How eye tracking data can enhance human performance in tomorrow’s cockpit. Results from a flight simulation study in FUTURE SKY SAFETY
Marcus Biella, Matthias Wies, Rebecca Charles, Nicolas Maille, Jim Nixon

To cite this version:
Marcus Biella, Matthias Wies, Rebecca Charles, Nicolas Maille, Jim Nixon. How eye tracking data can enhance human performance in tomorrow’s cockpit. Results from a flight simulation study in FUTURE SKY SAFETY. Joint AIAA and Royal Aeronautical Society (RaeS) Fall Conference on Modeling and Simulation for ATM, Nov 2017, LONDRES, United Kingdom. hal-02183334

HAL Id: hal-02183334
https://hal.archives-ouvertes.fr/hal-02183334
Submitted on 15 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
How eye tracking data can enhance human performance in tomorrow’s cockpit.

Results from a flight simulation study in FUTURE SKY SAFETY.

Marcus Biella¹, Matthias Wies¹, Rebecca Charles², Nicolas Maille³, Jim Nixon²

¹ DLR (German Aerospace Center), Institute of Flight Guidance, Braunschweig, Germany
² Cranfield University, Centre for Safety & Accident Investigation, Cranfield, UK
³ ONERA Systems Control and Flight Dynamics Department, Salon de Provence, France

Adaptive automation appears to be one of the next most challenging milestones in the field of aviation and cockpit development. This approach is aimed to be human centred (Billings, 1996) and has to be user-friendly in order to increase safety (Flightpath 2050).

The detection of human operators’ performance inevitably has to be the first step in this activity (Maiwald & Schulte, 2011). Here the edges of acceptable behaviour have to be identified well in time, which means that points where human performance deteriorates have to be found.

This has led to the notion of a Human Performance Envelope (Edwards, 2013, 2014; Graziani et al., 2016). Within this context a reliable on-line recording and analysis of operational data is needed. In a second step, these data will then set the trigger and recommend when a redistributions of tasks has to take place and a higher level of automation has to take over.

A good candidate for the on-line recording of pilots’ workload and situation awareness are eye tracking data (Biella et al., 2005; Manske, 2015). First results from a flight simulator study with ten flight crews in the project “Human Performance Envelope” of the programme Future Sky Safety funded by the European Commission’s Horizon 2020 initiative will be presented in this paper. Gaze fixation and fast eye movements were measured in order to gain online insight for the flight crew’s situation awareness level 1 “perception”. Furthermore, the pupil diameter was measured to identify pilot’s workload.

A set of 22 areas of interest was defined in DLR’s A320 flight simulator AVES. Four of them match simulator’s outside view while the other ones are dedicated to instrument panels. The eye-tracking system continuously evaluates from the pupil measurements and the head position which area of interest the operator is looking at. For the recording of the eye movements the comfortable Eye Tracking Glasses of SensoMotoric Instruments (SMI) were used.

The results of this study will show how to increase the Human Performance Envelope to improve both performance and safety.
1. Introduction
During the last decades, Human Factors research tried to prevent or recover the degradation of performance that could bring the pilot out of the flight safety limits (ICAO, 1994; Harris, 2011; Sherman, 2003; Wiener, 1985). However, previous human performance and error research have focused on the effects of a single factor on performance (e.g. Svensson et al., 1997; Loft et al., 2007). In contrast, little is known about both the potential interaction between different factors and the availability of technical resources, e.g., HMI and automation support, enabling pilots to be in a situation where they have sufficient cognitive resources to perform efficiently their tasks.

Therefore, a new concept has been introduced called Human Performance Envelope (HPE). This concept was originally proposed in the Air Traffic Management (ATM) domain (Edwards et al., 2012; Edwards, 2013, Edwards et al., 2014), but has been extended to Human Operators in the cockpit in the European project “Future Sky Safety: Human Performance Envelope”. The aim of the project is to define and apply the Human Performance Envelope for cockpit operations and design, and determining methods to recover crew’s performance to the centre of the envelope, and consequently to augment this envelope.

The Human Performance Envelope is to some extent a new paradigm in Human Factors. Rather than focusing on one or two individual factors (e.g. fatigue, situation awareness, etc.), it considers a range of common factors in accidents and maps how they work alone or in combination to lead to a performance decrement that could affect safety. The safe region on the envelope is bordered by markers, which can be measured and signalled, allowing the pilots to detect and recover, or enabling external agencies to prompt recovery, or allowing automation to kick in and take over.

![Diagram of the Human Performance Envelope](image)

**Figure 1: Factors of the Human Performance Envelope**

Simulator experiments with pilots were conducted to measure the different factors of the Human Performance Envelope, to identify performance decrements, and subsequently to establish the boundaries of the Envelope. Furthermore, the experiments were used to gain insights into how to help pilots to recover from performance decrements and how to re-design Human Machine Interfaces (HMIs) to support the pilots for recovery in order to protect the HPE and keep it within the boundaries.

In this article we describe the development of two scenarios that were specifically designed for these experiments to elicit differential levels of human performance. One scenario consists of an approach with a complex system failure. The other scenario consists of more controlled approaches that are related to specific elements of the human performance as described by the human performance...
envelope. During these scenarios eye tracking data was acquired. Aspects of the different types of analysis are presented here together with selected results. Finally, we describe ongoing simulation work with new cockpit interface concepts which has been supported by the eye tracking studies performed previously.

2. Development of scenarios
This section describes the two different styles of scenarios developed to stimulate different aspects of human performance during the capture of eye tracking data. Both scenarios were conducted in a full-motion Airbus A320 simulator called AVES (Air Vehicle Simulator) at DLR.

The simulator has a motion system with six degrees-of-freedom using an electric motion cueing system. The cockpit is a nearly complete replica of an Airbus A320-200 with IAE V2500 engines. It contains the full set of control and display elements in the glare shield, the pedestal and the front as well as the overhead panel. Navigational aid simulation is available for the complete area of Europe. The corresponding data is based on the Lido/FMS database provided by Lufthansa Systems. The flight dynamics simulation model provides flight dynamics that are comparable to an Airbus A320-200. However the dynamics are not completely validated with flight test data. A basic simulation of the aircraft systems, such as the fuel, the hydraulic and the electrical systems, is implemented in order to generate the main indications on the ECAM and the correct feedback in the overhead panel. The complete data exchange in the AVES simulator is centralized on the Interface Computer. The Interface Computer offers the possibility to record the entire data exchange including flight dynamic parameters, aircraft system states, auto-flight modes, manual control modes, flight envelope protection activations, all cockpit inputs and all variables sent to the cockpit displays. Additionally, video and audio data is recorded from five different cameras inside the AVES simulator.

Figure 2: Airbus A320 flight simulator AVES (Air Vehicle Simulator)

Scenario 1 was designed to elicit different levels of SA, workload and stress in an experimental setup. Scenario 2 was designed to provide a mission-specific scenario involving system failures during the approach and landing phases of flight. They will be described in more detail in the following sections.
2.1. Scenario 1
In this scenario the pilots were required to fly an ILS approach with manual control into one of two airports, Frankfurt (RWY 25L) or Hannover (RWY 27R). As an example, an approach to Frankfurt is shown below.

![Example of an approach into Frankfurt of scenario 1](image)

Scenario 1 consisted of 8 different runs, in which workload, stress and situation awareness levels were varied. A run lasted between 12 and 20 minutes, depending on the variation of the factors. The following listing shows the variation of the factors in the different runs.

- **Run 1:** Only basic task: low workload, low stress, low reduced SA (or high SA)
- **Run 2:** Medium workload
- **Run 3:** High workload
- **Run 4:** Very high workload
- **Run 5:** High stress
- **Run 6:** Highly reduced SA (or low SA)
- **Run 7:** Medium workload, medium stress, medium reduced SA (or medium SA)
- **Run 8:** High workload, high stress, highly reduced SA (low SA)

The following table shows, how the factors were varied.
<table>
<thead>
<tr>
<th>Workload increase</th>
<th>Turbulences (medium or high)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approach and RWY change during initial approach (between IAF and FAF)</td>
<td></td>
</tr>
<tr>
<td>Stress increase</td>
<td>Low fuel situation</td>
</tr>
<tr>
<td>Delay vectors during initial approach (between IAF and FAF)</td>
<td></td>
</tr>
<tr>
<td>Loud noise during final approach (between FAF and landing)</td>
<td></td>
</tr>
<tr>
<td>Situation awareness reduction</td>
<td>Low visibility</td>
</tr>
<tr>
<td>Localiser interference during final approach (between FAF and landing)</td>
<td></td>
</tr>
<tr>
<td>Wind shift from head to tail during final approach (between FAF and landing)</td>
<td></td>
</tr>
</tbody>
</table>

It was anticipated that compared with the baseline (run 1) a decrement in performance is observed in all the others. In run 4 (very high workload) we expected to identify the point where the limit of performance for each pilot is reached (called performance decrement limit), in other words, the point after which performance collapses and the pilot is no longer able to safely perform the tasks. The following pictures shows a general representation of the expected performance decrement between the first four runs, where only workload was manipulated.

![Figure 4: Expected performance decrement of run 1, 2 and 3 of scenario 1](image)

![Figure 5: Expected performance decrement of run 3, 5, 6, 7 and 8 of scenario 1](image)
2.2. Scenario 2

Scenario 2 was developed to proof the HPE concept in a more uncontrolled and longer scenario. In this scenario, pilots conducted a low visibility approach into Bremen airport. The following events took place which increase workload and stress, and reduce situation awareness:

- Approaching Bremen with standard fuel for 50 min remaining flight time
- Preparation of CAT1 approach RWY 27
- Go-around during ILS approach RWY 27 due to runway blockage
- During downwind ELEC AC BUS 1 Failure
- Slight wind change requiring a landing on RWY 09
- Constantly increasing workload due to the procedure and the time needed
- Decision making and handling of complexity under low fuel conditions
- CAT2 ILS approach RWY 09

The whole run lasted between 40 and 50 minutes. The technical failure that occurred during downwind and the runway and environmental conditions required the pilots to take very complex operational decisions which required very good situation awareness about the operational consequences of the technical failure in the given situation.

The scenario progressively increased HPE load factors to a maximum value during which performance observations and data collection were made using HPE-sensitive tools. The following picture shows a general representation of the expected performance decrement during the scenario. As can be seen, the variation of workload, stress and situation awareness was much more realistic and less controlled compared to scenario 1.

![Performance decrements](image)

2.3. Participants

Ten A320 first officers (FO) from a major European airline participated in the experiment. All were male, aged between 28 and 36 years (M = 31 years, SD= 3.28). On average they had 3125 hours (SD= 1557) in the A320, with an average of 678.7 hours (SD= 43) in the last month. They had a total flight experience ranging from 2250 to 7000 hours (M = 4045 hours, SD= 1569). The cockpit crew was complemented by a Captain from the same airline, who participated throughout the whole experiment.
3. Using the eye-tracking data

3.1. Eye-tracking technology

The use of eye-tracking devices is increasingly common in aviation experiments to evaluate new displays, including how information is selected and managed in modern glass cockpits. Changes in eye movement patterns can also be associated with attentional demands and hence workload. In addition, eye-tracking also provides an indirect measure of Situation Awareness (level 1) if one assumes that the pilots perceive everything they fixate.

To record the eye movements SMI Eye Tracking Glasses were used. SMI’s eye tracking technology provides binocular tracking up to a 120 Hz sampling rate. Combined with a high definition scene camera and automatic parallax compensation this ensures accurate data over all distances. The SMI BeGaze analysis software supports aggregation of eye tracking data over multiple participants and allows qualitative visualization as well as quantitative analysis of eye tracking data. Data and visuals such as heat maps or Key Eye Tracking Metrics can be exported for further analysis.

Figure 7: SMI eye-tracking glasses and example of point of regard measurement

The following picture shows the different Areas of Interest (AOI) that were defined.

Figure 8: AOIs (Areas of Interest) of eye-tracking measurement
3.2. Scenario 1

This section details two treatments of the eye-tracking data acquired in Scenario 1. The aim of this work was two-fold. Firstly, to understand movement of point of regard in relation to scenario events and secondly, to begin to understand pilot situation awareness (SA) in response to scenario events through detailed analysis of gaze behaviour. In support of these twin aims, we provide a detailed analysis of pilot point of regard across a run.

In the following figure a summary plot of percentage dwell time per group of AOIs shows the pilot differences across the whole run.

In run 8 of scenario 1, pilots were required to fly an ILS approach with manual control. The run starts with increased turbulence which remains throughout the whole run. Three events were introduced to increase stress levels. These events were low fuel, delay vectors and the sudden introduction of a loud noise. The low fuel is an issue from the start of the run. Delay vectors occur from the beginning of the run during initial approach - between the intermediate approach fix (IAF) and the final approach fix (FAF). The loud noise occurs during final approach (between FAF and landing) and lasts for approximately one and a half minutes.

Low visibility is an issue throughout the whole run, localiser interference occurs during final approach (between FAF and landing), and there is a wind shift, from head to tail during the final approach (between FAF and landing). These runs were designed to decrease situation awareness.

The analysis of the eye tracking data and the cockpit dialogue was able to identify how SA was shared between the captain and FO and how this was managed. In most cases, the Captain initiated cross checking with the FO. At a surface level this would indicate that the Captain had better SA than the FO. However, the FO spent the majority of their time focussed on their PFD, which may indicate a level of shared SA between the captain and the FO, with the FO being supported by their PFD. It was especially apparent in some situations that the FO was effectively ‘offloading’ their SA to the Captain, with the FO cross referencing information on their instruments when required. Although the eye tracking data cannot explicitly detect performance degradation or recovery strategy, it is able to indicate how the flight crew reacted at key points, for example, when the low fuel situation was realised. This resulted in a significant change in strategy for the flight team, as they then had to manage the low fuel situation. The realisation in the limited fuel level led to the FO and captain working together to establish the future state of the aircraft: Explicit evidence of level 3 SA was captured when the FO was required to project the amount of time remaining given the amount of fuel. The proactive approach of the Captain was different to the reactive approach of the FO. This can be observed through the analysis of the dialogue, supported by the eye tracking data. They managed to recover the situation by sharing information and crosschecking. A certain amount of cognitive processing was also required, in order to calculate the remaining fuel time. Their misalignment in views (the captain wanting to call emergency but the FO not agreeing) could have been due to a number of things; the Captain’s SA was being supported by external information being fed directly to him, in addition to observing the FO’s actions and monitoring the instruments. The FO’s SA was supported by the information being fed to him by the captain, along with his own instruments. They were using different information, which as a result built different mental models. It is difficult to envisage how this could be better supported by the interface, but this mismatch indicates that it could potentially be improved; whether this is by interface improvement or SOP changes will need additional analysis.

Evidence of comprehension was reached on more occasions, notably during the loud noise, when the FO was able to establish that there was nothing wrong with the aircraft and that the current situation was normal, by monitoring his instruments. In effect, the FOs SA was being supported by the instruments. For the remainder of the run, the FO effectively offloaded his SA requirement to the
captain, who, through communications with ATC and constant monitoring of the instruments may have a more accurate, holistic view of the state of the aircraft than the FO.

In the following figure a summary plot of percentage dwell time per group of AOIs shows the pilot differences across the whole run 8.

The stacked percentage plot in *Erreur ! Source du renvoi introuvable.* indicates a potential issue with the data quality of pilot 3 which suggests that the pilot spent 75% of their time looking outside the specified AOI’s. For pilot 6 this accounted for 30% of their time, whereas the other pilots spent between 4.5 and 15% of their time not focussing on any specified AOI. Pilot 6 was a poorer performer so the higher percentage could be an indicator of reduced SA throughout the run. This explanation is supported by the amount of time pilot 6 spent looking at other areas in and around the cockpit when compared to the other pilots, specifically the captain controls and outside the cockpit. This is particularly apparent when comparing pilot 6 to pilot 8, who was assessed by an expert as a good performer. Pilot 8 may have better level 1 SA since he was attending more to the key AOIs defined.

This proof of concept has demonstrated that this type of approach to eye tracking analysis can be valuable in giving us an insight into the SA of the eye tracking wearer. This enables us to make certain inferences about the information that is important, and what is comprehended and carried forward.

### 3.3. Scenario 2

The analysis of the eye-tracking data acquired in Scenario 2 is still under process. These data will be used to better understand if the three types of interfaces used in the project modify the type of information the pilot search for in response to scenario events and consequently improve or not his/her situation awareness.

In this scenario, the simulation started with a low full situation (but not critical) that will later severely restrict the possible decisions of the crew. Also, the understanding of the fuel status by the pilot is a critical cue and the eye-tracking data are used to better understand how this sudden awareness is made. With the first type of interface (A320 conventional displays), some pilots have an early detection of the fuel status (before the go-around) while other ones discover this important indication far later. Preliminary results on the gaze movements highlight two factors that differentiate these populations: (1) pilots who understand sooner the importance of the low fuel
situation are the ones who spent more time looking at this indicator before the go-around and, (2) pilots who dedicated more time to calculations with the electronic flight bag during the approach phase (so before the go-around) are part of the ones with the late understanding of the low fuel situation.

Also, even if this paper presents only preliminary results for the analysis of scenario 2 eye-tracking data, we are confident that this methodology will contribute to the better understanding of the impact of the three types of interfaces on the pilot performances in a realistic complex situation.

4. Next steps
We are currently engaged in the measurement of human performance associated with new cockpit concepts developed by the project. The eye tracking data captured in the previous simulations will be compared to the data captured from line pilots engaged in the simulation.

These second simulator experiments are conducted in the Avionics 2020 simulator of Thales. The flight dynamics simulation model of the simulator is as well comparable to the flight dynamics of an Airbus A320-200. Furthermore, the basic simulation of the aircraft systems, such as the fuel, the hydraulic and the electrical systems, is again implemented in order to generate the same indications on the ECAM as in the first experiments.

In the second simulator experiments, the same scenario 2 will be flow again, but newly developed HMI’s will be available to the pilot. These are subject to the evaluation. The comparison of the eye-tracking data of scenario 2 in the first simulator experiments with the eye-tracking data of scenario 2 of the second simulator experiments will allow us evaluate the effectiveness of the new HMI’s and will give us insight into whether the new HMI’s are able to augment the Human Performance Envelope and to support the pilots in situations where they reach or exceed the boundaries of the HPE.

Overall, there is no doubt that capturing eye tracking data can give us insight into where pilots are looking and can allow us to make inferences as to what information they are using to make decisions or to guide response. We will continue to use this technology to enhance human performance in future generations of aircraft which will be able to exploit the different kinds of display and control technologies that are now becoming viable for the use in safety critical systems.

Figure 10: Avionics 2020 simulator
Figure 11: Example of newly developed HMIs (Human Machine Interfaces)

References


