ADAPTIVE AUTOMATION BASED ON AIR TRAFFIC CONTROLLER DECISION-MAKING

M. IJtsma, C. Borst, G.A. Mercado-Velasco, M. Mulder, M.M. van Paassen
Delft University of Technology
Delft, the Netherlands

Through smart scheduling and triggering of automation support, adaptive automation has the potential to balance air traffic controller workload. The challenge in the design of adaptive automation systems is to decide how and when the automation should provide support. This paper describes the design of a novel mechanism for adaptively invoking automation support. Whereas most adaptive automation support systems are reactive in that they invoke automation support *after* controller workload has increased, the aim of the designed mechanism is to proactively trigger automation support *prior* to workload increases. To do this, the mechanism assesses the quality of air traffic controller's decisions. The designed adaptive automation system has been tested in a human-in-the-loop experiment. Results indicate that the adaptive support helps to increase efficiency and safety as compared to manual control. However, lower triggering thresholds (resulting in more frequent automation intervention) increased the frustration level of participants (as measured with NASA TLX) and decreased acceptance of the support.

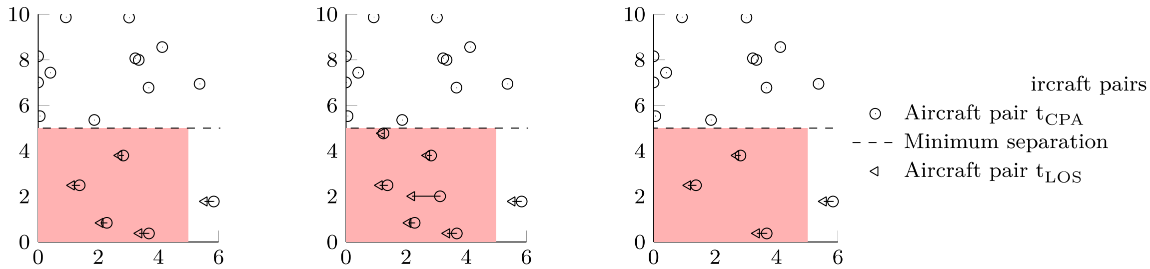
Currently, one of the main limiting factors toward increasing the airspace capacity is the workload of the Air Traffic Controller (ATCo) (Tobaruela et al., 2014). In managing ATCo workload, the concept of adaptive automation has the potential to balance workload between underload and overload. Contrary to static automation, adaptive automation does not operate at a single Level of Automation (LOA), but adapts itself by dynamically switching between multiple LOAs during operations. By smartly trading control between automation and the human controller, higher LOA support can be provided during times of high workload, while lower LOA support or manual control can be offered during low workload conditions.

Two main challenges in the design of adaptive automation systems are to determine *how* and *when* the automation should intervene. Firstly, a wide range of variables can be considered to trigger switches between LOAs. Secondly, whatever the type of triggering variable, different thresholds can be set at which to raise or lower LOAs. The goal here is to find a combination between the triggering variable and its threshold that yields the best possible timing of automation support (Parasuraman, Bahri, Deaton, Morrison, & Barnes, 1992).

Earlier studies have looked at a multitude of metrics for invoking automation support, for example the number of aircraft in an ATC sector (Hilburn, Jorna, Byrne, & Parasuraman, 1997) or electroencephalograms (EEGs) of the operator (Freeman, Mikulka, Scerbo, Prinzel, & Clouatre, 2000). Although these studies have shown promising results, there is room for developing "smarter" mechanisms that can adapt the automation support more reliably and intelligently. Instead of reacting to increased workload conditions, as most existing mechanisms do, novel triggering mechanisms can be designed that focus on preventing higher workload through the use of predictive measures. With these predictive measures, automation support can be invoked *prior* to the actual workload increase.

This study investigates a new combination of triggering variable and thresholds that can potentially prevent workload: decision-quality and the number of allowed "bad" decisions. Several studies have observed that a remarkable amount of traffic complexity is a direct result of the controller's wrong decisions and suboptimal control actions. For example, a study on operational data of Short-Term Conflict Alert (STCA) found that for one out of two STCAs, the controller implemented a resolution that triggered additional STCAs (Lillo et al. 2009). Similarly, a study on sector complexity found that in 120 out of 400 experiment runs participants had introduced one or more additional conflicts as a result of suboptimal control actions (Rahman, Borst, Mulder, & van Paassen, 2015). Detecting and preventing such "self-induced" complexity is expected to reduce workload considerably.

This study explores the use of adaptive automation to improve the quality of the controller's decisions and control actions. It is hypothesized that through the promotion of good decision-making, the adaptive automation system can help to balance the controller's workload. The study includes the design of a decision-based adaptive automation system for Air Traffic Control (ATC) and a subsequent experiment to test the effects of three different triggering thresholds – in terms of allowed "bad" decisions – on the workload, performance and situation awareness of ATCos.



(a) (b) (c)

Figure 1. Three visualizations of time to CPA and separation at CPA. Subfigures (a) the initial condition, (b) a bad decision (increase in number of conflicts) and (c) a good decision (decrease in the number of conflicts).

Adaptive Automation Design

The designed adaptive automation system assesses decision-quality based on an increase or decrease of the number of conflicting aircraft pairs in the airspace. For each aircraft pair in the controller's sector, the projected separation at Closest Point of Approach (CPA) is an indication of whether the two aircraft will lose separation when no action is taken. For conflicting aircraft pairs, the time to Loss of Separation (LOS) indicates the urgency of the conflict. For the adaptive automation system, an aircraft pair is a conflicting pair when their projected separation at CPA is less than 5 NM and the time to CPA is less than 5 minutes.

Figure 1 visualizes the projected separation at CPA and the time to LOS (or time to CPA for non-conflicting pairs) for all aircraft pairs in a particular sector using the Separation Monitor (Irfan, Bull, Clinch, & Pember, 2012). Aircraft pairs within the red shaded area are in conflict. When taking actions to resolve a conflict, the controller's resolution should move the concerned aircraft pair outside this critical area. Additionally, other aircraft pairs should not be moved inside the critical area as a consequence of the controller's action. In this latter case, the controller has induced secondary conflicts, which will then result in self-induced workload because these conflicts also need to be resolved. Thus, a "good" decision should reduce the number of aircraft pairs in the critical area of the separation monitor. Vice versa, a decision and subsequent control action that increases the number of aircraft pairs in the critical area can be considered a "bad" decision.

With the *how* of triggering adaptive automation support being the occurence of bad decisions, one needs to decide what the appropriate threshold is for switching to a higher level of support. This will determine *when* the support is provided: should additional support be provided as soon as one bad decision is being made or should the system be more lenient and only intervene when multiple bad decisions are being made. To explore the effect of this triggering threshold on the system performance, we tested three different thresholds for the adaptive automation in a human-in-the-loop experiment that is described in the next sections.

Although adaptive automation can employ any number of LOAs, we limited the adaptive automation aid here to two LOAs to reduce the risk of over-complicating the design and thereby confounding our experiment. Two LOAs were defined, a low and a high level:

- The low LOA is manual control with short-term conflict warnings and alerts (STCAs). Visual STCA warnings are provided 130 seconds prior to LOS events. Aural and visual STCA alerts are provided 60 seconds prior to LOS.
- At the high LOA, additionally to the STCA warnings and alerts, the automation provides advisories for resolving conflicts between aircraft. The resolution advisories are provided on a management-by-exception basis, i.e., the controllers have a fifteen-second timespan to accept or reject an advisory, after that the advisory is implemented automatically.

The algorithm for automated resolution advisories has been designed to follow ATCo best practices for three types of conflict geometries: overtaking, crossing and reciprocal. In order to limit the scope of the automation algorithm design, the experiment focused on the use of pure heading changes to resolve conflicts: participants could not give altitude or speed commands.

Table 1.

Independent Variables

Condition	Description	Hypothesis
AA1	Triggering support after any single bad decision	Early intervention, lowest workload, highest efficiency & safety
AA2	Triggering support after two bad decisions	Intermediate intervention, lower workload, higher efficiency & safety
AA3	Triggering support after three bad decisions	Late intervention, medium workload, medium efficiency & safety
MAN	Baseline: manual control without automation support.	Highest workload, lowest efficiency & safety

Note. One run was performed with full automation, for which no participants were required. This full automation run was used as a baseline performance condition, in which the automation solution is regarded as the optimal solution. In the following, this condition is referred to as condition AUTO.

Experiment Design

An experiment was conducted to study the effects of different triggering thresholds for invoking adaptive automation on controller workload, automation acceptance, efficiency and safety. In addition, the experiment is a test case for the effectiveness of the decision-based triggering mechanism in balancing workload and preventing self-induced conflicts.

Participants and task. Eighteen participants (2 female, 16 male, average age 27 years) were selected, consisting of students and staff members of the Faculty of Aerospace Engineering. All participants had some experience with ATC, through participating in courses and earlier experiments related to ATC.. Participants could interactively control aircraft using the mouse and keyboard and thus no radiotelephony was needed. Commands were implemented automatically, corresponding to a situation in which a data link is available to communicate vector clearances to aircraft. The task of the participants was to vector (the experiment was limited to heading clearances only, speed and altitude was fixed) aircraft to their designated sector exit waypoints, whilst preventing and resolving any conflicts between aircraft.

Independent and control variables. As the independent variable, the triggering threshold for invoking the adaptive automation was varied. Table 1 lists the different conditions. Control variables include the duration that automation support is active after it is triggered (during pre-experiment testing it was found that an automation duration of 30 seconds provided appropriate support), the expiration time of resolution advisories (set at 15 seconds as determined during pre-experiment testing), and the traffic scenario (to prevent the participants from recognizing conflict geometries from earlier experiment runs, the airspace was between experiment runs).

Procedure. The experiment consisted of the pre-experiment briefing, a training phase and a measurement phase. Breaks were held between the training and measurement phases and halfway the measurement runs. The training phase consisted of eight training runs, which allowed the participants to become familiar with the working of the simulator and the automation support. Here, the adaptive automation system was used as a training tool: when a participant could manage the traffic in these last few runs without triggering the automation support (i.e., making bad decisions), it was an indication that a baseline performance was met and that the participant was sufficiently trained. Finally, the measurement phase consisted of four runs, one for each experiment condition. A Latin square design was used to randomize the conditions and prevent carry-over effects in the measurements.

Scenario. A single traffic scenario was used for the measurement runs. This scenario consisted of three traffic flows, between which conflicts emanated at two intersection points. The intersections were chosen such that in order to prevent self-induced conflicts, any action to resolve conflicts at the first intersection required careful consideration of the traffic at the second intersection. Runs were 15 minutes long and the simulation was run twice as fast as real-time, to make sure participant stayed engaged with the task. At the start the scenario, there was a fade-in period of two minutes, to allow the participant to become familiar with the traffic situation while the traffic density gradually built up. There were nine conflicting aircraft pairs precoded in the traffic scenario.

Dependent measures. During the experiment runs, participants were asked for Instantaneous Self-Assessment (ISA) ratings of their subjective workload once every minute. A metric for airspace complexity, which has been shown to correlate with ATCo workload (d'Engelbronner, Borst, Ellerbroek, van Paassen, & Mulder, 2015), was used to gain more objective insight in the participants' workload. This metric consists of the relative number of heading and speed commands that will create a conflict, averaged over all aircraft in the airspace sector, and provides an indication of the average "solution-space" that is available to the air traffic controller; the higher the number, the smaller the solution-space is and the more likely it is that the participant is experiencing a high workload. Furthermore, performance indicators such as number of conflicts, number of control actions (for this measure the AUTO condition was used as baseline) and number of Short Term Conflict Alerts (STCA) were recorded during the runs. After each run, participants filled out Controller Acceptance Rating Scales (CARS), a NASA-TLX workload form and a short questionnaire.

Results and Discussion

Performance. Figure 2 shows the total number of aircraft pairs in the critical area of the separation monitor, defined by $t_{CPA} < 5$ and $d_{CPA} < 5$. With higher thresholds, there are notably fewer conflicts, which confirms that the automation algorithm has worked as intended. A repeated-measures ANOVA indicated that the difference between conditions is significant ($F(3,45) = 14.478, p < 0.001$). A post-hoc test revealed that the significance is found between conditions AA1 and AA3, conditions AA1 and MAN and conditions AA2 and MAN.

The minimum required number of control actions for the scenario was sixteen (as solved in the AUTO condition). None of the participants was able to solve the scenarios with this minimum number of control actions, requiring on average nine more control actions. The data did show a clear reduction in the number of implemented control actions with lower triggering thresholds, shown in Figure 3. A repeated-measures ANOVA indicated that this reduction is significant ($F (3, 45) = 8.091, p < 0.01$). A post-hoc test revealed that this effect is found between conditions AA1 and AA3 and conditions AA1 and MAN. Although the number of implemented control actions reduced with lower triggering thresholds, the total number of actions (which also includes accepting and rejecting advisories) appeared to be constant or even slightly increasing. A repeated-measures ANOVA indicated that there was no significant difference between the conditions ($F (3, 45) = 0.825, p = 0.487$).

Table 2 shows the number of STCA alerts that were encountered during the experiment runs. The mean ranks clearly show that there were fewer STCA alerts with lower triggering thresholds. Indeed, a Friedman test indicated that there was a significant difference between the conditions ($\chi2\ (3) = 15.593, p = 0.001$). Post hoc analysis with a Wilcoxon sign-ranked test applying a Bonferroni correction indicated that this difference can be found between AA1 and the three other conditions: AA2 ($Z = -2.646, p = 0.008$), AA3 ($Z=-2.887, p=0.004$) and MAN ($Z=-2.673, p= 0.004$); but not between other condition pairs.

Complexity and workload. The means of the ISA rating Z-scores over each experiment run are shown in Figure 3. With stricter automation it seems that the means of the Z-scores are reduced, which indicates a slight decrease in workload with lower triggering thresholds. A repeated-measures ANOVA indicated that this reduction is not significant ($F (2, 30) = 0.642, p = 0.592$). The total scores of the NASA-TLX ratings are shown in Figure 4. The data do not show a clear pattern as a result of the experiment manipulation, which contrasts with the ISA ratings. A Friedman test indicated that there was no significant difference between the conditions ($\chi2(3) = 4.208, p = 0.240$). A breakdown of the various components of the NASA-TLX score revealed that the stricter automation conditions have more workload originating from frustration, both in weighting as well as in rating.

From the questionnaire results, it appeared that frustration mainly originated from occasions in which the automation's advice interfered with the participant's own plans. This resulted in occasional "fights" between

Table 2

Medians and mean rank of the number of STCA alerts

Condition	Median (interquartile range)	Mean rank
AA1	0 (0 to 2)	1.72
AA2	2 (2 to 2)	2.56
AA3	2 (2 to 2)	2.84
MAN	2 (2 to 2)	2.88

Table 3

Medians and mean rank of CARS ratings.

Condition	Median (interquartile range)	Mean rank
AA1	7.5 (6.25 to 8)	1.84
AA2	7 (7 to 9)	2.00
AA3	8 (7 to 9)	2.16

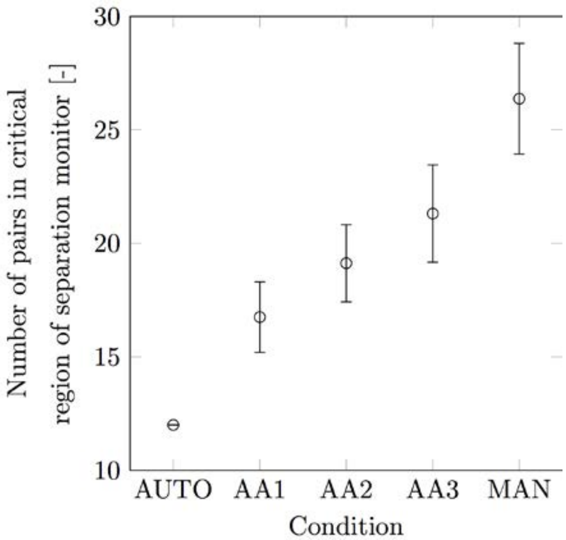


Figure 2. Total number of aircraft pairs in conflict, over duration of experiment run.

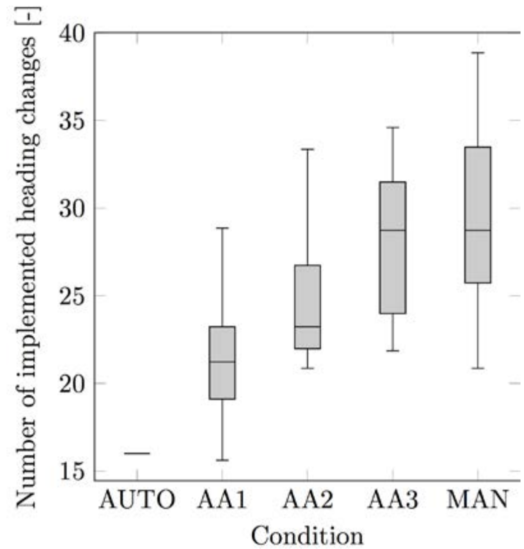


Figure 3. Number of implemented heading changes.

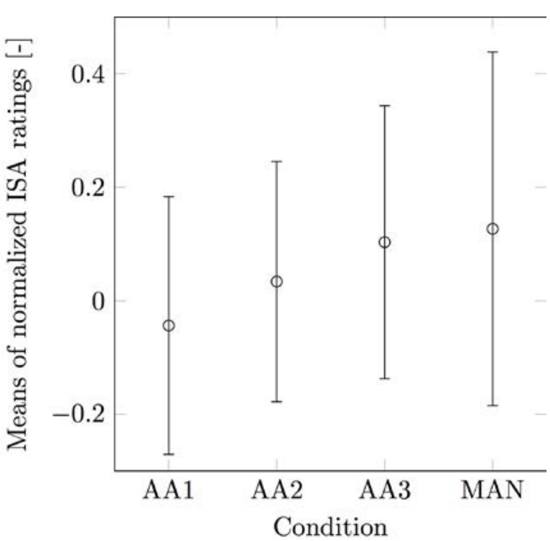


Figure 4. Means of normalized ISA ratings.

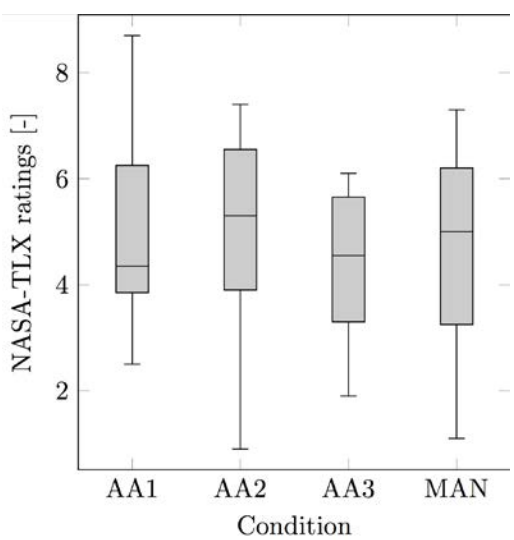


Figure 5. Total NASA-TLX ratings.

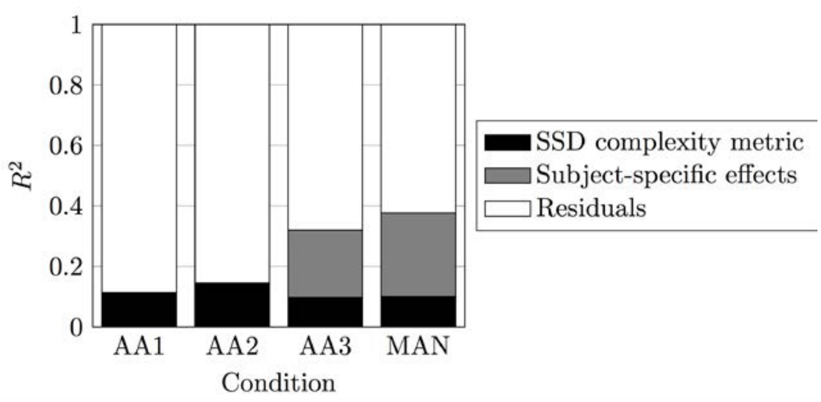


Figure 5. Explained variance of the linear mixed-effects model of airspace complexity

automation and participant. In these situations – even though sometimes automation proposed the better solution – the participant had a more elaborate plan than the automation, in that the participant had thought two or three control actions 'ahead'. In other words, possible workload gains were nullified by the automation's support missing the *intent* information of the human operator.

The complexity metric showed slight variations between conditions, but none were significant. In particular, condition AA3 showed a slightly higher and more variable complexity. A linear mixed-effects model with an intercept and random subject-specific effect was applied to test the correlation between ISA-ratings Z-scores and the complexity metric. Figure 5 shows the amount of variance in ISA-ratings Z-scores that can be explained by complexity and subject-specific effects. Approximately 10% of variance is explained by complexity, which is unaffected by the triggering threshold. For conditions AA3 and MAN, subject-specific effects start to play a more dominant role in the ISA-ratings indicating larger deviations in participants' performance.

Automation acceptance. Table 3 shows the medians and mean ranks of the CARS ratings. With lower triggering thresholds, fewer participants rated the automation at an eight or higher (corresponding to answering the question "is the system satisfactory without improvement" with YES). Instead, participants rated six or below more frequently. The mean ranks decrease with stricter automation, indicating a decrease in automation acceptance with lower thresholds. However, a Friedman test indicated that this difference is not significant ($\chi2(2) = 1.111$, $p = 0.574$).

Conclusions

The results from the experiment indicate that the designed adaptive automation system was effective in improving performance. With stricter automation, there were fewer self-induced conflicts and the overall system was more efficient and safe. However, subjective workload ratings indicated that with stricter automation, frustration of the controllers increased. This meant that although workload reduced on other aspects because of the automation support, workload reductions were nullified by higher frustration of participants. It was observed that automation with lower triggering thresholds reduced the acceptance of the automation support.

In conclusion, finding an optimal combination of triggering variable and threshold for adaptive automation, means that a trade-off must be made between performance, safety and workload. Whereas this research aimed to provide more insight in the design space for triggering mechanisms, future research can further explore the possibilities and effects of different interfaces, triggering variables and thresholds, and LOAs for adaptive automation.

Acknowledgements

The authors would like to thank the participants that volunteered to participate in this study.

References

D'Engelbronner, J.G., Borst, C., Ellerbroek, J., van Paassen, M.M. & Mulder, M. (2015). Solution-Space-Based Analysis of Dynamic Air Traffic Controller Workload. *Journal of Aircraft, 52*(4), 1146-1160.

Hilburn, B., Jorna, P.G., Byrne, E.A. & Parasuraman, R. (1997). The Effect of Adaptive Air Traffic Control (ATC) Decision Aiding on Controller Mental Workload. In Mouloua, M. & Koonce, J. M., (Eds), *Human automation interaction: Research and practice* (pp. 84–91). Mahwah, NJ: Lawrence Erlbaum Associates.

Irfan, M., Bull, M.J., Clinch, A.T. & Pember, S.J. (2012). Air Traffic Control.

Freeman, F.G., Mikulka, P.J., Scerbo, M.W., Prinzel, L.J. & Clouatre, K. (2000). Evaluation of a psychophysiologically controlled adaptive automation system, using performance on a tracking task. *Applied Psychophysiology and Biofeedback*, *25*(2), 103–115.

Lillo, F., Pozzi, S., Tedeschi, A., Ferrara, G., Matrella, G., Lieutaud, F., Lucat, B. & Licu, A. (2009). Coupling and Complexity of Interaction of STCA Networks. In *Proceedings on the EUROCONTROL Conference 8th Innovative Research Workshop & Exhibition* (pp. 1–8). Bretigny-sur-Orge, France: EUROCONTROL.

Parasuraman, R., Bahri, T., Deaton, J.E., Morrison, J.G. & Barnes, M. (1992). Theory and Design of Adaptive Automation in Aviation Systems. *Naval Air Development Center, Washington, DC, Tech. Rep.*

Rahman, S.M.A., Borst, C., Mulder, M. & van Paassen, M.M. (2015). Solution Space Diagram in Conflict Detection Scenarios. In *Jurnal Teknologi (Sciences & Engineering), 75*(8), 53–58.

Tobaruela, G., Schuster, W., Majumdar, A., Ochieng, W.Y., Martinez, L., & Hendrickx, P. (2014). A Method to Estimate Air Traffic Controller Mental Workload Based on Traffic Clearances. *Journal of Air Transport Management, 39*, 59–71.