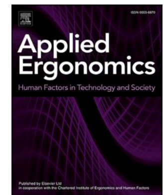
Applied Ergonomics 125 (2025) 104456

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo



Impact of automation level on airline pilots' flying performance and visual scanning strategies: A full flight simulator study

Mickaël Causse [a,*], Manuel Mercier [a], Olivier Lefrançois [a], Nadine Matton [a,b]

[a] Fédération ENAC ISAE-SUPAERO ONERA, Université de Toulouse, France
[b] CLLE, Université de Toulouse, CNRS, UT2J & ENAC, Toulouse, France

A R T I C L E I N F O

Keywords:
Automation levels
Pilot monitoring
Pilot flying
Visual scanning strategies
Aviation
Eye tracking
Complacency
Visual entropy
Mental workload

A B S T R A C T

Cockpit automation has brought significant benefits in terms of mental workload and fatigue. However, the way primary flight instruments are monitored by pilots may be negatively affected by the high confidence in systems. We examined the effects of automation level on mental workload, manual flight performance and visual strategies. Twenty professional pilots performed six landing scenarios at three levels of automation depending on flight director (flight path guidance) and autothrust (automatic management of the speed) engagements. Higher levels of automation increased flight performance and reduced mental workload, but were associated with a decrease in vigilance to primary instruments, particularly flight path indicators and engines' thrust. We also found that gaze entropy was sensitive to pilot role (pilot flying vs pilot monitoring) and automation level. These results confirmed the risks of adverse effects of automation on visual monitoring. Designing procedures for pilots to actively monitor automated cockpit systems should be encouraged.

1. Introduction

1.1. Cockpit monitoring and manual flying

Reviews of accidents and incidents in commercial aviation show that professional flight crews have been sometimes surprisingly unaware of deviations in basic flight parameters (Mumaw et al., 2019). This issue has been frequently attributed to insufficient or inappropriate visual monitoring of instruments (e.g., Lounis et al., 2021; Sarter et al., 2007; Van De Merwe et al., 2012). Since the majority of information in a cockpit is conveyed visually, and given that over 75% of pilot errors have been attributed to perceptual failures (Jones and Endsley, 1996), it is crucial for pilots to possess strong visual perception skills and the ability to accurately and quickly interpret the information presented to them.

Efficient monitoring is crucial, particularly during dynamic phases such as take-off and landing. Prompt adjustments must be made in response to any deviations, and this demands a collaborative effort of surveillance between the Pilot Flying (PF) and Pilot Monitoring (PM). The PF focuses on flying the aircraft while the PM, previously referred to as the pilot not flying (PNF), monitors the flight management, cross-checks actions of the PF, and carries out support tasks such as radio communications and running checklists. The critical importance of effectively monitoring the instruments led the FAA to replace the term PNF with PM, to emphasize the importance of this activity. At most airline companies, the role of PF or PM is interchangeable among the flightdeck pilots and is simply determined prior to initiating a particular flight segment. Despite this division of roles in the cockpit, both PF and PM are required to perform monitoring tasks, with some expected noticeable differences in their visual activities (Dehais et al., 2017; Mercier et al., 2022; Reynal et al., 2016) due to their respective duties. For example, during the approach/landing phases, while the PF generally monitors the attitude indicator or the outside world (if visibility permits) to monitor the flight path (Reynal et al., 2016), the PM distributes his gaze over other areas, such as the navigation display to monitor the aircraft's position or the speed indicator, to avoid any deviation to this vital parameter (Dehais et al., 2017).

Manual control of the aircraft has been described as a highly skilled task requiring continuous adjustments, mainly through adequate fine-motor inputs to flight controllers (Haslbeck and Hoermann, 2016). The pilot must control and monitor cross-coupled flight parameters and if needed, motor inputs are performed to correct parameters and reach the desired flight path. According to Ebbatson (2009), pilots develop a mental model that acts as a mechanism of mental projection, enabling

* Corresponding author.
E-mail address: mickael.causse@isae.fr (M. Causse).

https://doi.org/10.1016/j.apergo.2024.104456
Received 23 April 2024; Received in revised form 21 October 2024; Accepted 9 December 2024
Available online 14 December 2024
0003-6870/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

the anticipatory control of the systems. These mental models are simplifications of the actual systems (heuristics) that allow a pilot to quickly and accurately anticipate how the aircraft will behave. Consistently, monitoring has been considered as essentially sense making (Mumaw et al., 2020), a systematic observation and interpretation of the current state of the aircraft requiring integration of current inputs (mainly visual) with operational knowledge, which includes these mental models, and the generation of expected flight parameter values and aircraft behaviors. Despite this ability to anticipate and efficiently orient attention toward the relevant channel of information, performing efficient visual scanning remains complex. In accordance with the previously mentioned study (Jones and Endsley, 1996), failure to perceive critical information in the cockpit was identified by the National Transportation Safety Board (NTSB) as being involved in 84% of major accidents in the United States from 1978 to 1990 (NTSB, 1994).

1.2. Different levels of automation available for manual flying

Generally speaking, automation in modern cockpits contributed to improvements in flight safety by reducing pilot workload and fatigue (Lee and Seppelt, 2012). Yet, whereas lack of automation was problematic in the beginnings of aviation, growing role of automation now raises new challenges with experts pointing at risks associated with an over-reliance on automation (Bowden et al., 2021). The first risk associated with the use of high level of automation is the loss of situation awareness associated with pilots being « out-of-the-loop » (Endsley and Kiris, 1995; Gouraud et al., 2017) or unable to effectively monitor automated systems when required (Mumaw et al., 2001; Parasuraman et al., 1993). Second, over-relying on automation may also induce progressive loss of manual flying skills (Haslbeck and Zhang, 2017). Third, pilots may be prone to over-confidence (Antonovich, 2008) or automation complacency (Brown, 2016; Parasuraman and Manzey, 2010) that can result in an improper monitoring of flight instruments, decreasing their abilities to take-over in case of automation failure (Nikolic and Sarter, 2007) or to perceive abnormal parameters. In this sense, automation bias Wickens et al. (2015) has been defined as the tendency to follow automation information as a substitute for vigilant information seeking (Mosier and Skitka, 2018) at the expense of other cues. This can lead to errors when automation fails to detect problems or when humans inappropriately follow automation or announcements. According to Brown et al. (2016), automation complacency is a significant human factor concern requiring intensification of research to mitigate its contribution to aviation incidents and accidents.

Modern aircraft offer pilots the flexibility to choose the level of automation that suits their preferences. The autoflight system integrates the flight director (FD), a guidance system for the pilot (in manual flight) or the autopilot; the autopilot, an automatic system controlling the flight control surfaces to follow the flight director; and the autothrust, a system setting engine thrust automatically. When the autoflight system is fully engaged, the aircraft can take over the entire flight path, leaving the pilot with the task of monitoring the flight parameters. In this scenario, the pilot assumes a supervisory role, while the autoflight system handles tasks that were once exclusively performed by human pilots (Casner et al., 2014). However, the pilot can take manual control of the aircraft, and then he/she is responsible for actively manipulating the flight controls to manage the aircraft's flight path. Even in this situation of manual control, pilots have the possibility of engaging some level of flight guidance, with the FD, and automation, with the autothrust.

With FD engaged, command bars are displayed on the attitude indicator. The PF then controls the aircraft's flight path using the flight controllers and follows the indication of the FD to adjust the aircraft's flight path to match the flight path specified in the flight plan. This method of piloting is expected to reduce mental workload compared to flying without the FD, where the PF must rely on primary flight parameters to build and maintain situation awareness of the aircraft's state. Consequently, the use of the FD may change the pilot's monitoring behavior (Zaal et al., 2021), as the primary flight parameters may no longer be deemed necessary to fly the aircraft (see Fig. 1). However, it is not clear how the FD activation impacts the monitoring of "raw" flight path parameters. In addition, autothrust systems adjust the engine thrust and speed, relieving the pilot of this task, but at the risk of greatly reducing its propensity to check the aircraft speed and thrust level since it is automatically controlled by this system. Without the autothrust, the pilot controls the engines' thrust level manually, which "forces" him/her to continuously monitor the impact of his/her actions on the speed. FD and Autothrust can be activated independently or at the same time, according to the pilot's preferences.

Even if these two systems are extremely reliable, they can sometimes disengage unexpectedly or display erroneous data as a result of inadequate mode management (Silva and Hansman, 2015). For example, during landing, the aircraft can capture a false glide slope signal (an artifact of the signal that the aircraft can capture when flying well above the main glide slope), resulting in erroneous guidance from the FD, such as indicating that the aircraft must increase the pitch attitude instead of continuing the descent. If the pilot blindly follows FD indications, e.g., Delta Air Lines Flight 723 crash (NTSB, 1974), it could lead to a loss of control. Also, automation mode confusion, human-automation interaction breakdowns (Skraaning and Jamieson, 2023), or surprise and startle effects (Deniel et al., 2023) are a significant problem even when the automation itself is operating exactly as designed (Sarter and Woods, 1997). In a famous accident at San Francisco airport (Asiana Airlines Flight 214), the autothrust was disengaged because the Captain selected an inappropriate autopilot mode that, without his awareness, resulted in the autothrust no longer controlling airspeed and performance (NTSB, 2014). The crew did not notice the airspeed decrease, leading the aircraft to stall and crash on final approach at San Francisco International Airport. The primary means of airspeed awareness for flight crews are indications of the speed tape on the primary flight displays. It is very likely that they did not gaze at these indicators, probably expecting speed to be under control of the autothrust. Eventually, automation failures can also emerge from "propagating malfunctions" or "poorly integrated equipment" (Skraaning and Jamieson, 2023), as highlighted in events such as the recent B737 MAX accidents (AAIB, 2022; KNKT, 2019).

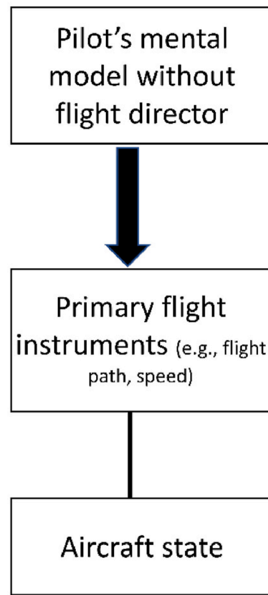
1.3. Objectives and hypotheses

We aimed to examine the influence of the level of automation on the flying performance and monitoring strategies in 20 professional airline pilots. We assessed 3 levels of automation depending on the activation of the FD and autothrust engagement. We predicted that the engagement of the FD and the autothrust 1) would result in lower mental workload and 2) better flying performance during the landing. However, with these systems engaged, 3) a shift of visual attention allocation away from the primary flight parameters (in particular speed, flight path, thrust level) towards the FDs may occur, especially for the PF, reflecting a high focusing of visual attention (eye fixations) on the FD and a decreased vigilance on the primary flight parameters. This increased time spent on FD should be more specific to PF since he/she is directly in charge of the control of the flight path of the aircraft.

2. Method

2.1. Participants

Twenty A320 qualified pilots including 10 Captains and 10 First Officers were recruited for this flight simulator experiment. All were males, with a mean age of 42 years (SD = 3.8) for Captains and of 29 years (SD = 2.7) for First Officers, and with a flight experience of respectively 11500 flying hours (SD = 1300 flying hours) and 3500 flying hours (SD = 340 flying hours). All were volunteers, unaware of the purpose of the study, and randomly assigned to another pilot to form



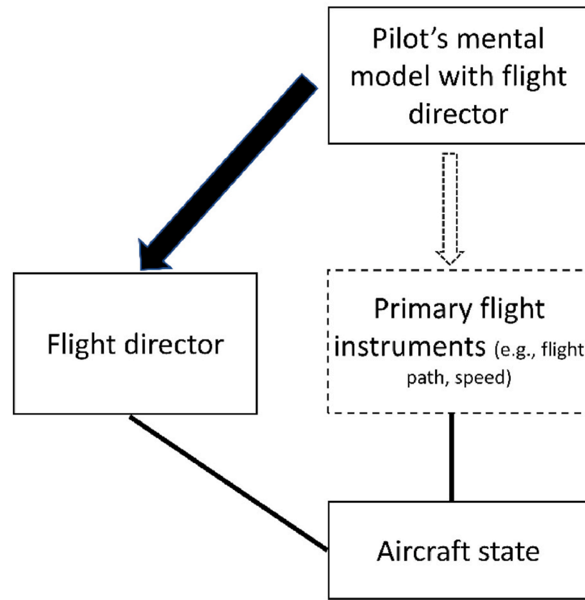


Fig. 1. Illustration of how the pilot might monitor instruments and control the aircraft without (left) and with (right) the flight director guidance. When activated, it is suspected that the pilot has a lower level of consciousness of the information displayed in the primary flight instruments since he/she might focus only on the command bars (overlaid on the attitude indicator) of the flight director. A similar logic is likely to apply to airspeed and engines thrust information when autothrust is activated.

a crew. They had all previously experienced landing at Toulouse Airport in their careers. This research complied with the Declaration of Helsinki and was approved by the CERNI (Ethics Committee of the University of Toulouse, France, IRB00011835-2020-03-03-210). Informed consent was obtained from each participant.

2.2. Flight scenarios

All 20 pilots performed six times the same flight, from take-off to landing (with an Instrument Landing System, ILS) at Toulouse airport (LFBO, runway 32R). They performed it alternatively as PF (actually flying the aircraft) and PM, with the three possible levels of automation (no use, partial use, high use), see Table 1. A total of 120 flight scenarios were examined. The eye-tracking and performance analyses were conducted on the approach and landing phase of the flight, which covered the altitude range between 2500 feet and the moment of touchdown. Weather conditions were instrument meteorological conditions (IMC), with a visibility higher than 550 m and a 15 knots crosswind. Each pilot began either as PF or PM randomly and went in a counterbalanced order through the three levels of automation in the same role. Subsequently, they switched to the other role and again went through the three levels of automation, also in a counterbalanced fashion (see Table 2).

Table 1
The six flight scenarios performed by each pilot.

As pilot flying (PF)	As pilot monitoring (PM)
No use of automation	No use of automation
Partial use of automation	Partial use of automation
High use of automation	High use of automation

Table 2
The three levels of automation used during the experiment, depending on flight director and autothrust activations.

	Flight director (FD)	Autothrust (A/T)
No use of automation	OFF	OFF
Partial use of automation	ON	OFF
High use of automation	ON	ON

The levels of automation consisted of modulating the activation of two systems: Flight Directors (FD) guidance and Autothrust (A/T) management. Both are automatic systems that are designed to assist the pilot in respectively providing flight path guidance and aircraft speed management (by automatically adjusting engines' thrust level). For each flight scenario, pilots were instructed to manually perform the approach (i.e., with autopilot disengaged) but with these different levels of automation.

The professional pilots tested in this experiment all had current Airbus A320 qualifications and flew the actual aircraft and the simulator on a very regular basis. Therefore, no familiarization with the simulator was required. At the beginning of the experiment, the pilots were informed that they would perform several flights at Toulouse airport and that the level of automation for the approach and landing would be specified just before the descent point at 2500 feet. They were also informed that the experimenter would act as an "air traffic controller", providing weather information and flight instructions.

2.3. Flight simulator

Experiments were conducted in a Thomson A320 level "D" qualified full-motion flight simulator, see Fig. 2. Level "D" is the highest standard, allowing pilot qualification training in the absence of any flight time requirement in the actual aircraft. It provides a collimated cross-cockpit display for the out-the-window world view; and reproduces all systems, communications devices, and glass cockpit displays, in a similar fashion to actual aircraft.

Four flight performance parameters were recorded during the approach/landing. Specifically, we measured lateral and vertical deviations (lower is better), speed (target = 138 knots), and height above runway threshold when reaching the runway threshold (target = 50 ft).

2.4. Eye-tracking and area of interests

Eye movements were recorded using two head mounted Pertech eye-trackers (accuracy .25°, sampling rate 50Hz). An eight-point calibration procedure was performed before the first flight scenario. Before each scenario, the quality of the calibration was checked by asking the participants to look at a series of instruments. If a measurement problem was detected at this point, a new calibration was performed. A total of five areas-of-interest (AOIs) were defined, see Fig. 3. Contrary to previous studies that consider the Primary Flight Display (PFD) as a single AOI (e.g., Zaal et al. (2021), we dissociated three important AOIs in the PFD: FD command bars (the attitude indicator with the FD displayed inside when activated), Speed instrument, and Flight path information that aggregates heading instrument and lateral deviation scale (i.e., localizer) as well as altitude instrument and vertical deviation scale (i.e., glide slope). In addition, we also defined the AOIs Engines thrust and Window. Percentages of fixation times were calculated to characterize pilots' attention distribution over the cockpit (Haslbeck and Zhang, 2017) with the aim to understand how FD and autothrust use impact the monitoring of primary flight parameters. As a complementary analysis we measured the complexity of pilots' eye movements with stationary gaze entropy (Krejtz et al., 2015) using the following equation: $H(x) = -\sum_{i=1}^{n}(p_i)log_2(p_i)$. According to Krejtz et al. (2015), a higher value of stationary entropy indicates that the subject distributes their visual attention more equally among AOIs. A lower value is observed when fixations tend to be focused on a smaller amount of AOIs. Tole et al. (1983) demonstrated that increases in pilot workload were associated with reductions in the complexity of their gaze patterns across a simulated flight instrument console. More recently, entropy has been shown to distinguish between baseline driving and conditions of high visual-spatial task load (Schieber and Gilland, 2008). It is thus interesting to examine whether entropy is sensitive to the level of automation. FD engagement may reduce visual complexity due to a high focusing of attention on the latter. Due to technical issues, eye-tracking data were missing for 10% of the no automation scenarios (4 out of 40) and 15% of the partial automation scenarios (6 out of 40). These missing data were replaced using mean imputation. No eye-tracking data were missing for the high automation scenario.

Fig. 2. Picture of the Thomson A320 flight simulator used for the experiments.

2.5. Procedure

The crew (Captain and the First Officer) was welcomed by the experimenters and completed the consent form and a demographic questionnaire asking for their age, flight experience, and current aircraft model. They were then accompanied into the flight simulator. They were equipped with the eye-tracker and performed the calibration. They then performed six flights, all experiencing the three levels of automation as PF and PM. After each of the flights, participants performed a subjective measurement of their perceived workload using only the raw mental demand subscale of the NASA-TLX (Hart, 2006) to reduce the experimental time. Each flight scenario lasted approximately 12 min. In this paper, we focused our analysis on the approach/landing phase, whose duration was approximatively 2 min. The whole experiment duration was approximately 2 h for each crew.

3. Results

3.1. Statistical analysis

One-way ANOVAs (3 levels of automation) were performed to examine flight performance and mental workload of PFs. Flight performance analysis excluded pilots who did not manage to stabilize the approach since they performed a go-around (no performance data for them). For eye-tracking data, a two-way repeated measures ANOVA (3 levels of automation x 2 pilots' roles) were performed on percentages of time spent on each AOI (Fig. 3). An additional two-way repeated measure ANOVA (3 levels of automation x 2 pilots' roles) was performed to analyze entropy levels variations. Post-hoc pairwise comparisons were performed using Tukey's HSD (honestly significant difference) test with p value adjustment for multiple comparisons. Normality and homoscedasticity were checked using Kolmogorov–Smirnov and Mauchly tests respectively.

3.2. Mental demand and flight performance according to automation levels

As expected, an increase in the level of automation was associated with a decrease in subjective pilot (PF) mental demand, $F(2,38) = 277.46$, $p < .001$, $\eta_p^2 = .49$. More precisely, mental demand decreased across each increase in level of automation (Fig. 4), with a higher subjective demand in the no automation ($M = 71.2$, $SD = 9.7$) vs partial automation ($M = 28.8$, $SD = 8.7$) ($t(29) = 14.59$, $p < .001$), and a higher subjective demand in the partial vs high automation ($M = 12.5$, $SD = 7.2$) ($t(29) = 7.28$, $p < .001$).

In general, increased automation level was associated with an increase in flight performances. We analyzed the number of unstabilized approaches. A go-around was performed when the pilot failed to stabilize the aircraft, which is the standard decision. As expected, the highest number of unstabilized approaches was observed in the no automation level. In this condition, 5 pilots out of 20 (25%) had to go-around, see Table 3.

In addition, increased automation level was associated with lower

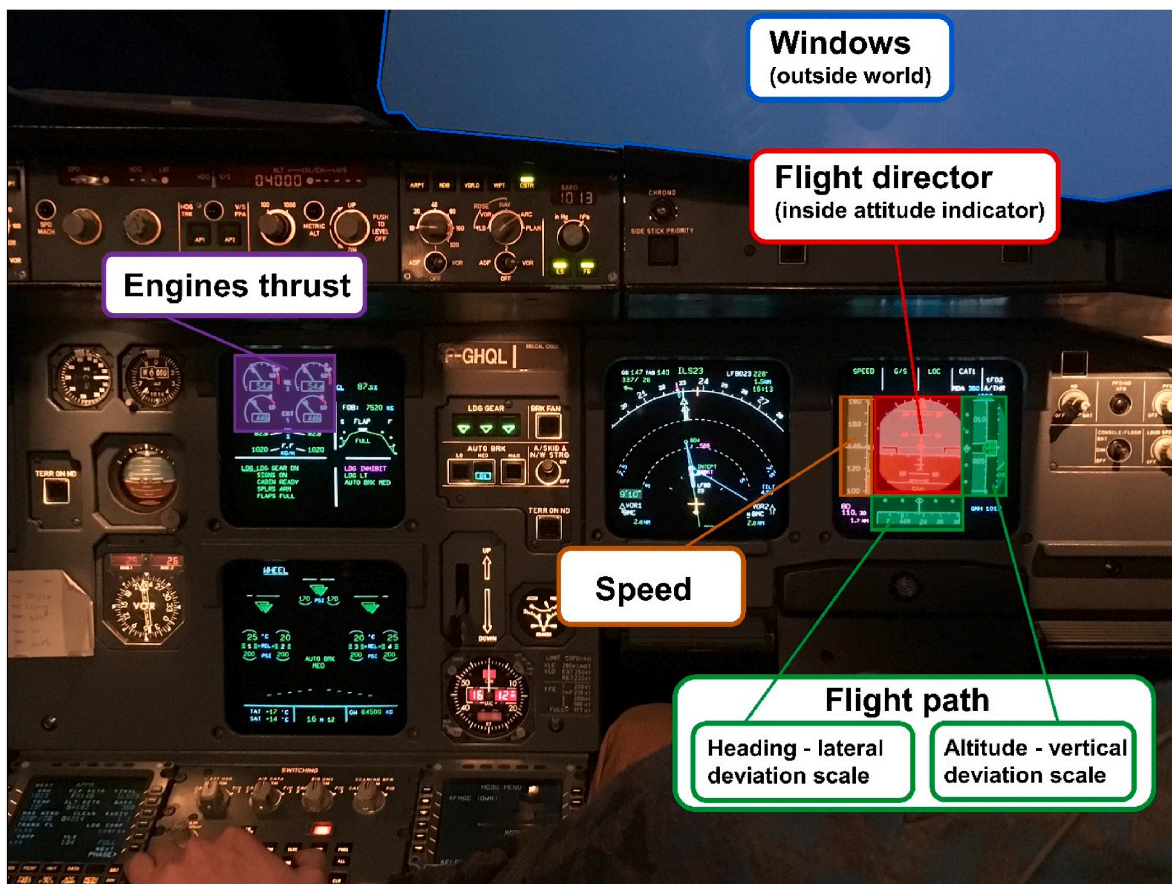


Fig. 3. Picture of the actual Thomson A320 flight simulator cockpit showing the five AOIs used in the experiment. The same AOIs were used on the left seat. FD is inside the attitude indicator, visual fixations to this instrument are assumed to be on the FD when engaged.

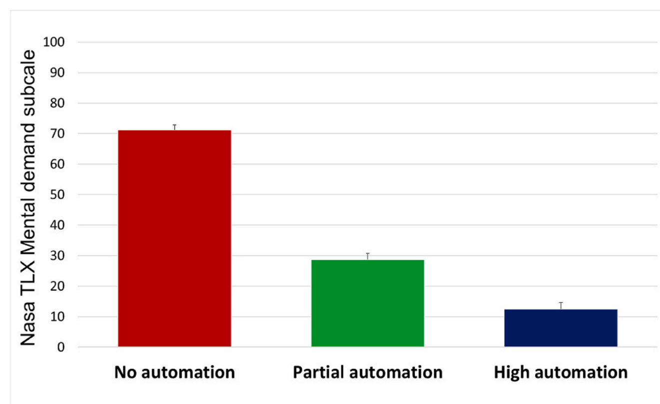


Fig. 4. Subjective mental demand score according to the three levels of automation. No = no use of automation; Partial = Flight director only; High = Flight director + Autothrust. Error bars are SEM.

Table 3
Percentage of unstabilized approaches leading to a go-around as a function of the automation levels.

Automation level	Percentage of unstabilized approaches
No use of automation	25%
Partial use of automation	5%
High use of automation	0%

lateral ($F(2,28) = 12.90$, $p < .001$, $\eta_p^2 = .48$) and vertical path deviations ($F(2,28) = 7.92$, $p = .002$, $\eta_p^2 = .36$). More precisely, path deviations were significantly smaller in high vs no automation for both lateral ($t(14) = 3.76$, $p = .006$) and vertical ($t(14) = 3.38$, $p = .01$) deviations, see Fig. 5.

We analyzed two additional flight parameters during the landing phase, namely speed and height above runway threshold when landing. Due to loss of data, in addition to the unavailable data from the pilots that performed a go-around, data from 2 pilots were missing in partial automation level and from 1 pilot in no automation level. Automation level had no significant effect on speed at runway, $F(2,24) = 4.96$, $p = .19$, $\eta_p^2 = .13$, see Fig. 6 (and Table S1 for descriptive statistics for performance parameters).

On the contrary, height at runway was impacted by automation level, $F(2,24) = 1.80$, $p = .016$, $\eta_p^2 = .29$. However, post-hoc comparisons only revealed a marginally lower height deviation in partial vs no automation level ($t(12) = 2.51$, $p = .066$), see Fig. 7.

3.3. Percentage of fixation times on AOIs according to automation levels

Main effects of pilot role were significant for all AOIs. PF were more focused on Window ($F(1,19) = 100.53$, $p < .001$, $\eta_p^2 = .84$) and Flight Director ($F(1,19) = 9.16$, $p = .007$, $\eta_p^2 = .33$), while PM were more focused on Speed ($F(1,19) = 18.61$, $p < .001$, $\eta_p^2 = .49$), Engines thrust ($F(1,19) = 78.73$, $p < .001$, $\eta_p^2 = .81$), and Flight path ($F(1,19) = 33.81$, $p < .001$, $\eta_p^2 = .64$), see Fig. 8.

The main effect of the level of automation was significant for Flight Director $F(2,38) = 6.99$, $p = .003$, $\eta_p^2 = .27$, Engines thrust, $F(2,38) =$

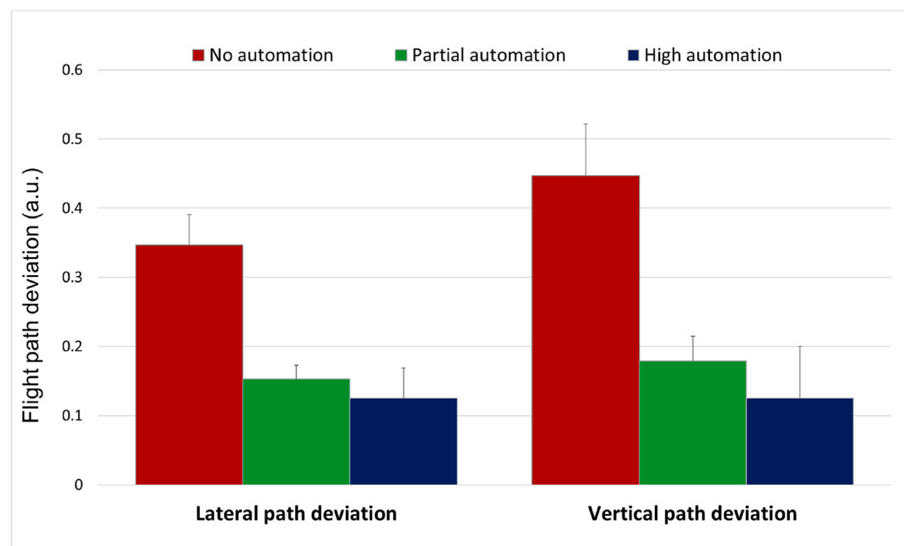


Fig. 5. Lateral and vertical path deviations according to the three levels of automation. Pilots who did not manage to stabilize the approach were not included. No = no use of automation; Partial = Flight director only; High = Flight director + Autothrust. Error bars are SEM.

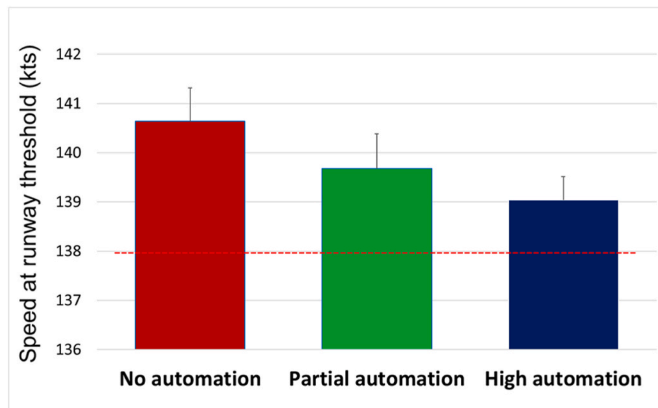


Fig. 6. Speed at runway threshold according to the three levels of automation. The red dashed lines indicate target height. Pilots who did not manage to stabilize the approach were not included (no data available). No = no use of automation; Partial = Flight director only; High = Flight director + Autothrust. Error bars are SEM. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

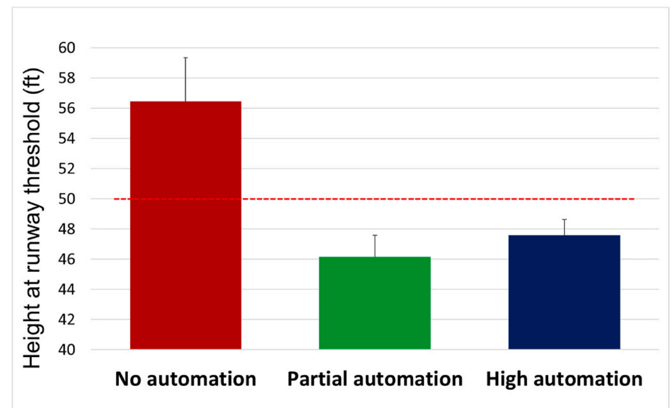


Fig. 7. Height at runway threshold according to the three levels of automation. The red dashed lines indicate target height. Pilots that did not manage to stabilize the approach were not included (no data available). No = no use of automation; Partial = Flight director only; High = Flight director + Autothrust. Error bars are SEM. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

3.81, $p = .03$, $\eta_p^2 = .17$ and Flight path, $F(2,38) = 14.09$, $p < .001$, $\eta_p^2 = .43$. Post-hoc tests showed less Flight director gaze time for no vs high automation ($t(19) = 3.60$, $p = .005$). Engines gaze time was higher in partial vs no use of automation ($t(19) = 3.46$, $p = .007$). Time spent gazing Flight path decreased with increased level of automation. More precisely, it was lower in partial/high vs no automation ($t(19) = 4.37$, $p < .001$, $t(19) = 4.00$, $p = .002$, respectively).

Importantly, we also found significant pilot role × level of automation interactions for Flight Director ($F(2,38) = 5.13$, $p = .01$, $\eta_p^2 = .21$), and Engines thrust ($F(2,38) = 4.65$, $p = .02$, $\eta_p^2 = .20$) Post-hoc tests revealed that PF gazed more at the Flight Director with high automation than without automation ($t(19) = 4.28$, $p = .005$) whereas gaze time on Flight Director did not change significantly for PM across the different levels of automation, see Fig. 8.

Time spent by PF gazing Engines thrust was lower in high vs partial automation ($t(19) = 3.57$, $p = .02$), which likely reflected the effect of the autothrust engagement. Interestingly, time spent by PM gazing at the Engines thrust did not decrease between partial and high automation ($t(19) = .12$, $p = 1$), and was higher in partial vs no automation ($t(19) = 3.29$, $p = .04$). All other main effects and interactions were non-significant (see Table S2 for descriptive statistics for percentage of fixation time on each AOI for both pilot roles).

3.4. *Visual entropy according to automation levels*

Main effect of pilot role was significant and showed that PM entropy was higher than PF ($F(1,19) = 38.38$, $p < .001$, $\eta_p^2 = .67$), see Fig. 9. The pilot role × level of automation interaction was also significant ($F(2,38) = 4.20$, $p = .023$, $\eta_p^2 = .18$). Post-hoc tests revealed that this higher entropy for PM compared to PF was significant in the no automation condition ($t(19) = 3.27$, $p = .040$) and in the high automation condition ($t(19) = 7.71$, $p < .001$). While PF entropy diminished in the high automation condition, it increased for the PM. The effect was close to significance in the partial automation condition ($t(19) = 3.05$, $p = .062$).

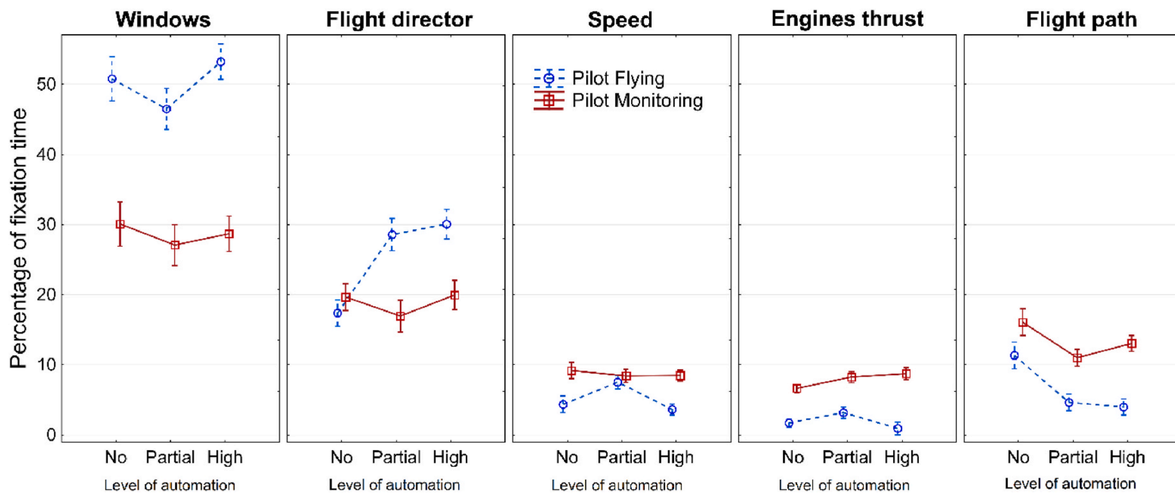


Fig. 8. Percentage of fixation time on the different AOIs for pilot flying and pilot monitoring across the three levels of automation. No = no use of automation, Partial = Flight director only, High = Flight director + Autothrust. Note that in the no automation condition, the Flight director bar commands are not displayed in the attitude indicator. Error bars are SEM.

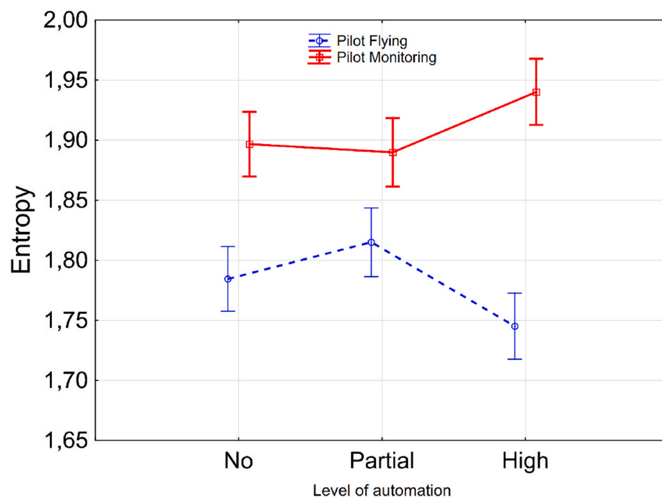


Fig. 9. Entropy for pilot flying and pilot monitoring across the three levels of automation. No = no use of automation; Partial = Flight director only; High = Flight director + Autothrust. Error bars are SEM.

4. Discussion

It was hypothesized that the engagement of the flight director and the autothrust would result in reduced mental workload and improved flight performance. The results confirmed the first hypothesis: higher levels of automation drastically reduced mental workload, confirming the benefits of automation when it relieves humans of some tasks (Young and Stanton, 2002). Logically, higher levels of automation were associated with an increase in flight accuracy and a lower rate of unstabilized approach (no go-around observed in the full automation condition), confirming the second hypothesis. Intuitive flight path guidance provided by the FD and the automatic management of the speed with the auto-thrust have made it possible to increase the landing performance: the flight path was significantly more accurate for the full (FD and A/T) and partial (FD only) automation levels compared to the scenarios without automation. We can speculate that the decrease of flying performance when flying without the help of the FD and autothrust may also somehow reflect an erosion of manual flying skills (Haslbeck and Hoermann, 2016). To confirm this, a future study could attempt to correlate recent individual manual flight time with flight performance without the assistance of the flight director. Interestingly, Kaber & Endsley (2004) indicated that adaptive automation could be helpful in regulating workload while maintaining operator's involvement.

Despite the anticipated benefits of automation, a third hypothesis predicted that higher levels of automation (with the flight director (FD) and auto-throttle (A/T) engaged) would lead to a shift in attention away from critical primary flight parameters (such as flight path and engine thrust) toward the FD. This shift would likely result in a lack of situation awareness, as primary flight parameters are essential inputs for level 1 of pilot situation awareness (Endsley, 1995), which involves the perception of elements in the current situation. We indeed observed that pilots reallocated their attention to the FD (inside attitude indicator) in the highest levels of automation, at the expense of a more direct monitoring of the primary flight parameters. This effect of automation appeared to be specific to PF, which demonstrated an increased time spent on the FD and a decreased time spent checking the engines' thrust level in full automation. Although this shift of attention is expected as the pilot delegates flight path calculation and speed management to respectively FD and autothrust, it reflects a change of pilot's representations from primary flight parameters in manual mode to flight guidance (i.e., FD) when flying with a high level of automation. Such a change could make pilots more vulnerable to losses of situation awareness, which is particularly sensitive when facing unreliable or inconsistent flight guidance (Endsley and Kiris, 1995) or automatic speed management.

The high fixation time on the FD shown by PFs at higher levels of automation in our study resemble attention tunneling behavior, known to be induced by the head-up-display (Wickens and Alexander, 2009). Whether that behavior is training-induced, task-induced, linked to automation complacency, a natural method to mitigate workload, or several of these possibilities at the same time is open to question and would justify further eye-tracking based research work. Regarding complacency, we must however be cautious as the definition of optimal sampling of very highly reliable systems is still an open question (Moray, 2003). Introducing an automation failure could have helped to estimate the impact of the reduced monitoring on flight path and engine thrust on the capacity to take over for automation. Pilots could be trained to practice more on manual flight during recurrent training and made aware of the fact that primary flight parameters should be glanced frequently to detect any discordance with automation behavior. Regarding engines thrust monitoring with autothrust engaged, there is accumulating evidence of the difficulty in maintaining attention focused on the activities of an automated system that nearly never fails (Casner

et al., 2014). A possibility is to raise awareness of the risks of automation, with training modules on the dangers of complacency in automation, discussing past incidents where lack of vigilance led to accidents. However, this solution needs to be evaluated further since active monitoring of automated systems must not be performed at the expense of other important instruments. Active vigilance (Sumwalt, 2003), can also be promoted to emphasize the importance of scanning the primary flight instruments at regular intervals, even when in automatic mode. This will help pilots maintain sustained attention to primary flight parameters. Pilots could also be trained more on emergency situations where they would have to disable automation and take manual control, enabling to stay engaged and practice the skills necessary to manage critical situations without automation.

We observed that PMs' visual strategies were generally more spatially distributed across the different instruments than PFs'. They looked more at the primary flight parameters, in particular flight path, speed and engines thrust. Interestingly, PMs' visual behavior remained relatively unchanged across the different levels of automation. In addition, entropy results showed that while PF's entropy diminished in the full automation condition, it increased for the PM. A higher value of stationary entropy indicates that the subject distributes more their visual attention among AOIs (Krejtz et al., 2015). These findings are important and show that PMs continue to build their situation awareness in high levels of automation, in particular with primary flight instruments as a primary source of information. They have thus a critical role in detecting anomalies in the automation. The PM appears to be less sensitive to the attention tunneling effect of the FD, because she/he is not actually engaged in the control of the aircraft flight path, thus she/he may be the first to detect an erroneous autoflight mode or an inappropriate guidance situation and should be able to react more quickly than the PF. Previous research has shown that gaze entropy decreases (and thus gaze dispersion) during periods of high mental workload in car driving (e.g., Goodridge et al., 2024) and aircraft piloting (Diaz-Piedra et al., 2019), in a fashion that reflects attention tunneling. On the contrary, we observed a reduction in entropy in PF when mental workload was low and automation was high. In our view, this supports the notion that our eye-tracking measures captured a complacency effect (rather than an effect of mental workload) as PF exhibited increased focus on the FD bar at the expense of other critical flight parameters, such as engine thrust.

The PM's lower sensitivity to attention tunneling further highlights the challenges associated with transitioning to single-pilot operations (SPO) and extended minimum crew operations (eMCO). The latter proposes the introduction of single-pilot flight deck operations for significant periods of time, especially during cruise periods. In these periods, the "additional" PM would no longer be there to act as a safety net in the event of failure of the automatic systems. Worst of all, it appears that the absence of the PM would lead to a change in scanning behavior: indeed, a recent study showed that the PF spent significantly more time scanning secondary instruments at the expense of primary instruments when flying alone (Faulhaber et al., 2022).

4.1. Limitations

All pilots participating in this study had prior experience with Toulouse airport, which is an airport without any particular challenge. While this setting allowed for controlled data collection, it would be valuable to extend this study to more complex scenarios involving airports with more demanding approaches or unexpected events. Such conditions would allow to test manual piloting skills and cockpit monitoring abilities in a more challenging context. One limitation of the study is the moderate number of pilots recruited. Increasing the sample could lead to more robust findings. However, recruiting professional pilots qualified for a specific aircraft and having access to a level D full flight simulator during numerous hours remains a logistical challenge. Lastly, while eye-tracking technology was used to gather data, it does present some biases. Notably, eye-tracker essentially provides the current fixation point, meaning that when pilots focus on the flight director bars in the center of the attitude indicator, they may still capture some information in peripheral vision on some primary flight parameters, in particular flight path and speed. However, such perception is likely to be imprecise, as it would only concern the detection of movements or changes on the indicators (Ramón Alamán et al., 2020). Additionally, engine thrust indications are located far from the flight director bars, meaning that the lack of fixation on this indicator, as observed in higher levels of automation in our study, could not be compensated by peripheral vision.

5. Conclusion

Our results show that the mental resources saved by the use of automation are in fact not directly reinvested in a more rigorous monitoring of the primary flight parameters, in particular when acting as pilot flying. This supports the idea of a risk of attention tunneling from the pilot flying to the flight director, suggesting a certain degree of automation complacency. The reduced monitoring of the automated tasks could be seen as a consequence of the pilots' rational choice to increase their attention towards the larger context of the cockpit at the expense of tasks that automation can reliably handle in the vast majority of cases. However, the entropy result contradicts this possibility in the case of the pilot flying, who showed a narrower visual attention at the higher level of automation. Overall, it confirms the sensitive situation of flying with a high degree of automation in case of failure, malfunction or inappropriate use of automation. Indeed, our results suggest that when flying with higher degrees of automation, pilot flying's mental representation is less likely to be built upon primary flight instruments. Without proper awareness of critical primary flight parameters, it is difficult for pilots to take over for automation if required. Pilots should be made more aware that those "raw" primary flight parameters need to be frequently sampled when high levels of automation are in use. Importantly, our results confirmed that the pilot monitoring is a key element of flight safety as he/she is much less affected by the level of automation and maintains a higher level of attention on primary flight instruments. The possible incoming transition to eMCO flight will necessarily raise issues of primary flight parameters monitoring. It is crucial to conduct experiments to check that eMCO flight will allow a comparable level of vigilance on primary flight instruments. Replicating this study with a single pilot could help address this issue. Finally, beyond the training solutions mentioned above, eye-tracking technology may also provide a powerful way to help pilots improve their gaze strategies, for example, by identifying ineffective visual behaviors (Lefrançois et al., 2021) and providing real-time feedback (Dubois et al., 2015; Lounis et al., 2020) to correct suboptimal gaze monitoring habits. By understanding these behaviors, instructors could develop specific exercises to help pilots adopt more effective gaze strategies.

CRediT authorship contribution statement

Mickaël Causse: Writing – original draft, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Manuel Mercier:** Writing – review & editing, Software, Formal analysis, Data curation. **Olivier Lefrançois:** Writing – review & editing, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Nadine Matton:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apergo.2024.104456.

References

Antonovich, B., 2008. The flight crew and automation. Journal of Aviation/Aerospace Education & Research 17 (3), 3. https://doi.org/10.58940/2329-258X.1453.
Bowden, V.K., Griffiths, N., Strickland, L., Loft, S., 2021. Detecting a single automation failure : the impact of expected (but not experienced) automation reliability. Hum. Factors, 00187208211037188.
Brown, J.P., 2016. The effect of automation on human factors in aviation. J. Instrument. Automat. Sys. 3 (2), 31–46.
Casner, S.M., Geven, R.W., Recker, M.P., Schooler, J.W., 2014. The retention of manual flying skills in the automated cockpit. Hum. Factors 56 (8), 1506–1516.
Dehais, F., Behrend, J., Peysakhovich, V., Causse, M., Wickens, C.D., 2017. Pilot flying and pilot monitoring's aircraft state awareness during go-around execution in aviation : a behavioral and eye tracking study. Int. J. Aerospace Psychol. 27 (1–2), 15–28. https://doi.org/10.1080/10508414.2017.1366269.
Deniel, J., Dupuy, M., Duchevet, A., Matton, N., Imbert, J.-P., Causse, M., 2023. An In-Depth Examination of Mental Incapacitation and Startle Reflex : A Flight Simulator Study, pp. 46–59.
Diaz-Piedra, C., Rieiro, H., Cherino, A., Fuentes, L.J., Catena, A., Di Stasi, L.L., 2019. The effects of flight complexity on gaze entropy : an experimental study with fighter pilots. Appl. Ergon. 77, 92–99.
Dubois, E., Blättler, C., Camachon, C., Hurter, C., 2015. Eye movements data processing for ab initio military pilot training. In: Neves-Silva, R., Jain, L., Howlett, R. (Eds.), Intelligent Decision Technologies. Springer International Publishing, pp. 125–135.
Ebbatson, M., 2009. The Loss of Manual Flying Skills in Pilots of Highly Automated Airliners.
Endsley, M.R., 1995. Toward a theory of situation awareness in dynamic systems. Hum. Factors 37 (1), 32–64.
Endsley, M.R., Kiris, E.O., 1995. The out-of-the-loop performance problem and level of control in automation. Hum. Factors 37 (2), 381–394.
Faulhaber, A.K., Friedrich, M., Kapol, T., 2022. Absence of pilot monitoring affects scanning behavior of pilot flying : implications for the design of single-pilot cockpits. Hum. Factors 64 (2), 278–290. https://doi.org/10.1177/0018720820939691.
Goodridge, C.M., Goncalves, R.C., Arabian, A., Horrobin, A., Solernou, A., Lee, Y.T., Lee, Y.M., Madigan, R., Merat, N., 2024. Gaze entropy metrics for mental workload estimation are heterogenous during hands-off level 2 automation. Accid. Anal. Prev. 202, 107560.
Gouraud, J., Delorme, A., Berberian, B., 2017. Autopilot, mind wandering, and the out of the loop performance problem. Front. Neurosci. 11, 541.
Hart, S.G., 2006. NASA-task load index (NASA-TLX); 20 years later, 50 (9), 904–908.
Haslbeck, A., Hoermann, H.-J., 2016. Flying the needles : flight deck automation erodes fine-motor flying skills among airline pilots. Hum. Factors 58 (4), 533–545.
Haslbeck, A., Zhang, B., 2017. I spy with my little eye : analysis of airline pilots' gaze patterns in a manual instrument flight scenario. Appl. Ergon. 63, 62–71. https://doi.org/10.1016/j.apergo.2017.03.015.
Jones, D.G., Endsley, M.R., 1996. Sources of situation awareness errors in aviation. Aviat Space Environ. Med.
Kaber, D.B., Endsley, M.R., 2004. The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. Theor. Issues Ergon. Sci. 5 (2), 113–153.
Krejtz, K., Duchowski, A., Szmidt, T., Krejtz, I., González Perilli, F., Pires, A., Vilaro, A., Villalobos, N., 2015. Gaze transition entropy. Trans. Appl. Percept. 13 (1), 1–20.
Lee, J., Seppelt, B., 2012. Human factors and ergonomics in automation design. In: Human Factors and Ergonomics, fourth ed. John Wiley & Sons, pp. 1615–1642.
Lefrançois, O., Matton, N., Causse, M., 2021. Improving airline pilots' visual scanning and manual flight performance through training on skilled eye gaze strategies. Saf. Now. 7 (4), 70. https://doi.org/10.3390/safety7040070.
Lounis, C., Peysakhovich, V., Causse, M., 2020. Flight Eye Tracking Assistant (Feta) : Proof of Concept, pp. 739–751.
Lounis, C., Peysakhovich, V., Causse, M., 2021. Visual scanning strategies in the cockpit are modulated by pilots' expertise : a flight simulator study. PLoS One 16 (2), e0247061.
Mercier, M., Lefrançois, O., Matton, N., Causse, M., 2022. Effects of the Automation Level on Gaze Behavior : A Full Flight Simulator Campaign with Professional Airline Pilots.
Moray, N., 2003. Monitoring, complacency, scepticism and eutactic behaviour. Selected papers from the second cyberspace conference on ergonomi cs, CybErg 1999 31 (3), 175–178. https://doi.org/10.1016/S0169-8141(02)00194-4.
Mosier, K.L., Skitka, L.J., 2018. Human decision makers and automated decision aids : made for each other?. In: Automation and Human Performance. CRC Press, pp. 201–220.
Mumaw, R.J., Billman, D., Feary, M.S., 2019. Factors that Influenced Airplane State Awareness Accidents and Incidents : CAST SE-210 Output 2 : Report 2 of 6.
Mumaw, R.J., Billman, D., Feary, M.S., 2020. Analysis of Pilot Monitoring Skills and a Review of Training Effectiveness.
Mumaw, R.J., Sarter, N., Wickens, C.D., 2001. Analysis of Pilots' Monitoring and Performance on an Automated Flight Deck, vol. 8.
National Transportation Safety Board, 1974. Aircraft accident report, Delta air lines, inc. Douglas DC-9-31, N975NE, Boston, Massachusetts, July 31, 1973" (PDF). National Transportation Safety Board (NTSB-AAR-74-3). https://libraryonline.erau.edu/online-full-text/ntsb/aircraft-accident-reports/AAR74-03.pdf.
National Transportation Safety Board, 1994. A review of flightcrew-involved, major accidents of US air carriers, 1978 through 1990. Safety Study Report No. PB94-97001.
National Transportation Safety Board, 2014. Descent below Visual Glidepath and Impact with Seawall, Asiana Airlines Flight 214, Boeing 777-200ER, HL7742, San Francisco, California (Aircraft Accident Report NTSB/AAR-14/01). (Accessed 6 July 2013).
Nikolic, M.I., Sarter, N.B., 2007. Flight deck disturbance management : a simulator study of diagnosis and recovery from breakdowns in pilot-automation coordination. Hum. Factors 49 (4), 553–563.
Parasuraman, R., Manzey, D.H., 2010. Complacency and bias in human use of automation : an attentional integration. Hum. Factors 52 (3), 381–410.
Parasuraman, R., Molloy, R., Singh, I.L., 1993. Performance consequences of automation-induced « complacency. Int. J. Aviat. Psychol. 3 (1), 1–23. https://doi.org/10.1207/s15327108ijap0301_1.
Ramón Alamán, J., Causse, M., Peysakhovich, V., 2020. Attentional Span of Aircraft Pilots : Did You Look at the Speed?, pp. 108–113.
Reynal, M., Colineaux, Y., Vernay, A., Dehais, F., 2016. Pilot Flying vs. Pilot Monitoring during the Approach Phase : An Eye-tracking Study, pp. 1–7.
Sarter, N.B., Mumaw, R.J., Wickens, C.D., 2007. Pilots' monitoring strategies and performance on automated flight decks : an empirical study combining behavioral and eye-tracking data. Hum. Factors 49 (3), 347–357.
Sarter, N.B., Woods, D.D., 1997. Team play with a powerful and independent agent : operational experiences and automation surprises on the Airbus A-320. Hum. Factors 39 (4), 553–569.
Schieber, F., Gilland, J., 2008. Visual entropy metric reveals differences in drivers' eye gaze complexity across variations in age and subsidiary task load, 52 (23), 1883–1887.
Silva, S.S., Hansman, R.J., 2015. Divergence between flight crew mental model and aircraft system state in auto-throttle mode confusion accident and incident cases. J. Cogn. Eng. Decis. Mak. 9 (4), 312–328.
Skraaning, G., Jamieson, G.A., 2023. The failure to grasp automation failure. J. Cogn. Eng. Decis. Mak., 15553434231189375 https://doi.org/10.1177/15553434231189375.
Sumwalt, R., 2003. Cockpit monitoring : using procedures to enhance crew vigilance. Professional Pilot 2–6.
Tole, J., Stephens, A., Vivaudou, M., Ephrath, A., Young, L., 1983. Visual Scanning Behavior and Pilot Workload. NASA.
Van De Merwe, K., Van Dijk, H., Zon, R., 2012. Eye movements as an indicator of situation awareness in a flight simulator experiment. Int. J. Aviat. Psychol. 22 (1), 78–95.
Wickens, C.D., Alexander, A.L., 2009. Attentional tunneling and task management in synthetic vision displays. Int. J. Aviat. Psychol. 19 (2), 182–199.
Wickens, C.D., Clegg, B.A., Vieane, A.Z., Sebok, A.L., 2015. Complacency and automation bias in the use of imperfect automation. Hum. Factors 57 (5), 728–739.
Young, M.S., Stanton, N.A., 2002. Attention and automation : new perspectives on mental underload and performance. Theor. Issues Ergon. Sci. 3 (2), 178–194. https://doi.org/10.1080/14639220210123789.
Zaal, P., Lombaerts, T., Mumaw, R., Billman, D., Torron, I., Jamal, S., Shyr, M.C., Feary, M., 2021. Eye-Tracking Analysis from a Flight-Director-Use and Pilot-Monitoring Study, p. 2995.