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: 5116

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 to trigger alerts in a flight simulator. In an experiment, ten certified pilots conducted two trials, with and without adaptive alerting. The experimental task involved tracking of flight parameters which were disturbed or changed at either random or previously known 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 the change of flight parameters could not be predicted by participants. Also, adaptive alerts reduced mean and variance of change detection duration. However, subjective ratings were mixed 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 showed that adaptive alerts improve performance and change detection, but mixed subjective ratings indicated 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 could identify situations of poor pilot monitoring and notify the pilot about relevant system states. The goal of this contribution is to evaluate an eye-tracking based SA metric, called *awareness deviation*, as a suitable trigger for an adaptive alerting system. Focus of this study is to demonstrate how this measure can be applied and evaluate the resulting adaptive system behavior in a representative task setting. Our hypothesis was, that this metric has a meaningful relationship to pilot errors, and we assumed, that the adaptive system improves monitoring performance compared to no assistance.

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), there are very few real adaptive systems deployed in operational aircraft flight decks because they are considered opaque and complex by aviation professionals (Dorneich et al., 2016). Nevertheless, the idea of computerized adaptive systems is attractive because they resemble the behavior of human assistants, comparable to a pilot-monitoring supporting a pilot-flying, where the pilot-monitoring has knowledge of the situation, the task, and the activity of the pilot-flying while they support based on their own initiative (Johnson et al., 2014). 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. In numerous aviation studies, monitoring was analyzed by gaze measurement in the cockpit using gaze metrics that capture the temporal, spatial and sequential dynamic of gaze behavior (Peißl et al., 2018; Ziv, 2016). With these metrics (e.g., fixation duration), 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 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 the task for which monitoring is evaluated. While this is no issue for standardized environments such as airliner cockpits, it might complicate the applicability for other environments. Finally, most studies focus on gaze behavior alone and ignore the situation and the aircraft state. Again, this is problematic when monitoring behavior is optimal for a standard situation but the situation itself is exceptional. We therefore argue that a measure is needed that incorporates both gaze and system states to identify situations, when monitoring support is required. The model of SA provides such a framework since it is dependent on the situation and the perception 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., SART) or the use of physiological and behavioral measurements, such as eye-tracking or EEG (Salmon et al., 2009), where only the latter two are suitable for online measurement. For example, (Vidulich & McMillan, 2000) used fighter pilot control inputs that work towards the mission goal to estimate SA. But most studies used eye-tracking to infer SA because monitoring is considered an important activity to gather SA (Zhang et al., 2020). In the context of aviation, this has been studied for air traffic control (Moore et al., 2010), UAV control (Ratwani et al., 2010), flight decks (Kilingaru et al., 2013; van de Merwe et al., 2012), construction (Hasanzadeh et al., 2018) or manufacturing (Dini et al., 2017).

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 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. This improved detection time of UAV malfunctions and intruders in their task scenario. 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 prevented performance decrements by guiding the operator’s attention to the most relevant parts of the task environment. In the context of aircraft cockpits, Lounis et al. (2020) developed a 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 ratings were not optimal due to false alarms. The authors concluded that the integration of flight parameters could 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 applicable for other task environments (e.g., driving or process control). Further, studies in aircraft cockpits did not incorporate the aircraft's 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 application of this novel measure in a SA adaptive system to improve monitoring in representative task setting.

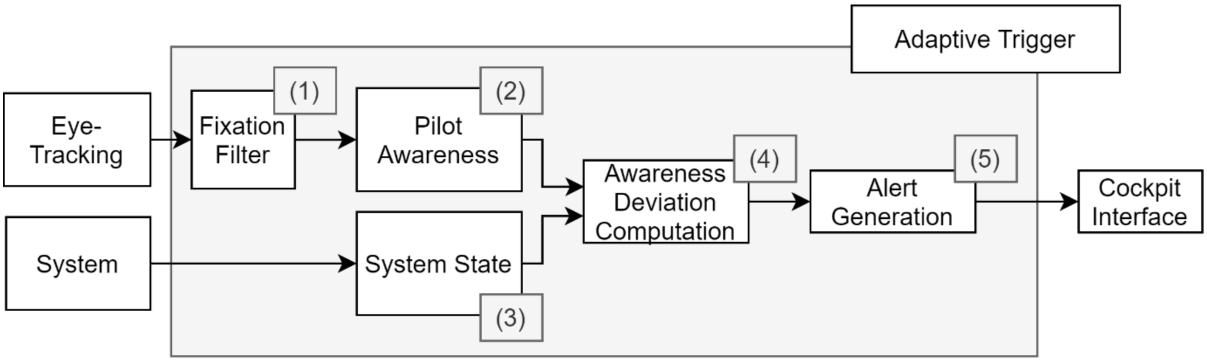
Awareness Deviation as a measure of SA

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 *awareness deviation* (AD), which describes the difference between the current state value and the assumed pilot’s awareness in each aircraft state. The eye-tracking based measurement of AD corresponds to the perceptional level of SA (Level 1), which 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 AD 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 a 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)).

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.

Awareness Deviation

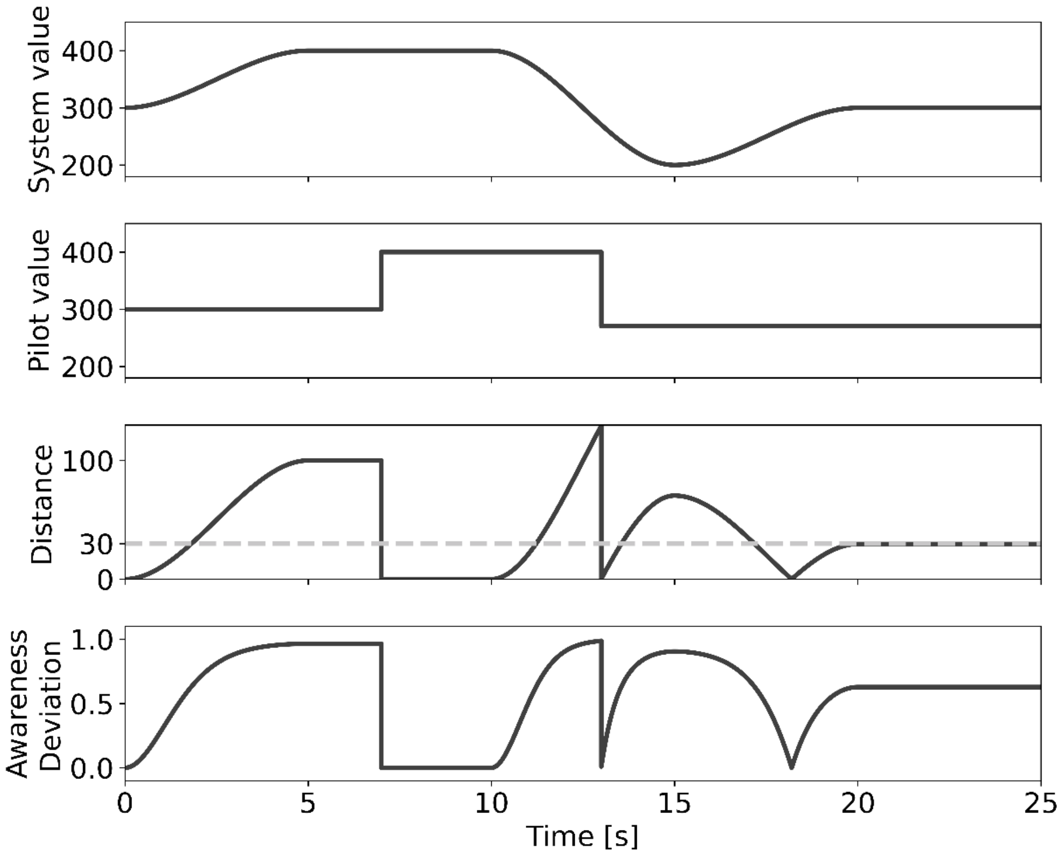
In the AD computation (see Figure 1 (4))., we compute a normalized difference between system state and pilot awareness. This difference is computed for every relevant system state, which can be represented in different forms of information, such as modes, figures, text, times, or position. For each of 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 4 in the appendix shows example functions for different categories. To normalize the distance for further processing, we apply the following formula to compute the AD:

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

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 AD while the change in AD 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), AD should not grow proportionally.

Figure 2

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



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

Figure 2 shows an exemplary course of the AD 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 AD are zero. When the speed changes, the distance between the pilot and system value increases. Accordingly, the AD 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 AD 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 AD. After that, the speed returns to 300 kts resulting in an AD of zero.

Alert

The AD for all states is passed to the alert generation (see Figure 1,5). When AD exceeds a limit for a predefined duration, an alert is created. We selected this limit to be .6 which indicates a distance in pilot and system value of approximately the normalizing constant. 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 check if there is a pilot error present before an alert is triggered. We decided not to filter the alerts to observe unaltered behavior of such an adaptive system independent of the specific experimental task

Algorithm

Table **5** in the appendix denotes the trigger algorithm for a single information value. The algorithm is executed in a loop, that continuously evaluates the AD and processes the delay logic. A normalizing constant c_n and a time until notification 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 contained three touchscreens and a Head-Up-Display (HUD). The setup is shown in Figure 4.

For eye-tracking, we used the commercially available four-camera system by SmartEye© (Smart Eye Pro 0.3 MP, best accuracy <0.5° under ideal conditions) which was connected to our simulation setup. The gaze was measured with a sampling frequency of 60Hz and was tracked across all displays including the HUD. We used the functionality of the commercial software to classify fixations, where the gaze angular velocity threshold was set to $\dot{\alpha} \leq 2\frac{\circ}{s}$.

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 varied between 500 and 29000 hours reflecting the difference between private and professional pilots.

Procedure

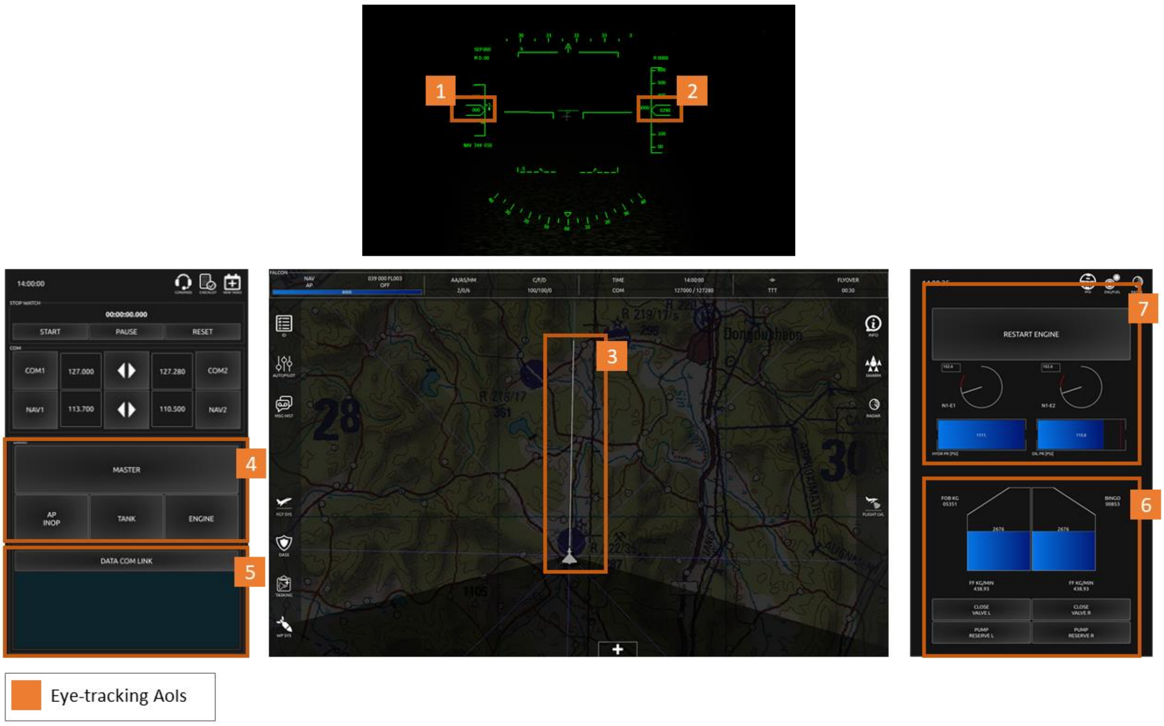
All participants received instructions about the simulator before the experiment. Each participant received a simulator training in the experimental task for approximately one hour. During this training, they encountered all aspects of the experimental task including emergency procedures as reaction to system warnings (e.g., fuel system failure). Then, the eye-tracking was calibrated with mean accuracy of $1.04°$ ($SD_{acc} = 0.72°$). After that, the participants conducted two half-hour 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. After the second trial, all participants answered a short questionnaire to evaluate the assistance system.

Tasks

In both trials, the experimental task included monitoring of continuous parameters and discrete states of the aircraft - the interfaces including task relevant AoIs are shown in Figure 4. The participants had three continuous tracking tasks without using an autopilot: 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, there were disturbances of either altitude or velocity triggered with an average frequency of 3.5 times per minute, however, there was no disturbance of both variables at the same time. 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: Areas of interest (AoI) marked in orange displayed the following information: (1) Altitude indicator, (2) Velocity 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 AD measurement was implemented for each of the five experimental tasks: route, velocity and altitude tracking as well as confirmation of a new message and a warning procedure. For the alert, we used a computer-generated voice to notify the pilot with keywords (for example "ALTITUDE", "SPEED"). To trigger alerts for route, altitude, and speed, we selected the normalizing constants c_n to be close to the tracking limits. For warning and message, $c_n = 1$ because these indications are binary ("there is a message" vs. "there is no message") and there is no meaningful numeric distance between these states. The trigger was reached when the distance between system and pilot state approximately reaches the value of this normalizing constant. When AD exceeded the trigger 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 1 contains the constants c_n and delay times for each information relevant to the experimental task.

Table 1

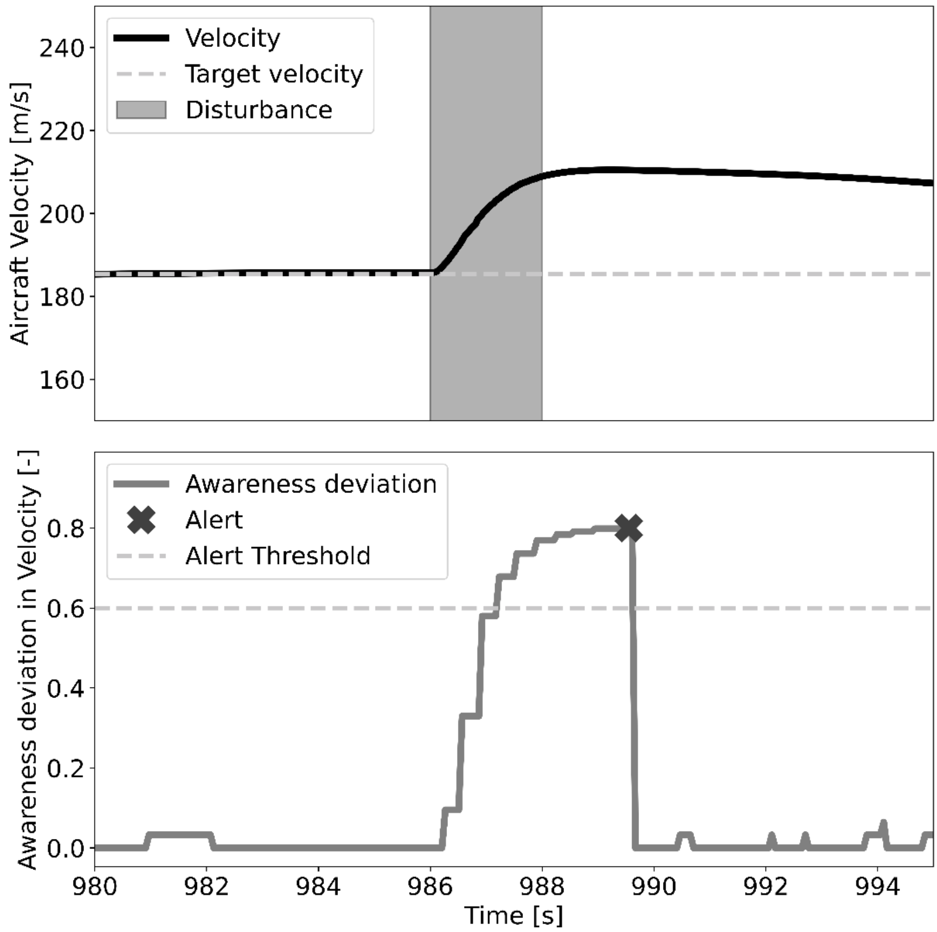
Constants for adaptive functions in altitude, velocity, route, message, and warning

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

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 AD in velocity. Due to the AD exceeding a threshold for more than 3 seconds, a vocal alert is triggered. The pilot reacts by looking at the velocity indicator which causes the AD 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 AD 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 limit, error was set to 0. Tracking limits were 200 ft, 15 knots and 2 NM for tracking of altitude, velocity, and route, respectively. During the trial, we logged data of AD, 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’s 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 $p=.05$.

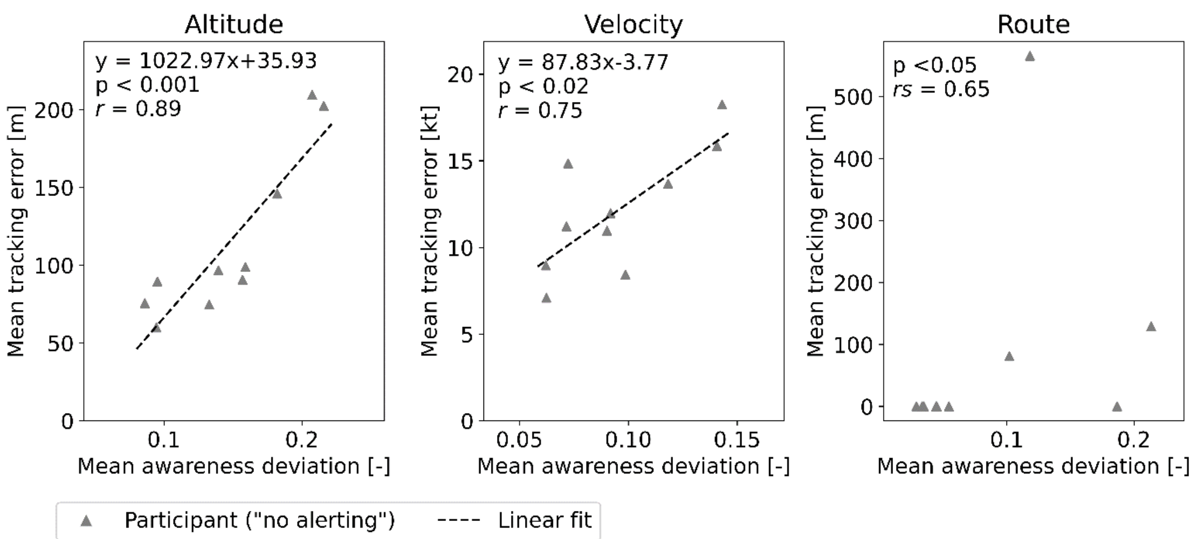
Results

Validation of operator measurement

To evaluate our approach to SA measurement, we compared AD 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. Figure 6 displays mean error and mean AD over the complete trial in all tracking tasks for the control condition. For altitude tracking, the correlation coefficient was $r = .88$ (significant, $p<.001$). For velocity, the data showed a strong positive correlation of $r = .75$ (significant, $p<.01$). These results reproduced 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 AD and error in route tracking. Of all participants, 60% did not commit any error in this task. However, the mediate positive correlation coefficient of $rs = .65$ is significant. The strong positive correlation clearly shows that the measure of AD is meaningful in tracking speed and altitude. A similar but weaker trend can be observed in route tracking.

Figure 6

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

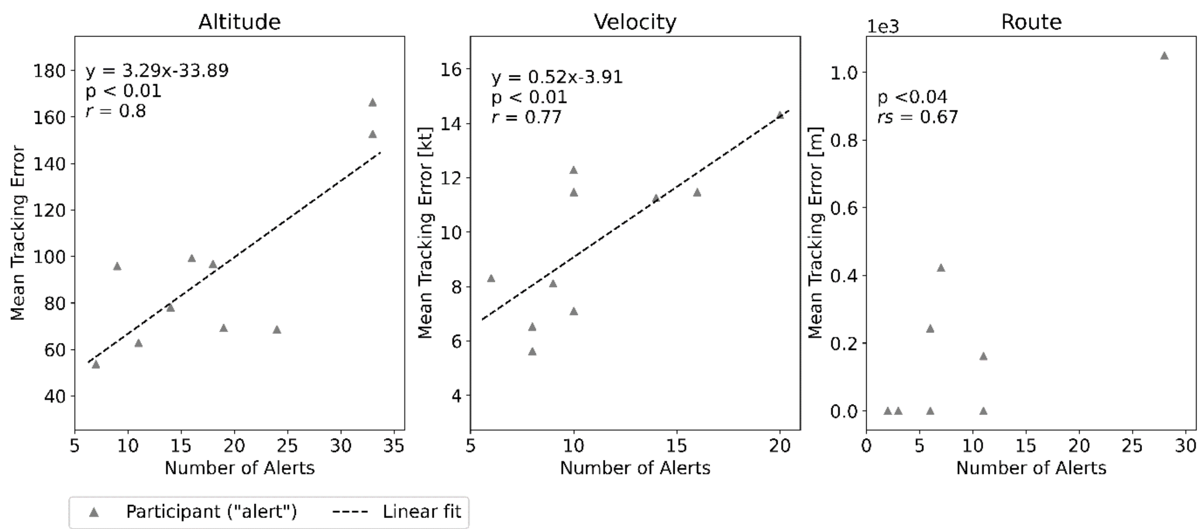


Relationship between error and trigger

In the adaptive condition, we compared the number of triggered alerts against the error rate. Results are shown in Figure 7. There was 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 was a mediate correlation of rs=.67 (p<.04) between number of route alerts and mean error in route tracking. Also, the route data showed that participants triggered alerts even when there was no error. These false alarms were triggered because the adaptive mechanism solely used the AD 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).

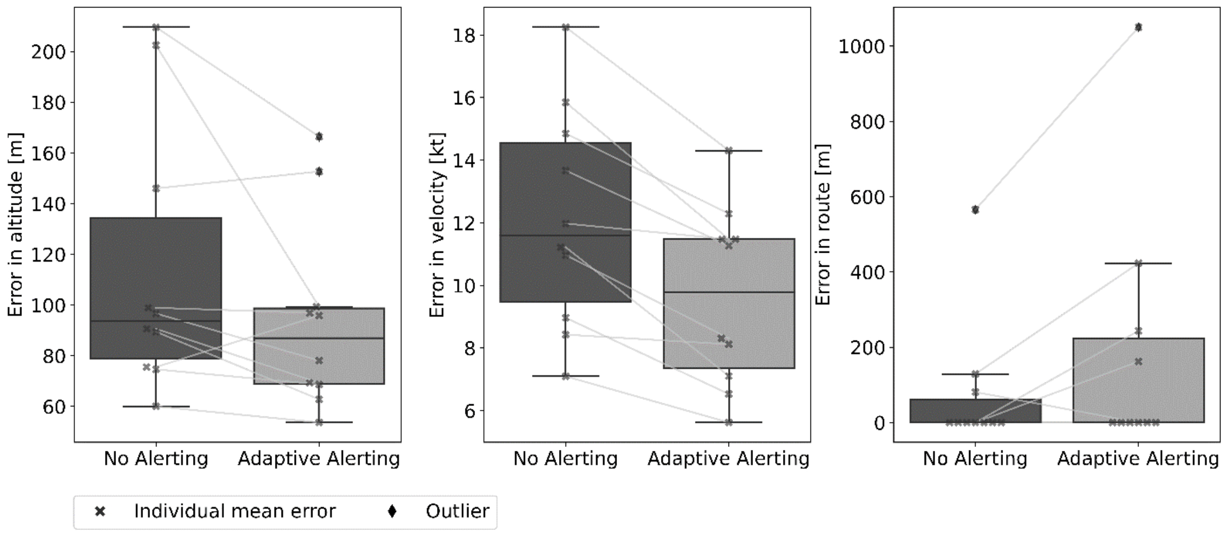


Comparison of performance

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, the error in route tracking increased with adaptive alerting. The increase in route tracking error was not significant (ps < .11). Average performance as displayed in Figure 8 reflected 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 the participants noticed an error in aircraft state after it has been induced.

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

Detection time was 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). Data in Table 2 shows mean (M) and standard deviation (SD) values, in addition, Figure 9 shows a letter-value box of the distribution. For altitude and velocity, the mean time until change detection slightly decreased in the adaptive condition. For route, there was a non-significant increase in mean time of detection.

Table 2

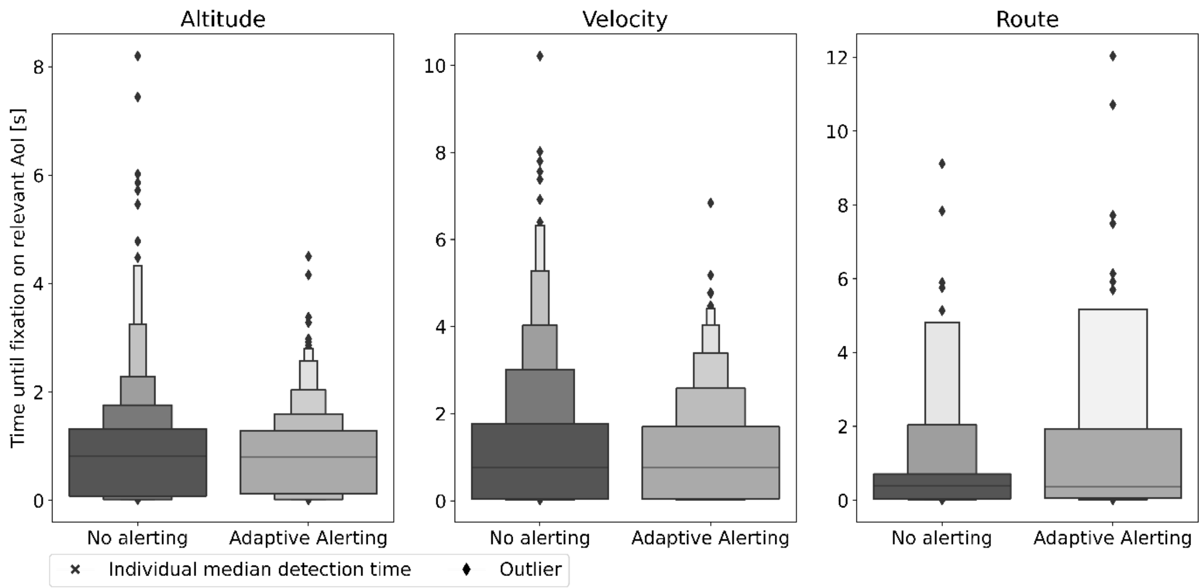
Detection times for tracking tasks.

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

For altitude and velocity, adaptive alerting also reduced the SD. The decrease between conditions was 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 samples with long detection times were reduced by the alerts triggered after AD measurement exceeding the threshold for more than the specified time delay. There was an opposite effect for route tracking. Although the median time for detection decreased, the SD increased, however, the change was not significant ($p<.09$). Note, that the median times were below the time delay for the trigger, therefore could not have been 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 3 shows the mean detection times for new messages and warnings. There was a decrease in both, mean and SD, in the experimental condition. This effect can mainly be attributed to the extreme samples visible in

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

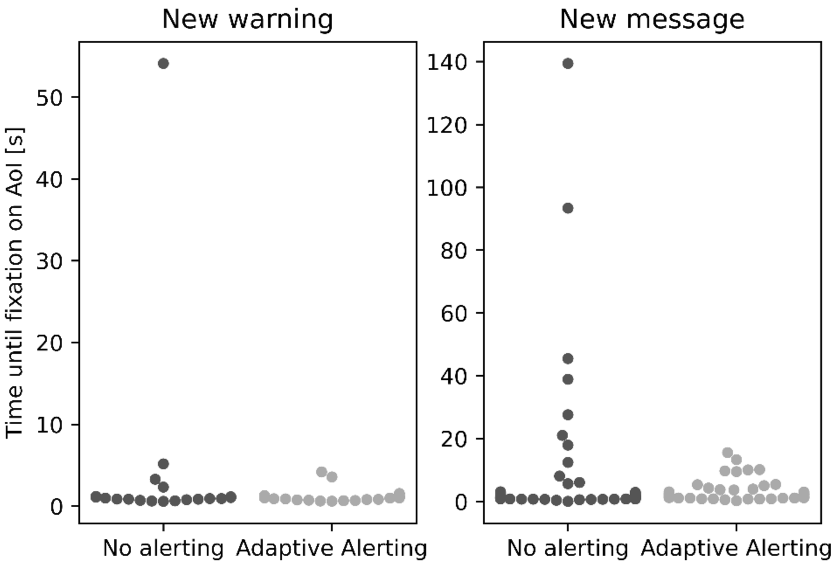
Table 3

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 new warning (left) and new message (right).

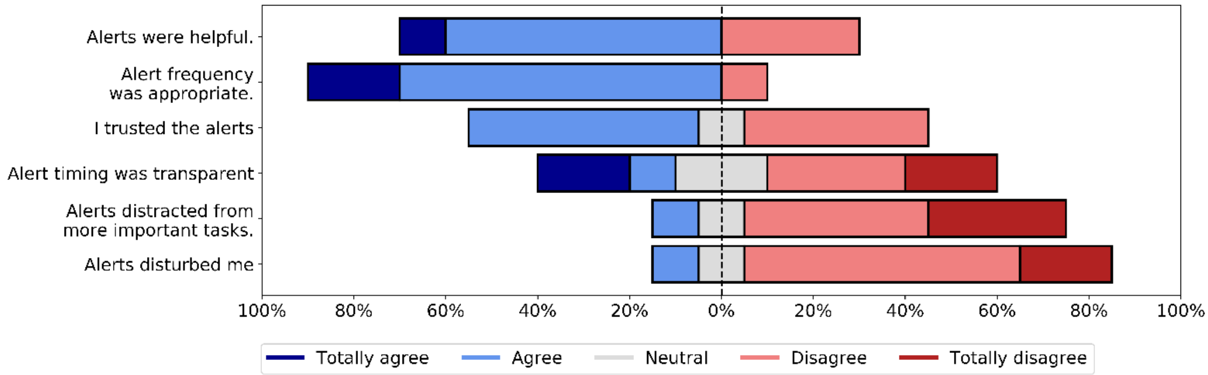


Subjective ratings

After the second trial, participants were asked to rate utility and transparency in a non-standardized questionnaire (see Figure 11). Overall results showed that most participants found the alerts to be helpful and triggered with an appropriate frequency. Also, most participants did not feel disturbed by the alerts or distracted from other tasks. However, mean participant rating showed low scores of trust and transparency. In a debriefing after the trial, some participants criticized that they did not always understand why an alert was triggered, especially when their flight parameters have been within the desired limits. The participants lost trust after a few of these false alerts. Another criticism was that auditive alerts should only be used for critical situations and not for a small deviation in flight parameters.

Figure 11

Subjective ratings of adaptive alerting (n=10).



Discussion

In summary, the measure of AD captured interesting properties about the participant's performance and results in the control condition showed that it is correlated with pilot error. Also, AD could be used to generate reasonable alerts in the experimental condition even without directly checking if an error is present. The alerts improved mean performance and reduced variance especially in those "unpredictable" tasks such as altitude tracking disturbed by virtual gusts. While results in tracking of velocity and altitude were consistent, route tracking performance deteriorated. Our first explanation is, that the alerts for velocity and altitude 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 ratings were mixed. We propose two explanations for this rating. First, the false alerts contributed to this negative perception, especially in route tracking where 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 mixed 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, participants should be trained on the adaptive system. Further, system design could be improved by considering models of transparency (Chen et al., 2018).

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 (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 the observed effects show that AD is a promising approach to measure SA. 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. Further, there was a high variance of participant experience and participants were not trained with the adaptive system, which could confound the performance results.

Conclusion

The presented SA-adaptive alerting was able to trigger useful alerts during the experiment. We discussed the reasons for mixed subjective ratings and proposed ideas to improve the design. A remaining challenge of the AD measurement is the robust estimation of pilot attention on a professional flight deck. 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 attended information is not trivial because of insufficient eye tracking accuracy, look-but-failed-to-see phenomena, or near-peripheral perception.

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 operators might change their behavior when they are supported by a behavior-adaptive system and how this would affect the efficiency of the whole human-machine-system.

Appendix

Table 4

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	$\begin{cases} 0.0 \text{ } if \ x_1 \equiv x_2 \\ 1.0 \ otherwise \end{cases}$	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 5

Pseudocode of trigger mechanism.

```
ADAPTIVE-ALERTING (information, c-norm, time-until-notification, threshold = .6)

        time-of-high-AD = NIL
        WHILE adaptive-alert-active
            pilot-value = LAST-FIXATED-VALUE-OF ( information )
            system-value = SYSTEM-VALUE-OF ( information )
            aw-deviation = COMPUTE-AW-DEVIATION ( pilot-value, system-value, c-norm)
            IF aw_deviation > threshold
                now = GET-TIME ( )
```

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

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 ratings of the adaptive support were 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

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

Chen, J. Y. C., Lakhmani, S. G., Stowers, K., Selkowitz, A. R., Wright, J. L., & Barnes, M. (2018). Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical Issues in Ergonomics Science*, *19*(3), 259–282. https://doi.org/10.1080/1463922X.2017.1315750

Dini, A., Murko, C., Yahyanejad, S., Augsdörfer, U., Hofbaur, M., & Paletta, L. (2017). Measurement and prediction of situation awareness in human-robot interaction based on a framework of probabilistic attention. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* Symposium conducted at the meeting of IEEE.

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

Hasanzadeh, S., Esmaeili, B., & Dodd, M. D. (2018). Examining the relationship between construction workers' visual attention and situation awareness under fall and tripping hazard conditions: using mobile eye tracking. *Journal of Construction Engineering and Management*, *144*(7), 4018060. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001516

Johnson, A. W., Oman, C. M., Sheridan, T. B., & Duda, K. R. (2014). Dynamic task allocation in operational systems: Issues, gaps, and recommendations. *IEEE Conference on Aerospace*, 1–15. 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

Kilingaru, K., Tweedale, J. W., Thatcher, S., & Jain, L. C. (2013). Monitoring pilot "Situation Awareness". *Journal of Intelligent & Fuzzy Systems*, *24*(3), 457–466. https://doi.org/10.3233/IFS-2012-0566

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

Moore, Kristin and Gugerty, Leo (2010). Development of a novel measure of situation awareness: The case for eye movement analysis. *Proceedings of the Human Factors and Ergonomics Society*, *54*.

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

Ratwani, R. M., McCurry, J. M., & Trafton, J. G. (2010). Single operator, multiple robots: An eye movement based theoretic model of operator situation awareness. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. http://ieeexplore.ieee.org/servlet/opac?punumber=5447157

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].

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

van de Merwe, K., van Dijk, H., & Zon, R. (2012). Eye movements as an indicator of situation awareness in a flight simulator experiment. *The International Journal of Aviation Psychology*, *22*(1), 78–95. https://doi.org/10.1080/10508414.2012.635129

Vidulich, M., & McMillan, G. (2000). The global implicit measure: evaluation of metrics for cockpit adaptation. *Contemporary Ergonomics*, 75–80.

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.