IMPERFECT AUTOMATION IN AVIATION TRAFFIC ALERTS: A REVIEW OF CONFLICT DETECTION ALGORITHMS AND THEIR IMPLICATIONS FOR HUMAN FACTORS RESEARCH

Lisa C. Thomas, Christopher D. Wickens, and Esa M. Rantanen
Institute of Aviation, Aviation Human Factors Division
University of Illinois, Urbana-Champaign

Automated warning and alert devices such as airborne collision avoidance systems (ACASs) represent a class of automation that is often found to be imperfect. The imperfections can be expressed as the number of false alarms or missed events. Most ACASs are constructed with a bias to prevent misses (which may have catastrophic consequences) and therefore, coupled with a low base-rate of conflict events, create high false alarm rates. In this paper, we review the adequacy of various CDTI warning algorithms that have been proposed and tested in addressing the false alarm issue, and the potential for multiple levels of alerting to mitigate the effects of false alarms on trust and reliance on the CDTI. We suggest new directions for future research, including evaluating the effects of false alarm rates on pilots' use of the CDTI, determining what strategies may enhance pilot tolerance of false alarms, and investigating the use of CDTI in conjunction with air traffic controllers.

INTRODUCTION

Automated warning and alert devices represent classes of automation that are often found to be imperfect (Pritchett, 2001; Stanton, 1994; Sorkin, 1988). The diagnosis of dangerous versus safe conditions is often ambiguous when dealing with uncertain information in a probabilistic world, particularly when the alerting system is forecasting *future* situations in uncertain environments. Such circumstances characterize airborne collision avoidance systems (ACASs) such as the traffic information and collision avoidance system (TCAS), which is in operational use today, or longer range planning systems, such as the Cockpit Display of Traffic Information (CDTI), which is still under development (Johnson, Battiste, & Bochow, 1999; Battiste & Johnson, 2002; Johnson, Jordan, Liao, & Granada, 2003).

In a general sense, the imperfection of any warning system can be expressed in signal detection terms as errors of misses (a true dangerous situation is not detected) or false alarms (a safe situation triggers a warning). Misses and false alarms (FAs) trade off against each other; an extremely sensitive system that almost never misses a potential conflict necessarily has a high false alarm rate (FAR). Because misses have potentially catastrophic consequences to aircraft crews and passengers as well as often negative legal implications to systems manufacturers, most warning systems are constructed with a bias to prevent misses, consequently increasing the FAR (Parasuraman, Hancock, & Olofinboba, 1997; Kuchar, 2001). The FAR can be quite high if the base rate of events to be detected is low (Krois, 1999). However, high FAR has significant negative repercussions too, and may lead to operator mistrust and consequent "disuse" of automation (e.g., Sorkin, 1988; Parasuraman & Riley, 1997). The effect of FAs on human performance is therefore the primary human factors issue associated with automated alerting systems.

However, relatively little research appears to have examined the relative consequences of FAs *versus* misses in influencing human trust and reliance on automated alerts. A few recent studies in the context of automobile warnings suggest that FAs may indeed be more degrading of trust than misses (Gupta, Bisantz, & Singh, 2001; Cotté, Meyer, & Coughlin, 2001). When an alarm is annunciated and directs the attention of the operator away from other tasks, and this alarm turns out to be false, the operator has wasted time and effort in dealing with it and is more likely to lose trust in a system that demands this extra effort. A miss, on the other hand, is by definition not annunciated and therefore the operator has spent no energy in dealing with it and is not likely to even know that a real event exists and was missed. Unless the operator is somehow prompted to determine whether the system missed some critical events, the operator is likely to maintain his/her initial level of trust in the system. It must be noted here that this discussion pertains only to human performance, that is, trust; although misses that remain unknown to the user do not erode trust, they are hardly desirable from the system performance perspective.

THE PROBLEM DEFINED

It should also be noted that the issue in conflict detection algorithms is not so much misses per se as it is delayed issuance of alarms. A system that detects conflicts based on continuously updated information about the location and trajectory of surrounding aircraft will always detect a conflict eventually. If the conflict actually exists, the evidence for it will eventually cross the critical threshold for an alert. We therefore define a miss by the conflict detection system as an alert that is produced at such a late time that the pilot has to take immediate action (if any action can be taken at all) to resolve the conflict.

Figure 1 provides a schematic illustration of the issues in selecting alarm thresholds for a CDTI, plotting the separation between two aircraft as a function of the passage of time during a potential conflict episode. Time 0 is some arbitrary time prior to the point of closest passage between the two aircraft, defined as the look-ahead time (LAT). The solid line shows the nominal prediction, illustrating the steadily decreasing distance to closest passage (DCP), followed by the

increase thereafter. The instant any trajectory crosses the minimum threshold of 3 (or 5) miles of separation, (or any other arbitrary separation distance) a formal *conflict* is defined. In the CDTI, the pilot should be alerted with a sufficient margin of time before conflict occurs so that she or he is able to non-aggressively maneuver in any of the three axes of flight to avoid it.

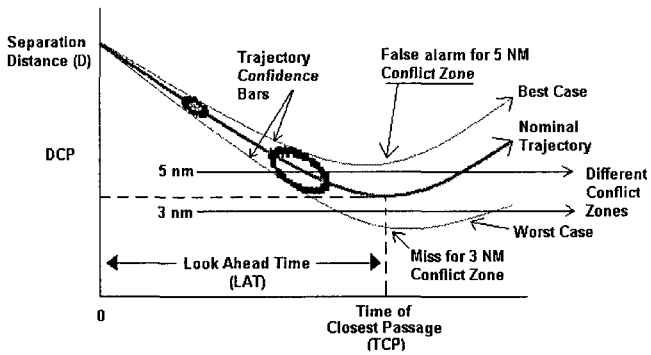

Figure 1. Representation of the evolving space and time aspects of a conflict.

The nominal trajectory represents the expected evolution if neither aircraft alters its speed or heading from that observed at time 0. However, such deterministic behavior is rarely observed. The two "eggs" in Figure 1 represent the anticipated variability in both speed and lateral position around the nominal trajectory (Magill, 1997). This variability increases with increasing time. These "eggs" can be thought of as confidence intervals (e.g., 90%). The lighter lines surrounding the nominal trajectory represent the confidence intervals on lateral separation, in which the "best case" line is the maximum predicted separation distance at closest passage, and the "worst case" line is the minimum predicted distance at closest passage. The growth of uncertainty over time represents the impact of winds or other factors that cannot be predicted with certainty.

Now consider a warning that might be given at time 0, defining an LAT to closest passage or to another event, such as a loss of separation. If, for example, the warning is based on the nominal trajectory for a 5 mile protected zone, and then a "best case" trajectory actually occurs, this would lead to a false alert. On the other hand, if the protected zone is 3 miles, no warning will be given if it is based on the same projected nominal trajectory, and if a "worst case" trajectory actually occurs the system has produced a miss (or at best, a delayed alarm).

The designer must decide whether to issue the warning based upon the nominal trajectory, or some worst case value (90%, 95%, etc.), by balancing the costs of delayed alerts ("misses") versus the costs of false alerts (Yang & Kuchar, 1997). Complicating the design issue further is the LAT. If the trajectory is deterministic, then any LAT will produce equal (and perfect) accuracy. Furthermore, if LAT is very short, accuracy can also be nearly perfect. However the growth of uncertainty with longer LATs, shown by the increasing range of confidence intervals in Figure 1, implies that the longer the LAT, the greater the tradeoff between late alerts and false alerts. Yet, as noted above, the LAT must be great enough to allow the pilot sufficient time to maneuver in a non-aggressive fashion.

As Figure 2 illustrates, the LATs can be categorized into three basic categories according to proposed use of the conflict detection system: Emergency, which generally requires immediate and often constrained actions (e.g., vertical maneuvers only) to resolve the detected conflict; Tactical, which allows the pilot enough time to consider several resolution options and then choose one to implement; and Strategic, which provides a significantly larger amount of time to create very slight modifications of the flight plan in order to avoid conflict with the least impact to the existing flight plan. TCAS' Resolution Advisories operate within Emergency LATs, which are expected to produce the highest hit rate but may still suffer the effects of FAs. The CDTI developed at UIUC uses an algorithm that provides 45 seconds of warning before loss of separation (see Alexander & Wickens, 2002). In both cases, the pilots are expected to take immediate action (usually a time-efficient vertical maneuver) to resolve the imminent conflict.

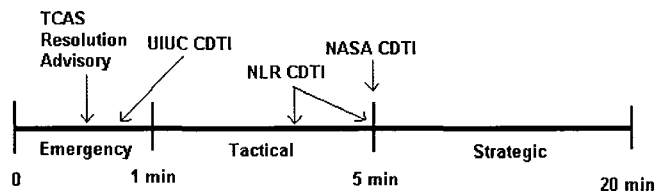

Figure 2. Representation of Look-Ahead Times

NLR and NASA have created CDTIs using algorithms that provide 3 to 5 minute LATs (see Hoekstra & Bussink, 2003; Johnson, Battiste, & Bochow, 1999). Pilots are alerted to a detected conflict, but have several minutes to determine the best course of action to resolve the conflict with minimal impact to flight characteristics such as the time schedule, fuel costs, and physical maneuvers available. When the pilots have more time to create conflict resolution plans, they can utilize maneuvers in any of the three flight dimensions (vertical, lateral, and airspeed), which in turn allows them to create more efficient (albeit more complex) resolutions. With a 3-5 minute LAT, however, the system is subject to both misses and false alarms depending on how accurately the algorithm predicts the trajectory. Algorithms with longer LATs have been evaluated (see Magill, 1997) for strategic flight planning use, but it is likely that with the increase in uncertainty at such long LATs the rate of both false alarms and misses will be prohibitively high and will not produce a useful tool when it comes to planning for projected conflicts.

Thus, as is evident from the analysis, the joint influence of the three parameters (the protected zone size, the LAT, and the assumptions about the growth of uncertainty with time) will affect the sensitivity of discriminating predicted conflicts from non-conflicts, and hence the extent of the tradeoffs between the two negative events of false alerts and misses or late alerts. We will review some empirical work pertaining to these factors and their human performance implications next.

LIMITATIONS OF CURRENT RESEARCH

In the process of gathering information on proposed conflict detection algorithms, we reviewed over 40 articles which contained one or more of the following: (1) a description of an algorithm, (2) analytical validation of an algorithm, or (3) validation of an algorithm by pilot-in-the-loop (PIL) simulations. This review revealed that very few of the algorithms have been validated in realistic free flight simulations with PIL performance data. For the purposes of this paper, we have chosen to illustrate six PIL studies (Table 1), which are representative of the type of studies that have been conducted on the different algorithms mentioned above, along with a breakdown of key characteristics of the studies.

The NASA studies that appear in the first three rows in Table 1 show the range of approaches in implementing and evaluating a CDTI containing a single conflict detection algorithm (Yang & Kuchar's 1997 algorithm), which detects conflicts for 5 NM protected zones. All of these studies used multi-level alerts and reported PIL performance data, but only one (Johnson et al., 1997) varied uncertainty growth parameters and none considered false alarms.

The fourth study (Wing, Barmore, & Krishnamurthy, 2002; see also Wing et al, 2001) is an investigation of a CDTI that incorporates features of two probabilistic algorithms (Yang & Kuchar's 1997 algorithm & NLR algorithm) to detect conflicts using different sources of information (state or intent), while the fifth study (Hoekstra & Bussink, 2003) implemented the NLR algorithm alone. Neither of these studies specified any uncertainty parameters nor manipulated FAR as an independent variable.

The final set of studies (from the University of Illinois) used a non-probabilistic algorithm developed at the University. These experiments are the only ones discovered that manipulated the protected zone and lateral uncertainty as independent variables. In addition, only Wickens, Gempler, and Morphew (2000) involved misses as an experimental variable.

The reviewed research shows that pilots can use ACAS technology to successfully aid in-flight separation and also that pilots report high subjective approval ratings of the availability of CDTI information. However, there are several major areas of research that have not yet been addressed by simulations of CDTI and conflict detection algorithms. The simulation-based validations reviewed here tended to be limited in scope with respect to consideration of a variety of conflict situations, alerting and traffic display characteristics, and conflict detection capabilities. Furthermore, the sample size has been generally small, potentially resulting in lack of statistical power in making strong general conclusions. While there has been some discussion of FA and delayed alarm rates (Yang & Kuchar, 1997; Kuchar, 2001; Hoekstra & Bussink, 2003), we have found only one study that has addressed "missed" conflicts (or delayed alerts) as a variable (Wickens et al., 2000), and none that have investigated FA effects on pilot preference and trust directly, much less manipulated FAR (as dictated by alarm threshold or LAT) as variables in a study.

Our analysis of the larger set of algorithm studies (from which Table 1 is derived) reveals that the three most critical variables for affecting the balance of FAs versus late alarms (depicted in Figure 1) are (1) LAT, as a longer LAT produces more FAs, (2) the size of the minimum separation boundary, where the larger the boundary, the more FAs produced, and (3) the assumptions that are made about the growth of uncertainty (see also Magill, 1997). Yet Table 1 reveals little consistency across these variables between studies (see in particular the "Uncertainty Growth Parameters" column), and no systematic manipulation of them in the PIL studies.

Some prior research has suggested that a key feature for mitigating the negative consequences of false or "nuisance" alarms is the capability of providing graded levels of alerting, such that the user would be less distressed if an alert at the lowest level of predicted danger proves to be incorrect (Sorkin, Kantowitz, & Kantowitz, 1988; St. John & Manes, 2002). As shown in column 5, the six studies described in Table 1 used multiple levels of alerting (between 2 and 5 alerting levels) to indicate the relative urgency of the alarm. However, none of these studies directly compared different numbers of alerting levels to each other within a single study. In sum, we have found no consistency in the implementation of the multiple level alerts across studies, and have found no studies that have investigated the *optimal* number of alert levels.

FUTURE RESEARCH ISSUES

Based on our review of ACAS literature, we will make several recommendations for future research. First, since it probably is not possible to determine a fixed threshold for an "acceptable" FAR due to the complexity of constructs such as trust and workload and the innumerable factors affecting them (see Parasuraman & Riley, 1997), as well as the diversity of the operational environments and settings in which alerting systems are used, research focus should be on the operators' tolerance for the inevitably high FAR and the role of training and system design in improving that tolerance. The FA tolerance could be increased by improving pilots' general awareness of the traffic situation on one hand, and the accuracy of their mental model of the algorithms of the collision alert system on the other. Second, since unaided humans are notoriously bad at estimating probabilities and making judgments based on probabilistic information (Kahneman & Tversky, 1984), the operators' performance could be improved by displaying probabilistic information to them in a form that is easy to perceive and understand and that can be readily used in their tasks, such as in the form of graded levels of alerting. Finally, the role of CDTI in the free flight environment will be drastically different from that of TCAS. It is hence crucial to examine its use in concurrence with ATC procedures and controllers' tasks. The congruence of planning and conflict detection algorithms of CDTI and ATC automation tools will have a substantial impact on the performance of both pilots and controllers.

SUMMARY AND CONCLUSION

In this paper, we have considered the adequacy of various CDTI warning algorithms that have been proposed and tested

in addressing the FA issue. We also noted the important distinction between testing the algorithm (software) itself, and testing the pilot's use of the algorithm in a conflict avoidance PIL simulation or in operational context. Finally, we described a framework for addressing the FA issue from the perspective of the pilot's decision-making when interacting with CDTIs.

It is apparent that the present research findings on the effects of FAs on human trust, workload, and performance in conjunction with ACAS technology must be considered in the light of the operational environment in which the systems are to be used. For example, the envisioned use of CDTI as a strategic planning tool with relatively long LAT will likely result in very different pilot responses to FAs than what has been found in immediate conflict avoidance settings. Such complex environments, however, place substantial demands to the design of experiments, which must manipulate or control all the relevant independent variables and accurately measure the dependent variables. In the latter category, what ultimately matters most is human performance, posing further challenges to the characterization and measurement of apposite parameters.

ACKNOWLEDGMENTS

This research was supported in part by a grant from the Federal Aviation Administration (no. 02-G-032) to the third author. The FAA technical monitor for this grant is Dr. William K. Krebs, AAR-100. The views expressed in this paper are those of the authors and do not necessarily reflect those of the FAA

REFERENCES

Alexander, A., & Wickens, C. D. (2001) *Cockpit display of traffic information: The effects of traffic load, dimensionality, and vertical profile orientation* (ARL-01-17/NASA-01-2), UIUC, Aviation Research Lab.

Battiste, V., & Johnson, N. (2002) An operational evaluation of ADS-B and CDTI during airport surface and final approach operations. *Proc. 46th Annual HFES Meeting.*

Cotté, N., Meyer, J., & Coughlin, J. F. (2001). Older and younger drivers reliance on collision warning systems. *Proc. 45th Annual HFES Meeting.*

DiMeo, K, Sollenberger, R., Kopardekar, P., Lozito, S., Mackintosh, M-A., Cardosi, K., & McCloy, T. (2002) *Air-ground integration experiment* (DOT/FAA/CT-TN02/06).

Gupta, N., Bisantz, A. M., & Singh, T. (2001). Investigation of factors affecting driver performance using adverse condition warning systems. *Proc. 45th Annual HFES Meeting.*

Hoekstra, J., & Bussink, F. (2003) *Free flight: How low can you go?* DASC Report.

Johnson, W. W., Battiste, V., Delzell, S., Holland, S., Belcher, S., & Jordan, K. (1997) Development and demonstration of a prototype free flight cockpit display of traffic information. *Proc. 1997 SAE/AIAA World Aviation Congress.*

Johnson, W. W., Battiste, V., & Bochow, S. (1999) A cockpit display designed to enable limited flight deck separation responsibility. *Proc. 1999 World Aviation Conference.*

Johnson, W. W., Jordan, K., Liao, M., & Granada, S. (2003) Sensitivity and bias in searches of cockpit display of traffic information utilizing highlighting/lowlighting. *Proc. 12th International Symposium on Aviation Psychology.*

Kahneman, D., & Tversky, A. (1984) Choices, values, and frames. *American Psychologist, 39*, 341-350.

Krois, P. (1999). *Alerting systems and how to address the lack of base rate information.* Unpublished manuscript.

Kuchar, J. K. (2001) Managing uncertainty in decision-aiding and alerting system design. *Proc. 6th CNS/ATM Conference.* Taipei, Taiwan.

Magill, S. A. N. (1997). Trajectory predictability and frequency of conflict avoiding action. *Proc. CEAS Free Flight Conference.* DERA.

Parasuraman, R., Hancock P. A., & Olofinboba, O. (1997). Alarm effectiveness in driver-centred collision-warning systems. *Ergonomics, 40*(3), 390-399.

Parasuraman, R., & Riley, V. (1997). Humans and automation: use, misuse, disuse, abuse. *Human Factors. 39*(2), 230-253.

Pritchett, A. (2001) Reviewing the role of cockpit alerting systems. *Human Factors and Aerospace Safety, 1*(1), 5-38.

St. John, M., & Manes, D.I. (2002) Making unreliable automation useful. *Proc. 46th Annual HFES Meeting.*

Sorkin, R. D. (1988). Why are people turning off our alarms? *Journal of Acoustical Society of America, 84*, 1107-1108.

Sorkin, R. D., Kantowitz, B. H., & Kantowitz, S. C. (1988). Likelihood alarm displays. *Human Factors, 30*(4), 445-459.

Stanton, N. (1994). *The human factors of alarm design.* Bristol, PA: Taylor and Francis.

Wickens, C. D., Gempler, K., & Morphew, M. E. (2000). Workload and reliability of predictor displays in aircraft traffic avoidance. *Transportation Human Factors Journal, 2*(2), 99-126.

Wickens, C. D., Xu, X., Helleberg, J, & Marsh, R. (2001) Pilot visual workload and task management in free flight: A model of visual scanning. *Proc. 11th International Symposium on Aviation Psychology.*

Wing, Adams, Duley, Legan, Barmore, & Moses, (2001) "Airborne use of traffic intent information in a distributed air-ground traffic management concept: Experiment design and preliminary results." NASA/TM-2001-211254

Wing, D. J., Barmore, B. E, & Krishnamurthy, K. (2002). *Use of traffic intent information by autonomous aircraft in constrained operations.* AIAA.

Yang, L. C. & Kuchar, J. K. (1997). Prototype conflict alerting system for free flight. *Journal of Guidance, Control, and Dynamics, 20*(4), 768-773.

Table 1. *Summary of six pilot-in-the-loop studies that incorporated different conflict detection algorithms into CDTIs for use in free flig* *imulations.*

Reference	Algorithm type	Attributes varied?	Uncertainty growth parameters	Multi-level alerts	t-in-loop mary	False Alert discussion?
Johnson, Battiste, & Bochow (1999)	Yang & Kuchar algorithm used in the CDTI	Protected Zone (PZ) = 5 NM, +/-1000 ft; intent info was either shared or unshared; 2 types of conflicts = true or "apparent"; initial altitudes were either separated or not separated	No error factors	3 levels: 1. Moderate threat in the future, or low immediate threat 2. Moderate immediate threat, high threat in the future 3. High near-future threat	ews flew either or low flight-deck onsibility arios; positive ect responses to lability of info, rs, display options	
Johnson, Battiste, Delzell, Holland, Belcher, & Jordan (1997)	Yang & Kuchar algorithm used in the CDTI	PZ = 5 NM, +/- 1000 ft vertical	Along-track error standard deviation (s.d.) of 15 knots, cross-track error s.d. of 1 NM; probabilities of turns or altitude changes included (4 turns/hr, 4 alt changes/hr)	5 levels: 1. Low: no alert/info only 2. Moderate: no alert info only 3. High: alert 4. Very high + Traffic advisory 5. Very high + Resolution advisory	rews flew either anced (included lution tools) or c CDTI condition; ew 8 scenarios w/ ast one conflict	Authors state explicitly that no false alarms were programmed into the scenarios
DiMeo, Sollenberger, Kopardekar, Lozito, Mackintosh, Cardosi, & McCloy (2002)	used Yang & Kuchar (1997) logic, overlaid on TCAS	PZ = 5 NM, +/- 1000 or +/- 2000 ft as appropriate (if ownship alt. > 29,500 ft, vertical separation is +/- 2000 ft)	Uncertainty parameters not specified (assumed to be same as described in Yang & Kuchar, 1997)	4 stages: • First two levels are CDTI-based • Last two levels are TCAS Traffic Advisory and Resolution Advisory	sets of /copilot flew 4 rent conditions, 3 s each	
Wing, Barmore, & Krishnamurthy (2002); Wing, Adams, Duley, Legan, Barmore, & Moses (2001)	Combination of Yang & Kuchar logic (state + intent) plus NLR/Hoekstra logic (state only)	PZ = 5 NM, +/-1000 ft; LAT = 5 minutes for state info, 8 minutes for intent info; 3 categories of conflicts = state-only, intent-only, and "blunder"	Uncertainty parameters not specified (assumed to be same as described in Yang & Kuchar, 1997)	3 alerting stages: 1. Inform pilot of potential conflict, no action required 2. Conflict detected, action by OS required 3. LOS has occurred, immediate action by OS required	ommercial pilots 3 en-route trials of 4 scenarios ical vs. strategic ation mode, low igh ops plexity)	
Hoekstra & Bussink (2003)	NLR state-based conflict detection algorithm	PZ = +/- 1000 ft, 5 NM en-route, 3 NM near terminal	Uncertainty parameters not specified - traffic generated by NLR software "Traffic Manager"	2 levels of urgency: 1. LOS occurs between 5-3 min before closest point of approach 2. LOS occurs less than 3 min to closest point	t sets of pilots en-route descent terminal oach trials with vs. high traffic, managed vs. free t terminal oach	False alarms are assumed to be minimized by setting LAT to 5 min (see Magill, 1997)
Wickens, Gempler, & Morphew (2002); Wickens, Xu, Helleberg, & Marsh (2001); Alexander & Wickens (2001)	Geometric linear extrapolation of vertical and curvilinear extrapolation of lateral (rate of turn).	PZ = lateral and vertical dimensions varied between experiments.	Wickens, Gempler, and Morphew introduced lateral uncertainty. Not quantitatively varied.	Two levels: 1. Predicted loss of separation within 45 seconds 2. Actual loss of separation.	s found useful predictor of e trajectory, and lict specific mation about t of closest sage.	Wickens, Gempler, & Morphew produced conflict misses; the only such study to do so.