

PILOT NON-CONFORMANCE TO ALERTING SYSTEM COMMANDS

Amy R. Pritchett
Georgia Institute of Technology
Atlanta, GA

R. John Hansman
Massachusetts Institute of Technology
Cambridge, MA

ABSTRACT

Instances of pilot non-conformance to alerting system commands have been identified in previous studies. Pilot non-conformance changes the final behavior of the system, and therefore may reduce actual performance from that anticipated. A simulator study has examined pilot non-conformance, using the task of collision avoidance during closely spaced parallel approaches as a case study. Consonance between the display and the alerting system was found to significantly improve subject agreement with automatic alerts. Based on these results, a more general discussion of the factors involved in pilot conformance is given, and design guidelines for alerting systems are given.

INTRODUCTION & BACKGROUND

Cockpit alerting systems provide an automatic means to assess hazards, evaluate if an alert is required to cue an action and, with some alerting systems, decide upon methods to prevent or resolve the hazardous condition. Some alerting systems are being given executive roles (i.e. designed with the implicit assumption that their commands will be executed quickly and precisely by the pilot.) For example, with the Traffic alert and Collision Avoidance System (TCAS), conflicting aircraft agree on their respective directions for an avoidance maneuver. The avoidance maneuvers are then calculated with the assumption that one aircraft will climb, the other descend. (RTCA, 1983)

The assumption that pilots will conform is not always valid. The factors involved with non-conformance are not well understood. Studies of currently operational alerting systems have identified non-conformance situations where pilots have delayed in responding to automatic alerts, or have executed different resolutions to the hazard than commanded by the automatic system. For example, pilot questionnaires on the use of TCAS II reported pilots intentionally did not follow commanded avoidance maneuvers in 24.7% of the cases where alerts and commands were given. (Ciemier et al, 1993) A similar tendency was noted with the Ground Proximity Warning System (GPWS). (DeCelles, 1991)

Many alerting systems are intended for conditions where pilots may not have enough information, enough time, or enough free attention to use high performance decision strategies. In these cases, pilot non-conformance may reduce the positive benefit in performance expected from the availability of the automatic commands, and does not allow the alerting system to relieve pilots of the work load of the decision-making tasks.

EXPERIMENTAL EVALUATION OF PILOT CONFORMANCE TO COLLISION AVOIDANCE COMMANDS DURING CLOSELY SPACED PARALLEL APPROACHES

Previous experiments identified potential problems with pilot non-conformance to TCAS commands during the task of collision avoidance during closely spaced parallel approaches. (Pritchett & Hansman, 1997) A need for an alerting system was identified, as shown by a significant decrease in the rate of loss of aircraft separation when TCAS was available. However, the full benefit of the alerting system was not realized due to pilot non-conformance; pilots did not conform to a modified TCAS alerting system in 40% of the approaches. These non-conformance cases resulted in a higher rate of loss of aircraft separation. Subjects appeared to use different criteria than the alerting system for reacting to a possible collision and selecting an avoidance maneuver, but these criteria do not always have adequate performance.

The task of collision avoidance during closely spaced parallel approaches provides a useful case study because of the measurable discrepancy between the types of alerting criteria apparently used by subjects and the higher performance alerting criteria designed for alerting systems. This experiment served as a preliminary investigation of methods of promoting conformance to alerting system commands through the explicit display of the criteria underlying automatically generated alerts. Two criteria were tested: the intuitive -- but lower performance -- Non-Transgression Zone (NTZ) criteria used by subjects in previous experiments, and a higher performance MIT criteria intended for use in alerting systems. In some cases, alert criteria were explicitly displayed to the pilot which

supported the timing of automatically generated alerts, creating consonance between the display and the alerting system. In other cases, the explicitly displayed alert criteria contradicted the timing of the automatically generated alerts, creating dissonance. For comparison, baseline conditions with no automatic alerts and/or no display of alert criteria were also tested.

Each run consisted of three sequential parts:

- *The Flight* The subjects were told they were flying an approach, and should indicate when the aircraft on a parallel approach was blundering towards them, as evidenced by a traffic display. In some cases, automatic alerts were given. Subjects were asked to use their best judgment; conformance to the automatic alerts was not mandated.
- *Certainty and Timeliness Ratings* The traffic display was blanked and subjects were asked to rate their certainty in their decision and, if an automatic alert had been given, the timeliness of the automatic alert.
- *Numerical Simulation of Avoidance Maneuvers* The simulator then projected the resulting miss distance between the intruder and of the subject aircraft resulting from avoidance maneuvers triggered by the subject's reaction, by the NTZ alert criteria, and by the MIT alert criteria. These numerical simulations were transparent to the subject.

The simulator used a Silicon Graphics Indigo 2 workstation for the displays and aircraft dynamics computations. A sidestick was connected for the flying task, and a mouse for the avoidance maneuver selection. Subjects controlled their progress, selecting further practice or commencement of the experiment runs.

In total, twelve subjects participated. Three held Certified Flight Instructor (CFI) ratings; six had some flight experience, and the remaining three were students without flight experience. No subjects were airline flight crew.

Three displays were tested. All were based on a moving map display, with a top-down view, track-up orientation, iconic presentation of the other aircraft's positions, and a text presentation of the other aircraft's altitude. All features of the traffic display were updated once per second, an update rate feasible with current technology.

- *Baseline Display*: Emulated the current TCAS traffic display, with an additional indication of the other aircraft's heading, as shown in Figure 1.
- *NTZ Criteria Display*: Added a graphic indication of a Non-Transgression Zone between the approaches, as shown in Figure 2. This criteria is consistent with subjects' reactions in previous experiments.
- *MIT Criteria Display*: Added a graphic indication of the alert criteria used by the prototype MIT alerting logic to the baseline display, as shown in Figure 3. The shape of this alert criteria changes with each update of information about the other aircraft, making it a potentially distracting feature.

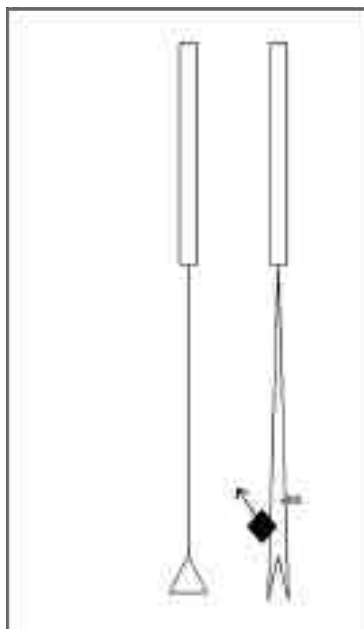


Figure 1 Baseline Traffic Display

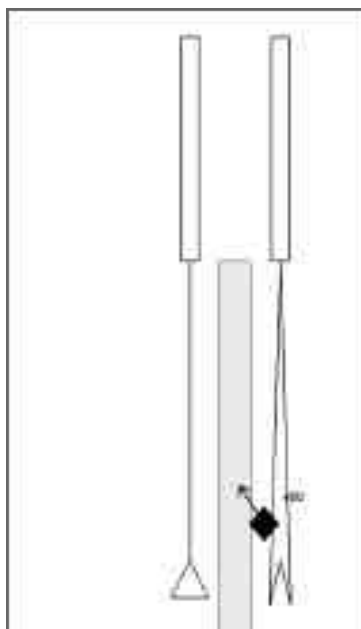


Figure 2 NTZ Criteria Shown on Traffic Display

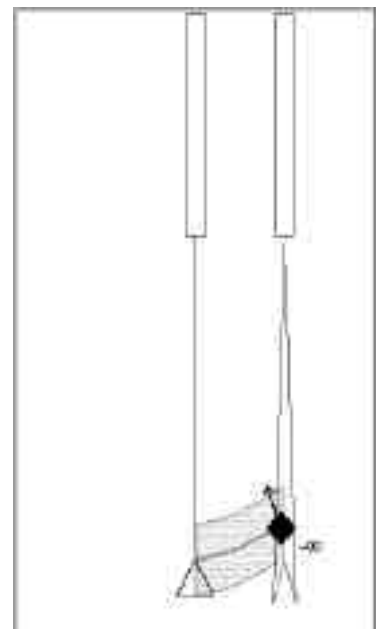


Figure 3 MIT Criteria Shown on Traffic Display

Three different automatic alerting conditions were used in the experiment:

- No automatic alerts were given to the subjects.
- Automatic alerts based on an NTZ criteria were given. This underlying criteria was the same as that shown explicitly on the NTZ Alert Criteria display.
- Automatic alerts based on the MIT prototype alerting logic were given. This underlying criteria was the same as that shown explicitly on the MIT Alert Criteria display.

Four scenarios were flown, in random order, within each test block. These scenarios were designed to test a variety of conditions. Half of the time, the NTZ criteria would generate a false alarm or trigger before the MIT criteria; in the other half of the cases the MIT criteria would trigger before the NTZ criteria.

The test matrix for this experiment was three dimensional, testing all combinations of displays, alerts and traffic conflict scenarios. Altogether, subjects completed 36 experiment runs, allowing for within-subject comparisons. The scenarios were flown in 9 blocks of four, where each block included all the runs for each particular display-workload combination. Paired-comparison statistical tests were used to analyze differences between conditions.

When no automatic alerts were given, the subject's reactions appeared to be strongly correlated with criteria shown explicitly on the display, as measured by the time difference between the subjects' reactions and when each of the alert criteria would have triggered. The mean values of these differences are shown in Figure 4. The average difference between the subject's response time and the time the NTZ criteria triggered is significantly different when the NTZ criteria is shown compared to when the baseline display is shown ($p < 0.01$). A similar effect is found for the MIT alert criteria, with a statistically significant difference between subject's reactions with the baseline display available and with the display of the MIT criteria ($p < 0.05$).

Combined display and automatic alert effects were also found. In general, consonance between the criteria on the display and the automatic alert reduced the difference in time between the subjects' reactions and the time when each type of automatic alerts were given, as shown in Figure 5. Responses to automatic alerts based on the MIT criteria were the quickest when the MIT criteria was explicitly shown on the display. In contrast, subjects' reactions varied the most from the time of the MIT criteria based automatic alerts when the dissonant NTZ criteria was explicitly displayed. However, because subjects' reactions to automatic alerts based on the MIT criteria were variable, statistical significance of these trends can not be proven. Subjects' reactions were significantly closer to automatic alerts based on the NTZ criteria when either alert criteria was explicitly shown on the traffic display. The mean difference in time between the subjects' reactions and the time of NTZ-based automatic alerts drops significantly from the runs with the baseline display ($p < 0.01$ & $p < 0.05$, respectively).

Several statistically significant effects were noted between cases with no automatic alerts and with each of the types of alerts. These effects correlate with subjects' responses to "How did the (automatic) alerts affect your decisions?" These responses indicate a tendency for the decision-making process to be affected in three ways:

- The automatic alerts may have been used as additional input to the subjects' reasoning.
- The automatic alerts may have served as a cue for the subjects to evaluate the situation.
- The automatic alerts may have given the subjects greater trust in their reactions when they coincided.

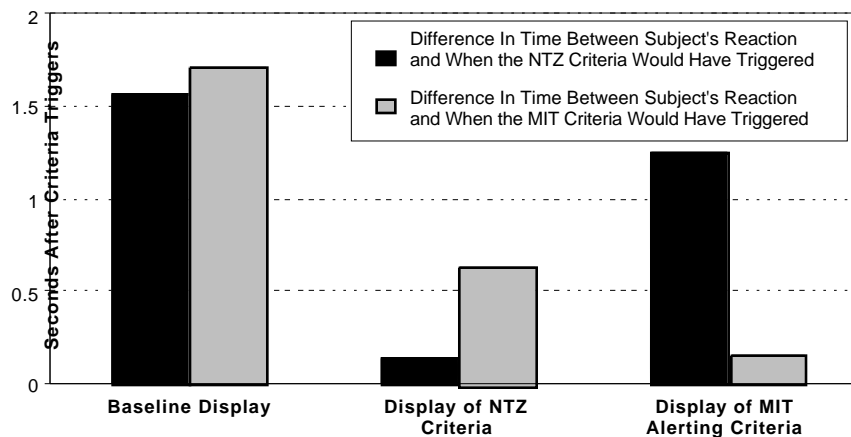


Figure 4. Mean Difference in Time Between When Subject's Reaction and the Two Different Alert Criteria Would Have Triggered, When Automatic Alerts Were Not Shown

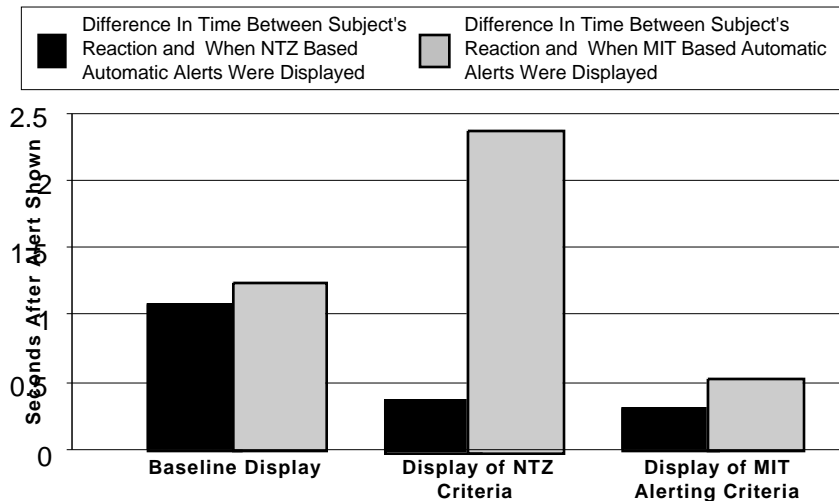


Figure 5. Mean Difference in Time Between When the Two Different Automatic Alerts Were Shown and Subjects Reacted

This results provide insight into the relative effects of automatic alerts and the explicit display of alert criteria, and highlight the importance of consonance between the displays and the automatic alerts. Practical considerations for the task of closely spaced parallel approaches require further study, however. Although benefits were found with the display of the MIT criteria, it did not completely meet the ultimate objective of enabling the subjects to consistently use strategies good enough to ensure collision avoidance. In addition, the display of the MIT criteria -- or a similar criteria -- may not be the final or best display to provide to pilots.

PILOT CONFORMANCE TO AUTOMATICALLY GENERATED COMMANDS

These results raise broader issues about pilot interaction with executive alerting systems. Alerting systems with executive roles are designed with the implicit assumption that pilots will execute the commands quickly and precisely. In cases of non-conformance, pilots instead elect to examine the situation, and execute a resolution to the hazard which may not resemble the commands. Their actions effectively change the role of the alerting system and of the pilots. The alerting system outputs to the pilots are treated as information sources instead as executive commands. The pilots may also consider information not used by the alerting system. The pilots then, through reconciliation of their own decisions and the commands given by the alerting system, decide on a hazard resolution.

The frequency with which pilots perform these re-evaluation and reconciliation processes may be higher than the measured non-conformance rate, which is noted only when the pilots' reactions differed from the alerting system's commands. When pilots follow the alerting system commands, it is unknown whether they are trusting them completely, or taking on the extra workload of evaluation and reconciliation, and then accepting the commands.

Two factors may contribute to pilot non-conformance: pilots may perceive a need to confirm the alerting system's commands, and then the pilots may disagree with the commands and elect to not conform.

The pilots' perceived need to confirm the alerting system's commands may involve several factors, including:

- The pilot may be concerned that the alerting system will fail to act as it should.
- The pilot may feel the alerting system can not consider relevant information or has different objectives.
- The pilot may place greater confidence in their own decisions than in the alerting system's.

Pilots' confirmation of the alerting system's commands may stem from a concern that the alerting system will fail to act as it should. These failures may be of two types, each of which have different implications. The first type of failure occurs if the alerting system either fails to identify a problem, or does not command sufficient action to remedy a problem. The decision of pilots to monitor for these types of failures may have several causes. First, the direct effects of these failures can have very high costs; for example, in the case of a collision avoidance system, this type of failure can have catastrophic results. Second, it may be difficult for pilots to develop confidence in the alerting system. Some alerting systems are designed to monitor for rare events. In this case, pilots will not see the alerting system perform correctly in enough instances to build up trust in the system.

Pilot concern about this type of failure has several implications. First, if the pilots are not confident that the alerting system will generate an alert when required, they may feel compelled to assess the situation regularly independent of the alerting system. Second, if the pilots feel the commanded resolution to the hazard is insufficient, they may feel compelled to make their own decisions about a resolution to the hazard, or they may execute a more severe version of the commanded resolution.

The second type of alerting system failure occurs when the alerting system generates unnecessary or overly conservative commands. When the alerting system is designed to prevent catastrophic events, variance in the sensor measurements and unpredictability in the system dynamics requires its reasoning to be conservative. While a conservative design helps ensure prompt, adequate reactions to dangerous situations, it also increases the frequency of false alarms and excessive commands from the alerting system. Although the alerting system is performing to specifications, false alarms may appear to the pilot as failures of the system. For most tasks, the resultant cost of this type of alerting system failure is comparatively low in any single event. However, this type of failure can have indirect, cumulative effects. First, they may degrade the pilots' trust in all information presented by the alerting system by making the alerting system's functioning appear spurious and unreasonable. Second, past experience with these failures reduces confidence that future alert or command is not also a second type failure.

The second factor in pilots' desire to confirm alerting system commands is a perception that, while the alerting system is functioning to its specifications, these specifications do not include knowledge of all information or have the same objectives as the pilots. For example, pilots indicated in a survey that they sometimes do not follow TCAS commands -- or turn them off -- in conditions where they visual contact with the other aircraft or have knowledge of the other aircraft's intentions through ATC communications. (Ciemier et al, 1993)

The third factor in pilots' desire to confirm an alerting system's commands is the relative confidence they place in the alerting system compared to their own decisions. When pilots have a high confidence in their own reasoning and a low confidence in the alerting system's reasoning, they are more likely to act upon their own reasoning and to confirm automatic commands. With a higher relative confidence in the alerting system, pilots will feel less of a need for confirmation of automatic commands. However, when pilots place equal confidence in the alerting system and their own reasoning, whether this confidence is high or low, the pilots' final actions can not be predicted.

If the pilots do not have confidence in the alerting system, they may attempt to confirm its alerts and commands. This confirmation process alone can cause a delay in the pilots' responses. If the pilots' assessments do not agree with the alerting system's commands, they may additionally execute different resolutions to the hazard. These experiment results illustrate how pilots' strategies may vary from the increasingly sophisticated logic being developed for alerting systems. A resulting mismatch between pilot decisions and alerting system commands may contribute to non-conformance.

Implementation of an executive alerting system is typically expected to increase system performance at some metric, such as an increased ability to resolve traffic conflicts, while eliminating the need for pilots to perform the alerting and decision-making sub-tasks. When pilots instead confirm the alerting system's alerts and decisions, they are effectively changing the role of the alerting system. In doing so, the anticipated benefits of the alerting system may not be fully realized.

First, by giving an executive role to the alerting system, it is expected that the pilots will be relieved of responsibility for some components of the alerting task. However, non-conformance to the alerting system's commands implies that the pilots are still executing some or all of the components of the alerting task. If the alerting system presents sufficient supporting information to make verifying its commands easy, then this workload may be small. However, if the alerting system's commands are difficult to understand, the reconciliation and decision-making tasks may be intensive or the pilots may choose to ignore the alerting system entirely.

In addition, if the pilot does not follow the alerting system immediately and/or does not execute its commands, the resultant system behavior can no longer be described by the pre-determined functioning of the alerting system and the performance of the system can be affected. Unlike the logic underlying the an alerting system, the algorithms pilots will use to formulate their own decisions and to reconcile their decisions with the alerting system's commands can not be predicted with certainty. Involvement of pilots in the decision making removes the ability to analyze the system behavior with the same degree of certainty. This variability may limit the extent to which the performance of the combined pilot-alerting system can be predicted during design and certification.

METHODS OF ENCOURAGING PILOT CONFORMANCE

In cases where pilot non-conformance may have detrimental effects, two possible methods of promoting pilot conformance can be envisioned. First, the alerting system's commands may be made mandatory for the pilots to follow. Although this method promotes pilot conformance, it also raises several issues. In reducing the role of the pilot to an un-informed control actuator, the anticipated benefits of having a pilot in the loop are lost, and the pilots' reactions will be slower. This role of the alerting system may also have difficulty being accepted. Finally, blocking off all relevant information may be impractical; pilots may use information from other sources as a basis for non-conformance.

The second method encourages informed decisions by the pilots by incorporating consonance between the pilot's displays and the alerting system's commands. In situations where the alerting system's commands are valid, this method promotes pilot conformance, while maintaining the benefits of a pilot in the loop in situations where pilots have better reasoning. As such, this method has two design objectives:

- To reduce mismatches between the pilots' decisions and the alerting system's commands, explicitly present the synthesized information implicit in the alerting system's algorithms.
- To make the task of reconciling the pilots' decisions and alerting system commands easier, explicitly present the *alerting thresholds and decision-making objectives* used by the alerting system's algorithms.

For example, the hazard assessment and alerting function implicitly contains intermediate steps. Given the current state of the system, the future behavior is predicted and the hazard level is calculated -- *synthesized information*. This synthesized information is then evaluated to determine the need for an alert; this determination is performed in alerting systems by comparisons to predetermined *alert thresholds*.

The results of the preliminary study described in this paper suggest positive benefits may be found towards encouraging informed pilot conformance by generating consonance between the pilot's displays and the underlying logic of the alerting system used to generate automatic alerts and commands. However, display-alerting system consonance requires that the underlying logic of the alerting system be *communicable* in a quickly understood form. Alerting systems are being proposed for operations which are very complex, require very specific types of performance, or involve many operators. In these situations, an alerting system may be required to reach the desired specifications, but rigorously encouraging pilot conformance through display consonance may not be possible because of the complexity of the alerting system's functionality. Such cases may represent a limit on the use of alerting systems, and on the types of operations which require these sophisticated alerting systems to extend the human pilot's abilities.

ACKNOWLEDGMENTS

This work was supported by the National Aeronautics and Space Administration/ Ames Research Center under grant NAG 2-716. The authors would like to acknowledge the following persons for their contributions to this study: Kevin Corker of NASA Ames for his insights; the graduate students in the MIT Aeronautical Systems Lab for their assistance in preparing the simulator; and the subjects who donated their time to participate in the experiment.

REFERENCES

- Ciemier, C.E., Gambarani, G.P., Sheridan, E.H., Stangle, S.C., Stead, R.P., & Tillotson, D.H. "The Results of the TCAS II Transition Program (TTP)", The ARINC Research Corporation, April, 1993
- DeCelles, J.L. "The Delayed GPWS Response Syndrome", Aviation Research and Education Foundation, Herndon VA, July 1991
- Pritchett, A.R. & Hansman, R.J. "Experimental Studies of Pilot Performance at Collision Avoidance During Closely Spaced Parallel Approaches", Presented at The 9th International Symposium on Aviation Psychology, Columbus OH, April 28-May 1, 1997
- Radio Technical Committee on Aeronautics (RTCA). "Minimum Performance Specifications for TCAS Airborne Equipment", Document No. RTCA/DO-185, Washington DC, September 1983