Situation Awareness Adaptive Alerting in an Aircraft Cockpit: A Simulator Study

Simon Schwerd and Axel Schulte

Institute for Flight Systems, University of the Bundeswehr, Munich

Author Note

We have no conflicts of interest to disclose. Correspondence concerning this article should be addressed to Simon Schwerd, LRT 13, Universität der Bundeswehr München, Werner-Heisenberg-Weg 39, 85577 Neubiberg. Email: simon.schwerd@unibw.de

Manuscript Type: Research article

Word count: 5105

Abstract

Objective The goal of this study was to develop an automated cockpit support system that is adaptive to the flight crew's situation awareness (SA) estimated by online gaze analysis.

Background Flight crew errors are often attributed to low SA. Online measurement of SA could be used to automatically guide the user's attention for the sake of fewer errors and better performance.

Method An eye-tracking based measure for SA was developed and used for the adaptive generation of alerts in a flight simulator. In an experiment, ten certified pilots conducted two trials with no and adaptive alerting. The experimental task involved tracking of flight parameters which were partially disturbed or changed at random times.

Results Our online estimation of SA showed a strong correlation with observed pilot performance. With adaptive alerts, the average performance increased in those experimental tasks, where a situational change could not be predicted by participants. Also, adaptive alerts improved change detection and reduced the number of outliers, where a change was not noticed for an exceptionally long time. However, subjective rating was poor due to low transparency and false positives.

Conclusion SA-adaptive support can improve change detection performance in typical tasks on the flight deck. For a greater acceptance, pilots should be trained to understand the adaption policy.

Application The presented adaptive system could be integrated in aircraft cockpits to reduce pilot errors.

Keywords: Monitoring, Eye tracking, Pilot behavior, Warning systems, Adaptive automation

Précis: In this study, we present the design and experimental evaluation of an alerting system adaptive to an eye-tracking based measure of situation awareness. The results of a flight simulator experiment show that adaptive alerts improve performance and change detection, but poor subjective rating indicates that transparency should be considered for further development.

Situation Awareness Adaptive Alerting in an Aircraft Cockpit: A Simulator Study

Human error remains a major contributor to aircraft accidents. Among different aspects, loss of Situation Awareness (SA) was identified as the most common factor causing hazard events in aviation (Kharoufah et al., 2018). To maintain SA, efficient monitoring of flight path, aircraft configuration, on-board systems and automation modes is a crucial task of the flight deck crew. This will become more important with increasing automation on flight decks.

During monitoring, the pilot continuously compares their expectation about the aircraft's state to the indications by cockpit instruments. When the pilot fails to efficiently monitor the relevant aspects of their environment, their mental picture of the situation can divert from the actual state resulting in poor decision making (Silva & Hansman, 2015).

The application of SA-adaptive systems in the cockpit could improve safety associated with monitoring activity. By observing the pilot's gaze movements, these systems can identify situations of poor pilot monitoring and notify the pilot about relevant system states. The goal of this contribution is to explore the design of such an adaptive system that acts on an estimation of pilot situation awareness. We developed a system that alerts the pilot when they fail to monitor a system state properly. Further, we conducted an experiment with pilots to evaluate this approach.

Adaptive systems to support monitoring

Adaptive systems change their behavior according to the needs of the user (Feigh et al., 2012). Although this concept is not new (Kaber & Endsley, 2004; Rouse, 1988), very few systems adapting themselves to the flight crew are deployed in aircraft cockpits. Generally, they are considered as opaque and complex by aviation professionals (Dorneich et al., 2016). Nevertheless, the idea of computerized systems adapting their behavior is attractive, because it resembles the behavior of human assistants (Johnson et al., 2014), comparable to a pilot-monitoring supporting a pilot-flying in an aircraft cockpit. Here, the pilot-monitoring has extensive knowledge of the situation, the task, and the activity of the pilot-flying while they support based on their own initiative. Adaptive systems have two similar traits:

1. An adaptive system **initiates** a change of its own behavior based on a pre-defined trigger mechanism.
2. For triggering the adaption, adaptive systems use context knowledge beyond what is necessary for the supported task.

Therefore, a system adapting to monitoring performed by the pilot requires the measurement of the monitoring activity during operation. In numerous studies, monitoring was analyzed by using gaze measurement in the cockpit (Peißl et al., 2018; Ziv, 2016). In these contributions, different gaze metrics were proposed capturing the temporal, spatial and sequential dynamic of gaze behavior. With these measures, researchers were able to find differences between monitoring behavior of expert and novice pilots (Lounis et al., 2021) or identify poor monitoring as a reason for automation surprises (Sarter et al., 2007).

There are drawbacks to the use of these metrics as triggers for adaptive monitoring support: First, these gaze metrics rely on averaging statistics, for example the fixation frequency on an area of interest (AoI). This is suitable to identify correlations between gaze and monitoring performance but fails to identify poor monitoring in highly dynamic situations. For example, a pilot might be efficient at monitoring for most of the flight but fails to notice one relevant element within a few critical seconds. Second, optimal values of the gaze metrics depend on the cockpit layout and task in which monitoring is evaluated. While this is no issue for rather standardized environments such as airliner cockpits, it might complicate the applicability for other operational environments. Finally, most studies focus on gaze behavior alone and ignore the situation and the state of the aircraft and the flight mission. Again, this becomes a problem when monitoring behavior is optimal for standard situations but the situation itself is exceptional. We therefore argue that we need a measure that integrates both gaze and situational aspects to identify situations, when monitoring support is required. The model of SA might provide such a framework since it is context-dependent a includes the perception and comprehension of task-relevant states (Endsley & Jones, 2012). There are different approaches to measure SA: freeze-probes (e.g. SAGAT), online probes (e.g. SPAM), post-trial questionnaires (e.g. SPAM) or the use of physiological and behavioral measurements (e.g. eye-tracking, EEG) (Salmon et al., 2009; Zhang et al., 2020). Online monitoring of SA is only possible with physiological and behavioral measurements and since monitoring is considered as an important activity to gather SA, many studies use eye-tracking to infer SA from gaze behavior. This also indicates, that monitoring and SA – at least on the perceptional level – are closely related (Billman et al., 2020).

So far, only a few studies investigated systems adaptive to measures of monitoring or SA. (Bosse et al., 2009) proposed a model which combines gaze measurement and display features to estimate the user’s attention. Based on the discrepancy between actual and desired attention, the saliency of task-relevant objects on a moving map display was adjusted in a naval warfare task. The authors found significant improvement in task performance comparing no support against adaptive support. In a similar study, (Fortmann & Mengeringhausen, 2014) developed an eye-tracking based adaptive interface supporting a UAV monitoring task. Based on a measure of SA, the saliency of relevant objects on the tactical map was adapted. (Schwarz & Fuchs, 2017) quantified the user state in different dimensions (workload and attention) and used a combination of physiological, behavioral and performance measures to identify critical states, which triggered five different adaption strategies for their task environment. Compared to adaption based on performance and workload measurement, SA-based adaption can prevent performance decrements by guiding the operator’s attention to the most relevant parts of the task environment. Therefore, undesired system states can be acknowledged before they become a time-critical problem. In the context of aircraft cockpits, (Lounis et al., 2020) developed an support system based on eye-tracking. By comparing pilot dwell times on different cockpit indicators to a database of standard gaze behavior, the adaptive system identified poor monitoring and triggered a vocal alert. Although the system was able to redirect attention towards critical flight instruments, no performance improvements were found, and subjective rating was not optimal due to false alarms. The authors concluded that the integration of flight parameters can further improve the usability of the system.

We conclude that recent studies showed the potential of SA-adaptive systems for supervisory control tasks and the feasibility of gaze-based assistance in the cockpit. However, there is no generic measure of monitoring performance or SA usable for other task environments (e.g., driving or process control). Further, studies in aircraft cockpits did not incorporate the aircrafts state into their measure. Therefore, the objective of this research is to contribute in two ways:

- The definition of a SA measure, that integrates eye-tracking and system state, and is transferable to different task contexts.
- The design and evaluation of an SA adaptive system to improve monitoring in an aircraft cockpit.

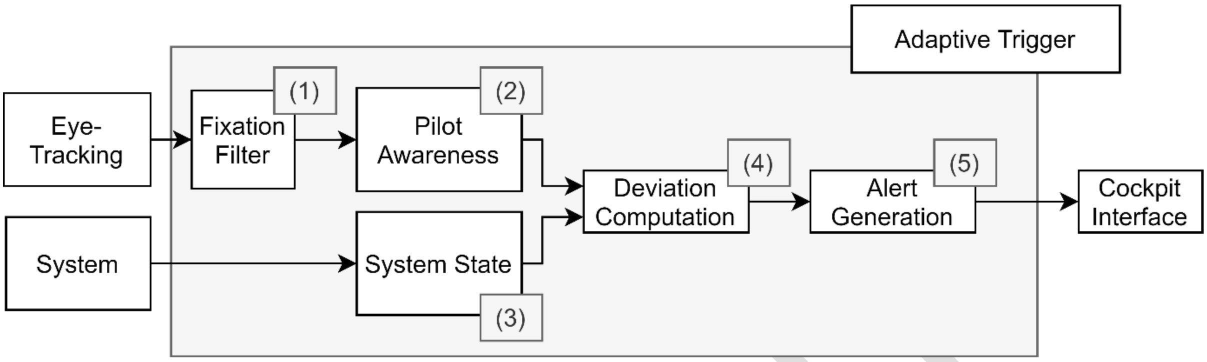
Deviation as a measure of SA

We envision a system, that tracks the pilot's awareness about the state of the aircraft and, given that information, guides the pilot's attention to relevant parameters when their SA is low. This measure should distinguish between different state information (e.g., altitude, position) and the associated awareness in this state. For this, we quantify SA in terms of an information-specific *deviation* in each aircraft state. Each deviation describes the difference between the current state value and the assumed pilot's awareness. The measurement of deviation is based on eye-tracking and was developed and evaluated in two prior studies (Schwerd & Schulte, 2020, 2021). Based on this measure we alert the pilot and guide the attention towards an unnoticed change of system state.

Figure 1 displays the full adaptive mechanism. There are two inputs to the adaptive trigger: the system state and eye-tracking measurement. The output is an information-specific alert for the pilot. In the following, we discuss the processing steps in greater detail.

Figure 1

Data flow of deviation measurement for specific information



Eye-Tracking Analysis

The eye-tracking system provides a data stream of gaze positions in the cockpit. Each sample can be classified as part of a fixation or a saccade. Based on the eye-mind hypothesis (Just & Carpenter, 1980), we assume, that the pilot only perceives displayed information when he fixates upon the relevant cockpit indicator. Given a fixation on a screen, we infer the corresponding AoI and extract the value which is displayed at the time of measurement (see Figure 1 (1)). Note, that this requires a detailed knowledge about the information displayed in every AoI. This is trivial for displays of numerical data, but more difficult for graphical displays encoding more than one information, for example a Primary Flight Display (PFD) representing pitch and roll at the same time. This concept is based on the eye-tracking analysis for pilot activity determination presented in (Honecker & Schulte, 2017).

Pilot and System Value

For further processing, the pilot's awareness is defined as the set of all system states associated with the value that was displayed at the most recent fixation of the pilot (see Figure 1 (2)). This set is continuously updated by the stream of gaze samples. Parallel to this, a second set contains all system states (Figure 1,3) associated with the current value. Similarly, this set is continuously updated with the current values of all states.

Deviation

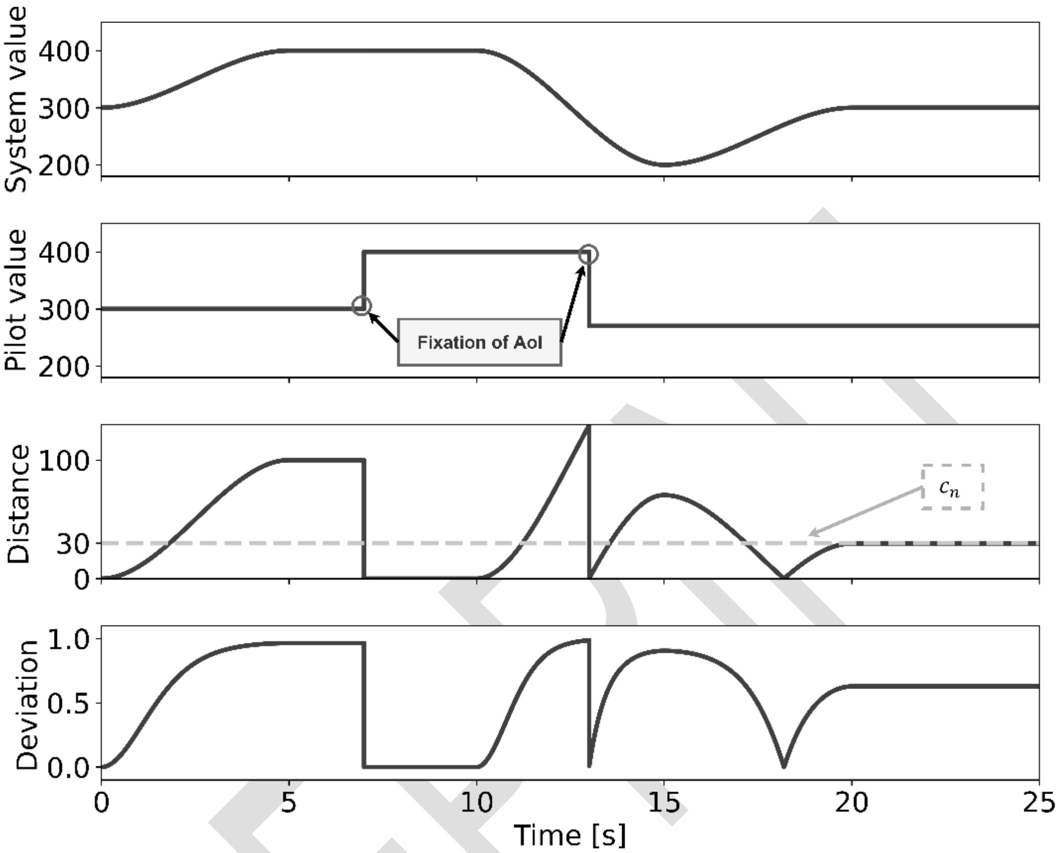
In the deviation computation (see Figure 1 (4))., we compute a normalized difference between system state and pilot awareness. This difference is computed for every relevant system state. Naturally, these states are represented in different forms of information, such as modes, figures, text, times, or position. For each of these forms, we defined a function $d(c_p, c_s)$ to compute the difference between pilot and system values c_p and c_s, respectively. Table 3 in the appendix shows example functions for different categories. To normalize the distance for further processing, we apply the following formula to compute the deviation:

$$f_{dev} = 1 - e^{-\frac{|d(c_p,c_s)|}{c_n}}, \tag{1}$$

where c_n denotes a normalizing constant, that quantifies the value, which should be characterized as a substantial distance for a given task. We choose an exponential function because a distance value $|d(c_p, c_s)|$ in proximity of c_n should imply a great deviation while the change in deviation should be damped when $|d(c_p, c_s)|$ grows significantly larger than c_n . The idea is, that when the distance between c_p and c_s is already greater than what is considered large in the task context (defined by c_n), deviation should not grow proportionally.

Figure 2

Example for course of values for system, pilot, distance function and deviation function.



Note: There are two fixations on the relevant AoI at $t = 7s$ *and* $t = 13s$*.*

Figure 2 shows an exemplary course of deviation for aircraft speed. Here, we chose a normalizing constant of $c_n = 30kts$. Initially, the speed is at 300 kts (indicated by system value) and the pilot is aware of this value (indicated by *pilot value*), therefore distance and deviation are zero. When the speed changes, the distance between the pilot and system value increases. Accordingly, the deviation increases within its normalized limits. The pilot's first fixation on the speed indicator updates his knowledge about a speed of 400 kts. When the pilot updates his knowledge, distance and deviation are set to zero. After 3 seconds, the speed changes again without the pilot's awareness. At 13s, the second fixation updates the knowledge about a current speed of 300 kts. The ongoing change in speed increases distance and deviation. After that, the speed returns to 300 kts resulting in a deviation of zero.

Alert

The deviation for all states is passed to the alert generation (see Figure 1,5). When deviation exceeds a limit for a predefined duration, an alert is created. The time delay prevents premature alerts where the pilot is not aware of some information state right after the state has changed. Note, that there is no function that checks if there is a pilot error present before an alert is triggered. We decided not to filter the alerts to observe the unaltered behavior of such an adaptive system.

Algorithm

Table **4** in the appendix denotes the pseudo code for the trigger algorithm of a single information value. The algorithm is executed in a loop, that continuously evaluates the deviation and processes the delay logic. A normalizing constant c_n , a time until notification and a trigger threshold must be selected for every task relevant information.

Experiment

Setup

Trials were conducted in a cockpit simulator resembling a generic fast-jet cockpit (see Figure 3). The participating pilots controlled the simulated aircraft with stick and throttle. For the display of task relevant information, the cockpit contains three touchscreens and a Head-Up-Display (HUD). The setup is shown in Figure 4. The cockpit design is custom but provides all common functions of a next generation military combat aircraft. For eye-tracking, we use the commercially available four-camera system by SmartEye© (Smart Eye Pro 0.3 MP) which was integrated into our cockpit setup and connected to our simulation software. The gaze is measured with a sampling frequency of 60Hz and can be tracked across all displays including the HUD.

Figure 3

Cockpit simulation environment with integrated eye-tracking system.

Participants

We conducted trials with 10 participants with different levels of flight experience. At minimum, the participant had a sailplane pilot license (SPL). Half of the participants were professional pilots. All participants were male with a mean age of 35.5 years ($\pm$13.2y). Their amount of flight hours varies between 500 and 29000 hours, which reflects the different levels of experience of the pilots (professional and private pilots).

Procedure

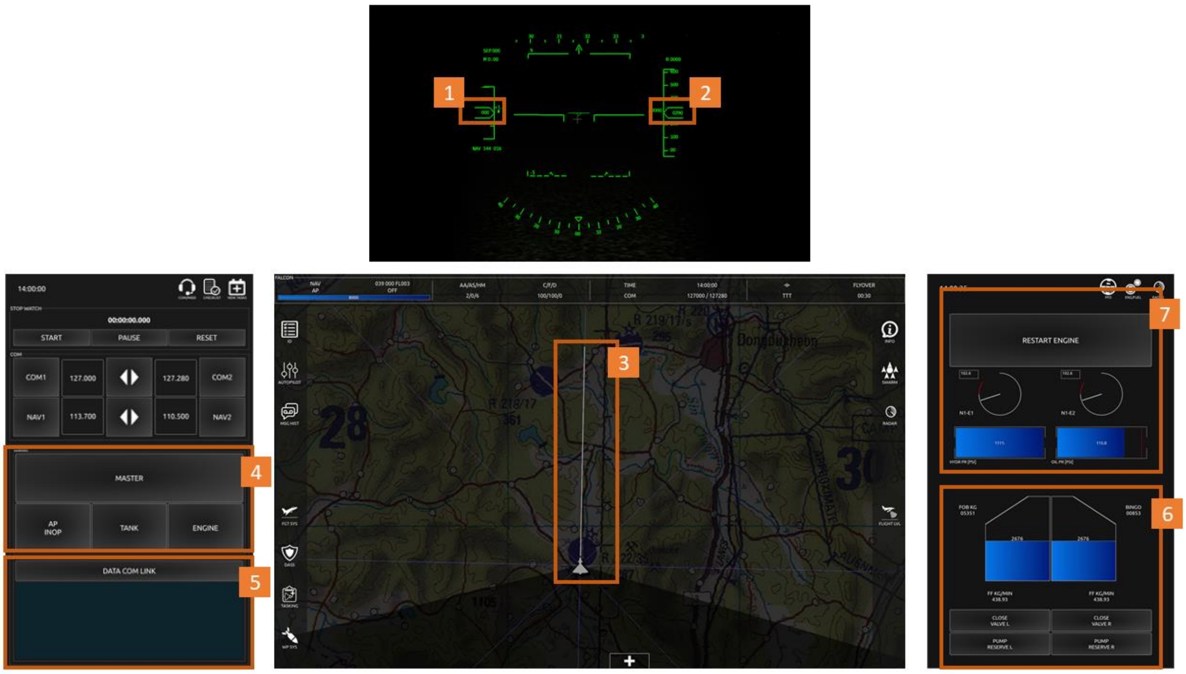
All participants received instructions about the simulator before the experiment. Each participant received a simulator training in the experimental task for approximately one hour. During this training, they encountered all aspects of the experimental task including emergency procedures as reaction to system warnings (e.g. fuel system failure). Then, the eye-tracking was calibrated. After that, the participants conducted two trials comparable in difficulty, where in the first trial no alerting was active and in the second SA-adaptive alerting was active. Participants were not briefed about the adaptive mechanism, but only told that there is an alerting system active in the second trial.

Tasks

Our experimental task included monitoring of continuous parameters and discrete states of the aircraft - the simulation interfaces including task relevant AoIs are shown in Figure 4. The participants had three continuous tracking tasks: First, they had to track a route displayed in their tactical map, where only the route leg to the next waypoint was visible. After reaching a waypoint, a new route leg appeared with an unknown heading (continuous state, Figure 4; Area 3). Second, the participant had to track a specified altitude and velocity, both displayed in the HUD (continuous states, Figure 4; Areas 1,2). The target altitude and velocity were specified and changed by a message at previously unknown times. A new message was indicated by a green light in the left screen (discrete state, Figure 4; Area 5). At unknown times, the aircraft altitude or velocity was disturbed by a simulated gust that changed the aircraft position or speed to a value beyond the target range. In both trials, disturbance of altitude and velocity was triggered 50 times. We simulated bad visibility to minimize the influence of spatial visual perception in the outside view. Therefore, change of velocity and altitude could only be recognized by looking at the displayed values. During each trial, two aircraft warnings were triggered, which were visible as red indications on the left screen (Figure 4; Areas 4). Participants had to react with trained procedures to these warnings. The procedures comprised the activation of system modes (Figure 4; Areas 6,7) or a change of thrust.

Figure 4

Simulator display setup with AoI (top: HUD, bottom: side and center display).



Note: The orange areas display the following information: (1) Altitude indicator, (2) Air speed indicator, (3) Position of aircraft and route to next waypoint, (4) warning system (red on new warning), (5) message system (green on new message), (6) fuel system with button controls for procedures and (7) engine system with controls for procedure

Configuration of adaptive mechanism

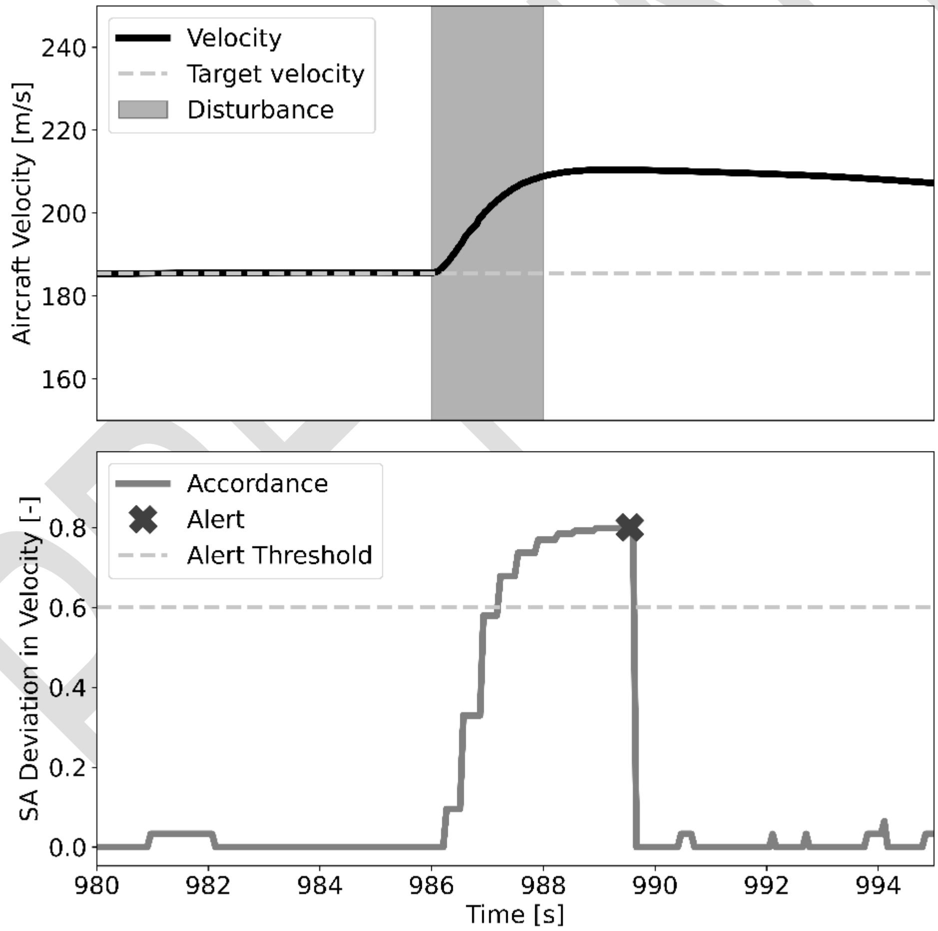
The adaptive mechanism was implemented for altitude, velocity, route, warning indication and message indication. We selected a deviation threshold of .6, because it indicates a deviation of approximately the value of the normalizing constant. For the experiment, these constants were selected to be close to the threshold values for the tracking tasks. When deviation exceeded the threshold, a time delay was defined for each parameter. As mentioned, this delay prevents instant alert on value changes and gives the pilot some time to react. Prior to the experiment, we tried to define delay values that balanced between premature and late alerts in the different tasks.

Table **5** in the appendix contains the normalizing constants c_n and delay times for each information relevant to the experimental task.

For the alert, we used a computer-generated voice to notify the pilot with keywords (for example “ALTITUDE”, “SPEED”). To illustrate the behavior of the adaptive system, Figure 5 shows an example of an alert. Here, a simulated disturbance changes the velocity of the aircraft (grey area). This goes unnoticed by the participant which results in an increasing deviation in velocity. Due to the deviation exceeding a threshold of .6 for more than 3 seconds, a vocal alert is triggered. The pilot reacts by looking at the velocity indicator which causes the deviation to drop to zero.

Figure 5

Example of awareness adaptive alert (here: "VELOCITY").



Note: This data is taken from the experimental results. The course of deviation is not smooth, because the sample frequency of system value updates is 2.5Hz.

Data Analysis

Task performance for the tracking tasks was quantified as follows: The mean error rate was computed by RMS of the difference between the target value and the actual value. When the actual value was below a threshold, error was set to 0. Tracking thresholds were 200 ft, 15 knots and 2 NM for tracking of altitude, velocity, and route, respectively. During the trial, we logged data of SA deviation, task performance, alerts, and gaze. We analyzed this data with Python Pandas (Reback et al., 2021) and conducted statistical tests with Python SciPy (Virtanen et al., 2020). To test for normality of the data, we used the Shapiro-Wilkson test. If the test indicated normality, we used Pearson's correlation coefficient and conducted a dependent t-test for the test of significance. Otherwise, we used Spearman's rank correlation coefficient and Wilcoxon signed-rank tests. We used Levene test for equality of variances. The effect was considered weak when $r<=.1$, medium when $.1<r<=.7$ and strong when $r>.7$. Significance level was set to $p=.05$.

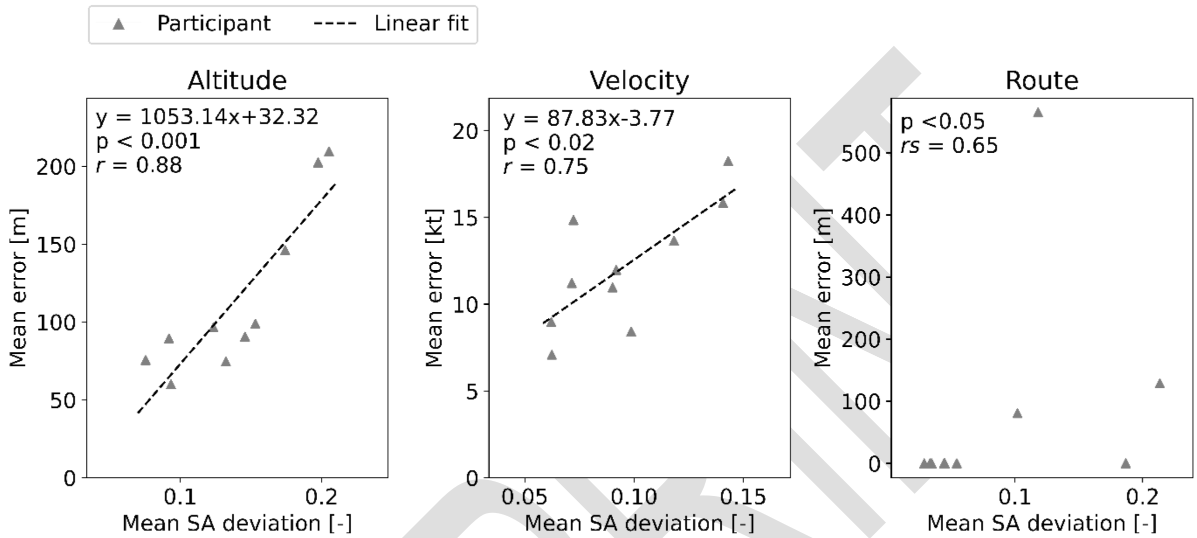
Results

Validation of operator measurement

To evaluate our approach to SA measurement, we compared SA deviation to pilot performance in the three tracking tasks. For this, we only analyzed data from the control condition without alerting to eliminate influence of the adaptive system on the measurement. Figure 6 displays mean error and mean SA deviation in all tracking tasks for the control condition. For altitude tracking, the correlation coefficient is $r = .88$ (significant, $p<.001$). For velocity, the data shows a strong positive correlation of $r = .75$ (significant, $p<.01$). These results reproduce findings from a previous study where the operator measurement has been experimentally validated in a similar task (Schwerd & Schulte, 2020). Further, Figure 6 shows mean deviation and error in route tracking. Of all participants, 60% did not commit any error in this task. However, the mediate positive correlation coefficient of $r_s = .65$ is significant. The strong positive correlation clearly shows that the measure of deviation is meaningful in tracking speed and altitude. A similar but weaker trend can be observed in route tracking.

Figure 6

Correlation between deviation and error in tracking for condition "no alerting" (n=10).

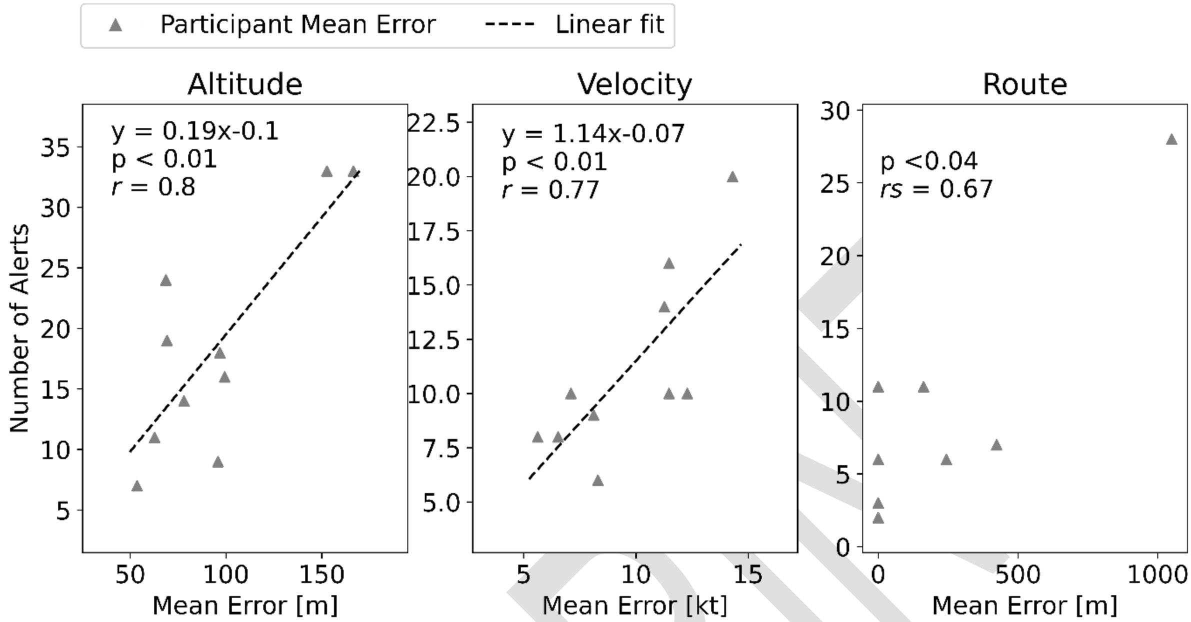


Relationship between error and trigger

In the adaptive condition, we compared the number of triggered alerts against the error rate. Results are shown in Figure 7. There is a strong correlation between the number of alerts and the error rate in altitude (r=.8, p<.01) and velocity (r=.77, p<.01). Similar to the control condition, some participants did not commit any error in route tracking, but there is a mediate correlation of rs=.67 (p<.04) between number of route alerts and mean error in route tracking. Also, the route data shows that participants triggered alerts even when there was no error. These false alarms were triggered because the adaptive mechanism solely used the SA deviation to trigger alerts but did not evaluate if there actually was an error.

Figure 7

Correlation between number of alerts and mean error (n=10).

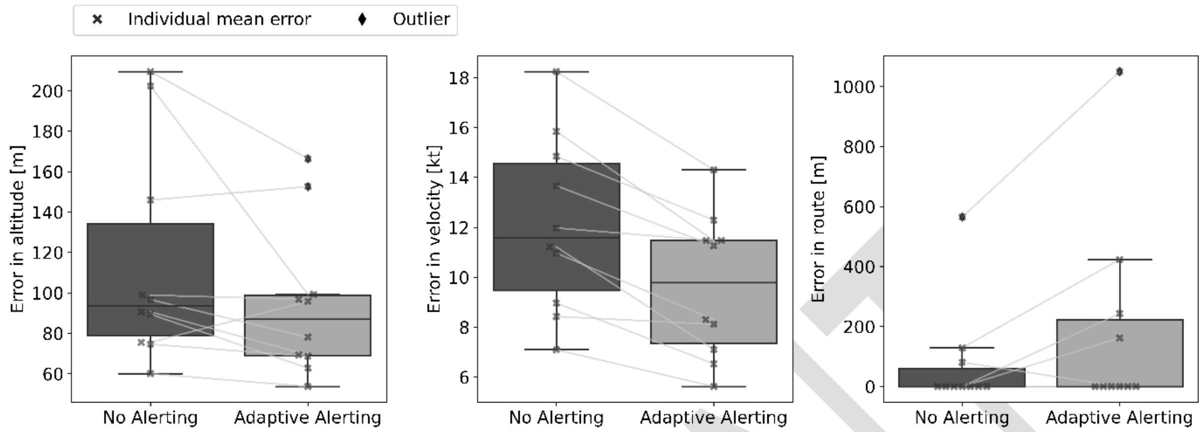


Comparison of performance

Figure 8 compares control and adaptive condition in the average tracking error. Errors in tracking both velocity and altitude decreased in the adaptive condition. For velocity error, data was distributed normally and the decrease in error was indicated as significant by the paired t-test ($t(9) = 5.53$, $p = .0004$). The change in altitude error was not significant ($ps < .084$). Opposed to velocity and altitude, error in route increased with adaptive alerting. The increase in route tracking error was not significant ($ps < .11$). Average performance as displayed in Figure 8 reflects two aspects of task performance: First, the participants ability to carry out the tracking tasks without any disturbance of the aircraft's altitude and velocity. Second, the ability to quickly notice a disturbance and return to the target value. The difference between both conditions can be analyzed in greater detail by examining the time until a disturbance is noticed after a gust.

Figure 8

Average error rate in the tracking tasks (n=10).



Note: "X"-markers indicate means of individual participant connected over two conditions by a grey line.

Detection performance

Data in Table 1 shows mean and standard deviation values for the detection times. Detection time is defined as the time between the change (disturbance or new waypoint) and the time until the participant first looked at the relevant AoI (see Figure 4). For altitude and velocity, the mean time until change detection slightly decreased in the adaptive condition. For route, there is a non-significant increase in mean time of detection.

Table 1

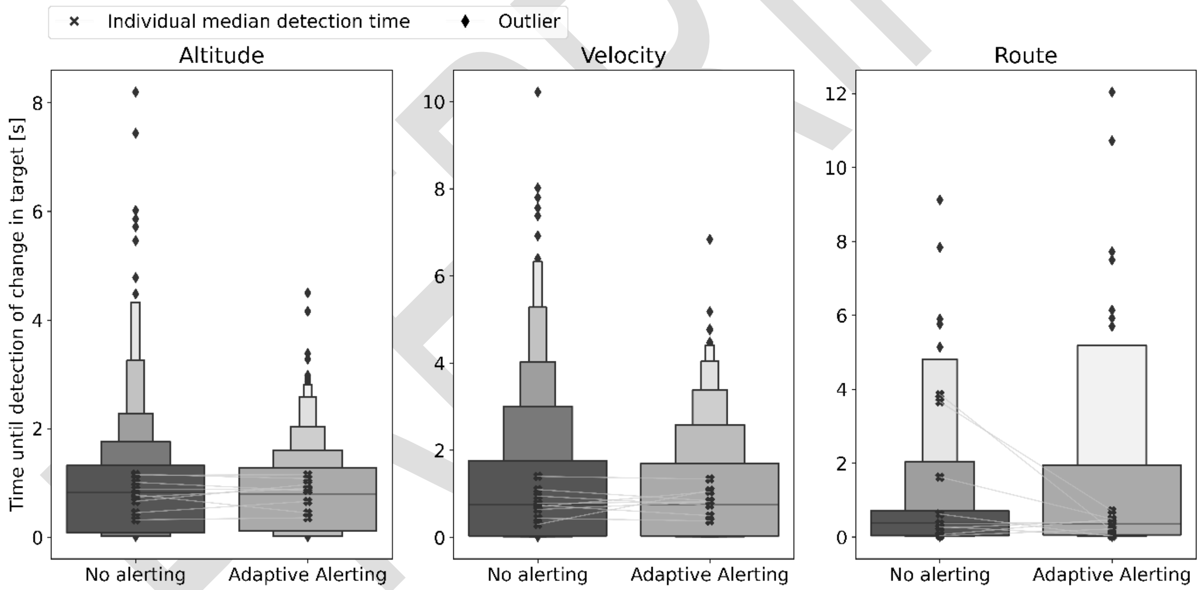
Detection times for tracking tasks.

Condition	Altitude Detection [s]		Velocity Detection [s]		Route Detection [s]	
	M	SD	M	SD	M	SD
No Alerting	.93	1.04	1.25	1.56	1.01	1.84
Adaptive Alerting	.83	.75	1.09	1.19	1.70	2.74

Figure 9 shows a letter-value box plot of the time until the participant detected a value change. For altitude and velocity, adaptive alerting reduced the number of outliers. The decrease of standard deviation between conditions is significant (altitude: $p<.01$, velocity: $p<.03$). This shows that the adaptive system could successfully identify situations, where the participants failed to notice a value change in a system state. The number of large outliers were reduced by the alerts triggered after deviation measurement exceeding the threshold for more than the specified time delay. There is an opposite effect for route tracking. Here, the median time for detection decreased but number of outliers as well as outlier values increased. However, difference in variance was not significant ($p<.09$). Note, that the median times are below the time delay for the trigger, therefore cannot be improved by the system.

Figure 9

Time until detection of disturbance in a target parameter.



*Note: "X"-markers indicate **medians** of individual participant connected over two conditions* by a grey line. *We used median instead mean because the median is less sensitive to outliers.*

Table 2 shows the mean detection times for new messages and warnings. There is a decrease in mean detection time in the experimental condition. However, this effect can mainly be attributed to the outliers visible in

Figure **10**. The plot also reveals that the main effect of the adaptive system is the reduction of outliers. The number of samples is small because warning only appeared two times during one trial and new messages arrived 5-6 times.

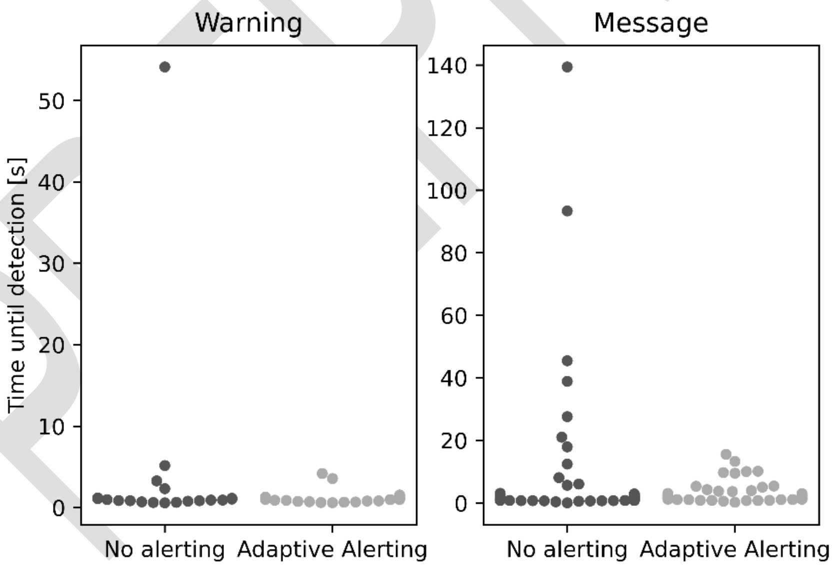
Table 2

Detection times for message and warning (N(Warning) = 20, N(Message)=54)

Condition	Warning Detection [s]		Message Detection [s]	
	M	SD	M	SD
No Alerting	3.96	11.86	7.93	22.49
Adaptive Alerting	1.23	.94	3.18	3.45

Figure 10

Time until detection of warning (top) and new message (bottom).

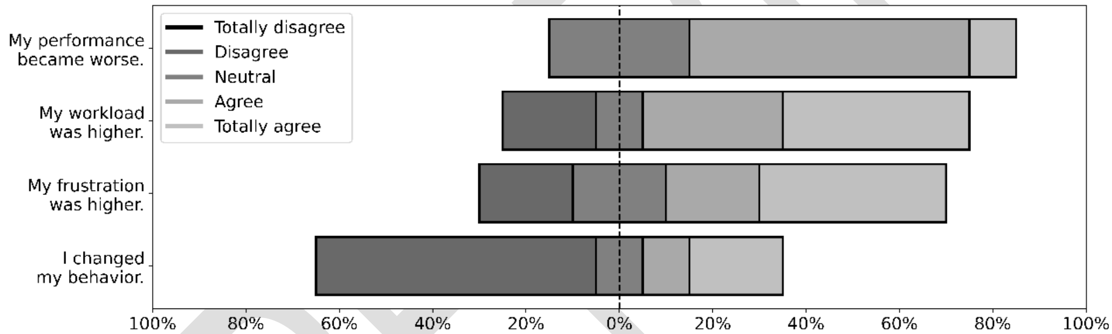


Subjective rating

After the second trial, the participants were asked to rate the adaptivity regarding their performance, workload, and frustration. Also, they answered whether they changed their behavior in the adaptive condition. Figure 11 shows the results in a Likert-scale. Overall subjective comparison of “no alerting” and “alerting” was negative towards the adaptive system. Most participant assumed that their performance became worse, and stated, that their workload and frustration was higher than without alerting. In post-trial discussions, many participants revealed that they were frustrated because the trigger mechanism was unclear to them. They lost trust in the system, when alerts were triggered even though the aircraft was within target parameters. Therefore, most participants reported that they did not change their behavior in the adaptive condition, which indicates that they tried to ignore the alerts.

Figure 11

Subjective Rating comparing performance, workload, frustration, and behavior of condition "no alerting" against "adaptive alerting" (n=10).



Discussion

In summary, the adaptive alerting improved the participant’s performance for altitude and velocity tracking. The measure of SA deviation in different system states proved to be useful in the context of the experimental task and was used to inform the alerting system. The median values of change detection were not increased, but situations of unnoticed changes were identified and reduced in the adaptive condition.

Results in tracking of velocity and altitude were consistent whereas route tracking showed an opposed trend. We explain this with differences between the tracking tasks. Altitude and velocity were disturbed externally at unknown times. Therefore, it was probable that a disturbance of altitude or velocity goes without notice of the participant. The same applies for messages and warnings triggered at unknown times. Compared with that, route changes were predictable since they always happened when a waypoint was reached. Because of this, most participants could track the route perfectly within the threshold. However, this does not explain, why route tracking became worse with adaptive aiding. Our first explanation is, that the alerts for velocity, altitude, warning, and messages distracted the pilots from tracking the route. Interestingly, we observed that the adaptive alerts increased stress for participants who were already under high workload. Another explanation could be, that pilots quickly lost trust in the route alert since it was sometimes triggered without an error (as indicated in Figure 7). This led to participants ignoring the route alert. We conclude two things: First, unpredictable changes of system states are better targets for adaptive alerting than predictable changes. Therefore, a predictable change of values should entail a greater delay until an alert is triggered, because there is a high probability that the pilot is aware of the change. Second, false positives are a critical issue for user trust and must be considered in the design. A similar issue has also been observed in (Lounis et al., 2020).

Although the performance of almost all participants increased, the subjective rating was extremely poor. We propose two explanations for this rating. First, the high number of false positives alerts contributed to this negative perception. Especially in the route tracking task, the alerts were often triggered unnecessarily. Second, pilots were not told the rationale behind the adaptive system before the experiment. This caused irritation when the pilots failed to understand why there is an alert. The combination of a high number of false positives and low transparency could be responsible for the very poor rating. This is in accordance with (Dorneich et al., 2016), where low transparency and predictability of adaptive systems on the flight deck was rated as the major risk to flight safety and acceptance by aviation professionals. Therefore, to increase acceptance and trust, participants should be trained on the adaptive system.

Additionally, the design of the alerts was highly intrusive for the multi-task experiment. The vocal notifications were disruptive to the participants’ workflow. This design might be reasonable to guide attention to safety-critical information but should be relaxed for less relevant information. In general, interruption management is an important aspect in the design of alert systems (McFarlane & Latorella, 2002). Alternatively, the system could have triggered visual alerts that increase saliency of changed parameters (comparable to (Fortmann & Mengeringhausen, 2014)) or delayed their alert until the participant finished his current task. However, the identification of a good moment of interruption must again be identified (Katidioti et al., 2016).

Limitations

There are limitations to the validity of the experiment. First, the training effect was not eliminated since the sequence of trials was “no alerting” followed by “adaptive alerting”. This could account for the improvements in performance. Yet we consider the observed effects to be strong enough to claim positive effects of the adaptive system. Second, there was no support system in the control condition. This is somewhat an unfair comparison, since any kind of support has a good chance of improving performance compared against no support. Third, there was a high variance of experience between the participants, which could have affected performance results.

Conclusion

In this contribution, we developed a SA-adaptive alerting system for an aircraft cockpit which improved performance and change detection in an experiment. We discussed the reasons for poor subjective rating and proposed ideas to improve the design. A remaining challenge of the SA measurement is the robust estimation of pilot attention. In the presented study, the different system values were uniquely allocated to the experimental task and corresponding AoIs were positioned with great distance. In a real cockpit, indicators are positioned closely (e.g. on a PFD) or integrate different information (e.g. symbols on a tactical map). Hence, identification of the attended information is not trivial because of insufficient eye tracking accuracy, look-but-failed-to-see phenomena or perception with peripheral vision.

In general, adaptive systems require more conceptual research if they were to be integrated in safety-critical environments such as aircraft. As already discussed by (Dorneich et al., 2016), the adaptivity must be transparent and provide tangible benefit to the operator. Most laboratory experiments – including this one – assume that the operator is not aware of the logic behind the adaption or at least not trained in the use of the adaptive system. This assumption is questionable because the integration of adaptive systems into safety-critical environments would entail training, that certainly includes this information. Therefore, it is not clear if self-adapting behavior of the operator might again change the efficiency in either a positive or negative way.

Appendix

Table 3

Distance functions for information categories.

Types of Information	Distance Function $d(x_1, x_2)$	
Numeric (Integer, Floating-point)	$x_2 - x_1$	Aircraft altitude, engine temperature, magnetic heading
Text		Text Message
Booleans	$\begin{cases} 0.0 \; if \; x_1 \equiv x_2 \\ 1.0 \; otherwise \end{cases}$	Gear fully extended.
Modes		Autopilot Hold/Acquire Status
Positional	Great circle distance in meter between x_1 and x_2	Aircraft Position, Airport Position

Table 4

Pseudocode of trigger mechanism.

```
ADAPTIVE-ALERTING (information, c-normalize, time-until-notification, threshold)
--------------------------------------------------------------------------------
        time-of-high-deviation = NIL
        WHILE adaptive-alert-active
            pilot-value = LAST-FIXATED-VALUE-OF ( information )
            system-value = SYSTEM-VALUE-OF ( information )
            deviation = COMPUTE-DEVIATION ( pilot-value, system-value, c-normalize)
            IF deviation > threshold
                now = GET-TIME ( )
```

```
        IF time-of-high-deviation == NIL
            time-of-high-deviation = now
        IF now – time-of-high-deviation > time-until-notification
            TRIGGER-ALERT ( information )
            time-of-high-deviation = NIL
    ELSE
        time-of-high-deviation = NIL
```

Table 5

Constants for adaptive functions in altitude, velocity, route, message, and warning

Information	Normalizing constant	Delay [s]
Altitude	200 [ft]	2
Velocity	30 [kt]	2
Route	3000 [m]	3
Warning	1 [-]	5
Message	1 [-]	5

Key Points

- This research article describes the design of a cockpit support system generating alerts adaptive to an eye-tracking based measure of SA.
- By comparing the pilot's perceived state to the actual system state, a deviation between the pilot's awareness and the situation can be computed, which is used to trigger alert for poorly monitored states.
- In an experiment, SA-adaptive alerts improved average performance and detection of critical changes, especially where changes could not have been predicted by the flight crew.
- Subjective rating of the adaptive support was poor due to low transparency and false alarms.
- SA-adaptive support can improve safety in aviation, but transparency must be considered in design and training.

References

Billman, D., Mumaw, R., & Feary, M. (2020). A Model of Monitoring as Sensemaking: Application to Flight Path Management and Pilot Training. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *64*(1), 244–248. https://doi.org/10.1177/1071181320641058

Bosse, T., van Lambalgen, R., van Maanen, P.-P., & Treur, J. (2009). Attention Manipulation for Naval Tactical Picture Compilation. In *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology* (pp. 450–457). IEEE. https://doi.org/10.1109/WI-IAT.2009.194

Dorneich, M. C., Rogers, W., Whitlow, S. D., & DeMers, R. (2016). Human Performance Risks and Benefits of Adaptive Systems on the Flight Deck. *The International Journal of Aviation Psychology*, *26*(1-2), 15–35. https://doi.org/10.1080/10508414.2016.1226834

Endsley, M. R., & Jones, D. G. (2012). *Designing for situation awareness: An approach to user-centered design / Mica R. Endsley and Debra G. Jones* (2nd ed.). CRC Press.

Feigh, K. M., Dorneich, M. C., & Hayes, C. C. (2012). Toward a characterization of adaptive systems: A framework for researchers and system designers. *Human Factors*, *54*(6), 1008–1024. https://doi.org/10.1177/0018720812443983

Fortmann, F., & Mengeringhausen, T. (2014). Development and Evaluation of an Assistant System to Aid Monitoring Behavior during Multi-UAV Supervisory Control. In C. Stary (Ed.), *Proceedings of the 2014 European Conference on Cognitive Ergonomics - ECCE '14* (pp. 1–8). ACM Press. https://doi.org/10.1145/2637248.2637257

Honecker, F., & Schulte, A. (2017). Automated Online Determination of Pilot Activity Under Uncertainty by Using Evidential Reasoning. In (pp. 231–250). Springer, Cham. https://doi.org/10.1007/978-3-319-58475-1_18

Johnson, A. W., Oman, C. M., Sheridan, T. B., & Duda, K. R. (2014). *Dynamic task allocation in operational systems: Issues, gaps, and recommendations: Big Sky, Montana, USA, 1 - 8 March 2014*. IEEE. https://doi.org/10.1109/AERO.2014.6836205

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354. https://doi.org/10.1037/0033-295X.87.4.329

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. *Theoretical Issues in Ergonomics Science*, *5*(2), 113–153. https://doi.org/10.1080/1463922021000054335

Katidioti, I., Borst, J. P., Bierens de Haan, D. J., Pepping, T., van Vugt, M. K., & Taatgen, N. A. (2016). Interrupted by Your Pupil: An Interruption Management System Based on Pupil Dilation. *International Journal of Human–Computer Interaction*, *32*(10), 791–801. https://doi.org/10.1080/10447318.2016.1198525

Kharoufah, H., Murray, J., Baxter, G., & Wild, G. (2018). A review of human factors causations in commercial air transport accidents and incidents: From to 2000–2016. *Progress in Aerospace Sciences*, *99*, 1–13. https://doi.org/10.1016/j.paerosci.2018.03.002

Lounis, C., Peysakhovich, V., & Causse, M. (2020). Flight Eye Tracking Assistant (FETA): Proof of Concept. In N. Stanton (Ed.), *Advances in Intelligent Systems and Computing. Advances in Human Factors of Transportation* (Vol. 964, pp. 739–751). Springer International Publishing. https://doi.org/10.1007/978-3-030-20503-4_66

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. https://doi.org/10.1371/journal.pone.0247061

McFarlane, D. C., & Latorella, K. A. (2002). The Scope and Importance of Human Interruption in Human-Computer Interaction Design. *Human–Computer Interaction*, *17*(1), 1–61. https://doi.org/10.1207/S15327051HCI1701_1

Peißl, S., Wickens, C. D., & Baruah, R. (2018). Eye-Tracking Measures in Aviation: A Selective Literature Review. *The International Journal of Aerospace Psychology*, *28*(3-4), 98–112. https://doi.org/10.1080/24721840.2018.1514978

Reback, J., McKinney, W., Jbrockmendel, van Bossche, J. den, Augspurger, T., Cloud, P., Gfyoung, Hawkins, S., Sinhrks, Roeschke, M., Klein, A., Terji Petersen, Tratner, J., She, C., Ayd, W., Naveh, S., Garcia, M., Patrick, Schendel, J., . . . H-Vetinari. (2021). *pandas-dev/pandas: Pandas 1.2.3* [Computer software]. Zenodo.

Rouse, W. B. (1988). Adaptive Aiding for Human/Computer Control. *Human Factors*, *30*(4), 431–443. https://doi.org/10.1177/001872088803000405

Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D., Ladva, D., Rafferty, L., & Young, M. (2009). Measuring Situation Awareness in complex systems: Comparison of measures study.

International Journal of Industrial Ergonomics, *39*(3), 490–500. https://doi.org/10.1016/j.ergon.2008.10.010

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. *Human Factors*, *49*(3), 347–357. https://doi.org/10.1518/001872007X196685

Schwarz, J., & Fuchs, S. (2017). Multidimensional Real-Time Assessment of User State and Performance to Trigger Dynamic System Adaptation. In D. D. Schmorrow & C. M. Fidopiastis (Eds.), *Lecture Notes in Computer Science. Augmented Cognition. Neurocognition and Machine Learning* (Vol. 10284, pp. 383–398). Springer International Publishing. https://doi.org/10.1007/978-3-319-58628-1_30

Schwerd, S., & Schulte, A. (2020). Experimental Validation of an Eye-Tracking-Based Computational Method for Continuous Situation Awareness Assessment in an Aircraft Cockpit. In D. Harris & W.-C. Li (Eds.), *Lecture Notes in Computer Science. Engineering Psychology and Cognitive Ergonomics. Cognition and Design* (Vol. 12187, pp. 412–425). Springer International Publishing. https://doi.org/10.1007/978-3-030-49183-3_32

Schwerd, S., & Schulte, A. (2021). Measuring the Deviation between Ground Truth and Operator Awareness in a UAV Management Scenario: An Eye-Tracking Approach. In *AIAA Scitech 2021 Forum.* American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2021-1579

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. *Journal of Cognitive Engineering and Decision Making*, *9*(4), 312–328. https://doi.org/10.1177/1555343415597344

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., . . . van Mulbregt, P. (2020). Scipy 1.0:

Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Zhang, T., Yang, J., Liang, N., Pitts, B. J., Prakah-Asante, K. O., Curry, R., Duerstock, B. S., Wachs, J. P., & Yu, D. (2020). Physiological Measurements of Situation Awareness: A Systematic Review. *Human Factors*, 18720820969071. https://doi.org/10.1177/0018720820969071

Ziv, G. (2016). Gaze Behavior and Visual Attention: A Review of Eye Tracking Studies in Aviation. *The International Journal of Aviation Psychology*, *26*(3-4), 75–104. https://doi.org/10.1080/10508414.2017.1313096

Author Biographies

Simon Schwerd received the M.Sc. (2016) in Mechanical Engineering from the Technical University of Munich Germany. He is currently pursuing his Ph.D. at the Institute of Flight Systems at the University of the Bundeswehr Munich. His research interests include flight crew cognitive state estimation and adaptive systems for aircraft cockpits.

Axel Schulte received Aerospace Engineering and Doctor of Engineering degrees in aerospace engineering from the University of the Bundeswehr Munich, Neubiberg, Germany, in 1990 and 1996, respectively. Since 2002, he has been a professor of aircraft dynamics and flight guidance, and since 2008, the Director of the Institute of Flight Systems, University of the Bundeswehr Munich. In 2010, he was the Visiting Professor with the Humans and Automation Lab, Massachusetts Institute of Technology (MIT). His research interests include human-autonomy teaming and cognitive automation in aviation.