Evaluation of a display for airborne separation in safety-critical situations

Marcus Arenius, Marc Schneider & Oliver Sträter
Department of Human and Organisational Engineering & Systemic Design
University of Kassel
Kassel, Germany
arenius@ifa.uni-kassel.de

Abstract—This article presents the results of a study conducted to assess the impact of an ATC information display for airborne separation on the situation awareness and performance of pilots. Pilots from a major German airline (n=10) flew a safety-critical flight scenario in a full flight simulator using the ATC-display. Their performance was evaluated by expert ratings of professional trainers from the airline and by means of eye-tracking performance data. The effect of the display on the individual components of the SA was assessed using Bravais-Pearson correlation. The results showed that there is a significant negative correlation (r(9) = -0.75, p < .05) between the fixations on the ATC-Display and performance in the safety-critical scenario. This finding is explained in terms of the associated situation awareness of the pilots. The results are discussed regarding the design conclusions and spatial arrangement options of the display inside the cockpit. Furthermore, this article discusses how these types of empirical design evaluations can be further improved by addressing the dynamics of human performance and how this type of research can be linked to results from human reliability research in order to improve the systematic of future studies.

Keywords: Human Factors, airborne separation, human-automation interaction, safety-critical events

I. INTRODUCTION

One of the main goals of the Single European Sky (SES) is to integrate the highly fragmented structures of European airspace step by step into functional airspace blocks (FAB's) in order to decrease the operational costs and increase the efficiency and capacity of the European airspace. The Free flight environment is envisioned to contribute in achieving these goals. In the free flight environment, aircrafts above a certain flight level decide themselves which route to fly, so that they can find the optimal and most economical route in order to save fuel, increase the payload and thus ultimately increase the total capacity of the airspace [1].

As evident from fig. 1, the Free Flight environment leads to a significant increase of mental effort for an air traffic controller, as dedicated airways are eliminated due to the fact that the aircrafts can choose their specific routes freely. This leads to high variability in trajectories for which centralized control by a human operator is unsuited. Therefore, a decentralized solution is preferred where the pilots, when they are within the limits of the Free Flight environment, are responsible for local air traffic control tasks (termed airborne separation) [2]. As this constitutes an additional task for the pilots, information sources are needed that facilitate the decision-making to ensure safe trajectory planning. This is a vital prerequisite if the Free Flight environment should function with constant or decreasing number of air traffic controllers. Ideally, these information sources should preserve or improve the target levels of safety already present in the system.

This is especially true when breakdowns of automation require human intervention to ensure system safety or conditions of work are otherwise not accounted for by the available technology and safety systems. In these cases, the responsibility for the safety of the overall system still remains at the hands of the human, and the information source should ideally support the human in managing these critical tasks. However, if the automation is not adapted to the performance of the human operators, situations may arise which could exceed the capabilities of the humans in these off-nominal conditions. In the literature these problems are related to the ironies of automation [3].

Typical issues include deficits in vigilance due to the lack of active participation in the system [4], not knowing what is happening in the system and what the automation will do (automation surprises [5]) and issues in trust of the operators in automation e.g. misuse or overreliance - when human operators use or rely on the automation when it is not warranted [6], [7] and disuse, when the automation is not used when it should have been [6]. These issues arise, when the automation does not provide support for good situational awareness (SA) by the pilots [8]–[11].

Therefore, in order to make statements about the quality of the display, it has to be determined to what extent it provides support for SA of the pilots, and what the associated impact on the safety of the system will be. This holds especially true for the large-scale technical innovations to be introduced in the context of the harmonization of European airspace, where new flight separation concepts affect a large body of actors and operators.
In order to provide further scientific input to this discussion, an empirical study was conducted in collaboration with a German airline, with the aim to assess the impact of the additional information display on the performance of pilots in safety critical events. Based on this, issues were identified that should be addressed in order to guarantee that the system stays safe and efficient. The results presented in this article originate from research described in detail in [13].

It should be noted, that several display concepts for airborne separation have already been developed over the recent decade (e.g. [12], [14], [15]). The display evaluated in this study constitutes a mock-up, that mimics core functionalities necessary for airborne separation. As such, it was developed based on reviews of existing design concepts which were integrated into the design shown in section II.A based on the feedback if the industry partner. As these (core) functionalities will be part of any design concept, this study demonstrates the benefits of testing the design concept for non-nominal conditions considered to be relevant by the industry in order to ensure that the display fulfills its purpose in specific safety-critical situations.

II. SITUATIONAL AWARENESS

Situational awareness (SA) is one of the most commonly applied concepts in human factors research [16]. It is defined as „the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future” [17, p. 5]. It can be divided into the three components: spatial awareness, systems awareness and task awareness [9].

Spatial Awareness includes all information and tasks required to control an aircraft in a specific situation. The aircraft system has six variables, the monitoring of which represents a burden on cognitive information processing, as they vary over time and are strongly interrelated [9]. The first three variables denote the movement possibilities of an aircraft on the three different axes. They include rollers on the longitudinal axis, the pitch of the transverse axis and the yawing around the vertical axis [18]. The other variables are the flight altitude, the lateral deviations from the flight path and the position along the flight route. The coordination of the six variables requires a high cognitive effort regarding the control movements, as their future impact on the behavior of the aircraft and the other variables has to be considered.

This constitutes a challenge for system design, as it has to be ensured that the pilot stays in the loop e.g. when a technical malfunction requires the takeover of functions from automation. If the pilot is decoupled from the automation, pilots may struggle with the situation assessment and takeover and safety critical situations may arise. This negative impact of decoupling should also be visible in the valid measures of the system awareness component of situational awareness.

The second component of SA is System Awareness. An aircraft has a variety of engineering systems to assist the pilot in order to reduce the demands imposed during the flight. The pilot does not have to be aware of all system parameters at each point in time, except when an unexpected event occurs [9]. In order to assess SA in empirical studies, indices have to be applied that show how human and system performance are related to the quality of SA. Different aspects of Situational Awareness can be measured with different indices (fig.2). The process indicators are logged by means of the system parameters of the full-flight simulator. Data related to the other indices is collected through the performance ratings by trainers, as well as eye-tracking data. The measurements, targeting the different indices are described in detail in the following section.

![Figure 2. Indices for measuring SA, adapted from [17]](image-url)
III. METHOD

A. Display

The mock-up display issues traffic warnings when the trajectories of two aircraft will lead to a violation of the respective TCAS protection volume in the near future. This is calculated based on the heading and speed of each aircraft in the environment. The display shows the aircrafts associated with the traffic warning in the upper part of the display and is updated based on the trajectories of the aircrafts. The lower part of the display shows the altitude of the aircrafts, for which the traffic warning was issued, in relation to the altitude of the own aircraft. The display generates three evasion options shown on the right-hand side of fig. 3, that were hardcoded and are described in detail in section II.D. The display is described in full detail in [13]. Due to space limitations, the display was placed at the location of the Flight Back (fig. 5).

B. Participants and Scenarios

A total of 10 male employed pilots participated in the study (n=10). Other demographic data could not be collected due to privacy restrictions of the participants. The scenario is a flight route of the standard training program by the airline and it is flown in the full-scope simulator that is used by the airline for regular checks of pilot performance. Thus, the pilots are familiar with the route and the simulator. The handling of the scenario itself requires a high level of SA, as a number parameters must be taken into consideration by the pilot in order to properly assess the situation and consequently to respond correctly and quickly (see section II.D for a detailed description).

The pilots start on route from Athens to Heraklion. After ~60 seconds, the pilot receives a traffic warning (TW) followed by an Engine Failure at the marked location on the map in fig. 4.

The location where the pilots receive the traffic warning is known for high traffic volume and therefore a likely location for actual occurrence of traffic warnings. During the traffic warning, the ATC-display shows four aircraft on the radar screen, with one of them being a potential conflict. The traffic conflict results from an approaching aircraft that is flying at the same altitude on the exact opposite course.

Following the traffic warning, the display shows three different options for solving the conflict: a climb to flight level 380, to turn right to heading 160° and to descend to flight level 300. The pilot may choose one of the options or ignore the display completely while assessing and handling the Engine Failure. The scenario ends when the aircraft is under control and in a safe state.

C. Eye-tracking

A portable eye-tracking system was used to record the gaze of the pilots on the information displays. Usually Areas of Interest are used to evaluate the gaze behavior and to locate the gaze in the experimental setting. The main Areas of Interest for the study were the Primary Flight Display (PFD), the Navigation Display (NAV), the Engine and Warning Display (E/WD), the Flight Management Computer (FMC) and the Flight Back were the ATC-Display was located.

D. Performance rating by experts

The evaluation of the performance of the pilots is done by a professional check pilot of the airline.
Three parameters were assessed:

- Number of critical flight situations
- The safety implications of the option for solving the traffic warning
- Total technical score, taking into account other parameters as well and the expertise of the rater

The number of critical flight situations describes the amount of flight conditions that exceeded the performance limitations. When a critical event such as an Engine Failure occurs, there are certain performance limits that should not be exceeded, as their violation results in a situation where the restoration of the planned flight is very difficult to achieve [19]. There is a one-sided loss in thrust of the aircraft and it slows down. In the case of the two engines, the aircraft will then roll around the longitudinal axis and so sooner or later it will turn over to the side of the failed engine. The warning "bank angle" indicates a cross-slope of ≥ 30° and thus constitutes an important limitation when handling the situation after an Engine Failure [19]. Therefore, the numbers of "Bank Angle" warnings are used as an indicator of the exceedance of performance limitations.

Since the cross-slope increase after a one-sided Engine Failure is a continuous process, bank angle warnings are another important indicator of the reduction of the SA, due to the fact that the pilots did not notice this gradual development into a safety-critical situation. In order to restore a safe flight status, the pilot has to coordinate several activities. On the one hand, the pilot must check the instruments for the system parameters, on the other hand he/she has to identify the failed engine as soon as possible. These two tasks require a great deal of attention as many different parameters play have to be considered [19]. The sound of the acoustic "bank angle" warning thus implies a slower reaction of the pilots, which in turn indicates reduced SA [20].

The second criterion describes the quality of the evasion option selected by the pilot to solve the traffic warning. Three options were available whose suitability to the situation (Engine Failure) varied. The options are denoted by qualitative categories. "Zero" means that the pilot has not committed to any specific alternative option. Since the traffic warning was followed almost immediately by an Engine Failure, restoring the status of the aircraft should come before the processing of traffic warning within the priority list.

The number "One" represents the worst option, a climb to a higher flight level. Due to the Engine Failure and the prevailing imbalance with respect to the thrust there is an increased angle of attack and lack of thrust, which means that the aircraft risks getting into the "stall"-condition which signifies a lack of airflow (that is required to keep the aircraft flying) [19].

The descent at 5000ft is denoted by "Two". Due to the technical situation in and around the aircraft, this is a preferable alternative, however, it is not encouraged from the economical perspective of the airline. The cost of a flight is greatly increased in lower altitude, since the fuel consumption increases considerably [19]. The best option would be the selection of a right-hand turn which is denoted with the number "Three". The difficulty with a right hand turn is the mentioned difficult condition regarding the imbalance in the thrust force. However, these effects are preferable to the climb option as the aircraft is less likely to enter the "stall"-condition [19].

The number of critical flight situations and the quality of selected options for handling the traffic warning were included in the overall assessment of the technical flying performance of pilots by the trainer. For this final score, other information was included as well, e.g. the handling of the situation and the quality of the conducted recovery efforts [13].

IV. RESULTS & DISCUSSION

A. Performance rating

Overall, the expert rating regarding the total performance score was done on a scale of one to five, with a five-for above-average performance and one for a below-average performance. The grade three represents an score in compliance with the specified standard [13]. The sample includes both above-average and below average ratings.

A total of four participants have exceeded the maximum cross-slope, two participants twice. Since the cross-slope increase after the failure of one engine is a continuous process, the warnings are another important indicator of the reduction of the SA.

Repeated occurrences of the bank angle warning, as present in two cases, constitutes an even stronger indicator, as the pilots are not able to restore the aircraft to safe state after the initial first bank angle warning.

The pilots also struggled with the correct assessment of the Engine Failure as evident from the delays in noticing the Engine Failure, which for one of the pilots accumulated to 10 seconds (fig.6). Furthermore, there is a significant negative correlation ($r(9) = -0.65, p < 0.05$) between the total score of pilot performance and the delay in noticing the Engine Failure. Thus, the longer it took for the pilots to notice the Engine Failure, the lower the total score for the scenario.

![Delay in noticing Engine Failure](image-url)

Figure 6. High delays in noticing the Engine Failure due to focus on ATC-display and associated Traffic Warning (TW).
This indicates, that the systems awareness aspect of situation awareness is negatively affected by the ATC-display as the pilots focusing on the ATC-display struggle to notice the change in vital system parameters associated with the Engine Failure [20]. The assessment and correct interpretation of the Engine Failure requires the pilot to redirect their attention from the traffic warning in the ATC-display to the primary flight information sources, which are given by the Primary Flight Display (PFD), the Navigation Display (NAV), the Engine and Warning Display (E/WD) [19].

This conclusion is further supported by the results from the analysis of eye-tracking data. There was a significant negative correlation between the fixation time on the ATC-display and the total rating for the pilot performance ($r(9) = -0.75, p < .05$)

Thus, a high fixation time on the ATC-Display is associated with a lower total score rating by the experts. The comparison of the cumulated fixation time on the Areas of Interest (AoI) over the total duration of the experiment by a pilot with a below average score and the pilot with an above average score illustrates this (fig. 7 & fig. 8).

The figures illustrate the prototypical eye-tracking behavior associated with good scores by the raters and below average scores. Good scores generally correlate negatively with fixation time on the ATC-display after the onset of the traffic warning and the Engine Failure. This is evident from the accumulation of fixation time on the Primary Flight Display (red line, PFD, fig. 8). Low scores are generally associated with the gaze distribution visible in fig. 7 where there is an increase in fixations on the ATC-display directly after the onset of the traffic warning (blue line, fig. 7). Following the priority principle of “fly-the-aircraft-first”, the Engine Failure requires immediate and full attention in order to avoid critical flight situations and to guarantee the safety of the flight.

Thus, these results also indicate that focusing on the ATC-display leads to decrements in task awareness, as the pilots who prioritize managing the traffic conflict do not have the resources to focus on the proper handling of the Engine Failure.

This is an important result, given the small sample of the experiment and has a high significance for the evaluation of the display.

This indicates, that the systems awareness aspect of the total rating for the pilot performance (red line, PFD, fig. 8). Low scores are generally associated with good scores by the raters and below average scores. Good scores generally correlate negatively with fixation time on the ATC-display after the onset of the traffic warning and the Engine Failure. This is evident from the accumulation of fixation time on the Primary Flight Display (red line, PFD, fig. 8). Low scores are generally associated with the gaze distribution visible in fig. 7 where there is an increase in fixations on the ATC-display directly after the onset of the traffic warning (blue line, fig. 7). Following the priority principle of “fly-the-aircraft-first”, the Engine Failure requires immediate and full attention in order to avoid critical flight situations and to guarantee the safety of the flight.

Thus, these results also indicate that focusing on the ATC-display leads to decrements in task awareness, as the pilots who prioritize managing the traffic conflict do not have the resources to focus on the proper handling of the Engine Failure.

This is an important result, given the small sample of the experiment and has a high significance for the evaluation of the display.

V. SUMMARY AND OUTLOOK

Airborne separation constitutes an important concept in the SES in the future. In order to implement this concept, the pilots have to be supported with adequate feedback information regarding their position and potential conflicts. A mock-up display incorporating the core functionality required for airborne separation has been developed. While the fit of the display to the information requirements of the pilots in nominal conditions is given [13], this study has extended the scope to non-nominal and critical situations.

The aim of this study was to evaluate the impact of the ATC-display on the situation awareness (SA) of pilots during the handling of a critical non-nominal situation during flight (Engine Failure). Although the sample size was small (n=10), significant effects were found. There is a significant negative correlation between the delay in noticing the Engine Failure and the total performance rating of the pilots for the scenario. This means that the higher the delay in noticing the Engine Failure, the worse the performance rating and vice versa. Given the high delays in noticing the Engine Failure (up to 10 seconds) this could lead to critical situations in real-life settings.

Furthermore, a significant negative correlation was found between the gaze dwelling time on the ATC-display and the total performance rating. The longer the pilots directed their visual attention on the display, the worse their resulting total performance score was and vice versa.

These effects are related to the negative impact of the ATC-display on the SA of some of the pilots. Some pilots exhibited low task awareness, as they prioritized the solving the traffic
warning over the checking of the primary flight parameters, which are essential to noticing the Engine Failure.

This in turn led to issues in system awareness, as they could not assess the Engine Failure as quickly as possible [9].

An approach for reducing the demands imposed by the ATC-display on the pilot could consist in a relocation of the display in the cockpit. The placement of the ATC-display at the location of the flight pack, at the left-hand side of the pilot, forces the pilots to lose peripheral vision on the other instruments when looking at the ATC-display. This could lead to the omission of visual cues associated with the onset of the Engine Failure that are displayed in the central information console. A possible solution could be the integration of the ATC-display in the NAV display with a switching option (fig. 9), however, this is associated with the risk of mixing of different task content, which may result to further deterioration of awareness task for other tasks [20].

Another option consists in decreasing the fixation time on the ATC-display by optimizing the layout of the display itself, in order to improve the fixation time necessary to extract the relevant information, e.g. [21].

The adequacy of placement and design options of the display will have to be evaluated in future studies with a larger sample to also evaluate whether the effects that were nearly significant will manifest. This is of particular interest as small samples commonly only allow for the detection of very strong effects. Furthermore, as the participants did not agree on the collection of valuable standard demographic data, the privacy concerns of the participants regarding demographic data have to be taken very seriously for the design of future studies, by providing additional assurance to guarantee the anonymity of the data - beyond the standard approach applied in this study.

A. Assessing the change of user concepts due to trigger events

One of the main results of the study is that a high visual focus of the pilots on the ATC-display during traffic warning and an Engine Failure is associated with decrements in performance.

These decrements in performance are associated with decrements in system and task awareness, as the pilots prioritize the task of conflict avoidance over the checking of aircraft flight parameters and therefore do not notice the Engine Failure.

Thus, the pilots focusing on heavily on the ATC-display exhibit a different user concept of the situation than the pilots that shift their attention to the immediate handling of the aircraft. Situational user concepts can be derived from continuous performance data that are indicative of cognitive processes. This can also be applied to eye-tracking data, where the application of non-metrical multidimensional scaling (NMDS, e.g. [22]) to properly transformed eye-tracking data can be used to derive user concepts [23]–[25]. This yields a representation that allows the analyst to look at the structure or underlying dimensions behind humans perception as reflected in a given metric (if the metric fulfills the criteria of reflecting similarities or dissimilarities) [26].

The following NMDS representation (fig. 10) shows changes in user concepts from a study conducted in healthcare. It shows changes in the user concept of anesthetists with technology and team members in a highly critical and dynamic situation during a simulated operation in a high fidelity simulator.

The left-hand NMDS-representation (fig. 10) shows the user concept of an anesthetist before the operating team was complemented with a senior team member and while she was being put under pressure by the team to diagnose the problem. Clusters of information sources are clearly visible (highlighted in red). These clusters reflect the user concept of the anesthetist during that particular time period: as one member of the operating team was pushing the anesthetist to actively diagnose the reason behind a dramatic decline in O₂-saturation of the patient, the gaze of the anesthetist often shifted between this stress-inducing team member and the display showing the vitality parameters of the patient. Furthermore, the gaze of the anesthetist shifted regularly between the body parts of the patient that can be used to determine signs of respiration (head and torso) and the health record of the patient, where the medical history and information of the patient is located which is critical for diagnosis.

However, this active role changed radically when the senior medical staff member entered the operating theatre and took control over the scenario (fig.10). The anesthetist adapted to this change by altering her monitoring strategy to a more heterogeneous and unfocused global monitoring of information.
sources, as evident from the approximate equidistance of the information displays in the NMDS representation. Thus, she exhibited a passive user concept that supplemented the senior medical staff member with information when requested.

Therefore, the NMDS representation provides a very parsimonious representation of coping strategies that human operators apply during work. If several measurements are done, gradually models of coping strategies can be constructed and set into relation to the outcome of the applied user concepts, that is, which user concepts that prove efficient and which user concepts that lead to adverse outcomes.

This approach can also be applied for the eye-tracking data that has been recorded in this study. NMDS representations can be generated both for the time before the traffic warning and Engine Failure and for the period after. These representations would make it possible not only to see the change of visual attention towards a single information source (e.g. the ATC-display) but as an expression of the change in total gaze ecology, that is, which information clusters that the pilots focuses on and in what way these are related to statistical effects on performance ratings.

B. Systematic for including safety-critical events in studies of human-automation interaction

A central challenge for any studies that aim at evaluating new technology for HMI consists in determining the scenarios for which to test the technology. This is especially true if the technology has to conform to the high safety standards of aviation. In order to demonstrate that the new technology does not introduce additional risks to the system, it has to be determined what risks the system is already exposed to. For the human element of the system, this is typically done in the context of Human Reliability Assessment (HRA).

Human Reliability Assessment is an essential part of system reliability assessment. The correct assessment a Human Error Probability depends on the correct modeling of the PSF (Performance Shaping Factors) which modify the Human error potential. Typical PSF formulated in literature can for instance be categorized into (negative) environmental influences, HMI, organization, procedures and task-complexity [27]. Usually event and accident databases, for instance based on the systematic analysis of accident reports, are used in order to identify psf that are associated with breakdowns in performance. In that sense, psf show the vulnerabilities of the system towards contextual conditions.

If an event and accident database has been generated, queries can be conducted targeting usual psf associated with a given task. This makes it possible to generate testing environments from event data.

In the study presented in this paper, a known technical malfunction (Engine Failure) was the psf that was translated into an empirical design. In order to improve the strategic focus of the display evaluation, queries could be done in event and accident databases to explore typical psf associated with performance breakdowns for specific tasks. Thus, if the new technology is applicable for short-haul flights, a query for the associated psf could lead to a multi-factorial research design, which would incorporate further relevant psf for the validation of the technology.

C. Next steps

Further research will demonstrate how changes of clustering of information sources by human operators can be detected dynamically from eye-tracking data and serve as a basis both for the evaluation of the quality of direct human-machine interaction and how these clustering technique can be used for task modeling based on actual performance, as task models are a cornerstone required for HRA.

Furthermore, the clustering technique will be made compatible to second generation HRA-methods [28] in order to improve the compatibility between performance data and accident and event data, as the transfer of knowledge between the domains seems promising.

REFERENCES

[8] N. A. Stanton, “Situation awareness: where have we been, where are we now and where are we going?,” Theoretical Issues in Ergonomics Science, vol. 11, no. 1–2, pp. 1–6, Jan. 2010.


