Pro eye tracker (Smart Eye, 2023) to track the pupil and eye gaze of the controller, i.e. where on the radar display the controller is looking. The Smart Eye Pro is a remote eye-tracking system that utilizes multiple infrared (IR) cameras (in this case 4) to provide 360-degree head and eye tracking capabilities. Its flexible camera placement and orientation options make it ideal for the ATC training environment which may use varying display setups. The IR cameras of the eye tracker were installed around the radar display and pointed towards the seating position of the controller.

Technical Infrastructure

As stated in Section 1, the INSPECT software consists of two standalone demonstrator applications: 1) a *live visualization tool* to support the instructor during a training session, and 2) a *debrief dashboard* to calculate and display a range of quantitative metrics from the data collected during the session, covering the competencies of the ACoPOS model. The high-level architecture of the INSPECT software is illustrated in Figure 5. The software architecture provides three functionalities:

1. The ability to collect, synchronize and analyze eye tracking data from the Smart Eye Pro and scenario data from the ATC training simulator, to obtain insightful metrics related to the competencies of the ACoPOS model: information perception (SA), workload management (WLM) and decision-making (PS&DM).
2. A debrief analytics dashboard to display the analysis results inside a graphical user interface.
3. An instructor live tool at the instructor position to display the eye gaze and scan pattern of the student in real-time on a copy of the radar screen of the trainee.

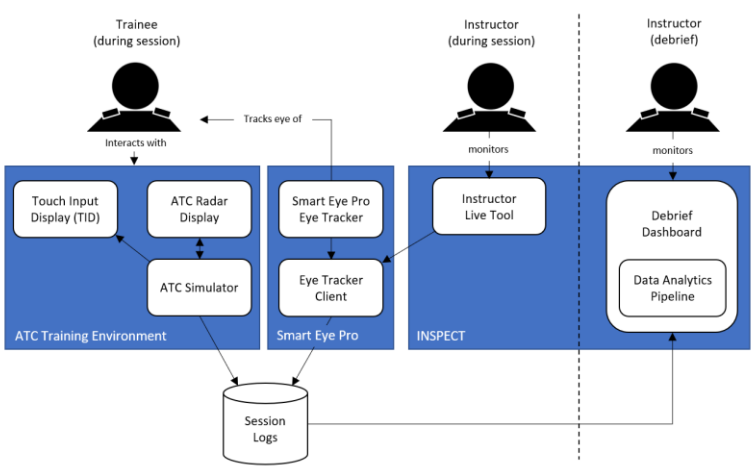

Figure 5: Container Diagram of the INSPECT Software Infrastructure

Instructor Live Tool

A live visualization tool was developed to support instructors with monitoring the scanning cycle of the trainee during a training session. To achieve this, the instructor live tool overlays the trainee's eye movements on top of the trainee's radar display duplicated at the instructor position. Moreover, to assist the instructor in identifying under-attended traffic, the live tool highlights infrequently observed flights and the QNH indicator, as illustrated in Figure 6. The QNH presents the altimeter setting, equivalent to the atmospheric pressure adjusted to sea level. When all aircraft are using the same altimeter setting, a safe vertical separation can be maintained.

To give the instructor insight into the trainee's scanning pattern, the eye gaze (i.e., the average eye movement of the left and right eye) is displayed as a circular blue marker on screen. This marker moves with the eye gaze of the trainee in real-time. Since it is desirable for ATCOs to actively monitor the airspace and repeatedly return to previously seen flights, the instructor live tool highlights the labels of the flights which are infrequently observed. Repeatedly highlighted labels on the radar screen thereby give the instructor an indication of the flights that are not monitored sufficiently by the trainee. By default, label highlights are triggered 30 seconds after the aircraft has last been seen; if the trainee's gaze wanders over a highlighted label, the colored highlight disappears.

Lastly, after constructive discussions with end-users, it was decided to include an option to flag important moments or events during the session by clicking on a flag icon at the top right corner of the instructor screen. These flags mark moments as important to allow one to easily refer back to these moments during the debrief (see *Debrief Dashboard*).

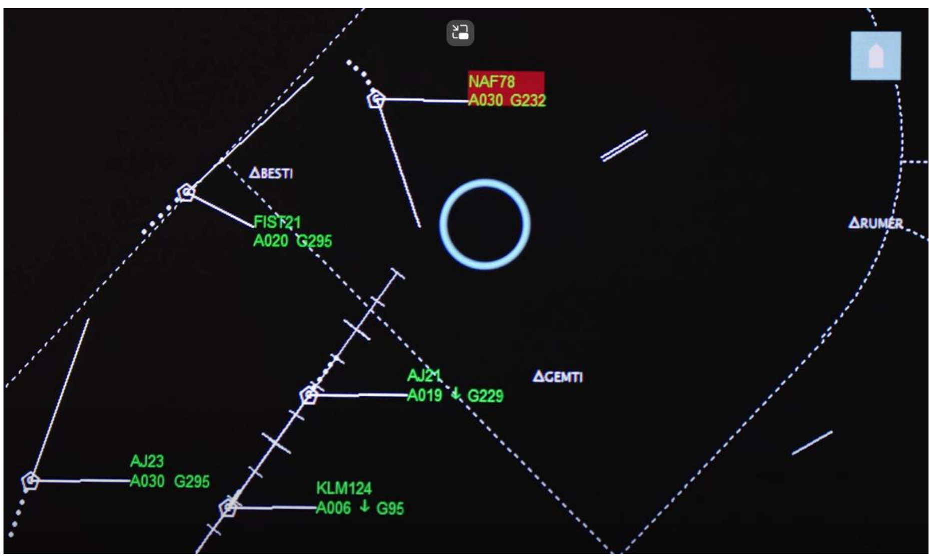

Figure 6: Snapshot of the Instructor Live Tool with Real-Time Eye Gaze and Label Highlighting of Flight NAF78 (Red = Unobserved). The Top-Right Shows a Flag Icon to Allow the Instructor to Key Moments During the Session.

Data Analytics

To populate a debrief dashboard with insightful metrics, an automated data analytics pipeline was developed to process the eye tracking and radar display data collected during each training session. The pipeline takes two primary data streams as input: (1) the eye gaze captured by the Smart Eye Pro eye tracker, responsible for tracking the trainee's eye movements on the radar screen, and (2) the flight data generated by the ATC training simulator, which provides contextual information about the simulated scenario and timestamped locations of flights on the radar display throughout the scenario. Both data streams are recorded continually in real-time during each session and are analyzed once the session is completed. A detailed breakdown of the content of each data stream is presented in Table 1.

Table 1. Definition of the Data Streams from the Smart Eye Pro Eye Tracker and ATC Training Simulator. N Denotes the Total Number of Aircraft Present in the ATC Scenario.

Eye gaze	Flight data (×N)	
UNIX Timestamp	UNIX Timestamp	Target screen coordinate
Gaze screen coordinate	Callsign	Flight level (FL)
Pupil diameter	Label screen coordinate	Under control (True/False)
Pupil quality (0 - 1)	Label extent (height, width)	Traffic type (in-, out-, crossing)

To realize the functional objectives of the software and strike an optimal balance between providing instructors with informative insights while avoiding information overload, a modular design was employed, following the ACoPOS model. Module I focuses on the competence Situational Assessment as a collection of non-technical skills, while Module II focuses on Workload Management as a non-technical skill, and Module III on Problem Solving & Decision-Making, which includes both non-technical and technical skills. The three modules are elaborated upon below.

Module I - Situation Assessment

As laid out in the ACoPOS model, one may distinguish between at least two distinct mental processes in the realm of situational assessment: *"perception of information"* and *"anticipation of the future situation"*. To support the instructor on the assessment of these processes, two metrics were developed. These metrics aim to quantify the degree of perception of flights on screen and the anticipation of flights entering one's sector.

Perception. The perception of the ATCO is calculated every 50ms during the scenario and is defined as the percentage of flights that have recently been seen, relative to the total number of flights that are under control at that time. The perception thus quantifies the distribution of attention over flights that are currently under control by the controller.

$$Perception = \frac{|\{observed\} \cap \{under\ control\}|}{|\{under\ control\}|} \qquad (1)$$

Here we consider an aircraft as 'observed' when a dwell (defined as a fixation of 150ms) has been registered over the aircraft or its corresponding textual label in the last 30 seconds. If there are multiple dwells within this period on the label (i.e., the label has been viewed more than once within 30 seconds), then this period is extended to 30 seconds after the last dwell ends.

Anticipation. Anticipation refers to how quickly a flight is noticed on screen before it enters one's sector. To quantify this, a measure of anticipation is calculated for inbound flights as the time difference between the first instance a flight becomes visible on screen and the moment a dwell is first registered on it, after which the flight can be considered 'observed' by the controller.

$$anticipation = t_{dwell} - t_{appear}, \qquad (2)$$

where t_{dwell} is defined as the time at which the flight was first dwelled on, and t_{appear} is defined as the first time the flight is shown on screen.

Module II - Workload Management

To analyze the controller's ability to manage high workloads, algorithms were designed that use eye gaze data to infer measures related to work- and task load.

Pupil dilation. In this study, relative pupil dilation was utilized as a physiological measure to assess the workload of ATCOs during a training exercise. Pupil dilation is a well-established indicator of cognitive load, as it is sensitive to changes in mental effort and attention, on the condition that other factors, such as illumination remain constant. When an individual is faced with a mentally demanding task, their pupils dilate to allow more light to enter the eye, facilitating increased visual processing and attention. As a result, rapid increases in pupil size, may be indicative of increased workload.

Communication frequency and duration. As frequency and duration of radio communication have been linked to work- and task load, we derive the density of incoming and outgoing communication of the controller. Communication frequency may then serve as an additional indicator and complementary measure of workload.

Module III - Problem Solving and Decision-Making

Conflict resolution. The ATCO's primary responsibility is to ensure the safety of air traffic by preventing *conflicts*, which are situations where aircraft come too close to each other. Formally, a conflict is said to occur when two aircraft breach the minimum separation requirements, which are typically defined as 3 to 5 nautical miles horizontally and 1000 feet vertically.[1] For an instructor it is beneficial to know when a conflict is expected to occur between two aircraft and when the controller took preventative action to resolve it, thereby measuring the efficiency of the decision made and the effective reaction time of the controller.

The conflict resolution system works by continuously monitoring the position, airspeed and heading of all aircraft within the airspace. It uses an extrapolation algorithm to analyze the trajectories of aircraft and identify any moment in the future, up to 20 minutes in advance, where two or more aircraft are projected to lose horizontal and vertical separation. Once a potential conflict is detected, the system calculates the future time at which the loss of separation was expected to occur, taking into account factors such as the aircraft's position, airspeed, heading and estimated flight level. It then determines when (and whether) a command was issued by the controller during the session which prevented the previously detected conflict from materializing.

Categorization of Scanning Patterns. ATCOs continuously monitor the situation in the airspace to detect and resolve potential conflicts as quickly and efficiently as possible. How someone solves such problems and makes decisions varies from person to person, and in the context of air traffic controllers it strongly depends on the ATCO's experience and the problem-solving phase they are in. Recent work (Fraga, Kang, Crutchfield, & Mandal, 2021) shows that each person prefers a different approach, but everyone scans the area in a systematic way, for example, using a circular, spiral, or linear pattern, depending on their problem solving phase.

In order to recognize the problem solving phase of the ATCO, we implemented an algorithm to recognize four modes of scanning behavior:

1. **Monitoring**: the trainee gazes at a series of aircraft repeatedly in a fixed order.
2. **ABAB**: the trainee alternates between two aircraft, where he/she looks back and forth between the aircraft at least three times.
3. **Tunneling**: the trainee focuses on a maximum of two aircraft for at least 45 seconds, failing to attend to the rest of the aircraft present on screen.
4. **Entropy**: no clear scanning pattern, thus none of the other scanning patterns are recognized. Alternatively, the trainee may be looking away from the screen to accept a call, interact with the TID, etc.

Debrief Dashboard

For the purpose of debriefing, the metrics above are displayed in a dashboard application (see Figure 7). The dashboard is organized into three tabs corresponding to the three aforementioned modules. Each tab shows a page with split-view layout, where the left side shows a visual presentation and graphs of the metrics obtained, and the right side shows the playback of the radar screen with the eye gaze as overlay. Both sides are synchronized in time; clicking a point of interest in any graph will jump the playback to that moment in time, and vice versa. This allows for examining specific moments of interest and reviewing which aircraft was observed at what moment. The right hand side of the

[1] A minimum separation of 5 nautical miles is required between two aircraft when only of the aircraft is under control of the ATCO; when both aircraft are managed by the same controller, a minimum separation of 3 nautical miles must be enforced.

dashboard also displays the transcriptions of the radio communication between the controller, pilots and supervisor, complementing the video playback. These transcriptions were obtained using Whisper speech recognition software (Radford, et al., 2022).

The left panel is dedicated to displaying metrics specific to each phase of the training session. By pressing one of three tabs, instructors can select the module of choice, and the corresponding metrics will be displayed as a function of time. This layout was deliberately designed in collaboration with an experienced ATC instructor through iterative end-user validation, ensuring that the dashboard meets the needs and expectations of its intended users. By providing a clear and comprehensive visualization of the trainee's performance, the dashboard facilitates effective debriefing and feedback, enabling instructors to identify areas for improvement and optimize the training process. That said, the user interface was carefully crafted as to not impose any judgement of the trainee's performance that cannot be made based on just those obtained metrics.

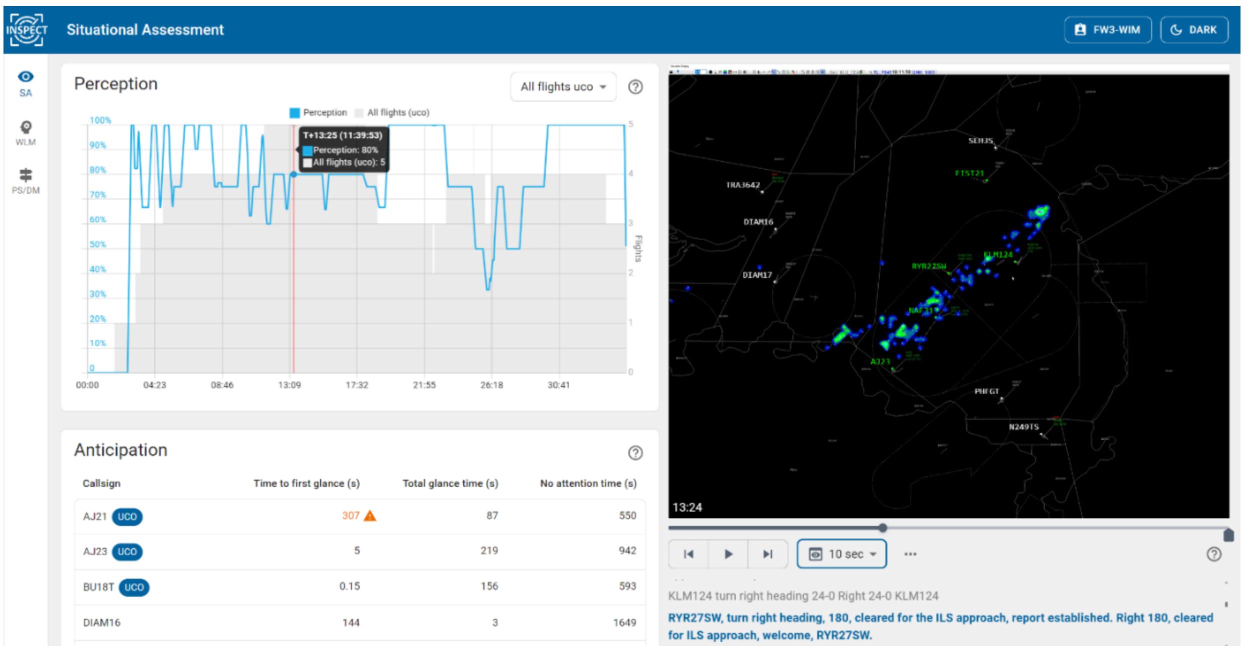

Figure 7: Dashboard Showing the Situation Assessment (SA) Page with Perception and Anticipation Metrics (left); the Right Side Displays a Playback of the Session with Communication Transcripts, Time-Synchronized with each Metric.

DEMONSTRATION OF CAPABILITIES

As part of the ongoing development of INSPECT, a preliminary examination was conducted on existing training sessions to assess the ability of the software to provide actionable insights. To test whether moments of reduced SA or high workload, relevant for guiding future training, are easily recognized from the metrics developed, a comparison between two training sessions was made.

To this end, two sessions were identified, initially used to develop the software, which showed distinctly different environmental contexts and skill levels, influencing the behavior of the participant:[2]

- **Session 1**: A session involving a beginning ATCO trainee (< 3 months experience) in which the training context yielded several moments during the session where the trainee was distracted and unable to focus on the approach control task at hand.

[2] Note that as several factors were varied at once between the two sessions, no conclusions must be drawn regarding the cause of variation in any given metric. Our demonstration of capabilities merely illustrates that our metrics display stark visual differences between sessions.

- **Session 2**: A session involving an experienced ATC instructor where no distractors were present during the training session.

Figure 8 (left) shows the information perception calculated for each of the two sessions. As can be seen, moments of reduced SA are easily recognizable for session 1, where the fraction of aircraft recently observed by the controller frequently drops below 0.5; in Session 2, flights are more consistently monitored, yielding a near-level perception envelope. In terms of pupil dilation, shown in Figure 8 (right), a similar pattern is observed, showing rapid changes in pupil size (i.e. pupil dilation) throughout Session 1, indicative of changes in workload; while pupil dilation remains more consistent throughout Session 2.

Note that, although the baseline pupil dilation also differs between sessions, with Session 1 having a much higher baseline pupil dilation than Session 2, this must not be attributed to workload as baseline pupil dilation is most likely caused by differences in the participants' baseline pupil size and/or ambient lighting conditions; baseline dilation cannot be meaningfully compared between sessions.

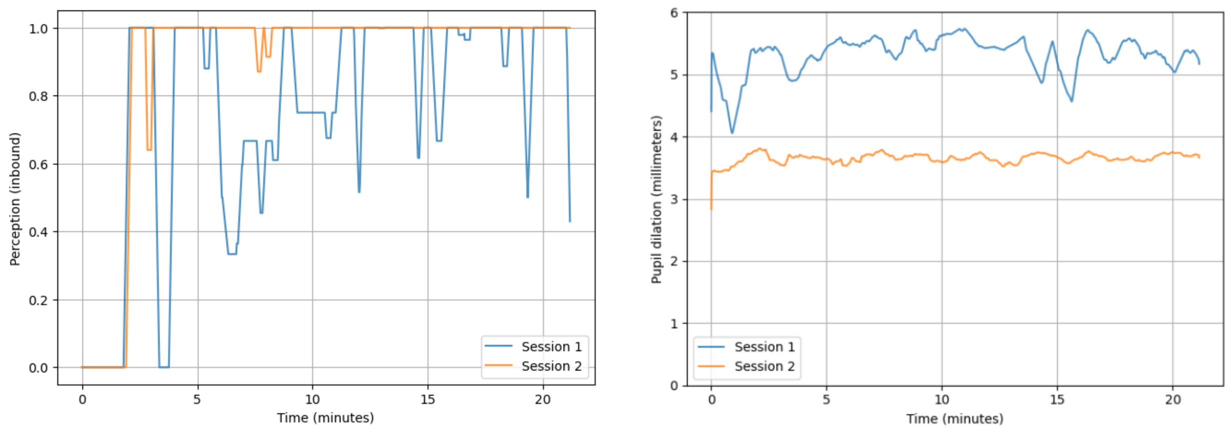

Figure 8: Perception (left) and Pupil Dilation (right) Curves for Two Separate Development Sessions

DISCUSSION AND FUTURE WORK

As highlighted, in ATC training, understanding the cognitive processes of trainees captures a vital component of improving the effectiveness of ATC training programs. To increase this effectiveness, INSPECT was designed with the goal of supporting instructors in coaching trainees in improving their non-technical skills. INSPECT implements an instructor live tool that shows the trainee's eye gaze on top of the radar display of the trainee, duplicated at the instructor station, allowing the instructor to follow the trainee's scanning cycles and identify under-attended traffic. Additionally, a debrief dashboard was developed to provide the instructor with an overview of metrics regarding the trainee's cognitive processes.

Visible differences were found between sessions of different ATCOs; although, the validity and reliability of these differences cannot be confirmed from a small demonstration. Despite this, the existence of visual differences support the expectation that one can use the metrics implemented in INSPECT to yield insights regarding the training, and identify moments of interest were SA was likely reduced and workload increased. However, in order to validate the utility of the calculated metrics, a validation study must be carried out. When the metrics in INSPECT are found to provide actionable insights into the cognitive processes of trainees in an operational training environment, the instructor and trainee will greatly benefit from them, as it will then become possible to pinpoint, with higher certainty, a trainee's strong and weak points.

While the use of eye tracking can aid in training of ATCOs, eye tracking has shown limitations that may need to be addressed. One of these limitations is the amount of time it takes to setup and calibrate the eye tracker for a single use. This is done by performing a series of tasks, including one-time setup and installation, and calibration of each IR camera, along with a head model of the trainee (if a model was not made already in an earlier session). After setup

and calibration, the simulator must be time-synchronized with the eye tracking system. Eye trackers with automatic calibration (e.g. Tobii Pro) could be a viable alternative to speed up the setup of the eye tracker in time-constrained situations where installation and calibration time ought to be minimized; since calibrating the Smart Eye Pro requires significant human effort, the user must ultimately decide when the eye tracker is calibrated well enough.
Moreover, in our research, the Smart Eye Pro has only been tested for one monitor; however, extending its use to two or more monitors, including the TID, would greatly enhance its strength, as ATCOs operating in a real ATC tower are required to keep an eye on multiple monitors at once. In the case of using multiple monitors, the Smart Eye Pro will have a significant edge over eye trackers that are capable of tracking only a single monitor.

Finally, we note that a trainee’s progression could be measured by tracking their metrics over time. For example, during a training session the pupil diameter should fluctuate less as the trainee is becoming more experienced. Another similar example is conflict detection, where an experienced trainee should be able to solve conflicts faster. An extension to the debrief dashboard would be to have the possibility to load multiple sessions into it, simplifying the tracking of progression. However, in order to achieve this, one would need access to a measure of workload that is insensitive to environmental factors that may influence the measurement of workload, such as varying levels of ambient light. What is more, visualising the actual understanding of the situation has not yet been realised. This does, however, require the definition of an ‘optimal’ approach to which the trainee’s performance can be compared, which is hard to define when it comes to handling air traffic, where multiple approaches can be considered appropriate.

Future work will focus on validating the metrics and algorithms implemented in INSPECT to ensure its effectiveness in providing ATC instructors with the necessary information to assess the non-technical cognitive skills of their trainees. A crucial factor will be a multi-factor experiment to systematically investigate how various skill levels and behaviors translate into our metrics. By doing so, we will be able to validate the system's accuracy, reliability, and relevance in capturing the complex interactions between trainees' cognitive skills and their performance in ATC scenarios. This comprehensive evaluation will ultimately inform the refinement and optimization of the INSPECT system, ensuring its utility for future adoption in existing ATC training environments.

CONCLUSION

This research has demonstrated the potential of instructor support tools for ATC training. A technology demonstrator named INSPECT was developed to support instructors in their coaching of non-technical skills by providing insights into the cognitive processes of trainees in military ATC training. By integrating eye tracking and data analytics into a simulated ATC training environment at Schiphol Airport, INSPECT aims to provide instructors a comprehensive overview of the trainee's cognitive processes, by calculating metrics concerning the perception of information, anticipation of inbound flights, mental workload, and visual scanning cycle. Our demonstrator supports the coaching of non-technical skills by integrating related metrics into a debrief dashboard. Furthermore, the system's ability to analyze commands and radio communication logs enables a range of objective metrics related to workload, task load, problem solving, and decision-making to be derived. The first demonstration of capabilities suggests promising applications of instructor support tools for enhancing the training of military ATCOs, ultimately contributing to improved ATC performance and safety.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to the ATC Training Center Polaris at Schiphol-Oost for their support in providing the training facilities to develop the software and funding of this research project.

REFERENCES

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science*, 1289--1292.
Fraga, R., Kang, Z., Crutchfield, J. M., & Mandal, S. (2021). Visual search and conflict mitigation strategies used by expert en route air traffic controllers. *Aerospace*, 170.
Jonasson, S. (2023). The criterion-related validity of pupil diamater as a measurement of mental fatigue in air traffic controllers.

Joseph, A. W., & Murugesh, R. (2020). Potential eye tracking metrics and indicators to measure cognitive load in human-computer interaction research. *J. Sci. Res*, 168--175.

Lee, Y. H., Jeon, J.-D., & Choi, Y.-C. (2012). Air traffic controllers' situation awareness and workload under dynamic air traffic situations. *Transportation journal*, 338--352.

Nordman, A., Meyer, L., Klang, K. J., Lundberg, J., & Vrotsou, K. (2023). Extraction of CD&R Work Phases from Eye-Tracking and Simulator Logs: A Topic Modelling Approach. *Aerospace*, 595.

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). *Robust Speech Recognition via Large-Scale Weak Supervision.* Arxiv.

Rodriguez, S. a. (2015). Pupillometry to assess air traffic controller workload through the mental workload model. In *Proceedings of the 5th international conference on application and theory of automation in command and control systems* (pp. 95--104).

Schuver-van Blanken, M., Huisman, H., & Roerdink, M. (2010). The ATC Cognitive Process & Operational Situation Model. *European Association for Aviation Psychology Conference.* Budapest, Hungary.

Smart Eye pro - Remote Eye Tracking System. (2023, May 24). Retrieved from Smart Eye: https://www.smarteye.se/smart-eye-pro/

Tole, J., Stephens, A. T., Harris Sr, R., & Ephrath, A. (1982). Visual scanning behavior and mental workload in aircraft pilots. *Aviation, Space, and Environmental Medicine*, 54--61.