

**DEVELOPING AND VALIDATING HUMAN FACTORS CERTIFICATION
CRITERIA FOR COCKPIT DISPLAYS OF TRAFFIC INFORMATION AVIONICS**

Federal Aviation Administration Research Grant No. 02-G-032

FINAL REPORT, AUGUST 23, 2002—DECEMBER 31, 2003

Esa M. Rantanen, Christopher D. Wickens
Xidong Xu, and Lisa C. Thomas
Institute of Aviation, Aviation Human Factors Division
University of Illinois at Urbana-Champaign
Willard Airport, Savoy, IL

January 6, 2004

EXECUTIVE SUMMARY

This report examines issues associated with traffic awareness and conflict detection and resolution applications, particularly the predictive and planning support offered by the Cockpit Display of Traffic Information (CDTI). However, inferences about future trajectories of ownship and traffic made by the CDTI logic will, by nature, be imperfect, leading to an apparent loss of reliability of the prediction/planning tools. This loss of reliability and its impact on pilots' behavior and cognition is the focus of this report. Three specific aspects of the CDTI are distinguished: Alerts, visual symbology, and planning tools.

The report draws from general human factors literature. The tradeoff between false alarms (FAs) and misses (or late alerts), how these tradeoffs affect trust and reliance in the alerting system, and how low base-rate events necessarily create a high false alarm rate (FAR), if high-cost misses are to be minimized, are examined. The report concludes with general recommendations emerging from the literature on solutions to mitigate the problems of FAs, with an emphasis on multi-level alerts. Also literature that has addressed the FA-late alert tradeoff and evaluated the CDTI alerts with pilot-in-the-loop (PIL) data was reviewed. Lessons learned from implementation of Traffic Alert and Collision Avoidance System (TCAS) were assessed and results of the analysis of ASRS reports from TCAS system are reported. From this review, guidance will emerge regarding what aspects of conflict situations make detection and response difficult from a pilot's point of view, and therefore should be targeted for inclusion in certification and approval. Also documents by RTCA and FAA were carefully reviewed for guidelines on alerts as well as for their origins, human factors data to back them up, and potentially overlooked issues. Finally, a full taxonomy of possible conflict scenarios was developed, from which—based of the literature reviewed—a subset of scenarios was detailed for bench testing. Finally, recommendations for evaluation of CDTIs during bench testing are offered.

Each section in this report is concluded with a summary, and where appropriate, a set of recommendations. Based on the review of general human factors literature on automated alerting systems in section 2, **Non-CDTI Alerting and Alarm Studies**, the following conclusions are drawn:

- CDTI Alerts are necessarily imperfect, given that the future of a probabilistic airspace is to be predicted.
- These alerts should produce more FAs than misses (late alerts) because of low base rate and the high cost of misses.
- While humans may lose trust with imperfection, they can learn to calibrate their reliance to the actual reliability of the system.
- Misses (late alerts) probably have a different consequence on trust and reliance than FAs, although more research is needed to understand this tradeoff.
- Human trust and reliance problems may be mitigated by training, status displays, and likelihood alarms although the optimum number of alarm states for likelihood alarms is not well established and research should address this issue.

A review of airborne conflict avoidance systems in section 3, **CDTI Conflict Detection Algorithms and Alert Properties**, yields these conclusions:

- Alert system performance is particularly degraded by three factors. Given the preference to create more FAs than late alerts, the following three factors tend to increase false alert rate (FAR): (1) Look-ahead time (LAT), as a longer LAT produces more FAs, (2) the size of the minimum separation boundary (PAZ), where the larger the boundary, the more FAs produced, and (3) the assumptions that are made about the growth of uncertainty.
- Pilot-in-the-loop research is strongly needed to evaluate the consequences to pilot performance of imperfect CDTI alerts as the three parameters of LAT, PAZ and uncertainty growth are manipulated, to assess the optimum value of these to support pilot trust, reliance and maneuvering options.

In section 4, **Airborne Alerting and Traffic Systems: Guidelines and Studies**, review of the lessons learned from TCAS development and implementation show that feedback on TCAS use and problems encountered—although generally qualitative rather than quantitative—have been adequately incorporated into system upgrades. Analysis of ASRS TCAS reports indicates that such reports are equally related to FAs as to misses, reflecting a decrease of FA-related reports since the early 1990s. There are also fewer negative comments about FAs than might be expected, suggesting that there is increasingly good pilot understanding of the causes and consequences of such events.

Negative consequences of FAs on operator trust or late alerts on restricting maneuver opportunities were not addressed in any of the RTCA and FAA documents reviewed. None of these materials provided source documents for assumptions on pilot response time, reasons or empirical basis for chosen reliability level criteria, or the source and magnitude of variability of future traffic trajectory, which would create alert imperfections

An exhaustive review of existing CDTI guidelines—most for near-term applications—affords several conclusions and recommendations:

- Several visual functions of CDTIs such as predictive leader lines or automatic decluttering (or restoring) provide implicit visual alerting and their behavior should be harmonized with auditory alerts.
- The role of time vs. space in traffic alerting should be addressed, since these are not always perfectly correlated in their contributions to risk.
- More focus should be placed on vertical representation of traffic.
- Failure modes analysis must be applied to CDTI systems to identify implications of non-transponder aircraft.

In section 5, **CDTI Non-Alerting Studies**, review of empirical pilot-in-the-loop (PIL) CDTI studies reveals the following conclusions, some drawn with less certainty than others, because they are based on sparse data:

- Pilots increase their preference for vertical maneuvers under increased time pressure because of both the increased time efficiency and their lower cognitive complexity. Presentation of the vertical axis graphically and unambiguously in a co-planar display also emphasizes the tendency to maneuver vertically.

- Pilots tend to avoid airspeed adjustments for conflict avoidance.
- Pilots tend to avoid complex multi-axis avoidance maneuvers.
- Vertical maneuvers tend to be less “safe” than lateral ones, as safety is operationally defined by avoiding a state of predicted conflict.
- Climbing/descending traffic challenges pilot visualization and safe maneuvering.
- Safety is decreased with a 3D display relative to a 2D co-planar display containing a vertical situation display (but not relative to a 2D plan-view display).
- Safety, as operationally defined by conflict understanding, is decreased with long TCPA, with differing relative speeds between ownship and traffic, and with non-orthogonal lateral geometry.
- Maneuver safety is challenged when the airspace is more constrained, due to the presence of other hazards (additional traffic, weather).
- There is a tradeoff between time versus fuel efficiency in the choice between vertical maneuvers (which favors time) and lateral maneuvers (which favors fuel efficiency).
- A final conclusion is that the parameter space for understanding “what makes conflict detection and avoidance difficult” remains less than fully populated with empirical data, particularly in high fidelity PIL simulations.

The above conclusions allow for identification of certain features that should be incorporated to challenge pilot use of alert-equipped CDTI’s to the utmost, so that assessment can be made under “worst case” scenarios. In addition to the detailed suggestions provided in section 6, **CDTI Evaluation Scenarios**, the following recommendations can be made for scenario development in general:

- The suggested scenarios should be considered as templates that guide consideration and combinations of various parameters; actual parameter values and constraints in the actual evaluation settings warrant careful consideration.
- The temporal characteristics of scenarios warrant careful consideration as well to determine the desired frequency and temporal spacing between multiple conflicts.
- Inclusion of unequipped, transponder-off or otherwise defined “distracter” aircraft in the scenarios must be determined separately.
- Since it is utterly possible that parameters that may not have been considered are nevertheless present in the actual evaluation environments. it is very important that all available data from the evaluations are collected and archived for retrospective analysis

Finally, in section 7, **Performance Measures in CDTI Literature**, general recommendations for measurement of both system and human performance can be made:

- All measurement should be preceded by a thorough task analysis.
- All measurement instruments, including human raters, should be properly calibrated.
- Explicit determination of the scales of measures must be done to ensure proper statistical treatment of the data and apposite conclusions drawn from them.

- Thorough review of the research literature establishing relationships between directly measurable task variables and covert cognitive constructs of interest (e.g., workload) must be done before making inferences about the latter on the basis of the former.
- Assumptions should be checked for violations and possible biases they may produce.

TABLE OF CONTENTS

Executive Summary	i
Table of contents.....	v
Acronyms and Definitions.....	ix
1. Background.....	1
1.1. Historical Overview	1
1.2. Scope of the Report	1
1.2. Outline of the Report.....	3
1.2.1. 2. Non-CDTI Alerting and Alarm Studies.....	3
1.2.2. 3. CDTI Conflict Detection Algorithms and Alert Properties	3
1.2.3. 4. Airborne Alerting and Traffic Systems: Guidelines and Studies	3
1.2.4. 5. CDTI Non-Alerting Studies.....	4
1.2.4. 6. CDTI Evaluation Scenarios.....	4
1.2.6. 7. Performance Measures in CDTI Literature.....	4
2. Non-CDTI Alerting and Alarm Studies	5
2.1. Overview.....	5
2.2. Discrete Alerting in CDTI	5
2.3. Imperfect Automation Research	7
2.3.1. The Response Criterion Must Be Low.....	7
2.3.2. Humans Lose Trust in and Reliance on Imperfect Automation.....	7
2.3.3. Humans Can Become Calibrated to Automation Imperfection.....	7
2.3.4. Trust and Reliance are Different Concepts.....	8
2.3.5. Misses May Effect Trust Differently From False Alarms.....	8
2.3.6. Imperfect Automation Costs May Be Mitigated by Display Features.....	8
2.4. Conclusions.....	9
3. CDTI Conflict Detection Algorithms and Alert Properties.....	11
3.1. Overview.....	11
3.2. Definitions	11
3.2.1. Conflict	11

3.2.2. Near Midair Collision (NMAC).....	11
3.2.3. Collision.....	11
3.2.4. Conflict Detection.....	12
3.2.5. Conflict Resolution.....	13
3.3. Algorithms.....	13
3.3.1. Algorithm Type.....	13
3.3.2. Information Used by the Algorithm.....	14
3.3.3. Conflict Criteria.....	14
3.3.4. Alerting Levels and Symbology.....	17
3.4. Limitations in the Reviewed Literature.....	18
3.5. Conclusions.....	19
4. Airborne Alerting and Traffic Systems: Guidelines and Studies.....	21
4.1. Overview.....	21
4.2. TCAS Lessons Learned.....	21
4.2.1. Information from TCAS Developers.....	21
4.2.2. Review of TCAS Literature.....	23
4.3. ASRS TCAS Report Analysis.....	24
4.3.1. General.....	24
4.3.2. TCAS Development Timeline.....	25
4.3.3. TCAS Transition Program.....	26
4.3.4. Limitations of Data Extracted from ASRS Reports.....	27
4.3.5. Method.....	27
4.3.6. Results: Qualitative Observations.....	28
4.3.6.1. Pilot-Related:.....	28
4.3.6.2. ATC-related:.....	29
4.3.6.3. System-related:.....	29
4.3.6.4. ATC-Air Crew Interaction:.....	30
4.3.6.5. Air Crew and ATC Conflict Assessment:.....	30
4.3.7. Results: ASRS Report Analysis.....	30
4.2.8. Summary.....	34
4.4. Documents and Guidelines Concerning Alerts.....	36
4.4.1. Assumptions of FAA Concern for CDTI Approval.....	36

4.4.2. Review of Non-CDTI Documents.....	36
4.4.2.1. FAA AC-20-131A, TCAS Certification. (A).....	36
4.4.2.2. MITRE Safety study: McLaughlin, 1997. (B).....	36
4.4.2.3. MITRE safety study: Chemiavsky, 1997. (C).....	37
4.4.2.4. FAA TSO-C147, Technical Standard for Traffic Advisory System Airborne Equipment. (D)	37
4.4.2.5 EGPWS Certification document N8110.64. (E).....	37
4.4.2.6 EGPWS Advisory Circular 25-23. (F).....	37
4.4.2.7 Honeywell EGPWS flight test document. (G)	37
4.4.2.8 Krois (1999b) (H)	38
4.4.3. Review on CDTI-Specific Documents, with Information Pertaining to Alerts. ...	38
4.4.3.1. RTCA DO-243 (1998). (I)	38
4.4.3.2. RTCA DO-239 (1997). (J)	38
4.4.3.3. SAE-G-10. (K).....	38
4.4.3.4. Draft FAA Advisory Circular ADS-B. (L)	39
4.4.3.5., Uhlarik, J., & Comerford, D. A. (2003) (M).....	39
4.5. CDTI Non-Alerting Guidelines	39
4.5.1. Overview.....	39
4.5.2. Visual Alerting	40
4.5.3. Traffic Highlighting and Decluttering.....	40
4.5.4. Vertical Representation.....	40
4.5.5. Flight Deck Integration.....	41
4.5.6. Failure Modes Analysis	41
4.6. Conclusions and Recommendations.....	41
5. CDTI Non-Alereting Studies.....	43
5.1. Overview.....	43
5.2. Time Pressure and Inherent Maneuver Preferences.....	43
5.2.1 Time Pressure Effects on Maneuver Tendency.....	43
5.2.2. Time Pressure and Maneuver Choice Effects on Safety.....	45
5.2.3. Time Pressure and Maneuver Choice Effects on Efficiency	46
5.3. Conflict Geometry Effects.....	46
5.3.1. Geometry Effects on Maneuver Tendency	46
5.3.2. Geometry Effects on Safety	46

5.3.3. Geometry Effects on Efficiency	47
5.4 Display Support Effects.....	47
5.4.1 Predictive Displays Are Useful	48
5.4.2 Vertical Representation Effects.....	48
5.4.2.1. Vertical representation effects on maneuver tendency.....	48
5.4.2.2. Vertical representation effects on maneuver safety.....	48
5.4.2.3. Vertical representation effects on maneuver efficiency.....	49
5.5 Conclusions.....	49
5.5.1. Maneuver Tendencies:.....	49
5.5.2. Maneuver Safety:.....	49
5.5.3. Maneuver Efficiency:	50
6. CDTI Evaluation Scenarios	51
6.1. General.....	51
6.2. General Guidelines for Scenario Development	51
6.3. Candidate Scenarios	52
6.3.1. Taxonomy of Conflict Geometries	52
6.3.2. General Guidelines	53
6.3.3. Specific Suggestions for CDTI Evaluation Scenarios.....	55
6.4. Summary.....	58
7. Performance Measures in CDTI Literature	60
7.1. General.....	60
7.2. Overview of the Measurement Problem.....	60
7.3. A Taxonomy of Measures	61
7.3.1. Criteria	61
7.3.2. Measurement Scales	63
7.4. Summary.....	64
7.5. Recommendations	64
Appendix	76

ACRONYMS AND DEFINITIONS

ACAS	Airborne Collision Avoidance System; ACAS is the ICAO standard for TCAS; used as a generic term (e.g., for both TCAS and CDTI)
ADS-B	Automatic Dependent Surveillance-Broadcast;
AGL	Above Ground Level
ARINC	Originally known as Aeronautical Radio, Inc. (incorporated on December 2, 1929).
ASRS	Aviation Safety Reporting System
ATC	Air Traffic Control
ATCRB	Air Traffic Control Radar Beacon
CA	Conflict Angle: The clockwise angle between ownship and intruder trajectories
CAS	Collision Avoidance System
CD	Conflict Detection.
CD&R	Conflict Detection and Resolution
CDTI	Cockpit Display of Traffic Information
CFR	Code of Federal Regulation
Conflict	Predicted converging of aircraft in space and time, which constitutes a violation of a given set of separation minima. (ICAO)
CP	Conflict Prevention
CPA	Closest Point of Approach (RTCA, 1997)
CR	Conflict Resolution
DCPA	Distance to the Closest Point of Approach.
EDCPA	Estimated Distance to the Closest Point of Approach
EGPWS	Enhanced Ground Proximity Warning System
ERB	Estimated Relative Bearing
ETCPA	Estimated Time to the Closest Point of Approach.
FA	False Alarm; for example, an alert for a conflict, which, in fact, was not a conflict
FAA	Federal Aviation Administration
FAR	False Alarm Rate; the number of false alarms divided by the total number of alarms
FMS	Flight Management System
HMD	Horizontal Miss Distance: The horizontal range between two aircraft at the point of closest approach

ICAO	International Civil Aviation Organization
LAT	Look-Ahead Time
MOPS	Minimum Operational Performance Standards
NASA	National Aeronautics and Space Administration
NLR	National Aerospace Laboratory. NLR, is an independent non-profit research institute based in the Netherlands that carries out contract research for national and international customers.
NMAC	Near Midair Collision
PAZ	Protected Airspace Zone: A zone about an aircraft defined by the lateral and vertical separation standards.
PIL	Pilot-In-the-Loop;
RA	Resolution Advisory
RAT	Route Advisory Tool
RB	Relative Bearing: The clockwise compass direction from ownship to intruder
RTCA	Radio Technical Commission for Aeronautics
RVSM	Reduced Vertical Separation Minima
SAE	Society of Automotive Engineers
TA	Traffic Alert: Notification of traffic, which is expected to violate some separation criteria.
TCAS	Traffic Alert [and] Collision Avoidance System
TCPA	Time to the Closest Point of Approach. RTCA uses True TAU: Actual time to closest point of approach assuming straight flight of the intruder.
TRACON	Terminal RAdar CONtrol; the air traffic control facility providing radar service to aircraft arriving and departing one or more airports within the terminal airspace.
TSO	Technical Standard Order
URET	User Request Evaluation Tool.
VMD	Vertical Miss Distance: The relative altitude between own and intruder aircraft at closest point of approach (RTCA, 1997).
VSI	Vertical Speed Indicator
VSL	Vertical Speed Limit

1. BACKGROUND

1.1. Historical Overview

The idea of airborne Collision Avoidance Systems (CAS) is old (see Larsen, 1967; Kulikowski, 1967) and the development history of viable technological solutions parallels rare but catastrophic occurrences of midair collisions (Wiener, 1989). For example, after the 1986 midair collision between an Aeromexico CD-9 and a private single-engine piper aircraft over Cerritos, California, the U.S. Congress passed a law requiring the Federal Aviation Administration (FAA) to mandate the use of a Traffic Alert and Collision Avoidance System (TCAS; see 14 CFR § 121.356). Since 1993, all aircraft with more than 10 seats have been required to be equipped by TCAS, which is based on the Air Traffic Control Radar Beacon (ATCRB) technology and signals sent by aircraft transponders in response to TCAS interrogation signals. Two types of TCAS are currently in use: TCAS I, which provides pilots with traffic advisories (TAs) of potentially threatening traffic, and TCAS II, which also provides for resolution advisories (RAs) instructing the pilots to take an appropriate vertical evasive maneuver.

The concept of a CDTI (Cockpit Display of Traffic Information) has a long history of research, dating back over two decades (e.g., Hart & Loomis, 1980; Palmer et al., 1980; Jago & Palmer, 1980; Abbott et al., 1980; see Rehmamr, 1982 for early summary). More recently, the free flight concept has accelerated CDTI research at NASA Ames Research Center (Battiste & Johnson, 1998; Johnson, Battiste, & Bochow, 1999; Johnson et al., 1997; Johnson, Liao, & Tse, 1999; Livack, McDaniel, Battiste, & Johnson, 1999), Netherlands Research Laboratory (Hoekstra & Bussink, 2002; Van Westrenen & Groenweg, 2003), the FAA (Prinzo, 2003), and some universities (Wickens, Gempler & Morphew, 2000; Alexander, Wickens & Merwin, in press). Unlike TCAS, CDTI relies on satellite-based Automatic Dependant Surveillance-Broadcast (ADS-B) and Traffic Information Services (TIS) technologies.

1.2. Scope of the Report

This document examines issues associated with applications that include traffic awareness and conflict detection and resolution. Such applications are critically dependent upon the effective predictive and planning support that can be offered by the CDTI. Importantly, both prediction and planning require automation, embedded within the CDTI logic to make inferences about future trajectories of ownship and traffic. Such inferences will, by nature, be imperfect, leading to an apparent loss of reliability of the prediction/planning tools, and this loss of reliability, and its impact on pilots behavior and cognition, will be the focus of this report. In particular, we distinguish between three aspects of the CDTI: alerts, visual symbology, and planning tools.



Figure 1.1. The NASA CDTI (downloaded from <<http://human-factors.arc.nasa.gov/ihh/cdti/cdti.html>> on December 18, 2003).

- (1) Alerts, whether auditory or visual, are designed to convey discrete information regarding whether the aircraft has crossed some threshold boundary of danger: for example, entering the protected zone of another aircraft. There may be multiple stage alerts, each one representing an increasing danger or risk level. These can be wrong if they trigger an alert when there is (or will be) no danger, or when they fail to trigger an alert when there is (or will be) danger.
- (2) Visual symbology in such forms as leader lines, or vectors depicting the point of closest approach offers an explicit display of the inference of future behavior; a display which may prove to be inaccurate. For example, a straight leader-line, predicting a straight course, will be an unreliable predictor if the pilot chooses to turn.
- (3) Interactive planning tools (Layton, Smith & McCoy, 1994; Johnson, Battiste & Bochow, 1999), embedded within the CDTI also must make explicit assumptions about the behavior (or range of possible behaviors) in order to recommend courses of action, or to portray the possible consequences of various courses of action.

Across all three uses, the impact of the potential unreliability of future predictions will be examined as well as how these should guide the construction of predictive and alerting tools for the CDTI applications in conflict prevention. This document will (1) provide an outline of the report, (2) discuss general issues in the human factors of alerts, (3) discuss the properties and sparse pilot-data of existing CDTI alerting prototypes (4) present data from non-CDTI alerting studies, and lessons learned from parallel systems, like TCAS, (5) review CDTI studies that have looked at other (non-alerting) evaluations of pilot performance, (6) briefly overview the existing CDTI guidelines that do exist, (7) present a taxonomy of scenarios for evaluating CDTI alerting logic, and (8) provide some recommendations for performance assessment in approving CDTI alerting logic.

1.2. Outline of the Report

1.2.1. 2. Non-CDTI Alerting and Alarm Studies.

This section will draw from the more general human factors literature on the alarm-false alarm issue, examining how this literature informs us about the tradeoff between false alarms and misses (or late alerts), how these tradeoffs affect trust and reliance in the alerting system, and how low base-rate events must create a high false alarm rate, if high-cost misses are to be minimized, and conclude with general recommendations emerging from the literature on solutions to mitigate the problems of alarm false-alarms, with an emphasis on multi-level (i.e., “likelihood”) alerts.

1.2.2. 3. CDTI Conflict Detection Algorithms and Alert Properties

This section will include a version of the Thomas, Wickens, and Rantanen (2003) paper, focusing on literature that has (a) addressed the false alert-late alert tradeoff. (b) evaluated the CDTI alerts with pilot-in-the-loop data, and (c) combined (a) and (b). This is a very small set. This section also includes tables summarizing the studies, conclusions for the research that needs to be conducted, and it will provide a basis for recommending what sort of conflicts will challenge the algorithms, and hence should be targeted for inclusion in bench test certification requirements.

1.2.3. 4. Airborne Alerting and Traffic Systems: Guidelines and Studies

This section contains three subsections; specifically, (4.1.) Lessons Learned: This section focuses primarily on the lessons learned from the field testing of TCAS and will include much of the document prepared following Wickens’ return from interviews at MITRE and ARINC (Document 4/30/03), coupled with notes from an excellent paper review on the design of TCAS 7 and references to TCAS TTP newsletter reports; (4.2.) TCAS ASRS analysis: This section reports the results of the ongoing analysis of ASRS reports from TCAS system, including all TCAS-related reports submitted to ASRS between years 1991 and 2003; (4.3.) Documents and guidelines: This section will highlight recommendations made and data provided regarding the desired and obtained reliability levels of airborne alerting systems, focusing on TCAS and, to a lesser extent, EGPWS.

1.2.4. 5. CDTI Non-Alerting Studies.

This material will emphasize on how trajectory geometry for long-range conflicts (i.e., CDTI, not TCAS) affects pilot maneuver choice (climb descend, turn, speed adjustment), and maneuver safety. From this review, guidance will emerge regarding what aspects of conflict situations make detection and response difficult from a pilot's point of view, and therefore should be targeted for inclusion in certification and approval. This section also reviews guidelines offered by RTCA DO-243, TSO-C147, and a forthcoming document from Volpe, regarding other features not pertaining to discrete alerts that should be included in CDTIs. Emphasis will be given to those features that are most relevant to traffic situation awareness (e.g., traffic map rendering, leader lines, altitude depiction, decluttering). We carefully review these guidelines for anything they say about alerts as well as for their origins, human factors data to back them up, and potentially overlooked issues.

1.2.4. 6. CDTI Evaluation Scenarios.

This section develops a full taxonomy of possible conflict scenarios, following the guidance of those derived by McLaughlin (1997) for TCAS, and expressed in the framework of certification documents for GPS (Wright, year) and Electronic Flight Bags (Chandra, year). A matrix depicting the creation of a large number (e.g., several hundred) of possible scenarios will be presented, as these are created by orthogonally combining lateral, vertical and speed features, and combining these in turn with operating conditions (traffic density, weather, system reliability). From this matrix, a subset of roughly 12 scenarios will be selected for bench testing, with emphasis on why this subset was chosen on the basis of the literature reviewed in Sections 3 and 5 above. Of these 12, 3 will be selected and provided with concrete geographically specified parameters (e.g., a specific sector of Indianapolis ARTCC, with climbing/descending, transitioning traffic).

1.2.6. 7. Performance Measures in CDTI Literature.

While we will not provide a full list of all aspects of performance to be assessed (as such is provided, for example, by Chemiavesky, 1997, for bench testing TCAS), we will offer recommendations for assessment of test pilot evaluation of CDTI alerts during bench testing.

2. NON-CDTI ALERTING AND ALARM STUDIES

2.1. Overview

In this section we show how signal detection theory is applied to alerting systems, we review the literature from human factors research on imperfect alerting/alarm systems, and we draw conclusions from this research on human performance effects and on the efficiency of display mitigation strategies.

2.2. Discrete Alerting in CDTI

The goal of alerting algorithms within a CDTI is to provide an automation assistance that will alert the pilot that some threshold of danger with a conflict aircraft, detected by the automation, has been exceeded. As such, and as with all detection systems (Pritchett, 2001), performance of either the alerting algorithm itself, or the algorithm-pilot together, can be represented in the context of the signal detection theory (Green & Swets, 1988; Wickens, 2001).

The theory postulate that there are two distributions of danger or risk state, one when a “signal” (future conflict) does not exist, and another when it does (Figure 2.1). Because of the impossibility of perfectly predicting future trajectories, these two distributions overlap, making it hard to define any alert criterion that will always discriminate dangerous from non-dangerous states. Errors will occur. The two states and two responses yield four outcomes as shown in Figure 2.2.

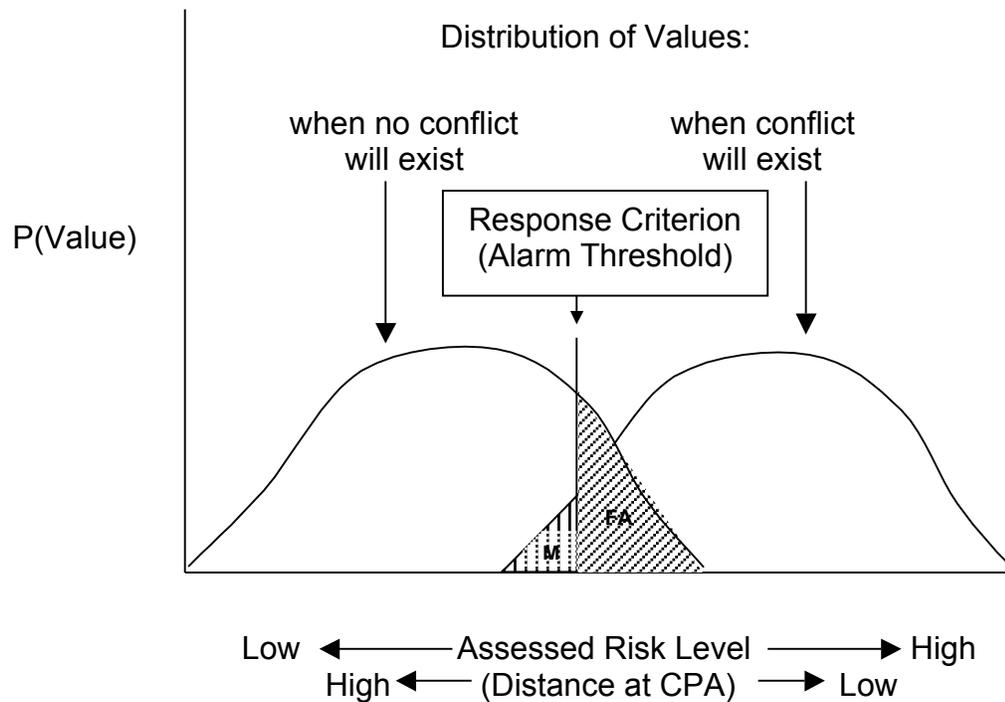


Figure 2.1. Signal detection theory representation of CDTI alarm threshold. FA=false alert, M=miss.

		State of the World	
		Conflict (will exist) "Signal"	Safe separation "Noise"
Agent Response	"Conflict" (alert)	"Hit"	"False Alarm"
	"No conflict" (no alert)	"Miss" (Late Alert)	"Correct Rejection"

Figure 2.2. A signal-detection matrix of alerts vs. state of the world.

In Figure 2.2, the responding agent can either be the output of the CDTI, which issues an alert or not, or the output of the pilot who consults the CDTI alert, and then either decides that a dangerous situation does or does not exist. Whether applied to the algorithm, or to the pilot-algorithm system, there are several common elements that characterize this representation.

1. In both cases, a "miss" is typically not a true miss, but rather, a delay in issuing or responding to the alert. Hence a miss can be defined here as an alert that is not issued by a particular time before the closest point of approach (CPA).
2. In both cases, real errors (false alarms, misses) do occur (reliability \ll 1.0), because of the impossibility of making perfect predictions in the uncertain airspace, an issue that will be discussed further in Section 3 (CDTI conflict detection algorithms). Formally, reliability can be defined by $1 - P(\text{error})$, such that $P(\text{error}) = [\text{FA} + \text{Miss}] / \text{total events}$. (Note that in the following discussion, we refer to the **imperfection** of the automated alert system, rather than the "**unreliability**" of the system. The latter term has a negative connotation that we prefer not to use, when dealing with systems that must cope with future environmental uncertainty, and hence must of necessity occasionally fail to fulfill their predictions).
3. In both cases, the nature of the errors (false alerts vs. misses) can be traded off, by varying the response criterion. In formal signal detection theory terms, this is referred to as "beta", set either high, such that there are many misses (late alerts) and few false alarms, or low, such that there are many false alarms, but few misses. The Beta in Figure 2.1 is set low. In the treatment of alarms, the response criterion is sometimes referred to as the "**threshold**", which can be varied in its "sensitivity". A very sensitive threshold

corresponds to a low beta. (The term “sensitivity” is ambiguous in discussing alarms, since, in signal detection theory terms, sensitivity refers directly to the reliability of the alarm system, not to the response criterion. Therefore in describing this concept in the context of alert systems, we will refer to a low or high threshold).

2.3. Imperfect Automation Research

In our current effort, the dynamics of the pilot response to changes in parameters of the alarm system (its reliability, and its threshold – low or high) is of particular interest. How does the pilot respond to decreases in reliability? And are these responses different, for increases in false alarm rate caused by a lower threshold, than for increases in miss (or late alert) rate caused by a higher threshold? No CDTI research has apparently examined these issues in systematic empirical studies. In the current section, therefore, we summarize six relevant conclusions or findings from research in other domains of imperfect automation.

2.3.1. The Response Criterion Must Be Low.

It is by now well established that when alarm systems are to alert users of important events, which occur infrequently (low base rate), and for which misses are of high cost, it is essential to set a low threshold (low beta), in order to guard against the negative consequences of the misses (Parasuraman, Hancock, & Olofinboba, 1997; Wickens, Lee, Liu, & Gordon-Becker, 2004). The inevitable consequence of this decision is a high false alarm rate. Such high false alarm rates have been well documented to occur in other fielded alert systems, such as the ATC URET system (Krois, 1999a,b; Wickens et al., 1998), and may be as high as 50%.

2.3.2. Humans Lose Trust in and Reliance on Imperfect Automation.

By now, there is a solid history of research in the loss of trust by human operators with automation that is imperfect or unreliable (e.g., Parasuraman & Riley, 1997; Lee & Moray, 1994; Lee & See, in press; Sheridan, 2002; Wickens & Xu, 2002). The dynamics of this loss often starts with an overtrust in automation that appears perfectly reliable until it exhibits its initial “failure” (or error). Such overtrust is expressed as “complacency” and will often lead to a serious problem at the time of the first failure, as the user has ceased to pay attention to what the automation is doing (or, in the case of alerting systems, ceased to monitor the “raw data” processed by the automation algorithm; Yeh, Merlo, Wickens, & Brandenburg, 2003; Metzger & Parasuraman, 2001; Parasuraman, Molloy, & Singh, 1993). However, following the initial human discovery of the imperfection (e.g., following the first failure), the operator may sufficiently lose trust in the automation, that s/he ceases to rely upon it if there are other avenues to process the raw data or do without the automation (Yeh & Wickens, 2001; Lee & Moray, 1992, 1994; Sorkin, 1988). This loss of trust and reliance, can be supported anecdotally, by the frequency of non-response to TCAS resolution advisories that was observed, following the frequent false alerts of earlier TCAS systems (Ciemier et al., 1993; see Section 4 of the current report).

2.3.3. Humans Can Become Calibrated to Automation Imperfection.

In spite of the loss of trust that may result following initial failures, there is also ample evidence that operators can become appropriately calibrated to the imperfection level, given enough feedback, and understanding of the sources of imperfection (Wickens & Xu, 2002; Yeh

et al., 2003; Maltz & Shinar, 2003; Dixon & Wickens, 2003; Wickens, Gempler, & Morphey, 2000; Moray, 2003). Such calibration can be defined in many different ways, such as the appropriate allocation of attention between the raw data that is integrated by the alarm, and other competing tasks.

2.3.4. Trust and Reliance are Different Concepts.

Trust is a subjective state of the believability of automation devices (Lee & See, in press). Reliance, on the other hand is an objective measure of the extent to which automation advice is followed (or automation control is deployed). When workload is low, the two are often tightly coupled (Wickens & Xu, 2002). However when workload is high, there may be many instances in which imperfect automation is not fully trusted (the operator is aware of its errors), but is relied upon nevertheless (e.g., used), because the resources saved (by using the automation) can be more valuably deployed in the performance of other tasks (Dixon & Wickens, 2003).

2.3.5. Misses May Effect Trust Differently From False Alarms.

Regarding alarm systems in particular, the literature remains somewhat ambiguous regarding whether systems that produce misses (or late alerts) generate a greater loss of trust (or reliance) than those that generate false alarms, given an equivalent overall reliability level between the two (Dixon & Wickens, 2003). For example Maltz and Shinar (2003) found that increasing system false alarm rate tended to disrupt reliance more (and led to poorer combined human-system performance) than did increasing system miss rate. On the other hand, Dixon and Wickens (2003) found that a miss-prone system led to a greater loss of trust than did a false alarm-prone system, and while both kinds of automation degradation were harmful to human-system performance in a multi-task environment, each led to qualitatively different kinds of performance decrements. Hence the issue of which kind of automation error is “worse” for human performance remains unresolved, in spite of the fact that these must be traded off in their frequency as alert thresholds are varied.

2.3.6. Imperfect Automation Costs May Be Mitigated by Display Features.

There appear to be at least three solutions available to mitigate possible costs, to human trust and reliance, associated with alert imperfection. First, as noted above, training (as to the inevitable sources of imperfection, and to its frequency) may allow operators to better calibrate their trust and reliance, although research on imperfect automation training is scarce.

Second, it is important to provide a clear display of the raw data upon which automation (e.g., an alert algorithm) bases its judgment. In the TCAS system, this is accomplished by the traffic display. For the CDTI, it is accomplished by the CDTI display itself, but these raw data may be enhanced by portraying perceptually, various parameters upon which the alerting agent is basing its judgment (e.g., predictive paths, time until, and distance at PCA; Alexander, Wickens, & Merwin, in press; Wickens, Gempler, & Morphey, 2000). This prescription (to present raw data) derives from data which suggest that human performance with automation-based command displays (which tell people only what to do) appears to be less error-tolerant than automation status displays (which inform them of what the current situation is, and the future is likely to be; i.e., status displays present the “raw data”; Sarter & Schroeder, 2001; Parasuraman, Sheridan, & Wickens, 2000). Error tolerance by pilot users is important here, because of the above-mentioned inevitability of error-commission by traffic prediction algorithms.

Third, there is some evidence that the costs of automation errors can be mitigated by using **multi-level alerts**, portraying the degree of seriousness of the situation at more than two levels (danger present or absent; alarm on or off). This is the concept of the “likelihood alarm” developed by Sorkin, Kantowitz, and Kantowitz (1988), and is inherent in such alerting systems as the current homeland security system (red/orange/amber, etc), the current TCAS system (resolution advisory is more serious than traffic alert), and in some proposed versions of the CDTI alerting logic (Johnson, Battiste, & Bochow, 1999; see Section 3). The logic of providing multi-level alerts is straightforward: if more levels are allowed, then when the alerting system is wrong, the nature of its failures will be perceived by the human user to be less catastrophic, than when a 1 level alert is employed, and hence, the loss of trust (and of reliance) will be less as well.

However, surprisingly, there has been very little research on how 2 (or greater) level alerts compare to single level (on-off) alerts. Sorkin, Kantowitz, and Kantowitz (1988) observed some benefit for a 2 level likelihood alarm, but this was only in performance of a concurrent task (reflecting reduced workload), not in increasing the sensitivity of detecting the alarm-driving signal itself. St. John and Manes (2002) appear to be the only investigators who have established the benefits of more levels of alerts (6 vs. 2) in an alarm system on detection of the alarm-driving variable itself. They found, however, that the advantage of the higher number was neutralized to the extent that the alerting system was imperfect. No other studies appear to have systematically examined this variable. In particular, it does not appear to have been evaluated in a traffic conflict detection system.

An important issue that remains unresolved is the source and weighting of system state variables that should drive a multi-level alarm which varies only on a single scalar variable of “danger”. In conflict alerting systems, these sources may include the time until PCA is reached (shorter time _ more alerting), the actual distance (or time) of separation at PCA (smaller distance _ more alerting), or the projected uncertainty (or certainty) that a single danger threshold will be crossed. Finally, likelihood alarms, by definition, define at least two levels of “errors” for the purpose of computing reliability. For example, a “false alarm” may either be a “small false alarm” (if for example an uncertain conflict is predicted and none occurs) or a “large false alarm”(if, for example, a certain conflict is predicted, and none occurs). Similar logic applies to misses. This complexity makes the logic of fuzzy signal detection theory important for assessing the quality of automation alerting (Parasuraman, Masalonis, & Hancock, 2000).

2.4. Conclusions

1. CDTI Alerts must be imperfect, given that the future of a probabilistic airspace is to be predicted.
2. These alerts should produce more FAs than misses (late alerts) because of low base rate, and the high cost of misses.
3. While humans may lose trust with imperfection, they can learn to calibrate their reliance to the actual reliability of the system.
4. Misses (late alerts) probably have a different consequence on trust and reliance than FA, although more research is needed to understand this tradeoff.

5. Human trust and reliance problems may be mitigated by training, status displays, and likelihood alarms although the optimum number of alarm states for likelihood alarms is not well established and research should address this issue.

In the following section we consider in detail the properties of CDTI algorithms that may impact its reliability.

3. CDTI CONFLICT DETECTION ALGORITHMS AND ALERT PROPERTIES

3.1. Overview

Automated warning and alert devices such as the CDTI 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 airborne collision avoidance systems (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 section, we define some important terms, and then 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 identify the parameters of CDTI alerts that appear to impose the greatest effects on FA rate.

3.2. Definitions

3.2.1. Conflict

A conflict is defined as the loss of the minimum required horizontal and vertical separation distance between two aircraft. The minimum separation distance from an aircraft defines its protected airspace zone (PAZ). The current radar-based horizontal separation minima are 5 nautical miles (nm) for en route environment and 3 nm for terminal areas (FAA 7110.65M, Air Traffic Control, 2000). There are smaller minima to be applied in special circumstances (e.g., simultaneous approaches to closely-spaced parallel runways) but these are beyond the scope of the present discussion. Current standard vertical separation minima are 1,000 ft below 29,000 ft and 2,000 ft above 29,000 ft. These minima, too, are changing with gradual implementation of reduced vertical separation minima (RVSM) standards, which allow a 1,000 ft vertical separation minima to be applied also above 29,000 ft. It should be noted that it has not been decided what type and what volume of airspace will be used under various circumstances in the free flight environment.

3.2.2. Near Midair Collision (NMAC)

A near midair collision (NMAC) is an incident associated with the operation of an aircraft in which a possibility of a collision occurs as a result of proximity of less than 500 feet to another aircraft, or a report is received from a pilot or flight crew member stating that a collision hazard existed between two or more aircraft (FAA, the National Aviation Safety Data Analysis Center).

3.2.3. Collision

A conflict should not be confused with a midair collision, which is defined as an actual impact (Johnson, Battiste, & Bochow, 1999). Johnson, Battiste, and Bochow (1999) discussed the change in philosophy from the Traffic Collision Avoidance System (TCAS) to the CDTI designs being proposed. TCAS was designed to detect imminent collisions and provide (literally) last minute advice for emergency avoidance maneuvers when air traffic control has failed to maintain separation between aircraft. However, CDTIs are designed to provide more advanced notice. It provides , more detailed, strategic information with a potential range of 120 nm

compared to the 40 nm maximum range of TCAS. This larger range and more precise traffic information can support earlier detection of projected conflicts, and hence enable earlier avoidance maneuvering so that no loss of separation will occur (Johnson et al., 1999).

An important distinction between the technologies behind TCAS and CDTI must be made, however. TCAS is based on mode-C transponders on board aircraft that transmit coded pulses including the aircraft's altitude upon interrogation by the on-board TCAS equipment. Hence, TCAS can "see" all mode-C transponder-equipped aircraft but not those not having a transponder. This is not a problem, however, as mode-C transponders are mandatory equipment in certain airspaces [Look up the exact regulation]. The CDTI, on the other hand, is based on ADS-B technology and other possible sensor and surveillance systems, and presently there are no regulations concerning aircraft equipage. Therefore, until ADS-B equipment can be mandated, which in turn depends on the reliability and cost of such systems, CDTI cannot be used for collision avoidance.

3.2.4. Conflict Detection

Conflict detection is the process of comparing the flight paths of ownship and other aircraft to determine whether the predicted point (both distance and time) of closest approach of the two aircraft is less than the prescribed minimum separation distance as defined above (e.g., 5 nm radius, ± 1000 ft vertical). Detection of a possible conflict is a four-dimensional task that involves projecting not only the future spatial locations of two aircraft, but estimating the time at which each aircraft will reach the point of closest approach. To illustrate, two planes may occupy spatial locations less than the minimum separation distance apart (e.g., have intersecting 3-D trajectories), but if they pass through these locations at sufficiently different times, no conflict will occur. CDTIs can support this task by providing graphical flight path information in the form of predictor lines, which can be visually compared to determine if they will intersect spatially and temporally in the future (Johnson et al., 1999). In theory, pilots can visually scan and compare the aircraft symbology on the screen to determine the risk of conflict, although how well they can perform this task while simultaneously performing other aviating tasks must be addressed by empirical research. The above example depicts a scenario where the level of automation is low (i.e., automation gathers and display relevant information) or moderately low (i.e., it performs the necessary calculations and displays the projected trajectories of potential intruder aircraft) and the higher-level cognitive functions of actual conflict detection and resolution formulation are left to the pilot (c.f., Sheridan, 1997).

Research on CDTIs increasingly involves the inclusion of different forms of automatic conflict detection alerting systems, most of which are based on algorithms developed independently and applied to CDTI and air traffic control displays (e.g., Yang & Kuchar, 1997; Paielli & Erzberger, 1997; Magill, 1997; see also Kopardekar, Sacco, & Mogford, for a comparison of CD&R algorithms). Although currently certified CDTIs do not include automatic conflict detection, it seems likely that future versions of the CDTI will include automated support for this activity (Rantanen, Wickens, Thomas & Xu, 2003). Since automatic systems are rarely 100% reliable, or perfectly calibrated to detect all conflicts and produce no false alarms (see Section 2 of this document and Yang & Kuchar, 1997, and Feldman & Weitzman, 1999, for discussions of the false alarm issue), it may also be important for pilots to be able to perform conflict detection tasks despite the presence of an automatic system. Hence, in addition to the pilots' conflict detection performance unaided by automation, as is the case with near-term CDTI

implementation, it is important to consider pilots' abilities to evaluate whether an alerted situation will result in a true conflict, requiring the pilot to create a resolution, or whether the situation does not actually represent a true conflict and can be considered a false alarm

3.2.5. Conflict Resolution

Conflict resolution is the process of developing, entering, and executing a maneuver to avoid a detected conflict. CDTIs that have been approved by the FAA and on the market as of the time of this report's publication did not contain any features for directly interacting with the flight plan, but there has been research to investigate the usefulness of such a feature (Johnson et al., 1999; Thomas & Johnson, 2001; Winterberg, Mulder, & van Paassen, 2003). NASA's Route Altering Tool (RAT) is one such feature; it allows direct manipulation of the flight path through a drag-and-drop interface (Thomas & Johnson, 2001; Uhlarik & Comerford, 2003). The pilot can use the computer mouse to select a point along the ownship's flight path and "stretch" the flight path from its existing trajectory to a new one, thereby creating a conflict avoidance maneuver. This new flight path would automatically be submitted to the FMS for implementation. Some algorithms have been developed to aid the pilot in generating the most safe and efficient maneuvers, either by suggesting alternatives or by automatically choosing the "best" alternative without pilot input (e.g., Irvine, 1998; Mondoloni & Conway, 2001; see Kuchar & Yang, 2000, for a review; see also Johnson et al., 2003, for a comparison of algorithm-derived vs. pilot-created resolutions).

Research has shown that some conflict geometries are more difficult to detect and resolve than others, as discussed in Section 5 of this document. It may be that even the addition of a spatial ambiguity resolution tool such as an interactive viewpoint will be insufficient to improve detection and resolution performance for all possible situations. It may thus be advisable to include an automated conflict detection and alerting feature to aid the pilot in determining whether another aircraft is a threat or not.

In addressing the issue of conflict alerting as it is used in automated conflict detection, it must first be broken down into its constituent parts. Conflict alerting is a product of calculations based on (1) a particular algorithm, which (2) takes in certain information, (3) compares it to some given criteria, and (4) produces some form of alerting symbology or audio signal to the pilot which indicates the calculated severity of the risk. Each of these parts will be considered in detail next

3.3. Algorithms

3.3.1. Algorithm Type

There have been many proposed algorithms, which fall under two general categories: deterministic and probabilistic (see Kopardekar et al., 2001, for a review). Although both types of algorithms continuously compare pairs of aircraft (ownship and other aircraft) to determine whether a conflict exists, the main difference is in the amount of uncertainty that is considered in calculating the risk of a conflict. Deterministic algorithms use state information (current position, airspeed, and trajectory of an aircraft) gathered from signals from traffic aircraft. CDTI studies at the University of Illinois using alerting have used a type of deterministic algorithm (e.g., Alexander & Wickens, 2002; Wickens, Helleberg, & Xu, 2001). This information is used to create a single geometric flight path against which the ownship's flight path is compared. If the

flight paths are projected to have a point of closest approach that is less than the minimum distance required by flight rules, then a conflict (or loss of separation) is calculated to occur. Probabilistic algorithms generally use a combination of state information and intent information (based on the filed flight plan of an aircraft), and formally introduce uncertainty into the calculations (see Yang & Kuchar, 1997, for a description of a probabilistic algorithm). Uncertainty can be due to wind conditions, measurement error of current state and trajectory, or any number of factors which cannot be precisely accounted for ahead of time. Probabilistic algorithms compare the projected flight paths but allow for some variability in the accuracy of the projections, such that a resultant conflict is based on the probability that the planes will actually occupy the projected points in space and time (Yang & Kuchar, 1997). Our review has identified these properties of the different algorithms, which are summarized in Appendix A.

3.3.2. Information Used by the Algorithm

The type of information used by the algorithms is a function of what is available and what is required to do the calculations. State information can be ascertained by radar projections, while intent information is gathered from a formal flight plan and assumes that the current state of the aircraft matches that of the flight plan. Consequently, the system's performance is dependent on both the reliability and timeliness of intent information broadcasted in case the filed flight plan is amended in-flight.

3.3.3. Conflict Criteria

The criteria for determining a conflict, or loss of separation, are determined by the protected zone of each aircraft and the look-ahead time (Thomas et al., 2003). As we have noted above, the protected zone is the area around an aircraft defined by minimum separation standards that may not be penetrated by the protected zone of another aircraft. The look-ahead time is the maximum amount of time before a projected conflict (or time-to-contact) that the alerting algorithm can determine whether a problem exists (Magill, 1997; Thomas et al., 2003). Sometimes look-ahead time is used interchangeably with time-to-contact (or tau), but it is also important to understand the distinction between these: Time-to-contact is the actual time until loss of separation occurs which will decrease continuously as a conflict evolves, while LAT is a pre-set parameter of an alerting algorithm which is based on the time-to-contact). For example, an algorithm with a 5-minute look-ahead time determines whether a conflict exists within 5 minutes of the projected point of minimum separation, while a 10-minute look-ahead time algorithm provides up to 10 minutes of warning before a projected conflict. While more warning time may appear to be beneficial, it should be noted that **the information will be degraded more by uncertainty with increased look-ahead time, and to guard against late alerts the threshold should be set lower in the face of this increased uncertainty, resulting in increased numbers of false alerts (or detected conflicts that turn out to be non-conflicts)**. Figure 3.1 illustrates this problem. Essentially, the problem with determining an effective look-ahead time is that it must be a compromise between the number of false alerts generated by heavily uncertain information and the amount of time needed in order to execute an avoidance maneuver in the case of a real conflict (Magill, 1997; Thomas et al., 2003; Rantanen et al., 2003).

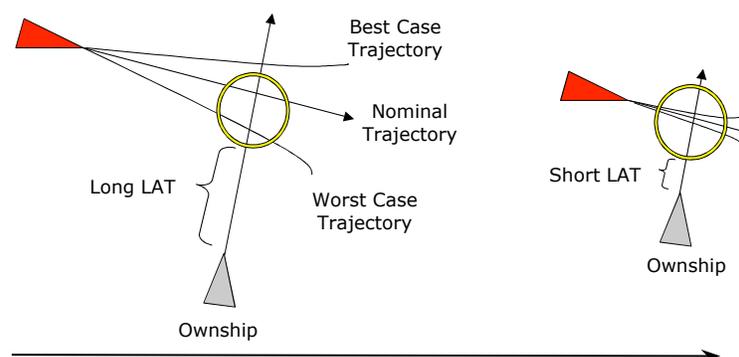


Figure 3.1. *The impact of look-ahead time (LAT) on false alarm rates. With the long LAT on the left, there will be greater uncertainty of traffic position at CPA.*

It will be recalled from Section 3 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 3.2 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 (i.e., look-ahead time). The middle line shows the nominal prediction, illustrating the steadily decreasing distance to the closest point of approach (CPA), 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. The CDTI should alert the pilot 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.

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 “circles” in Figure 3.2 represent the anticipated variability in both speed and lateral position around the nominal trajectory (Magill, 1997). This variability increases with increasing time. These “circles” can be thought of as confidence intervals (e.g., 90%). The lines on both sides of the middle nominal trajectory represent the confidence intervals on lateral separation, in which the “best case” line is the maximum predicted separation distance at CPA, and the “worst case” line is the minimum predicted distance at CPA. The growth of uncertainty over time represents the impact of winds or other factors that cannot be predicted with certainty.

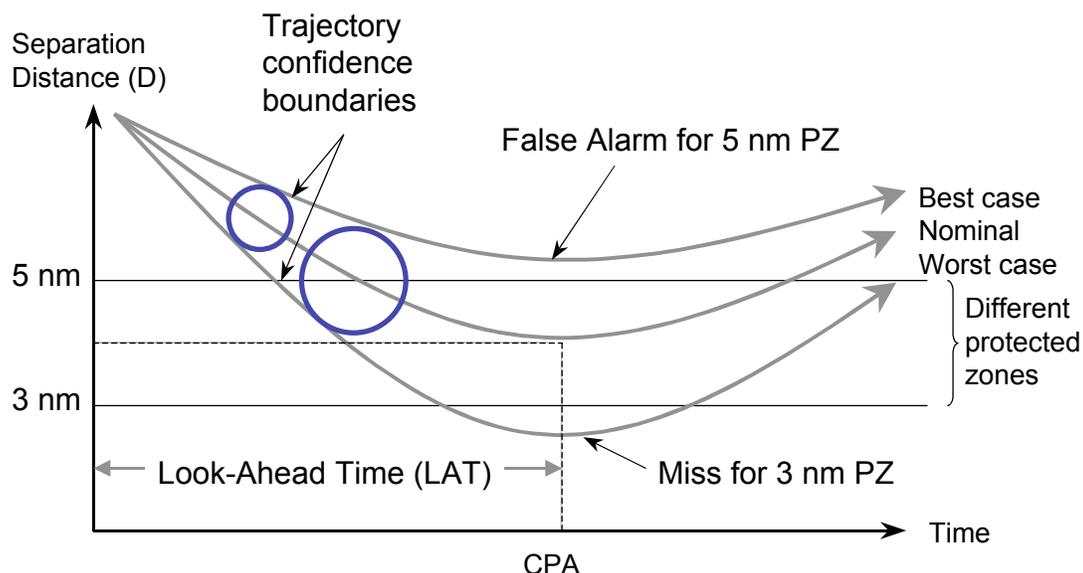


Figure 3.2. Schematic representation of the evolving space and time aspects of a conflict

Now consider a warning that might be given at time 0, defining a LAT to CPA 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 (PAZ), 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 3.2, 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.

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 the University

of Illinois at Urbana-Champaign 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.

The Netherlands Research Laboratory (NRL) 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 allow 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.

Research involving pilots using CDTIs to detect and resolve conflicts has typically used tactical and emergency-tactical LATs in the alerting logic. However, there have not been many studies that directly compare conflict and detection performance under different LATs within the same CDTI or the same study. As we discussed in Section 5, increasing time pressure produces different conflict avoidance maneuver choices. It was suggested above that the dividing line between high and low time pressure may fall along the division between LATs greater than one minute (Tactical) and less than one minute (Emergency-Tactical). It is possible that varying time pressure and maneuver options as a function of the length of the LAT will produce different conflict detection/resolution behaviors. Thus, as is evident from the analysis, the joint influence of the three parameters (the PAZ size, the LAT, and the assumptions about the growth of uncertainty with time) will affect the sensitivity of the alert system in discriminating predicted conflicts from non-conflicts, and hence the extent of the tradeoffs between the two negative events of false alerts versus misses or late alerts.

3.3.4. Alerting Levels and Symbolology

The alerting symbolology takes somewhat different forms in different CDTI designs (for three examples, see Johnson et al., 1999; DiMeo et al., 2000; Hoekstra & Bussink, 2003). The choice of alerting color and the number of different alerts and their meanings are not held constant across CDTI developers. A common decision in alerting is to provide more than two alerting levels (no alert, alert), typical of the likelihood alarm discussed in Section 2.3.6, so that pilots may determine how probable (given a probability calculation) or how serious (given the time or proximity to conflict) a conflict is based on the level of alert. A caveat to providing alarms with multiple stages is that the stages must be meaningful and provide information that can be easily extracted without the operator's full attention (Sorkin et al., 1988; Woods, 1995).

Most aviation alerting systems rely on two or more levels or stages of alerting to provide additional reliability information to the pilot (see Thomas, Wickens, & Rantanen, 2003, for a review). For example, TCAS II uses three levels: No alert, traffic advisory (TA), and resolution advisory (RA). However, there is a wide variety of implementation of alerting levels (anywhere

from 2 to 5, including some hybrid systems which use several alerting levels from one system and several from another in tandem), and to date there have not been any studies specifically designed to determine if there is an optimal number of alerting levels.

3.4. Limitations in the Reviewed Literature

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 3.1), 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 full list of studies reviewed is provided in Appendix A.

The NASA studies that appear in the first three rows in Table 3.1 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 FA rate 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 Morphey (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 3.1 is derived; see Appendix A) reveals that **the three most critical variables for affecting the increase of FAs or**

late alarms (depicted in Figure 3.2) are **(1) LAT, as a longer LAT produces more FAs, (2) the size of the minimum separation (PAZ) boundary, where the larger the boundary, the more FAs produced, and (3) the assumptions that are made about the growth of uncertainty where the more rapid uncertainty growth with time increases the FA rate for a given LAT on PAZ boundary.** Yet Table 3.1 reveals little consistency across these variables between studies (see in particular the “Uncertainty ” 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 3.1 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.

3.5. Conclusions

1. Alert system performance is particularly degraded by three factors. Given the preference to create more FAs than late alerts, these factors tend to increase FA rate: (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.
2. Pilot-in-the-loop research is strongly needed to evaluate the consequences to pilot performance of imperfect CDTI alerts as the three parameters of LAT, PAZ and uncertainty growth are manipulated, to assess the optimum value of these to support pilot trust, reliance and maneuvering options.

Table 3.1.: Summary of six pilot-in-the-loop studies that incorporated different conflict detection algorithms into CDTIs for use in free flight simulations.

Reference	Algorithm	Attributes Varied	Uncertainty	Multi-level alerts	PIL summary	FA discussion?
Johnson, Battiste, & Bochow (1999)	Yang & Kuchar algorithm used in the CDTI	PZ = 5 NM, +/- 1000 ft; intent info shared or not; 2 types of conflicts (1) true or (2) "apparent"; initial altitudes separated or not	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	8 crews flew either high or low flight-deck responsibility scenarios; positive subject responses to availability of info, colors, display options	
Johnson et al., 1997)	Yang & Kuchar algorithm used in the CDTI	PZ = 5 NM, \pm 1000 ft	Along-track error SD 15 knots, cross-track error SD 1 NM; probabilities of turns or altitude changes (4 /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	10 crews flew either enhanced (included resolution tools) or basic CDTI cond.; all flew 8 scenarios w/ at least one conflict each	Authors state explicitly that no false alarms were programmed into the scenarios
DiMeo, et al., 2002.	Yang & Kuchar (1997) logic, overlaid on TCAS	PZ = 5 NM, \pm 1000 or \pm 2000 ft (if ownship alt. > 29,500 ft)	Uncertainty parameters not specified (assumed to be same as in Yang & Kuchar, 1997)	4 stages, (1-2) CDTI-based, (3-4) TCAS Traffic Advisory and Resolution Advisory	Four sets of pilot/copilot flew 4 different conditions, 3 times each	
Wing et al., 2002, 2001.	Combination of Yang & Kuchar logic (state + intent) plus NLR/Hoekstra logic (state only)	PZ = 5 NM, \pm 1000 or \pm 2000 ft; LAT = 5 minutes for state info, 8 minutes for intent info; 3 cat. of conflicts, (1) state-only, (2) intent-only, and (3) "blunder"	Uncertainty parameters not specified (assumed to be same as 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	16 commercial pilots flew 3 en-route trials each of 4 scenarios (tactical vs. strategic operation mode, low vs. high ops complexity)	
Hoekstra & Bussink (2003)	NLR state-based conflict detection algorithm	PZ = \pm 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	8 sets of pilots flew en-route descent and terminal approach trials with low vs. high traffic, and managed vs. free flight terminal approach	False alarms are assumed to be minimized by setting LAT to 5 min (see Magill, 1997)
Wickens, et al., 2001, 2002; Alexander & Wickens, 2001	Geometric linear extrapolation of vert. and curvilinear extrapolation of lateral (rate of turn).	PZ = lat. and vert. dimensions varied between experiments.	Wickens, Gempler, and Morphew (2000) introduced lateral uncertainty. Not quantitatively varied.	Two levels, (1) predicted loss of separation within 45 seconds, (2) actual loss of separation.	Pilots found useful both predictor of future trajectory, and conflict specific information about point of closest passage.	Wickens, Gempler, & Morphew (2000) produced conflict misses; the only such study to do so.

4. AIRBORNE ALERTING AND TRAFFIC SYSTEMS: GUIDELINES AND STUDIES

4.1. Overview

In this section, the development of TCAS Guidelines is reviewed, as TCAS is the most closely related system to the CDTI, by summarizing the lessons learned in its development from personal interviews with developers, and from reviewed literature. An analysis of ASRS TCAS reports is provided and specific documents reviewed offering guidelines about TCAS, CDTI and EPGWS to ascertain the manner in which they treat the reliability of their alerting components. Finally, we review CDTI documents to identify other human factors guidelines that are indirectly related to the alerting functions.

4.2. TCAS Lessons Learned

4.2.1. Information from TCAS Developers

This section summarizes the information gained from Chris Wickens' meetings at MITRE and ARINC on April 30, 2003, and his discussions with Dan Tillotson and Gary Gambarani: of ARINC, who were involved in the human factors (pilot and controller) input to various revisions of TCAS, as well as meeting with Dave Domino, John Helleberg, and, most importantly, Andy Zeitlin from MITRE. Andy Zeitlin was the historical developer of most of the TCAS algorithms.

Both Dan Tillotson and Gary Gambarani outlined the "Dallas bump" as a seminal event in evolution of algorithms to address TCAS false alert problems. The response to this appears to be typical of the informal feedback loop whereby FA are addressed, that is, pilot complaints and concerns get expressed either "through the grapevine" or slightly more formally, through the TCAS Transition Plan (TTP) Newsletter (a concept lauded in the FAA ATC Automation books, Wickens et al., 1997, 1998), and these concerns bubble up with sufficient frequency that actions are taken to trigger development of a new version of TCAS. In the case of the Dallas bump, it triggered the development of version 6.04, which was said to be the most significant change across TCAS versions, because of its non-linear projection of intent. Dave Domino (pilot) particularly lauded the changes from 6.04 to 7, as of great value to pilots. Other examples of concerns that caused revisions to algorithms were better tracking of lateral separation in TCAS version 7, which now computes lateral passage, and better handling of "reversals" whereby an initially commanded vertical maneuver must be "reversed" because one aircraft does not comply.

Many of these changes, and others, were based on observations of concerns, but not any formal collection of false alert data; there does not seem to be any record of the frequency of these, other than ASRS data, which are addressed in a subsequent section. Personnel at both ARINC and MITRE are clearly concerned about the impact of false alarms on pilot trust and compliance, although there is no explicit evaluation thereof. It was mentioned that the changes to address the Dallas bump could produce a "miss" (or "late alert problem") if a plane that was assumed to be leveling off in fact did not, and busted the altitude. They were not however aware of these problems. Version 7 of TCAS is anticipated to be the last upgrade.

Considerable effort has gone into the linguistic issues in the oral RA's. Researchers said that, ideally, the linguistic oral alert will primarily function as a signal to look at the visual vertical

speed indicator (VSI) display. However, sometimes pilots apparently maneuvered solely on the basis of the oral alert, and given some human factors deficiencies in the formatting of the words, maneuvered incorrectly. Best example of this is “reverse climb” command, when a reversal was required. Some pilots followed the “Chris Webber” syndrome, and only attended to the positive meaning of this statement “climb”. Now the alert system says, “notice climb” (or something like that), which is designed to warn pilots to attend to new information on the visual VSI display.

Look ahead time (LAT) of RA was decided based on data from a tech center simulation (Barry Billman, 1970s), which estimated a 5 second pilot response time, plus an assumed vertical maneuver time, necessary to achieve a climb (or descent) rate, necessary to avoid penetration of a small zone around the aircraft. LAT of the visual Traffic Advisory added 10 seconds (total = 30 sec) to provide some (10 sec) warning of upcoming RA.

For TCAS, the uncertainty of future prediction is not due to pilot intent or winds (as in CDTI algorithms), but is entirely based on variability of estimation of current state of the two aircraft. To the extent that there is variability in state, there will be variability in state trend, and it is the trend information that is used to anticipate collision. Extrapolating this trend variance forward, and being “conservative” to avoid misses, is what produces the FA rate. ADS-B has the potential to transmit directly rate information, and hence exchange (between aircraft) more accurate trend information.

Suggested solutions to TCAS FA problems include:

1. Algorithm re-design (the current approach)
2. Airspace procedure modifications. For example, in geometries similar to that triggering the “Dallas bump,” ensure that there is now a lateral offset between the path of the climbing and the overflying airplane.
3. Reducing vertical speed and maneuver aggressiveness of FMS. The more rapid vertical speed changes of the automated aircraft render it more difficult to predict its vertical behavior (higher bandwidth, shorter span of prediction is possible).
4. Training, training, training. This point was highlighted by ARINC, referring to the great variety of training in TCAS across airlines. There is no requirement for on line simulator training (hands on), and such training that is done in simulators, often focuses more on “routine” RA responses, rather than the off-normal ones (i.e., “reversals”).
5. Controller training now appears to be adequate. French researchers at CUNA have developed an effective PC simulation for controller TCAS training.
6. Shared air ground information. The “Boston experiment” about giving controllers TCAS information on their displays, led to a discontinuation of this concept. Only in Japan do controllers see TCAS information. (Not clear why this is. Cultural differences?).

All people interviewed agreed that the issue of alerts on the CDTI is important to be addressed, in terms both of avoiding any potential conflicts with TCAS, and in terms of applying lessons learned. All also agreed that there are certain critical features of difference between the two systems that would prevent wholesale applications of design issues from one to the other, for example, tactical vs. strategic nature of the systems, avoidance versus planning, sensor-data information sources, and Mode S vs. ADS-B technologies.

4.2.2. Review of TCAS Literature

RTCA/DO-185A (RTCA, 1997) provides version 7.0 of the TCAS II logic and sets forth the minimum operational performance standards for TCAS II equipment. The version 7.0 improves the system's surveillance performance and modifies the interference limiting algorithms to account for aircraft densities near airports. This change directly addresses the problem with high FAR in terminal airspaces and is accomplished by selecting progressively smaller protected volumes as a TCAS enters low altitude or terminal airspace, known as sensitivity level control. Changes have also been made to preclude repeated TAs against the same target. Other changes include credibility checks to ensure integrity of the data being supplied to the collision avoidance logic and additional, more thorough tests for installed equipment (RTCA, 1997). The nominal advisory time before closest approach used by TCAS varies from 15 to 35 seconds in this version, illustrating the nature of TCAS as a last minute (literally!) collision prevention system. Human factors considerations are reflected in displaying and announcing all positive RAs (CLIMB, DESCEND, INCREASE CLIMB/DESCENT, MAINTAIN CLIMB/DESCENT) as corrective RAs, with a green "fly-to" indication being shown on the RA display. Positive RAs are only permitted to weaken to negative (VSL 0 fpm) RAs (LIMIT CLIMB and LIMIT DESCENT to 0 fpm RAs) once adequate separation has been achieved. This logic behavior reflects the observed behavior of pilots who will generally cause own aircraft to level off rather than initiating a maneuver towards a threat even when the logic calculates that the maneuver is safe (p.137). No references are given to the source of this information.

TCAS selects intruder aircraft for generating traffic advisories by using algorithms analogous to those used in threat detection but with larger thresholds. The traffic advisory normally precedes the resolution advisory in time. The logic classifies intruding traffic with four levels of priority to support symbology indicating their priority and to ensure that the highest ones are displayed. The highest priority is given to a threat causing a resolution advisory to be generated. All targets causing resolution advisories are displayed, even if there is no bearing estimate currently available. The next priority is given to other potential threats, including targets without altitude reports. The next priority is given to traffic satisfying a proximity test but not a tau-base detection test. The lowest priority is given to other traffic that is under track, but which does not meet any of the criteria for the previous classifications (RTCA, 1997; p. 138). False alert rate is not explicitly considered, but a false track rate is specified as the probability that a Mode S transponder-equipped aircraft track exists in the absence of an actual Mode S aircraft will be less than 0.1% (p. 31). False track is defined as a track for which an aircraft does not exist and can be the result of multipath interference, synchronous garble, fruit, noise, etc.

Issues discovered in the TCAS Transition Program (TTP) pertain to TCAS hardware, software, and human factors. Hardware issues identified included TCAS/mutual suppression bus interface, erroneous Mode C replies, 1030/1090 MHz interference, and illegal Mode S address. In our taxonomy these issues pertain to the system input variables, but no systematic documentation on the impact of them on system output or human performance variables was found. Software/logic issues included coordinated altitude crossing TAs, low altitude TAs and RAs, low altitude-descend RAs, RAs during parallel and crossing runway operations, and high vertical rate encounters. These issues pertain to the intermediate/system output variables in our taxonomy, but again no systematic documentation of the impact of these issues on human performance/behavioral variables was found. Finally, most issues directly related to human

factors pertain to training and design of procedures. Pilot training issues were addressed in the FAA AC120-55A (1993) and in AC120-55B (2001), which specifies flight crew TCAS qualifications, TCAS training program requirements, and operational use practices.

A good review of the features that have been incorporated into version 7, and the rationale for their incorporation is provided in Love (1998). A good description of the treatment of human factors issues in the early stages of TCAS development is provided in Chappell, 1990, while a good overview of human factors issues on TCAS related to air traffic control is provided in Wickens et al. (1998).

In summary, addressing human factors issues related to TCAS implementation has paralleled gains in operational experience with the system. Complete audit of these measures was outside the scope of this project, but given the different operational use of TCAS and CDTI, it is questionable whether the TCAS lessons learned would be applicable to CDTI beyond methodological and procedural problems in addressing them.

4.3. ASRS TCAS Report Analysis

4.3.1. General

The ASRS database has been used before to study TCAS-related incidents, most recently by Bliss (2003; see also Bliss, Freeland, & Millard, 1999), but several factors limit these studies' usefulness to fully gauge the nature and magnitude of the problem. Most important of these factors are the sampling method used to choose reports for analysis and the classification of reports. Both have been addressed in this part of the project and will be elaborated on below.

Bliss searched 50 most recently submitted reports in 20 categories for alarm-related keywords and acronyms; from a resulting total of 1,000 reports 273 were chosen for further analysis. Although these reports involved many different alarm systems, TCAS II was found to result in the highest number of false alert reports, either due to the system's too high sensitivity or crews' disagreement with the RA. Ironically, TCAS II was also associated with most missed alerts of all alerting system reports—also attributed to the system's sensitivity, in this case too low sensitivity.

Bliss (2003) used aircraft gross weight as a criterion for classification of the reports; this, however, does not adequately represent the alarm systems required for different kinds of aircraft. A better system would have been to classify the reports by the type of operations, that is, by the FAR part under which the pilot filing the incident report was operating. This is important because the FAA mandates different equipment for part 91 (general aviation), part 135 (commuter and air taxi) and part 121 (air carrier) operations. It is hence quite expected that part 121 or 135 operators would file more incidents pertaining to TCAS than part 91 operators simply because the former are mandated to use the system. Another major limitation in this paper was that the author did not indicate the time period from which the data were obtained but only reported the download date (i.e., the end of the data collection period). It would have been useful to know the start date, too, as well as to view the number of reports on a timeline. Such a timeline should further be correlated with introduction of new avionics and the FAA regulations concerning them (e.g., see CFR § 121.356).

To overcome these limitations and to gain a more comprehensive and detailed understanding of the TCAS-related problems encountered by flight crews in the operational environment,

another review of ASRS reports was undertaken under the auspices of this project. The analysis of TCAS-related reports submitted to the ASRS system was performed based on the so-called “landmark years” of TCAS development and implementation. These years were identified by a review of literature on the TCAS development for both technological and regulatory milestones. Also literature on TCAS operational experiences and evaluations was reviewed. In addition to the reporting year, the ASRS reports were classified according to the signal detection theory outcomes as hits, misses, and false alerts, and further whether the event was viewed as positive, negative, or neutral by the reporting pilot. This section will present the timeline of TCAS development and implementation, the findings from the TCAS Transition Program review, and results of the ASRS report analysis.

4.3.2. TCAS Development Timeline.

Both regulatory material and technological and operational advancements were included as milestones in TCAS development timeline. This timeline, and in particular the so-called “landmark years” of 1991, 1992, 1993, and 1994, was used as an organizational framework for the later analysis of the ASRS reports.

- 1981–89: TCAS development and testing began (FAA, 1990)
- 1989: Based on a Congressional Mandate (Public Law 100-223), the FAA issued a rule effective February 9, 1989 that required the equipage of TCAS II on airline aircraft with more than 30 seats, by December 30, 1991. Although not in the law, the FAA rule required airline aircraft with 10-30 seats to have TCAS I installed by February 9, 1995. Public Law 101-236, which was signed on December 15, 1989, amended P.L. 100-223 to (1) permit the FAA Administrator to extend the deadline for TCAS II fleet wide implementation to December 30, 1993 and (2) require the FAA to conduct a 1-year operational evaluation of TCAS II beginning no later than December 30, 1990. It was expected that 15-20 % of the U.S. fleet would be equipped with TCAS II by this date (FAA, 1990).
- 1990: Delivery of 5300+ TCAS units to the airlines by the summer (FAA, 1990). The TCAS Transition Program (TTP) began with first in-service flight. The initial logic is version 6.0 (Klass, 1994).
- 1991: Fokker Aircraft delivers first aircraft where pilots are expected to use pitch cue instead of VSI rate when responding to RAs (Mecham, 1991).
- 1992: Logic 6.04 was developed to mitigate the false alarm rate. Delta Airlines reported an 80 % reduction in RAs (MITRE, 2002)
- 1993: All air carrier aircraft with 30+ seats were required to be equipped with TCAS II. Aircraft with 10-30 seats were required to utilize TCAS I (MITRE, 2002). The TTP is extended to December 1993 due to operational problems (Vickers, 1992). It subsequently was continued through the initiation of the logic 7.0 (TTP Newsletter, 1996).
- 1994: A new version of the collision avoidance logic designated Version 6.04A was developed, including a modification to the resolution logic that eliminated altitude crossing RAs when the intruder and/or TCAS levels off with 1,000 feet of altitude separation (MITRE, 2002; Miller et al, 1994).

The FAA published Airworthiness Directives requiring its implementation in all TCAS II units prior to December 31, 1994. Airlines began equipping their fleets with the new 6.04A logic. Tested with computer simulations, the 6.04 logic RAs generated below 2500 ft AGL were reduced by 65 %, go-arounds by 77 %, and frequency of “bump-ups” by 44 %. Also, there was a 48 % reduction in false alarms for properly separated VFR traffic (Miller, et al, 1994). United Airlines began testing of TCAS for use in altitude changes on transoceanic flights (Aviation Week, 1993).

- 1996: FAA and industry began evaluation of RA Downlink at Boston’s Logan International Airport. (TTP Newsletter, 1996).
- 1997: The CAASD unit of MITRE completed version 7 of the TCAS logic (MITRE, 2002). It was subsequently approved by the RTCA standards committee and the FAA, and is the version that all new aircraft will be equipped with. It has also been adopted by the International Civil Aviation Organization (ICAO) as the international standard. Initially, version 7 will be installed on aircraft serving European and some other countries. American carriers who fly to these countries will have to upgrade from 6.04A to 7 on their aircraft used on international routes, and can voluntarily upgrade the equipment already on their US fleets. The version 7 logic is expected to reduce RA’s by at least 20% more than the previous version. This is expected to be the final logic for TCAS. (MITRE, 2002)
- 2001: Eurocontrol mandated TCAS-type systems (called ACAS; airborne collision avoidance system) using the new 7.0 type software beginning Jan 1, 2000. ACAS will be required for aircraft outfitted for more than 30 passengers and cargo transports with a take off weight of more than 15,000 kg (33,000 lb) (Miller et al, 1994).

4.3.3. TCAS Transition Program

Issues discovered in the TCAS Transition Program (TTP) pertain to TCAS hardware, software, and human factors (Tillotson & Gambarani, 1993). Hardware issues identified included TCAS/mutual suppression bus interface, erroneous Mode C replies, 1030/1090 MHz interference, and illegal Mode S address. These issues can be classified as system input variables; however, no systematic documentation on the impact of them on system output or human performance variables was found. Software/logic issues included coordinated altitude crossing TAs, low altitude TAs and RAs, low altitude-descend RAs, RAs during parallel and crossing runway operations, and high vertical rate encounters. These issues may be classified as intermediate/system output variables, but again no systematic documentation of the impact of these issues on human performance/behavioral variables was found. Finally, most issues directly related to human factors pertain to training and design of procedures. Pilot training issues were addressed in the FAA AC120-55A (1993) and in AC120-55B (2001), which specifies flight crew TCAS qualifications, TCAS training program requirements, and operational use practices.

In summary, addressing human factors issues related to TCAS implementation has paralleled gains in operational experience with the system. Give the complexity of the problem (i.e., airborne collision detection and avoidance) and the technological solution (i.e., TCAS) to it, it is very difficult to pinpoint any particular lessons to be learned, or what—in retrospect—could and

should have been done differently, although the importance of training of both pilots and controllers interacting with the system is manifest in the reviewed literature. Complete audit of the measures in response to the emergent problems with TCAS was outside the scope of this project, however. Furthermore, given the different operational use of TCAS and CDTI, it is questionable whether the TCAS lessons learned would be applicable to CDTI beyond methodological and procedural problems in addressing them.

4.3.4. Limitations of Data Extracted from ASRS Reports

The ASRS system is a trove of invaluable insights and information about complex systems in their operational context by first-hand accounts of the systems' users. Analysis of these reports, however, presents some thorny challenges for several reasons. First, there is no way of telling how many incidents involving a TCAS really occur annually, or how representative of potential problems the number of reports is. Second, the format of the report is free, which leaves no possibility for filling in missing information in retrospect; hence, the analyst is at the mercy of the reporters' meticulousness in providing all the relevant details about the event. Consequently, retrospective analysis is vulnerable to the analysts' biases and carefulness, which makes comparison of separate analyses tenuous and necessitates repeat analyses to be performed to research different questions that may arise in time. Third, it is very difficult to establish causal relationships between the TCAS development milestones and ASRS reports, as it is impossible to know the rate airlines have upgraded their TCAS equipment. There were no information in the reports about the particular version of TCAS involved in the incident, and often even TCAS II equipment were referred to simply as TCAS. Finally, one must be very cautious about normalizing the data, due to the fact that no "true" maximum is known. Therefore, it is crucial to bear in mind that all percentages are calculated from the reported incidents and to view these percentages against the actual number of reports for the given year.

4.3.5. Method

The ASRS database was searched using TCAS or TCASII as keywords; of the resulting 9,000+ reports, all reports from the landmark years of 1991–1994 were reviewed. In addition, reports from year 2002 were reviewed to gain an understanding of present-day situation. These years include a total of 3,825 reports. It was deemed necessary to review all reports from the chosen years rather than some sample, as no sampling method was found to be adequate for these kinds of data. As the reports were reviewed, they were classified according to the signal detection theory outcomes as hits, false alerts, and misses. Furthermore, each report was classified as positive, negative, or neutral.

If a report suggested pilots used the information from TCAS for decision-making or to enhance situational awareness, or if the pilot did not specifically mention anything negative, the report could be considered positive. Usually it was easy to classify the reports as positive or negative, as many pilots were either clearly "venting" their frustration with the system or praising it for saving their lives. Reports that were merely factual were classified as neutral. Similarly, if the pilot reported that in his or her opinion there was no conflict but TCAS issued a TA or an RA nevertheless, the report was classified as false alert. As was mentioned before, the analyst can only rely on the reporting pilot's perception of the event; it is not possible to determine whether a conflict in fact existed at the time of the alert. Hits were classified in an identical manner, as were misses. In the latter case, however, it must be kept in mind that it

appears very unlikely that an unaided pilot would detect a conflict missed by TCAS; a much more probable scenario is that both the pilot and TCAS would miss a conflict, which in that case would of course go unreported. Miss reports was evaluated under the assumption that TCAS was operating under normal parameters and limitations. Thus, if a TCAS equipped aircraft encountered a non-transponder intruder the report was not included nor was it classified as a miss. However, events in which TCAS II was known to be inoperative or written-up for maintenance were included, because (1) in all such scenarios the system was not disconnected or placarded “inop” and so was still active in the work environment, (2) such cases could present valuable evidence to pilots' over-reliance on the system as most reports on such scenarios indicated continued reference to the system despite knowledge of possible problems and (3) maintenance write-ups are often subjective and are viewed as such by most pilots and could easily be ignored.

After the reports were reviewed and classified as described above, analyses were performed at three levels. First, the total number of reports submitted by each year was tallied. Second, the proportions of hits, false alarms, and misses were calculated from the total number of reports submitted in each year. Finally, proportions of positive, neutral, and negative reports as well as those reported by ATC were computed from the total number of reports in the previous classes (i.e., hits, misses, and false alerts). Before presenting results from these analyses, a number of observations from the review of the reports will be described.

4.3.6. Results: Qualitative Observations

These observations are based collectively on the 3,825 narrative ASRS reports reviewed for this project.

4.3.6.1. Pilot-Related:

1. Pilots overuse TCAS; they often waited for the system to issue an RA before taking evasive action. This contributed to closer separation at the time evasive action was taken than if they had acted earlier.
2. The system was overwhelmingly considered a distraction in the cockpit, especially during critical phases of flight in the terminal area with numerous TAs and RAs while departing or arriving at an airport. This would happen despite the fact the RAs are supposed to be disengaged automatically at 1,000 ft AGL. Pilots have reported annoying false RAs and TAs while on short final, and while setting up for approaches.
3. Due to this distraction, pilots have “busted” altitude and heading clearance limits while sometimes missing radio calls due to the loud volume of the alerts.
4. Despite such complaints, there were many reports of pilots being extremely satisfied with the system and crediting it to preventing a near or actual midair collisions. Positive remarks were definitely the majority in all of the years analyzed.
5. Pilots reported being out of the control loop due to the heavy automation in the cockpit. Oftentimes the crew would begin to establish their situational awareness only after the alert or warning appeared.
6. Better CRM and task management in the cockpit would help solve the altitude and heading deviation problem. Upon receiving the warning, most of the time both pilots (or

all three) would have their heads outside of the aircraft attempting to acquire the intruder visually. Track and altitude deviations could be kept to a minimum if at least one pilot had their head inside the aircraft monitoring the systems. The task management issue at hand is the fixation pilots and crew tended to have towards the TCAS intruder. Other important aspects of the flight were immediately relegated to secondary tasks once the TA or RA happened.

7. Some pilots disliked the procedures regarding mandatory compliance with an RA. This is possibly a result of the high false alarm rate.

4.3.6.2. ATC-related:

1. There have been multiple misunderstandings with ATC. It seemed that controllers felt the implementation of TCAS was due to mistrust in their capabilities or an attempt to replace the human by a computer. Therefore, controllers disapproved of pilots responding to RAs and readily reported altitude and heading deviations that were a result of RA response.
2. Controllers exhibited a lack of familiarity with TCAS procedures and the system itself. This problem was subsequently mitigated through training interventions and TCAS familiarization flights.
3. Several reports made from foreign operating areas, particularly Spanish-speaking areas, revealed a lack of competence. Some pilots felt that TCAS should be mandatory for all flights in those areas. Some of these controllers did not even use English on all radio calls, which decreased pilots' situational awareness.
4. TCAS made it difficult in some circumstances for controllers to communicate with aircrews, resulting in missed clearances.
5. Some controllers had unrealistic expectations for pilots in terms of reacting to the TCAS. Some mentioned that they would like pilots to query them before responding to an RA. Most of the time, however, and especially in the situations where the RA was legitimate, there is no time for that.
6. Some controllers viewed TCAS as an excuse for pilots to deviate from a clearance without permission. However, controller acceptance of TCAS into the National Airspace System has improved over time.

4.3.6.3. System-related:

1. The volume of the alerts and advisories is overwhelmingly regarded as much too loud. Passengers in the first few rows of coach seating have reported being able to hear the TCAS alerts.
2. A large amount of false alarms were a result of aircraft climbing to an altitude at least 1000 ft lower than the airplane receiving the alert. If the TCAS were connected to the altitude alerter, the false alarm rate might drop considerably.
3. The false alerts and advisories in the terminal area are numerous. Some TCAS units were even issuing RAs and TAs for traffic that was on the ground. Upon receiving these alerts and advisories, pilots were distracted during a critical phase of flight and a few had chosen to disengage TCAS under such circumstances. These situations, known as the

“Dallas Bump”, were drastically reduced following the implementation of software upgrades. Furthermore, pilots were encouraged to utilize the AIM method of climbing to an altitude by reducing the rate of climb for the last 1,000 ft before the desired altitude.

4. There were quite many encounters in 1993 with VFR traffic flying 500 ft above or below ownship that TCAS II alerted as conflicts, although adequate regulatory separation existed. These were classified as hits, as they were within TCAS parameters; however, controllers and pilots alike were annoyed by these events.

4.3.6.4. ATC-Air Crew Interaction:

1. Different algorithms of different conflict detection systems sometimes created disagreement as to whether an actual conflict exists. One controller submitted an ASRS report wondering if their airspace planning should be structured for the shortcomings of TCAS RAs.
2. The ATC community’s attitude towards pilot response to RAs during the initial period of system implementation supports the need for the downlink of TCAS-related information. In numerous cases, controllers resisted pilots’ response to RAs by instructing them to disable TCAS in the terminal area or by discouraging responses on behalf of an RA. Many expressed disappointment in pilots trusting the system more than their instructions.
3. Whatever influence ATC had on pilot response to RAs was neutralized by regulatory action, which took controllers further out of the sphere of influence in terms of pilot responses to TCAS. The effect controllers had on pilot RA response cannot be determined conclusively from the ASRS reports. The motive behind TCAS was to create a checks-and-balances system of collision avoidance, so this regulatory action was appropriate on many levels. On the other hand, it may have removed the controller too far from the control loop of keeping aircraft separated.

4.3.6.5. Air Crew and ATC Conflict Assessment:

1. Pilot and controller reaction to and anticipation of a TA or an RA can be attributed to the perceived probability of conflict base rate in certain airspace. The actual conflict base rate is typically very low (Feldman and Weitzman, 1999).
2. The pilot’s perception of this conflict base rate is manifested in their reaction to the alert given the airspace they are flying in. This would translate to pilots being likely to express their perception of reduced conflict potential in various different ways, particularly disengaging the system or ignoring its indications.
3. Controllers would express their disapproval of TCAS in the discouragement of aircrew relying on TCAS. On the other hand, pilots’ approval of TCAS and a perception of higher conflict probability would be manifested in their trust of the system by allowing it to augment their SA by accepting its indications and using it to anticipate developing situations. Of course individual experiences with the TCAS system influence how a pilot will utilize the information that is presented in the cockpit.

4.3.7. Results: ASRS Report Analysis

The number of reports filed annually rose steadily through the first three years of TCAS implementation (1991–1993), cresting at 1,045 reports in 1993, but declined in 1994. Further

decline was evident in 2002, with only 371 TCAS-related reports total (Figure 4.1.). It is important to keep these trends and totals in mind when interpreting further results.

Figure 4.2. depicts the proportions of false alarms, hits, and misses out of the total number of reports filed during the selected years. The proportion of false alerts reported has declined since 1993 while the proportion of hits reported has steadily increased. The former may reflect the introduction of logic 6.04 in 1992 and logic 6.04A in 1994. However, one must be very cautious about inferring causal relationships from these data, as it is not known at what rate airlines had upgraded their equipment or what particular version of TCAS software was involved in the reported incident. Also, the proportions may reflect pilots' changing biases in reporting particular incidents. Hence, it may be that as pilots became familiar with TCAS and learned to anticipate TAs and even RAs, they were less likely to report false alerts (i.e., they had become false alert tolerant). Similarly, as they had become more appreciative of the technology, they may be more likely to report incidents when TCAS had "saved the day," or hits. Again, however, it is important to consider the total number of reports, as depicted in Table 4.1.

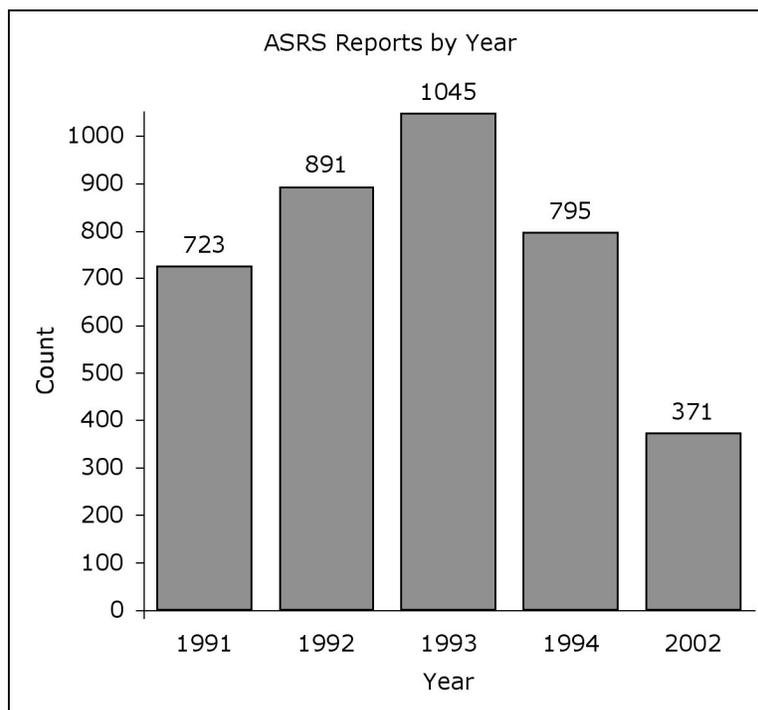


Figure 4.1. All TCAS-related incidents reported to the ASRS during the "landmark" years of TCAS development

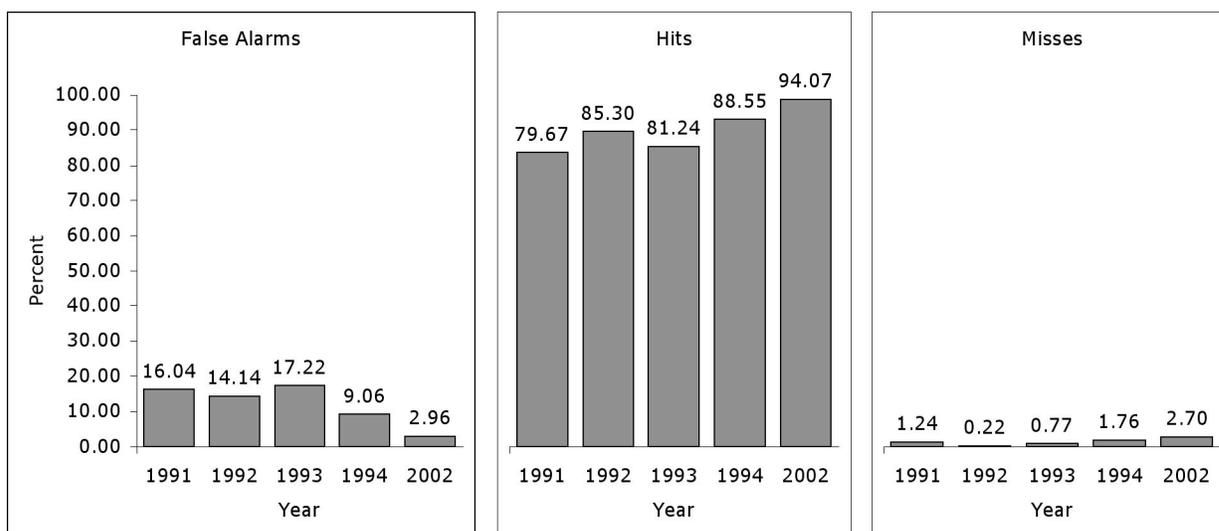


Figure 4.2. Proportion of false alarms, hits, and misses of the ASRS reports filed in the landmark years. The proportion of false alerts has declined since 1993, likely due to the introduction of logic 6.04A in 1994. The results may also reflect pilots' disposition towards filing ASRS reports for different events.

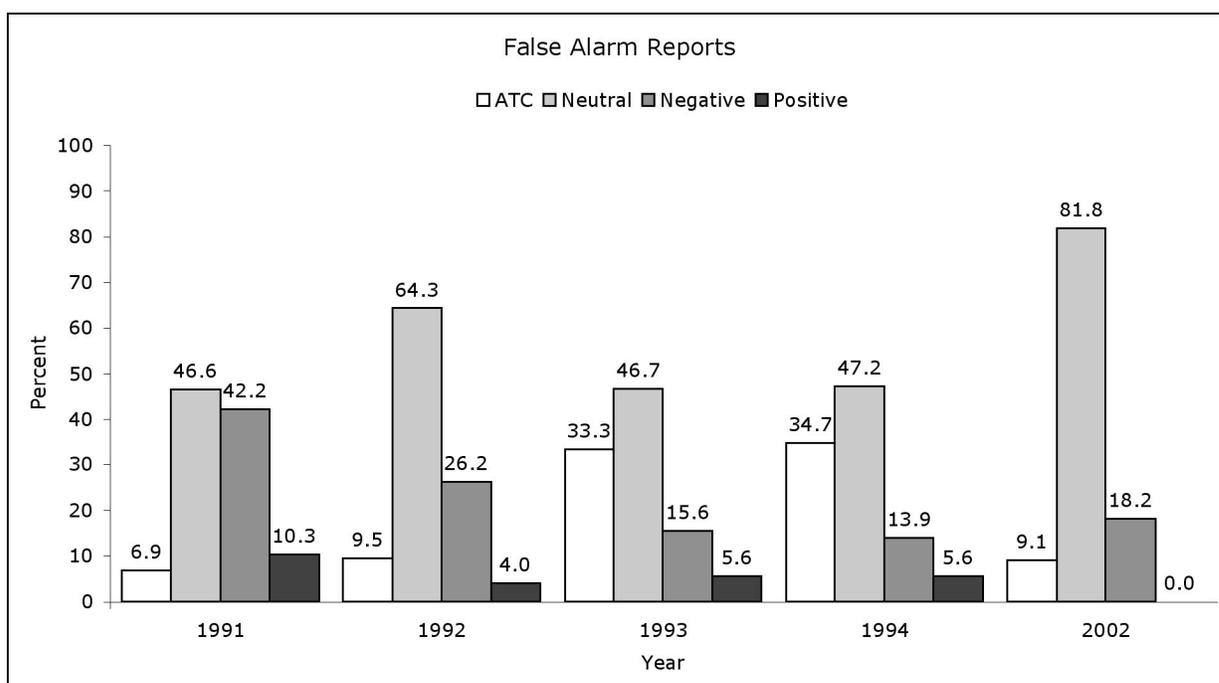


Figure 4.3. False alarm reports by year and by proportions of those reported by ATC as well as neutral, negative, and positive pilot reports.

The next level of analyses offers an interesting view to pilots' attitudes towards TCAS. In Figure 4.3, are depicted proportions of ATC reported as well as neutral, negative, and positive pilot reports. While false alerts are commonly thought of as detrimental to pilots' attitude and usage of TCAS, it is important to notice that in most years there were significantly fewer negative reports than neutral reports, and even in 1991, when TCAS was introduced with its original logic, there were more false alerts viewed neutrally than negatively. It is also interesting to note that a substantial proportion of false alert reports were positive in tone. It might be assumed that in these cases pilots were generally aware of the situation and were perhaps anticipating the TA; also, reviewing pilots' positive comments about false alerts reflect an overall positive attitude towards TCAS, that is, even if the reported incident was a false alert, such occurrence did not dampen their general enthusiasm about the technology.

The proportions of neutral, negative, and positive reports of all hits reported to the ASRS follow a predictable pattern (Figure 4.4.). A vast majority of the reports reviewed were factual (and hence classified as neutral), but the proportions of positive hit reports were substantially larger than that of negative ones for every year. The latter typically involved cases where the pilot disagreed with the RA, had visual contact with traffic and felt that he or she could have solved the conflict better, or when the pilot was frightened or distracted by the TA and/or RA. In other words, the complaint was usually made about the manner the system functioned, not because it worked as intended.

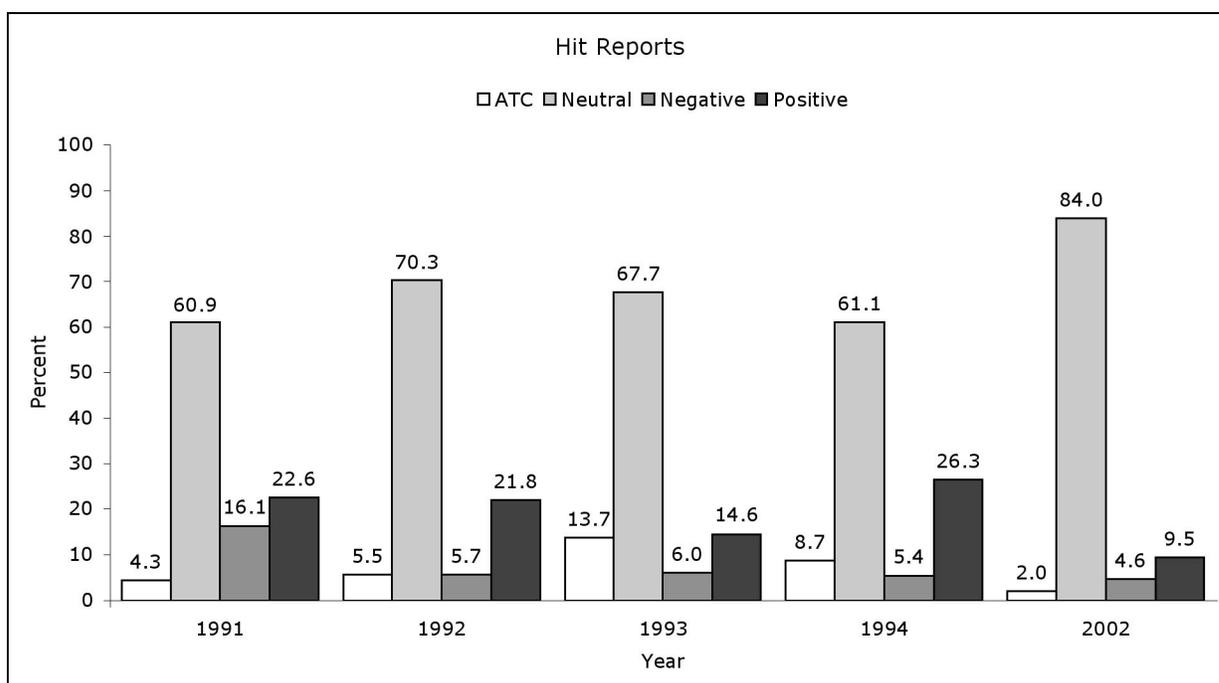


Figure 4.4. Hit reports by year and by proportions of those reported by ATC, as well as neutral, negative, and positive pilot reports. The results are predictable in the sense that positive reports

outnumber negative ones; the latter usually involved complaints about the particular RA suggested by TCAS or it interfering with cockpit tasks and routines.

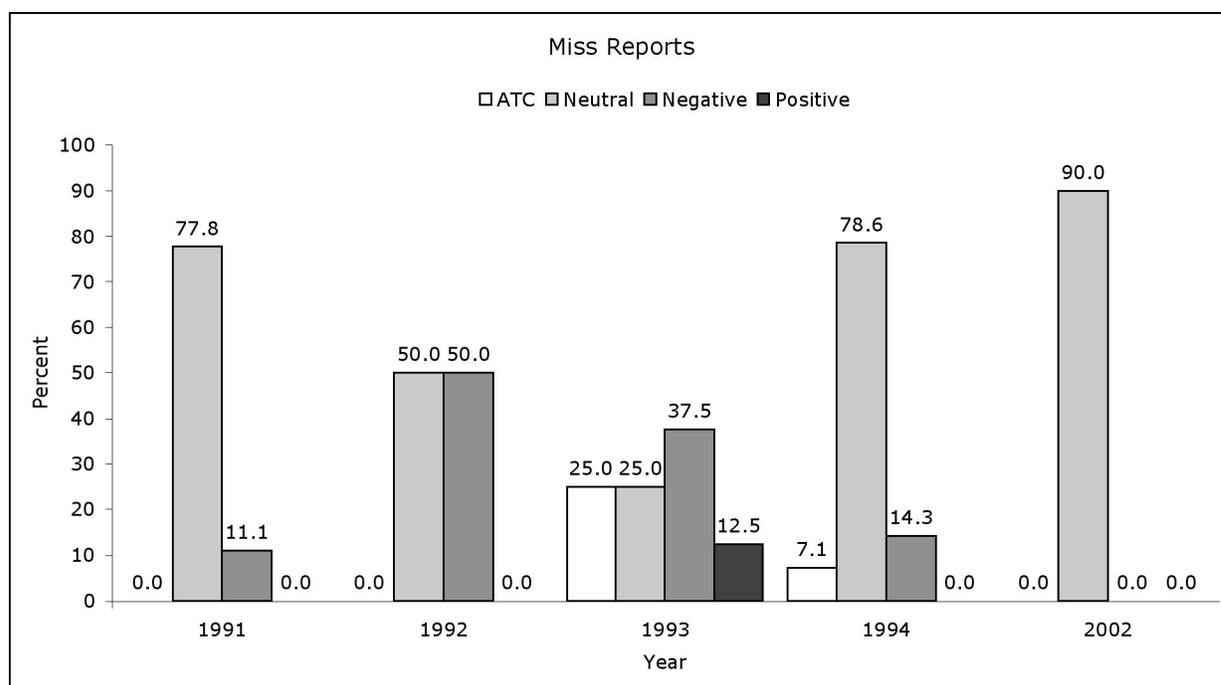


Figure 4.5. Proportions of miss reports by year and by proportions of those reported by ATC as well as neutral, negative, and positive pilot reports. The total number of miss reports was very small, however; for example, in 1992 only 6 such reports were filed.

Finally, reports involving missed conflicts are depicted in Figure 4.5. Interpretation of these results must be moderated by the very small number of such reports across all reviewed years. These cases were often reported with crews acquiring visual confirmation of an aircraft at close range. Situations where the intruder did not have a transponder or TCAS was inoperative were not regarded as a miss. The proportions of neutral, negative and positive reports in this case in particular should be evaluated against the total number of reports, as depicted in Table 4.1. below.

4.2.8. Summary

The trends in ASRS data appeared to be consistent with the data and results from the TCAS Transition Program (Tillotson & Gambarani, 1993). Given the limitations detailed in section 4.2.2., quantitative inferences drawn from ASRS data must be approached with caution. However, as should be apparent from section 4.2.4., qualitative analysis of the reports can be extremely valuable in revealing potential problems early on and by suggesting directions for further research. The most important contribution of quantitative analysis of the data was insights into the reporters' attitudes towards TCAS, in particular the fewer-than-expected negative comments about false alerts.

In terms of “lessons learned” from TCAS implementation, a few additional conclusions can be drawn. First, training and regulations should be standardized for controllers and aircrew and implemented prior to operational use of the system. The TCAS training and regulatory interventions was clearly created and modified on an “as needed” basis. The situation was similar to creating the rules of a game after play has begun. The regulatory and training systems were not prepared for the operational implementation of TCAS. Second, TCAS’ operational use has strongly supported the need for a downlink between flight decks and ATC. These issues are clearly generalizable to introduction of any new technologies in the operational environments, including CDTI.

Table 4.1.

The total number of TCAS-related ASRS reports for the landmark years of TCAS development and implementation, broken down by the signal detection outcome and whether the report was filed by ATC and the pilot’s perception of the reported incident (neutral, negative, or positive). Note that the totals may include reports that were not classifiable or were classified in more than one way (e.g., because they included both positive and negative comments).

Year		FA	Hit	Miss	N/A	Total
1991	Total	116	576	9	20	723
	ATC	8	25	0	1	35
	Neutral	54	351	7	6	419
	Negative	49	93	1	10	153
	Positive	12	130	0	2	144
1992	Total	126	760	2	3	891
	ATC	12	42	0	0	54
	Neutral	81	534	1	3	619
	Negative	33	43	1	0	77
	Positive	5	166	0	0	171
1993	Total	180	849	8	8	1045
	ATC	60	116	2	0	178
	Neutral	84	575	2	7	668
	Negative	28	51	3	1	83
	Positive	10	124	1	0	135
1994	Total	72	704	14	5	795
	ATC	25	61	1	1	88
	Neutral	34	430	11	4	479
	Negative	10	38	2	0	50
	Positive	4	185	0	0	189
2002	Total	11	349	10	1	371
	ATC	1	7	0	0	8
	Neutral	9	293	9	0	311
	Negative	2	16	0	0	18
	Positive	0	33	0	0	33

4.4. Documents and Guidelines Concerning Alerts

Below is an outline of information extracted from documents regarding TCAS 2 and EGPWS certification, most relevant to the system reliability and false alert rates of such systems, as well as those pertaining to CDTI design. We consider how these documents are relevant to certification of the CDTI for applications 9, 10, and 14 from RTCA DO-259 (Applications focusing on conflict and collision SA and avoidance maneuvers; RTCA, 2000).

4.4.1. Assumptions of FAA Concern for CDTI Approval

1. Establishing design standards and recommendations (e.g., color, modes, alerting logics for CDTI)
2. Performance of CDTI: (a) System (algorithms) on bench tests, using scenarios that could generate high error rate, and (b) pilot- in- the- loop in flight tests, using scenarios found to be problematic from above.
3. Primary goal of CDTI (for RTCA DO-259 application areas; RTCA, 2000) is to (a) avoid misses (late alerts), (b) minimize FA (to avoid unwanted maneuvering and reduced trust), (c) support safe but efficient strategic maneuvers (all 3 axes of flight: vertical, lateral, speed), and (d) integrate seamlessly with TCAS and potentially also terrain and weather displays.

4.4.2. Review of Non-CDTI Documents. Each document is identified by a title and a unique letter code.

4.4.2.1. FAA AC-20-131A, TCAS Certification. (A)

This document describes different scenarios for TCAS testing (conflict geometries). It specifies that FA rate should be “improbable” but offers no exact value. It also specifies that “incorrect RA” (which includes both false and missed or late RA’s) should be “improbable” with a number of 10^{-5} per flight hour en-route, and 10^{-4} per hour in TRACON environment. The document provides a denominator of conflicts-per-hour, which allows inference to be made of the failure/encounter rate of no more than (approximately) 10^{-3} failures/encounter (same for TRACON and en-route environments). There is no specification of where the values come from, nor is there a breakdown of how failures should be partitioned into FA versus late alerts (e.g., misses at a particular LAT value).

4.4.2.2. MITRE Safety study: McLaughlin, 1997. (B)

The purpose of this study was to show the predicted benefits of TCAS-7 on conflict avoidance and conflict behavior, relative to earlier TCAS version 6.04. A sample of 4300 conflict encounters that would have triggered an RA had TCAS been in place at the time, were drawn from a data base of flight behavior during the 1990–1991 time period (Owen, 1993). Assumptions were made about flight path uncertainty, by adding “jitter” to position estimates. Encounters were also classified by a number of different features, known possibly to influence the effectiveness of the algorithm (e.g., whether traffic or ownship was level or transitioning before or after the point of closest approach). For each of these encounter classes, algorithm performance of the different TCAS versions was calculated in terms of a “risk ratio,” which was the ratio of near midair collisions (NMACs), with TCAS to those without TCAS. From these

statistics it is impossible to deduce any “false alarm” statistics, because the criterion is simply that of NMAC separation standards (1000 lateral and 200 vertical feet), not the typical LOS protected zone (5 miles and 2000 feet). The author demonstrates how the value of the safety parameter is influenced by the different classes of geometry, and provides a useful taxonomy of these classes.

4.4.2.3. MITRE safety study: Chemiavsky, 1997. (C)

Using the same 4300 encounters described in (B), Chemiavsky (1997) provides a bench test comparing TCAS 7 with TCAS 6.04 performance, as reflected in a variety of performance metrics. Importantly, some of these correspond to “misses” (or late alerts), and one specifically corresponds to False Alerts (RA that would have resulted in a safe encounter with no TCAS). Probabilistic behavior was induced by adding “jitter” to position estimations, and by adding variability to pilot RT (although author does not specify how much variability). Each encounter was subjected to 10 Monte-Carlo runs. Performance metrics revealed approximately 4,000—5,000 FAs out of the 43,000 runs. This would suggest an approximately 10% FA rate, shown to diminish from TCAS-6.04 to TCAS-7. The miss rate data are expressed as the occurrence of NMACs, which are constant (and very low) across both systems.

4.4.2.4. FAA TSO-C147, Technical Standard for Traffic Advisory System Airborne Equipment. (D)

This appears to be umbrella document guiding specification of TCAS as well as CDTI alerting functions. It provides considerable detail regarding minimum standards for traffic depiction, alert characteristics, etc., and specifies minimum of two levels of alerting (prior to RA). The document specifies allowable use of either spatial or time (τ) separation but appears to make no assertions regarding false alarm rates of algorithms, except as expressed in this quote: “The acceptability of these aural annunciations must be reviewed during flight test. The following factors at a minimum must be evaluated for acceptability: quantity of unsolicited annunciations...” (Appendix 1 3.1.2a(3)).

4.4.2.5 EGPWS (Enhanced Ground Proximity Warning System) Certification document N8110.64. (E)

This document specifies an allowable failure rate (10^{-4} ; we assume the denominator is per/flight hour) but it does not specify how this value should be allocated to false alarms versus “misses” (late alerts).

4.4.2.6 EGPWS Advisory Circular 25-23. (F)

This AC describes flight test requirements and considerations for flight test. It appears that these requirements are those that would be “pushing the envelope” of EGPWS performance (e.g., “worst case” scenarios to challenge the algorithms).

4.4.2.7 Honeywell EGPWS flight test document. (G)

This report documents the fault tree and failure modes effects analysis done on EGPWS. It shows an example of a fault tree from which is derived an estimated false annunciation of a pull up warning of 2.06×10^{-5} per flight, or 2.94×10^{-6} per flight hour. It does not appear to contain pilot performance data.

4.4.2.8 Krois (1999b). Human Factors Assessment of URET Conflict Probe False Alarm Rate. (H)

The paper by Krois documented actual FA rate and miss rate in operation of URET system for Air Traffic Control. FA rate was high (approximately 50%), but miss rate very low, thus demonstrating (a) the appropriate criterion setting for low-base-rate events (conflicts), that will sacrifice high false alarms to avoid misses (see Section 2), (b) the necessary high error rate (50% FA) for a system that is making projections the long LAT characteristics of URET. This 50% rate, with the long (20 min) LAT of URET, can be compared with the 10% rate estimated by Chemiavsky, with the shorter LAT of TCAS, and the fact that this latter lower rate is estimated based on resolution advisories, which are issued at smaller separations than the loss of separation used in URET. As noted in Section 3, long LATs and large PZs are the ingredients for high FA rates.

4.4.3. Review on CDTI-Specific Documents, with Information Pertaining to Alerts.

4.4.3.1. RTCA DO-243 (1998). (I)

This document, prepared by SC-186, appears to be an advanced version of a MOP (the latter which does not yet appear to exist), providing good human factors design guidelines. It is focused on near term applications for CDTI, provides many design guidance options for the visual symbology, including leader lines (predictors), and automatic decluttering, but offers very little in terms of “best practices” prescriptions. No effective treatment of alerts, false alerts, or levels of reliability are provided. Discussion of traffic alerts (in their Section 2.1.1.2.1) refers to TIS MOPS (RTCA DO-239, 1997) and specifies that if there are multiple alerts, they should be prioritized by time-to-closest approach (TCPA). The document specifies that it is a “stand-in” until the MOPS gets written.

4.4.3.2. RTCA DO-239 (1997). (J)

This document covers MOPS for TIS, not yet acquired, and a draft form labeled MOPS for CDTI. It does not appear to contain much for warnings, but has good coverage of basic visual symbology guidelines, consistent with the SAE-G-10 below, and RTCA DO-243 above. It only says “alerts are recommended”, but does not present discussion of false alarm issue, LAT, etc. Only guidance provided states to “be consistent with TCAS.

4.4.3.3. SAE-G-10. (K)

This document, titled “Human Interface Criteria for CDTI Technology” (SAE ARP 5365, Date?) offers some of the most comprehensive “aerospace recommended practices,” regarding human factors of CDTI design, and references much of the older as well as more recent human factors literature. The document specifies that the CDTI should “alert the loss of separation,” although it does not specify what that separation is. It is explicit in addressing the false alarm and “nuisance” alert issues, as well as the distinction and tradeoff between these and misses. It prescribes “eliminate false alerts, minimize nuisance alerts” and describes the critical importance of early warnings, issued in time that a range of maneuvers can be chosen. It alludes to the importance of trust (confidence) and the degradation of confidence with more false alerts as discussed here in Section 2. It provides a good description of the visual graphics display options, including dimensionality (2D-3D), predictor information, decluttering, altitude coding, overlay, and consistency with other systems. It is the best reference to human factors research literature of

any document. No quantitative numbers offered however (e.g., maximum FA-rate, preferred LAT).

4.4.3.4. Draft FAA Advisory Circular ADS-B. (L)

This document, titled “Airworthiness and operational approval considerations for traffic surveillance” and its Appendix U: “Human Factors” contains discussion of visual symbology parallels much of that within RTCA DO-243 and SAE-G-10. Treatment of alerts does not address false alerts or LAT and only vague references are made to prediction and future uncertainty. Appendix M of this document also contains suggestion for mitigating the effects of “rogue aircraft” (transponder off, or transponder failed).

4.4.3.5., Uhlarik, J., & Comerford, D. A. (2003). Information requirements for traffic awareness in a free-flight environment: An application of the FAIT analysis (DOT/FAA/AM-03/5). (M)

This document describes a human performance analysis of the factors that will influence pilot performance and the pilot’s use of the CDTI in a free flight environment. Thus it is the most future-looking FAA document regarding the far-term applications of the CDTI. While it does not contain guidelines for design, it does contain implicit recommendations by identifying factors most likely to influence pilot use of the CDTI (weather, piloting skills, time of day, terrain, ownship state, workload level and time-pressure). The analysis highlights the importance of training, and the document contains a good description of NASA’s CDTI system. There appears to be little content that addresses the alert issue.

4.5. CDTI Non-Alerting Guidelines

4.5.1. Overview

As noted in Section 4.3, there are some vague, but not precise guidelines for CDTI alerts, emanating from RTCA DO-243 and –239, SAE-G-10, and the draft FAA Advisory Circular for ADS-B (see Sections 4.4.2.1., 4.4.2.2., 4.4.2.3., and 4.4.2.4., respectively). A review of these documents reveals considerably more guidelines that pertain to other “non-alerting” functions of the CDTI, as would be expected, given that the CDTI has been approved for certain non-alerting uses in its near term applications (e.g., RTCA DO-239 and -243). However, some of these guidelines bear indirectly on the alerting functions for the following reasons:

1. Just as alerts represent the explicit prediction of an imperfect automation system that there will (could) be a conflict at some point in the future, so various visual features of the CDTI, such as leader lines or future ground tracks make implicit assumptions about the future trajectories of traffic and ownship—assumptions that could also prove to be incorrect. These visual features may or may not map into discrete categorical levels, signaled for example by color changes, or by the sudden appearance of a previously decluttered aircraft, which has crossed some threshold level of danger. We refer to this as “visual alerting”.
2. Decluttering tools implicitly are designed to remove “less threatening” aircraft at a criterion, along a continuum inherent in the CDTI logic (and, as noted, designed to present them when they exceed a presumed threat level). We refer to this as “implicit alerting.”

RTCA DO-243, -239, and SAE-G-10, referenced in Section 4.3 (as I, J, and K, respectively), all contain various recommendations that bear on these issues of visual and implicit conflict alerting. We summarize some of the important categories of recommendations below:

4.5.2. Visual Alerting

Document I (RTCA DO-243 sometimes referred to as CDTI MOPlet) makes reference to leader lines, with a particular value of 60 seconds. This implicitly suggests a threshold of certainty, regarding the aircraft's future behavior within the next 60 seconds. Document J (RTCA DO-239 Draft) makes reference in 2.2.2.2.8.1 to a lateral trend vector, asserts that it is optional, and that if a predicted time is incorporated (i.e., by the length of the vector), it should be invariant across differences in display scale (the term "should" is meant to be advisory, in contrast to "shall" which signals mandatory). However, the document does not suggest what the predicted length should be. Document K (SAE G-10, ARP) provides a more elaborate discussion of options for presenting various visual indications of predictive information in Sections 7.1.2 and 9.1.2.1, but again, provides no indications of how long those predictors should be (in terms of LAT). The document provides the suggestion for discrete alerting logic, when "non-transgression zones" are entered, although not specifying the size of such zones.

4.5.3. Traffic Highlighting and Decluttering

These issues addressing implicit alerting are discussed by RTCA DO-243 (Sections 2.1.1.2.6 and 2.14, respectively), and SAE-G-10 (Section 9.1.1.5). Much of the discussion of decluttering focuses on the manual interaction necessary to achieve it (e.g., use a minimum number of keystrokes necessary to accomplish), as well as the risk variable that should be used to define the level of relevance (or prioritization) of traffic to be depicted. This risk variable is typically stated to be a combination of time to contact (TCPA) and distance. (e.g., RTCA DO-243, 2.2.2.2). In this regard, the recommendations follow the guidance of TSO 147 (TSO for Traffic Alerting Systems). There are no specific target values of time or distance cited for crossing these threshold risk levels of implicit alerting. The issue of clutter is closely related to the issue of range of coverage. A larger range display, within the same physical display space, will incorporate more traffic in a smaller region, hence amplifying the negative effects of clutter.

4.5.4. Vertical Representation

A somewhat surprising omission from RTCA DO-243 (I) is any recommendations regarding the nature of representation of vertical information in spite of the well-documented influence of that representation on maneuver choice (see Section 5.3). Options for symbolic-alphanumeric, or graphic representations are offered without guidance for which is better, or how the latter should be offered. This omission is potentially important, given current concerns in other aspects of cockpit design for vertical situation displays. The SAE-G-10 document (K) goes into some depth in its discussion of graphic displays, 3-D displays, and plan view Vertical Situation Displays. These are important because of their value in supporting conflict awareness, as well as their implicit role in guiding the choice of conflict avoidance maneuvers (lateral vs. vertical; see Section 5.3 of the current report).

4.5.5. Flight Deck Integration

All documents describe the importance of integrating, or “harmonizing” the CDTI with TCAS. Section 8 of SAE-G-10 (K) provides a more elaborate discussion of harmonization with other displays, including the roles of consistency and overlay, and the consequences of overlay (e.g., traffic and weather) for clutter.

4.5.6. Failure Modes Analysis

An important facet of CDTIs is the analysis of possible reasons for “failures”, and failure likelihoods. This analysis is contained to some degree in Appendix D of RTCA DO-239 (J), describing the differences between the surveillance sources that drive the CDTI, and in Document L, the draft FAA Advisory Circular for ADS-B Appendix L provides a good hazard assessment analysis of different causes of misleading or missing information on a CDTI. This information is particularly important in two respects. First, failure modes (such as false alerts) obviously impact trust. Second, system aspects that have low failure rates may engender “complacency” such that when the occasional failure does occur, safety might be compromised (see Section 2).

4.6. Conclusions and Recommendations

1. TCAS developers have done a good job of incorporating feedback, and problems into upgrades. Although this feedback has generally been qualitative, rather than quantitative.
2. ASRS TCAS reports reveal that such reports are equally related to false alerts as to misses, reflecting a decrease of FA-related reports since the early 1990s. There are also fewer negative comments about False Alerts than might be expected, suggesting that there is good pilot understanding of the causes and consequences of such events.
3. Regarding false alerts and trust, a review of the documents reveals that, with the exception of SAE-G-10, none of the reviewed documents appear to describe the negative consequences of false alerts on operator trust, nor late alerts on restricting maneuver opportunities. The SAE-G-10 document does not offer specific performance tolerance levels, nor report the findings of pilot-in-the-loop studies of traffic alert reliance, either. None of the materials provide source documents for assumptions on pilot response time (3 seconds), reasons for particular reliability level criteria (e.g., 10^{-5} per flight hour), nor the source and magnitude of variability of future traffic trajectory, which would create alert imperfections (e.g., Magill, 1997). Specifically, we conclude from these documents that:
 - a. There is no separate specification of allowable FA rate vs. miss rate for any of the systems. Documents only specify failure rate (10^{-x} /flight hour). These prescriptions do not generally take into account the low-base-rate requirement to reduce the alerting threshold (see Section 2).
 - b. FA rate is requested to be “improbable.” In SAE-G-10, there is a prescription to “eliminate false alerts, minimize nuisance alerts.”
 - c. Actuarial data suggest these false alert rates may be around 10% (for TCAS), 50% (for URET).

- d. Some guidance is provided on the nature of flight tests (for TCAS, but not for CDTI).
 - e. Documents do not generally provide any guidance on pilot trust, nor empirical basis for why particular specified reliability rates are chosen as desirable.
4. Regarding the visual properties of CDTIs, we conclude from the document review that:
- a. Several visual functions of CDTIs, like predictive leader lines or automatic decluttering (or restoring) provide implicit visual alerting and their behavior should be harmonized with auditory alerts.
 - b. A concern that should be addressed is the role of time vs. space in traffic alerting, since these are not always perfectly correlated in their contributions to risk.
 - c. More focus should be places on vertical representation of traffic.
5. Failure modes analysis must be applied to CDTI systems to identify implications of non-transponder aircraft.

5. CDTI NON-ALERTING STUDIES

5.1. Overview

In the following, the results of pilot-in-the-loop evaluations of CDTI effectiveness on pilot conflict detection, traffic-avoidance maneuver choice, and maneuver effectiveness are summarized. The section does not address studies that have addressed discrete CDTI alerting, but focuses instead on those factors that influence the above-mentioned performance variables. From this research emerges a valid picture of what aspects of a conflict situation present the greatest challenge for the pilot's use of the CDTI, and therefore should be reflected in candidate scenarios. Figure 5.1 presents an overview of several factors that influence CDTI use, as well as those that measure the effectiveness of that use. Some, but far from all, of these factors have been captured in the pilot-in-the-loop research reported. This research represents a range of experimental scenarios from those conducted in part task simulations to data collected in full mission simulators. Results of this research are described in the present section.

5.2. Time Pressure and Inherent Maneuver Preferences

It appears that time-pressure is an important driver of the kinds of maneuvers that pilots choose to use with a CDTI (maneuver tendency). In order to understand these effects, the current section will also consider some of the inherent properties of maneuvers on the three different axes defined by the vertical (climbs vs. descents), lateral (turning) and longitudinal (slowing or speeding) that can be used to avoid a loss of separation. Two fundamental aspects of the current airspace should be kept in mind in considering this material: (a) current FAA regulations (CFR 91.113) dictate procedures only for lateral maneuvering; (b) TCAS directs only vertical maneuvers.

5.2.1 Time Pressure Effects on Maneuver Tendency

Early CDTI research carried out in commercial aircraft simulations of large aircraft revealed mixed evidence as to the kinds of maneuvers pilots inherently chose in traffic avoidance scenarios using a CDTI, with some evidence for lateral preference (Smith, Ellis, & Lee, 1984; Ellis, McGreevy, & Hitchcock, 1987), and some results showing a vertical preference over lateral maneuvers (Abbott et al., 1980; Chappel & Palmer, 1983). Importantly Palmer (1983) found that increased time pressure eliminated a lateral preference. Palmer's finding is consistent with a set of studies carried out at University of Illinois, involving GA pilots with light aircraft handling dynamics, and relatively shorter (less than 2 minute) look-ahead time for the CDTI. **All of these studies found a strong preference for vertical maneuvers over both lateral and airspeed maneuvers** (Wickens, Gempler, & Morphew, 2000; Merwin & Wickens, 1996; O'Brien & Wickens, 1997; Alexander & Wickens, 2001, 2002; Wickens, Helleberg, & Xu, 2002; see Merwin, Wickens, & O'Brien, 1998; Alexander, Wickens, & Merwin, in press, for summary reviews).

A second conclusion, drawn from all of the Illinois studies, is that pilots **avoided making airspeed maneuvers**. Only one commercial aircraft study (Chappel & Palmer 1983) revealed that pilots preferred airspeed adjustment to avoid conflicts. Other commercial aircraft studies "froze" airspeed, so this was not an option available to pilots.

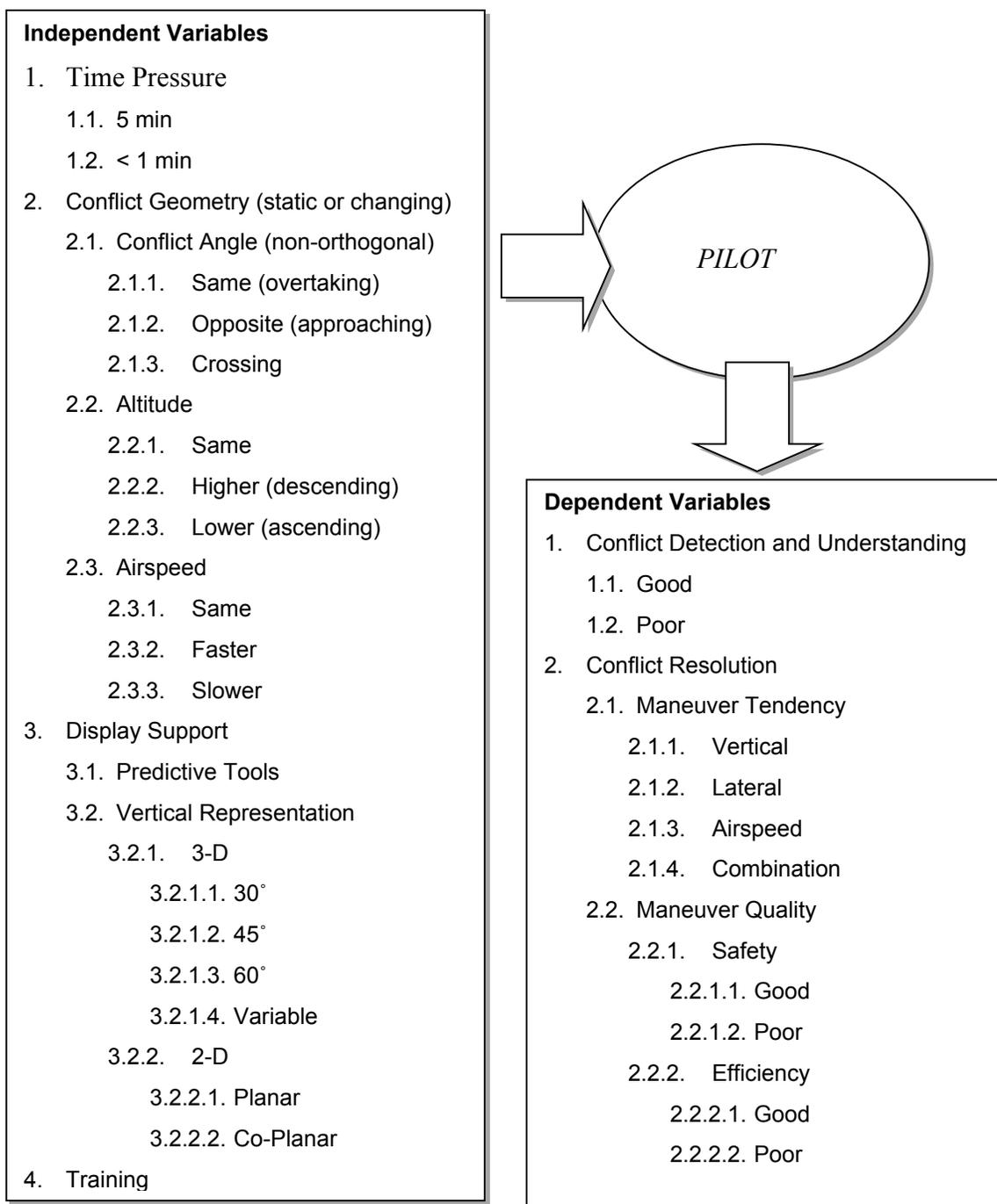


Figure 5.1. A model of four major classes of independent variable that affect pilots' ability to detect and resolve airborne conflicts.

The observations of a general tendency to use the vertical and avoid both lateral and airspeed maneuvering, when under time pressure, is consistent with two characteristics that differ across the three axes:

1. Analysis by Krozel and Peters (1997) clearly reveals the difference in the time effectiveness of the three, in which vertical maneuvering can more rapidly avoid a loss of separation than can airspeed maneuvering, and airspeed maneuvering in turn is more time effective than lateral maneuvering. This ordering is consistent with the ordering of dynamic system control order (Wickens, 1986), in which vertical maneuvering is a 2nd order control of position, airspeed is a 2nd order control with a lag, and lateral is a 3rd order control. Higher control orders impose greater lag to change position.
2. Higher control orders impose greater cognitive complexity on pilots trying to visualize the effectiveness of the maneuvers (Wickens, 1986). Thus there is higher complexity of the lateral than the vertical maneuvering. Furthermore, the future implications of speed adjustments are difficult to visualize on a CDTI since the display represents space, not time, but the latter is directly affected by speed.

According to this analysis, as time pressure increases pilots will be influenced to avoid maneuvers that take longer to accomplish and also require higher cognitive workload to visualize. Hence, the increasing preference of vertical over airspeed and lateral maneuvers as time pressure increases can be well understood.

A final conclusion drawn from the studies above is that **pilots tend to avoid complex dual axis maneuvers**.

5.2.2. Time Pressure and Maneuver Choice Effects on Safety

In CDTI research, “safety” has typically been defined operationally as either poor performance in detecting conflicts, an actual loss of separation, or a loss of separation that is predicted to occur within a particular LAT. Because the first of these is rarely assessed in empirical research, and the second is extremely rare, most conclusions regarding safety implications of the CDTI are based on the third criterion: a safer CDTI is one that supports less time in a state of “predicted conflict” within some TCPA.

The most consistent conclusion that appears to emerge from the research that has measured safety is that **vertical maneuvers (induced by high time pressure) are less safe than lateral maneuvers** (Wickens, Helleberg, & Xu, 2002; Alexander & Wickens, 2001, 2002), leading to greater time in a state of predicted conflict. On the surface there appears to be a contradiction here in that, as described above, because of its lower order, vertical maneuvers can be accomplished later (shorter TCPA) than lateral, and therefore they are more flexible. Hence the greater loss of safety (time in a state of predicted loss of separation) observed in empirical research may be a result of the fact that pilots initiate vertical maneuvers later (shorter TCPA) than lateral, offsetting their speed advantage.

A second conclusion, offered with slightly less certainty is that, within the vertical axis, descents are less safe (more time in predicted conflict) than climbs (Alexander & Wickens, 2002). Less certain still are the safety differences between different classes of lateral maneuvers (turn left, turn right, turn toward and turn away). Review of the literature suggests that few solid conclusions can be drawn here.

5.2.3. Time Pressure and Maneuver Choice Effects on Efficiency

Efficiency can be assessed in two ways: (a) the time required to return to an original path following a maneuver; (b) the fuel burn cost of the maneuver. In this respect the different criteria have different implications. Because of the higher control order, lateral maneuvers will involve greater time for a maneuver to return to the original path than will vertical maneuvers (less time) or airspeed maneuvers (a gain in time if acceleration is chosen, a loss in time if deceleration is chosen; 3D flight path is not effected, although 4D navigation is disrupted). In contrast to time efficiency, it is assumed that both vertical maneuvers and speed maneuvers will require increasing and decreasing engine thrust, which could increase fuel burn in a way that lateral maneuvers will not. **It is not clear how the two criteria of time efficiency and fuel efficiency may be traded off in pilot maneuver selection**, nor how this tradeoff will be evaluated by other agents in the National Airspace (e.g., ATC, AOD).

5.3. Conflict Geometry Effects

As identified in Figure 5.1, there are a vast number of possible geometries that can be created by a single traffic aircraft, relative to ownship's flight trajectory. One of the most important features may be whether the traffic is maintaining a steady vector (lateral or vertical) or is in a state of change (turn, or changing airspeed or vertical speed). These changing conditions do not appear to have been examined by research. We focus this section therefore exclusively on constant vector geometry, although we address changes in Section 6.

5.3.1. Geometry Effects on Maneuver Tendency

There appears to be very little research on how the conflict geometry affects maneuver choice tendency. There is a general tendency for pilots to adhere to a turn right tendency, all other characteristics being equal, and when the traffic is level and at ownship altitude, as consistent with FAR 91.113. Concerning vertical behavior, one study (Wickens, Helleberg, & Xu, 2002), reported a **strong tendency for pilots to maneuver in the opposite direction of the traffic: when the traffic was descending from above, pilots climbed, and vice versa.** (This tendency hastens the resolution of the conflict). A weaker effect was observed that when traffic was approaching pilots tended to climb rather than descend.

5.3.2. Geometry Effects on Safety

This is an important causal sequence because it bears on the question of which scenario geometries may be most critical for examining safety-relevant use of the CDTI (see Section 6). Several effects have been observed in the literature, which are ordered below in terms of our confidence in their validity.

- Vertical geometry: **Changing altitude traffic (ascending, descending) is less safe than level traffic.** Importantly, this effect has been observed multiple times (Merwin & Wickens, 1996; O'Brien & Wickens, 1997; Alexander & Wickens, 2001, 2002; Wickens & Helleberg, 1999; Wickens, Gempler, & Morphew, 2000; von Westrenen & Groenweg, 2003), even with CDTIs that present a clear vertical situation display (see Section 5.4.2 below).

- Lateral geometry: **Longer TCPA makes detection and understanding more difficult.** This effect is simply based on the greater difficulty that pilots have predicting effects over longer time spans and/or at slower speeds (Xu, Wickens & Rantanen, in preparation¹), and the fact that greater time spans typically span greater spatial regions on a CDTI, requiring more scanning (Remington et al., 2001).
- Lateral geometry: **Different relative speeds between ownship and traffic make detection and understanding more difficult.** This effect (Xu et al., in preparation) can be attributed to the fact that with equal velocities between ownship and traffic, spatial separation on a CDTI is always perfectly correlated with time separation (Tau), thereby rendering the conflict evolution quite easy to visualize. Any changes in relative speeds, therefore, negatively impact the pilots' ability to visualize
- Lateral geometry: **Conflict angle effects.** The conflict angle is the difference in heading between ownship and traffic. When this angle is either 0 (parallel overtaking), 180 (parallel approaching), or 90 (crossing), the conflict appears to be easy to understand (Xu et al, in preparation). However, for non-orthogonal overtaking and approaching angles (e.g., right approach 10-60 degrees and 120-170 degrees, respectively), conflict understanding becomes more difficult. Research results also suggest that overtaking geometries are more challenging than approaching angles (Thomas & Johnson, 2001; Johnson et al., 2003; Wickens et al., 2002; Scallen et al., 1997); However, one study has shown the opposite effect (Merwin & Wickens, 1996), and the difference in difficulty that has been more frequently observed may be attributed to the fact that overtaking geometries will typically involve a longer TCPA, because of the small difference in relative speeds of the aircraft. As noted above, longer TCPA makes detection more difficult. When relative speed is controlled (Xu et al., in preparation), there appears to be no difference in pilot prediction of the conflict situation.

5.3.3. Geometry Effects on Efficiency

The effects of geometry on maneuver efficiency are not well understood. They may be indirectly predicted from added data pertaining to Section 5.3.1. That is, if particular geometry induces vertical rather than lateral maneuvering, these efficiency effects could be predicted from the analysis presented in 5.2.3 (maneuver effects on efficiency). However, as noted above, there appears to be little evidence for a causal association between conflict geometry and maneuver axis choice (lateral, vertical, speed).

5.4 Display Support Effects

There are a series of display guidelines and recommendations, summarized in Section 4, pertaining to issues such as decluttering and range of coverage. We focus here on two of the most important of these that have received fairly extensive pilot-in-the-loop evaluation: the value of prediction, and the representation of vertical information.

¹ The term "predicting" is operationally defined in Xu et al., to be the score of a pilot on a test of projecting the DCPA, TCPA and the vector from ownship to the conflict to the traffic, at the PCA. This was derived to be the most accurate assessment of a pilot's "conflict awareness."

5.4.1 Predictive Displays Are Useful

One CDTI study (Wickens, Gempler, & Morphew, 2000), demonstrated empirically the benefit of both predictor displays of ownship and depiction of the future conflict geometry, on both performance and workload. As such, this study confirms the value of prediction that has been well established in other domains of research (see Wickens, 1986; Wickens, Gempler & Morphew, 2000 for a review). However, less certain is how long the span of prediction or LAT should be, for which benefits will be preserved. An important consideration in this design parameter is the fact that prediction across longer LATs will become less reliable and, as discussed earlier in this report, lower reliability will lead to problems, to the extent that pilots follow the guidance of an unreliable (or imperfect) prediction (Wickens et al., 2000). It is not known at what LAT the cost of unreliability begins to dominate the benefit of prediction.

5.4.2 Vertical Representation Effects

The vertical representation issue addresses how altitude is represented on a CDTI. Most CDTI versions present a 2D plan view display with altitude coded symbolically (e.g., by data tags, or arrows indicating climbs/descents). However, altitude may also be presented in a separate vertical situation display (VSD) in a suite that we call the **2D coplanar display**, or as a single **3D Display**, in which ownship is typically presented from above, behind, and slightly offset to the right (Ellis et al., 1987; Alexander, Wickens, & Merwin, in press).

5.4.2.1. Vertical representation effects on maneuver tendency.

There is by now a strong trend revealed by collective findings, suggesting that **the more graphically and linearly the altitude dimension is rendered, the more likely pilots are to chose vertical maneuvering**. This trend is revealed by contrasts of tendencies between experiments and conditions that have used only plan view displays (only symbolic vertical), have used 3D displays (vertical display is rendered, but is inherently ambiguous), and have used coplanar displays (vertical display is unambiguous). Across these three display types, the choice of vertical maneuvers becomes progressively more likely as the vertical is present graphically and unambiguously, independent of geometry (Ellis et al., 1987; Merwin & Wickens, 1996; Wickens, Helleberg & Xu, 2002; Alexander & Wickens, 2001, 2002). The reader should note, in evaluating this conclusion that the conclusion in Section 5.2.2 revealed that vertical maneuvers were inherently less safe than lateral maneuvers.

5.4.2.2. Vertical representation effects on maneuver safety.

A strong replicable trend is that **3D displays support safer maneuvering than do 2D plan view displays (Ellis et al., 1987), but less safe maneuvering than do 2D coplanar displays**. The latter effect (2D coplanar safer than 3D) is observed both when “safety” is defined in terms of conflict detection ability (Merwin & Wickens, 1996), and time-in-predicted conflict (Merwin & Wickens, 1996; Alexander & Wickens, 2001, 2002). Both aspects of performance appear to be degraded because of the ambiguity of trajectory estimation inherent in collapsing the 3D volume of airspace onto a 2D viewing screen, a conclusion consistent also with research on 3D displays for air traffic control (Wickens, Miller & Tham, 1995; May, Campbell & Wickens, 1995).

5.4.2.3. *Vertical representation effects on maneuver efficiency.*

To the extent that a better representation of altitude (2D coplanar > 3D > 2D plan view) renders vertical maneuvering more likely, then the effects of this representation on maneuver efficiency can be predicted through the analysis presented in Section 5.2.3 (tendency effects on efficiency). The only study to have directly examined how vertical rendering effected the efficiency of maneuvers within an axis (Alexander & Wickens, 2002), did not reveal strong effects between the two co-planar views. However, the study did reveal that within the coplanar display, whether the VSD was represented from behind or beside ownship, affected the amount of lateral deviation from the flight path when a lateral maneuver was chosen, with the “beside” VSD orientation leading to greater departures (less efficiency).

5.5 CONCLUSIONS

In conclusion, it should be noted that the total number of effects originally reported in the studies summarized above, is far more numerous than those actually summarized here, since we have restricted ourselves to either those effects that have been replicated across multiple studies, or, in a few cases, those effects observed in a single study of large magnitude (statistical and effect size) and not contra-indicated by results from other studies. The reader will also note that sometimes causality between the independent variables of geometry & vertical display on the one hand, and the dependent variables of safety and efficiency on the other, is difficult to firmly establish. This difficulty results because an influence of the former set on the latter may be either direct (e.g., 3D displays provide more time in predicted conflict), or may be indirectly mediated by choice tendency (e.g., a traffic climbing geometry induces a descending maneuver tendency that is inherently less safe). Only the studies carried out by Alexander and Wickens (2001, 2002) have explicitly tried to disentangle direct from indirect influences.

Given these qualifications, the following conclusions nevertheless appear to be warranted from the existing data:

5.5.1. *Maneuver Tendencies:*

1. Pilots increase their preference for vertical maneuvers under increased time pressure because of both the increased time efficiency and lower cognitive complexity of vertical maneuvers.
2. Pilots tend to avoid airspeed adjustments for conflict avoidance, perhaps because of the greater difficulty in visualizing the effectiveness of such maneuvers.
3. Pilots tend to avoid complex multi-axis avoidance maneuvers.
4. The tendency to maneuver vertically is enhanced to the extent that the vertical axis is presented graphically (rather than symbolically) and unambiguously in a co-planar display (rather than a 3D display).

5.5.2. *Maneuver Safety:*

1. Vertical maneuvers tend to be less “safe” than lateral ones, as safety is operationally defined by avoiding a state of predicted conflict.

2. Vertically changing traffic (climbing/descending) presents a greater challenge for pilot visualization and safe maneuvering.
3. Safety is decreased with a 3D display relative to a 2D **co-planar** display (but not a 2D plan-view display).
4. Safety, as operationally defined by conflict understanding, is decreased with long TCPA, with differing relative speeds between ownship and traffic, and with non-orthogonal lateral geometry.
5. Maneuver safety is challenged when the airspace is more constrained, due to the presence of other hazards (additional traffic, weather).

5.5.3. Maneuver Efficiency:

1. There is a tradeoff between time versus fuel efficiency in the choice between vertical maneuvers (which favors time) and lateral maneuvers (which favors fuel efficiency).

A final conclusion is that the parameter space for understanding “what makes conflict detection and avoidance difficult” remains less than fully populated with empirical data, particularly in high fidelity pilot-in-the-loop simulations. Nevertheless, the above conclusions allow us to identify certain features that should be incorporated to challenge pilot use of alert-equipped CDTIs to the utmost, so that assessment can be measured under “worst case” scenarios. These features will be the focus of Section 6.

6. CDTI EVALUATION SCENARIOS

6.1. General

Before suggesting conflict scenarios for CDTI evaluation it is critical to distinguish between three different levels of automation and tasks pilots may perform with a CDTI:

1. Pilots are prohibited from maneuvering on their own; the CDTI provides for situation awareness information but ATC maintains responsibility for safe separation and maneuver choice.
2. Pilots use the traffic information provided by the CDTI to plan and execute maneuvers to avoid or resolve conflicts; CDTI only provides alerting for potential conflicts
3. The CDTI also suggests a maneuver for the pilots to execute.

It is clear that the level of automation and pilots' task greatly influence the make-up of the evaluation scenarios. For example, if maneuvering is allowed, other aircraft should be placed in the scenario in such a manner that they effectively constrain available conflict avoidance maneuvers, whether chosen by pilots or automation; if no maneuvering is allowed, criticality of such traffic is greatly reduced.

6.2. General Guidelines for Scenario Development

The general approach to scenario selection is as follows. In Figure 6.1, we present three generic properties of scenarios. On the vertical axis is a dimension representing the multitude of possible geometries, defined by lateral, vertical and speed behavior of both ownship and intruder, as depicted in Section 5, Figure 5.1., and below, in Figure 6.2. and Table 6.1. This axis will be unpacked below.

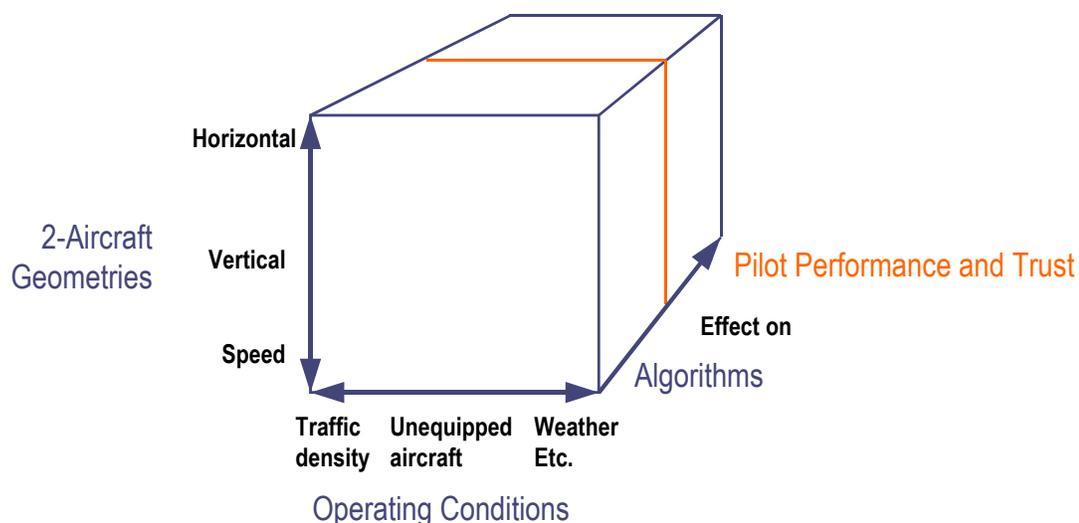


Figure 6.1. Dimensions for CDTI evaluation scenario development.

On the horizontal axis is represented a set of operating conditions that could impact performance of the CDTI, as well as performance of the pilot using the CDTI. How much of an impact these conditions may have will be, as noted above, dependent upon the use for which the CDTI is being approved. If, for example, it is to be approved for maneuver choice in a free-flight environment, or for pilot maneuver recommendation, in a collaborative decision making (CDM) environment, then the density and proximity of additional traffic becomes relevant. As another example, if it is to be operated in a mixed fleet environment, in which some aircraft are not equipped with the necessary transponders to convey full information to the CDTI, this factor should be considered.

On the depth axis we represent two important factors to be considered in approval: performance of the algorithm itself which, as we noted, may be degraded by uncertainty in trajectory extrapolation and by long LATs, and performance and trust of the pilot using the algorithm. Both of these factors should be taken into account when selecting scenarios for evaluation.

The three-dimensional matrix formed in Figure 6.1 avails a very large number of potential scenarios that could be chosen, given in particular, the number of geometries available by combining lateral, vertical and speed behavior. Our approach will be to recommend a restricted number of these, perhaps 12 for “bench testing” of the algorithms, based in particular on geometries that will provide serious challenges to the sensitivity of the algorithms to discriminate true conflicts from separation. Pilots will oversee the bench tests and be asked to provide a variety of ratings, regarding performance. We will also recommend a still smaller set of these scenarios for actual flight testing. In the following sections we describe particular geometries that we believe are most challenging.

6.3. Candidate Scenarios

6.3.1. Taxonomy of Conflict Geometries

There are innumerable possible conflict situations. However, conflict geometries can be classified into a manageable number of categories, from which scenarios for systems evaluation and experimental research can be chosen in a systematic manner. One such classification scheme is presented in Figure 6.2. below. This system involves all horizontal conflict geometries and it is based on the FAA course definitions for ATC separation purposes (FAA, 2000). The labels of angles (obtuse, right, and acute) are those frequently appearing in research literature.

This taxonomy is expanded in Table 6.1. by including also the vertical, longitudinal (i.e., speed) and temporal dimensions to it. The latter refers to situations where the aircraft in (potential) conflict commence climb or descent, or level off after climbing and descending, or change heading during the conflict situation. The candidate scenarios for CDTI evaluation suggested later in this section are based on this taxonomy in addition to the literature reviewed in earlier sections.

In these taxonomies it should be noted that directly head-on or overtaking situations, or directly opposite or same headings, are special cases and have very different implications to pilot performance than converging conflict angles. Therefore, conflict angles are sometimes referred to as 110° – 170° and 10° – 70° elsewhere in the report.

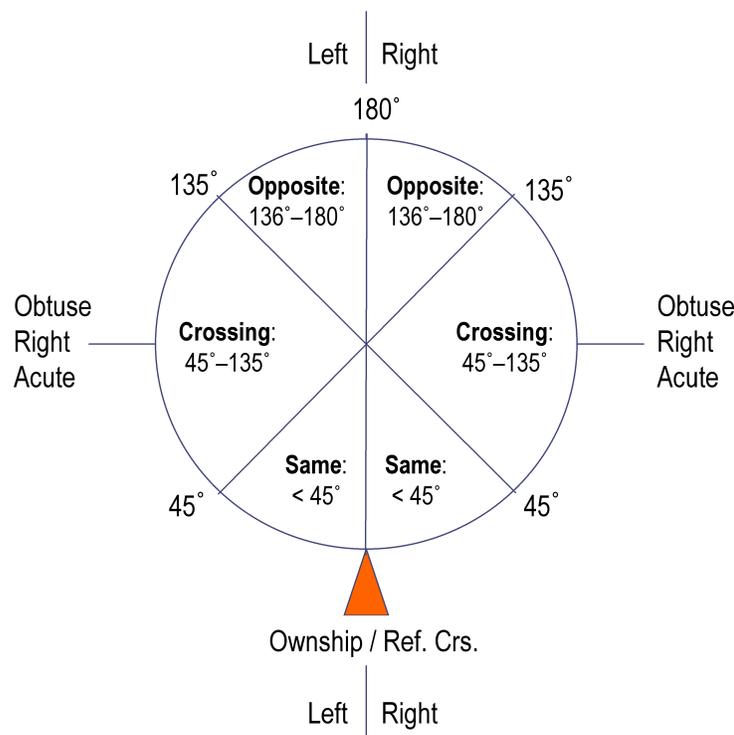


Figure 6.2. Classification of horizontal conflict geometries, based of the FAA course definitions. The labels of the angles are those frequently used in research literature.

6.3.2. General Guidelines

The general guidelines for candidate scenarios are based on what we deem “worst case scenarios” for both the alerting algorithms and to humans to detect conflicts accurately and in a timely manner, based on the literature reviewed in this report (Sections 3 & 5). From the alerting algorithm point-of-view, factors that may conspire against accuracy are long LAT and large separation standards (Magill, 1997) on one hand, and high rates of change of relevant parameters (e.g., vertical speed, heading) on the other (based on experiences with TCAS). Scenarios that would test the algorithm’s ability to suppress false alerts should therefore include turns as well as changes in vertical speeds of the aircraft (start climb or descent, or level off after climb or descent).

The reviewed literature suggest the following conflict geometries to be most difficult for humans in terms of conflict detection and resolution (the numbering refers to Table 6.1.):

- 1.1.1. Same (10-70°) but not parallel heading (Scallen et al., 1997; Thomas & Johnson, 2001, Johnson et al., 2003).
- 1.1.2.1. Opposite (110°–170°) but not parallel headings, intruder approaching from left (Merwin & Wickens, 1996)
- 2.1.3. Ownship descending (Alexander & Wickens, 2002)
- 2.2.2. Intruder climbing, and

- 2.2.3. Intruder descending (Merwin & Wickens, 1996; O'Brien & Wickens, 1997; Wickens & Gempler, 1998; Von Westrenen & Groeneweg, 2003)

Table 6.1.

An expanded taxonomy of all possible two-aircraft conflict scenarios.

<ul style="list-style-type: none"> 1. Horizontal <ul style="list-style-type: none"> 1.1. Constant <ul style="list-style-type: none"> 1.1.1. Same (< 45°) <ul style="list-style-type: none"> 1.1.1.1. Left 1.1.1.2. Right 1.1.2. Opposite (136°–180°) <ul style="list-style-type: none"> 1.1.2.1. Left 1.1.2.2. Right 1.1.3. Crossing (45°–135°) <ul style="list-style-type: none"> 1.1.3.1. Left <ul style="list-style-type: none"> 1.1.3.1.1. Obtuse angle 1.1.3.1.2. Acute angle 1.1.3.1.3. Right angle 1.1.3.2. Right <ul style="list-style-type: none"> 1.1.3.2.1. Obtuse angle 1.1.3.2.2. Acute angle 1.1.3.2.3. Right angle 1.2. Changing <ul style="list-style-type: none"> 1.2.1. Ownship <ul style="list-style-type: none"> 1.2.1.1. Turn towards intruder 1.2.1.2. Turn away from intruder 1.2.2. Intruder <ul style="list-style-type: none"> 1.2.2.1. Turn towards ownship 1.2.2.2. Turn away from ownship 	<ul style="list-style-type: none"> 2.2. Intruder <ul style="list-style-type: none"> 2.2.1. Level 2.2.2. Climbing <ul style="list-style-type: none"> 2.2.2.1. Constant 2.2.2.2. Changing <ul style="list-style-type: none"> 2.2.2.2.1. Start climb 2.2.2.2.2. Level-off 2.2.3. Descending <ul style="list-style-type: none"> 2.2.3.1. Constant 2.2.3.2. Changing <ul style="list-style-type: none"> 2.2.3.2.1. Start descent 2.2.3.2.2. Level-off 2.3. Vertical Speed <ul style="list-style-type: none"> 2.3.1. Ownship = Intruder 2.3.2. Ownship < Intruder <ul style="list-style-type: none"> 2.3.2.1. Constant 2.3.2.2. Changing <ul style="list-style-type: none"> 2.3.2.2.1. Increase 2.3.2.2.2. Decrease 2.3.3. Ownship > Intruder <ul style="list-style-type: none"> 2.3.3.1. Constant 2.3.3.2. Changing <ul style="list-style-type: none"> 2.3.3.2.1. Increase 2.3.3.2.2. Decrease
<ul style="list-style-type: none"> 2. Vertical <ul style="list-style-type: none"> 2.1. Ownship <ul style="list-style-type: none"> 2.1.1. Level 2.1.2. Climbing <ul style="list-style-type: none"> 2.1.2.1. Constant 2.1.2.2. Changing <ul style="list-style-type: none"> 2.1.2.2.1. Start climb 2.1.2.2.2. Level-off 2.1.3. Descending <ul style="list-style-type: none"> 2.1.3.1. Constant 2.1.3.2. Changing <ul style="list-style-type: none"> 2.1.3.2.1. Start descent 2.1.3.2.2. Level-off 	<ul style="list-style-type: none"> 3. Speed <ul style="list-style-type: none"> 3.1. Ownship = Intruder 3.2. Ownship < Intruder <ul style="list-style-type: none"> 3.2.1.1. Constant 3.2.1.2. Changing <ul style="list-style-type: none"> 3.2.1.2.1. Decelerate 3.2.1.2.2. Accelerate 3.3. Ownship > Intruder <ul style="list-style-type: none"> 3.3.1.1. Constant 3.3.1.2. Changing <ul style="list-style-type: none"> 3.3.1.2.1. Decelerate 3.3.1.2.2. Accelerate

In addition, results from a recent experiment (Rantanen, Wickens, Thomas, & Xu, 2003; Xu, Wickens, & Rantanen, in preparation) suggest that the following conditions add to the difficulty of unaided humans to make accurate conflict predictions: (1) Increasing DCPA, TCPA, and HMD, and reducing relative speed, made conflict understanding more difficult (i.e., reduced

absolute DCPA estimate accuracy). Based on these findings, specific guidelines for two-aircraft conflict scenarios can be derived:

- **Horizontal:** Same (10-70°) OR opposite (110°-170°) headings, AND intruder approaching from left-hand side, AND passing behind ownship, AND horizontal miss distance = large minimum separation (5 nm) + 0.1 nm (= 5.1 nm) for a no-conflict scenario OR minimum separation – 0.1 nm (= 4.9 nm) for a conflict scenario.

Note that these guidelines are well justified by empirical research, except the minimum separation (or separation violation) distances, which are somewhat arbitrary. These values can be set according to the level of accuracy required from the tested equipment.

- **Vertical:** Ownship descending OR commencing descent AND intruder descending OR climbing OR commencing descent OR climb within LAT, vertical miss distance = large minimum separation (2,000 ft) + 50 ft (= 2,050 ft) for a no-conflict scenario OR minimum separation – 50 (= 1,950 ft) for a conflict scenario.

Note that as was the case above, the minimum separation and separation violation distances can be chosen to reflect the desired accuracy of the equipment to be demonstrated.

- **Temporal:** Maximum LAT AND maximum feasible (i.e., depending on the type of aircraft and operation) vertical speed AND commencement of climb OR descent, and commencement of turn towards OR away within LAT but before minimum separation is reached. Different relative speeds between traffic and ownship.

Again, there exists some empirical support for these guidelines (primarily from Magill, 1997); the exact values will depend on the particular specifications for the equipment tested (i.e., LAT) and the operational environment in which the test is conducted.

Additionally, it should be noted that in contrast to conflict geometries that have been evaluated, as reviewed in Section 5.2., and which involve constant trajectories, the geometries recommended here involve a change in trajectory of ownship, or intruder, or both some time before the closest point of approach but within the LAT.

Finally, it is noteworthy that some aspects of conflict scenarios that make them difficult for algorithms make them easier for unaided humans, and vice versa. In particular, high relative speeds seem to make conflict prediction easier for humans than very low relative speeds, but might challenge the algorithm making conflict determinations. However, it is important to acknowledge that as far as is known, no empirical data exists that would allow determining the limits or scope of human vs. automatic conflict detection ability across all possible conflict geometries and situations.

6.3.3. Specific Suggestions for CDTI Evaluation Scenarios

In the following, 6 sample scenarios are described. These scenarios involve all characteristics described above, but naturally additional scenarios can be created by combining these characteristics differently and by varying the continuous variables. The scenarios are presented in Tables 6.2–6.7; the first 4 tables involve scenarios with constant trajectories and therefore lack the change elements. Tables 6.6. and 6.7. are derived from these by including changes in both the ownship's and intruder's trajectories, represented on row 3. The tables specify the necessary parameters for scenario generation; the notation convention is as follows:

Subscript $_0$ refers to ownship, subscript $_1$ to intruder

Δx = change in heading, $-$ = to the left, $+$ = to the right

Δy = change in vertical speed, the sign denotes direction, $-$ = descent, $+$ = climb

Δz = change in airspeed, the sign denotes direction, $-$ = decrease, $+$ = increase

s = speed

a = altitude

h = horizontal minimum separation

v = vertical minimum separation

HMD = horizontal miss distance

VMD = vertical miss distance

Table 6.2.

Scenario 1: A no-conflict scenario with HMD larger than minimum horizontal separation, ownship level, intruder descending.

<u>Ownship</u>	<u>Intruder</u>
1. Horizontal:	1. Horizontal: Overtaking, same heading ($10-70^\circ$), approaching from the left-hand side, HMD = $h + 0.1$ nm passing behind ownship
2. Vertical: Level at altitude a_0	2. Vertical: Descending at rate $-y_1$ through altitude a_0 , passing ownship at VMD = $v - 50$ ft
3. Longitudinal: Speed s_0	3. Longitudinal: $s_1 > s_0$ (overtaking)

Table 6.3.

Scenario 2: A no-conflict scenario with VMD larger than minimum vertical separation, ownship level, intruder climbing

<u>Ownship</u>	<u>Intruder</u>
1. Horizontal:	1. Horizontal: Overtaking, same heading ($10-70^\circ$), approaching from the left-hand side, HMD = $h - 100$ ft, passing behind ownship
2. Vertical: Level at altitude a_0	2. Vertical: Climbing at rate $+y_1$ through altitude a_0 , passing ownship at VMD = $v + 50$ ft
3. Longitudinal: Speed s_0	3. Longitudinal: $s_1 > s_0$ (overtaking)

Table 6.4.

Scenario 3: Ownship descending, horizontal conflict with intruder. Note that horizontal conflict means that it is the loss of horizontal separation that should trigger the alert, hence, horizontal separation must be lost after vertical separation is lost. Accurate timing of conflict scenarios is therefore of utmost importance.

<u>Ownship</u>	<u>Intruder</u>
1. Horizontal:	1. Horizontal: Opposite heading $110-170^\circ$, approaching from the left-hand side, HMD = $h - 0.1$ nm, passing behind ownship
2. Vertical: Descending at rate $-y_0$, through altitude a_1 , passing intruder at VMD = $v + 50$ ft	2. Vertical: Level at altitude a_1
3. Longitudinal: Speed s_0	6. Longitudinal: $s_1 < s_0$
4. Temporal: Vertical separation lost at time t_0	7. Temporal: Horizontal separation lost at time $t_1 > t_0$

Table 6.5.

Scenario 4: Ownship climbing, vertical conflict with intruder. Note that vertical conflict means that it is the loss of vertical separation that should trigger the alert, hence, vertical separation must be lost after horizontal separation is lost. Accurate timing of conflict scenarios is therefore of utmost importance.

<u>Ownship</u>	<u>Intruder</u>
1. Horizontal:	1. Horizontal: Opposite heading ($136^\circ-180^\circ$), approaching from the left-hand side, HMD = $h + 0.1$ nm, passing behind ownship
2. Vertical: Climbing at rate $+y_0$, through altitude a_1 , passing intruder at VMD = $v - 50$ ft	2. Vertical: Level at altitude a_1
3. Longitudinal: Speed s_0	3. Longitudinal: $s_1 > s_0$
4. Temporal: Vertical separation lost at time t_0	4. Temporal: Horizontal separation lost at time $t_1 < t_0$

Table 6.6.

Scenario 5; a no-conflict scenario with HMD larger than minimum horizontal separation, ownship climbing, intruder commencing descent.

<u>Ownship</u>	<u>Intruder</u>
1. Horizontal: Ownship turning to the left (in front of intruder) Δx_0 degrees.	1. Horizontal: Opposite heading (136° – 180°), approaching from the left-hand side, HMD = $h + 0.1$ nm, passing behind ownship
2. Vertical: Climbing at rate $+y_0$, through altitude a_1 , passing intruder at VMD = $v - 50$ ft	2. Vertical: Climbing at rate $+y_1$
3. Longitudinal: Speed s_0	3. Longitudinal: $s_1 = s_0$
4. Temporal: Turn at time $t_0 < \text{LAT}$	4. Temporal: Commencing descent at time $t_1 = t_0$

Table 6.7.

Scenario 6; a conflict scenario with VMD smaller than minimum vertical separation, ownship commencing descent, intruder climbing and turning toward ownship.

<u>Ownship</u>	<u>Intruder</u>
1. Horizontal:	1. Horizontal: Overtaking, same heading ($< 45^\circ$), approaching from the left-hand side, turning right Δx_0 degrees, HMD = $h + 0.1$ nm passing behind ownship
2. Vertical: Commencing descent at rate $-y_0$, through altitude a_1 , passing intruder at VMD = $v - 50$ ft	2. Vertical: Climbing at rate $+y_1$
3. Longitudinal: Speed s_0	3. Longitudinal: $s_1 > s_0$ (overtaking)
4. Temporal: Turn at time $t_0 < \text{LAT}$	4. Temporal: Turn at time $t_1 < \text{LAT}$

6.4. SUMMARY

Note that the above scenarios represent an incomplete (i.e., not fully factorial) sample of what can be deemed as the most challenging conflict geometries for both unaided humans and CD&R algorithms, based on the literature reviewed. The sampling is, however, systematic in a sense that all possible variations within the tables' cells occur at least once. Obviously, many more scenarios can be created by different combinations of these parameters, and the final scenarios to be tested in simulations or in actual flight will depend on many other factors present in such testing environments. Hence, a set of recommendations and caveats will be detailed:

1. The suggested scenarios should be considered as templates that guide consideration and combinations of various parameters, and into which actual parameter values will be inserted according to the desired scenario characteristics and constraints in the actual evaluation settings.
2. It is also crucial to understand that the above scenarios only concern the vertical axis of the dimensions of scenario development depicted in figure 6.1. Multiple intruders and conflicts can be easily created by combining any number of intruders from the basic scenario templates into a single evaluation scenario.
3. The temporal characteristics of such large scenarios warrant careful consideration, however, to determine exactly whether multiple conflicts occur simultaneously or what might be the desired frequency and temporal spacing between conflicts.
4. Inclusion of unequipped, transponder off or otherwise defined “distracter” aircraft in the scenarios must be determined separately.
5. The dimension of algorithm will be determined by the actual equipment to be tested.
6. It should be understood that the above scenario templates have been created based on vague specifications of the ultimate system evaluation settings. It is therefore possible that certain parameters that may not have been considered are nevertheless present in the actual evaluation environments. Therefore, it is very important that all available data from the evaluations are collected and archived for retrospective analysis of all parameters present in the settings and for determination of their impact on the evaluation outcomes.

7. PERFORMANCE MEASURES IN CDTI LITERATURE

7.1. General

All scientific research and subsequent engineering applications are dependent on methods of measurement (Chapanis, 1959). From the previous section it is clear that development of evaluation scenarios for CDTI testing only accounts for one half of the approval process. The second half involves developing appropriate measures for the system's as well as its users' (i.e., pilots') performance assessment, and setting criteria for their evaluation. This section provides a general overview of the measurement problems associated with CDTI evaluation, a review of measures used in the CDTI research literature, a taxonomy of measures applicable to the task at hand, and a set broad of recommendations for CDTI evaluation efforts.

7.2. Overview of the Measurement Problem

Direct observation of pilot (and system) performance by a proficient non-flying observer usually meets the validity and reliability requirements for performance evaluation for training and proficiency assessment. The method has some important prerequisites, namely the expertise and skill of the evaluator, efficient usage of standardized checklists where all the items to be evaluated are explicitly defined, and training of the evaluators to achieve reasonable inter-rater and intra-rater reliability. However, use of subjective evaluations of system and pilot performance by a pilot-rater or a non-flying observer has several drawbacks. A human observer/rater may not be able to provide sufficiently accurate quantitative data or record observed performance data at a sufficient frequency for research purposes, due to the limitations of human observation capabilities. This is particularly the case in observation of simultaneous events. . Another widely used method of measurement subjective measures where pilots themselves rate their performance and workload. This method is relatively easy to use and inexpensive, and expert subject-evaluators can also readily detect and process task-related information that would otherwise require vast amounts of objective data to be recorded, stored, coded, reduced, and analyzed to yield useful measures (Wickens, Gordon, & Liu, 1998). However, evaluations by subject-raters suffer from many of the same problems as those by an outside observer, that is, they depend on the expertise and experience of the subject-raters and their ability to make absolute and comparative judgments. Furthermore, concurrent measurement is often very intrusive and interferes with the task at hand, and measures taken after the task are subject to changing perceptions and decay of memory

Although sound and valid arguments have been made both for (e.g., Hennessy, 1990) and against (e.g., Kosso, 1989; Scheffler, 1967) the use of subjective measurements, in the flight deck R&D domain objective measures are highly desirable. Valid and reliable objective evaluation methods are particularly desirable both in conjunction with high-fidelity, realistic simulations and in operational settings, where they can be collected and analyzed concurrently and unintrusively during the task, and subjected to data mining techniques afterwards to detect trends in the system's performance, before any possible problems are manifested as incidents or errors.

There is, however, a substantial gap between measures such as described above and measures of real interest, that is, measures of system safety and efficiency and pilot workload, situation awareness, and performance. It is therefore important to distinguish between what can be termed direct and indirect measures. The latter are based on primary measures but make inferences on variables that were not directly measurable (e.g., workload or performance) (Rantanen & Nunes, 2003). Equally important is to make a distinction between measurement of system performance and the measurement of an individual pilot or a flight crew. System measures are defined in system terms (i.e., efficiency, safety, and delays). Although they are greatly influenced by human performance, they are usually insufficient in the measurement of the performance of an individual operator. Identified task performance, human activity, errors, omissions, physiological and biochemical indices, and subjective assessment are possible measures of individual operators (Hopkin, 1995).

7.3. A Taxonomy of Measures

To gain a better understanding of the measurement issues in flight deck R&D domain, taxonomy of measures is proposed (c.f., Rantanen & Nunes, 2003). The main division of measures (see Table 7.1.) is between direct and indirect measures. Direct measures are defined here as those that can explicitly measured. Examples of such measures include a direct observation of a pilots' actions, measurement of a response latency, or count of conflicts in a in a given time interval. Indirect measures are those that cannot be measured directly but must be inferred from directly measurable variables. For example, certain actions of a pilot may be indicative of his or her performance, response latency can be used to make inferences on some covert cognitive processes, and a number of aircraft in a sector can be used to signify sector complexity.

7.3.1. Criteria

Within the class of direct measures, the next sublevel is created by differentiating between subjective and objective measures. The method for making this distinction is identification of objective criteria, a prerequisite for meaningful measurement (Meister, 1989). An example can be seen in Table 8.1: Note that observer rating appears under both subjective and objective categories. The classification is here based on an objective criterion against which the observer bases his or her judgment. For example, instrument proficiency check flight evaluation form contains several items with explicit criteria (e.g., altitude must be maintained within ± 100 ft from prescribed). However, rating pilot's workload or situation awareness provides much latitude for a subjective assessment.

Another example of the importance of criteria in classification of measures is the differentiation between what is termed here primary and secondary measures. Primary measures are those that are measured directly, (e.g., count of conflicts, HMD). Secondary measures are those derived from primary measures, for example an average number of traffic in a sector, its variance, or range. In the case of the average, the criteria are implicit (the time duration or interval during which the aircraft were counted and the number of samples) but nevertheless have an impact on the eventual measure. The role of a criterion is explicit in the case of error measures. The term "error" clearly presupposes some threshold value, or outcome of an action, that separates correct action from an incorrect or erroneous one. In other words, without explicit criteria no actions could be classified as errors

Table 7.1.

A taxonomy of measures for CDTI evaluation. The table is populated by the measures used in the literature reviewed for this report.

-
1. Direct Measures
 - 1.1. Subjective Measures
 - 1.1.1. Self-Rated

NASA-TLX mean score (Alexander & Wickens, 2001, 2002; Merwin & Wickens, 1996; Wickens, Gempler, & Morpew, 2000), Subjective safety, workload, and SA rating (DiMeo, Kopardekar, Ashford, Lozito, & Mackintosh, 2001)
 - 1.1.2. Observer-rated

Conflict resolution strategies (DiMeo, Kopardekar, Ashford, Lozito, & Mackintosh, 2001)
 - 1.2. Objective Measures
 - 1.2.1. System
 - 1.2.1.1. Count

Number of conflicts (Scallen, Smith, & Hancock, 1996)
 - 1.2.1.2. Frequency

Frequency of maneuver types (Wickens, Helleberg, & Xu, 2002), Air-Ground and Air-Air communications (DiMeo, Kopardekar, Ashford, Lozito, & Mackintosh, 2001), Conflict detection rate (Merwin & Wickens, 1996), Conflict rate (Merwin & Wickens, 1996), False alarm rate (Merwin & Wickens, 1996)
 - 1.2.1.3. Delay
 - 1.2.1.4. RMSE

Gaussian statistical model of rms prediction errors (Paielli & Erzberger, 1997)
 - 1.2.2. Human
 - 1.2.2.1. Observer-Rated

Accuracy for monitoring event-detection (Wickens, Gempler, & Morpew, 2000), Monitoring accuracy for the out-of-window event (Wickens, Gempler, & Morpew, 2000)
 - 1.2.2.2. Cognitive

Number of data input errors (Scallen, Smith, & Hancock, 1996), Proportion of predicted conflicts (Merwin & Wickens, 1996), RMSE from prescribed airspeed (Wickens, Gempler, & Morpew, 2000), RMSE from prescribed altitude (Wickens, Gempler, & Morpew, 2000), RMSE from prescribed heading (Wickens, Gempler, & Morpew, 2000), RMSE of the maneuver away from the commanded trajectories on each of the 3 controlled axes (Wickens, Gempler, & Morpew, 2000), RT for conflict absence judgment (Merwin & Wickens, 1996), RT for the monitoring event-detection task (Wickens, Gempler, & Morpew, 2000), RT to the out-of-window event (Wickens, Gempler, & Morpew, 2001), Safety measures, RTs to out-of-window event, monitoring accuracy of out-of-window event, time to first maneuver (Wickens, Gempler, & Morpew, 2002), Sensitivity measure A' (Merwin & Wickens, 1996)

Table 7.1. continued:

A taxonomy of measures for CDTI evaluation.

1.2.2.3. Psychophysiological
2. Indirect Measures
2.1. System Performance
2.1.1. Safety
Conflict probability (Irvine, 2001), Lateral separation in conflict resolution (DiMeo, Kopardekar, Ashford, Lozito, & Mackintosh, 2001), Mean % time spent in actual conflict (Wickens, Gempler, & Morpew, 2000), Monte Carlo sims run on state information, deriving the probability of conflict (Yang & Kuchar, 1997), Proportion of actual conflicts (Merwin & Wickens, 1996), success in avoiding conflicts (Alexander, & Wickens, 2002; Ellis, McGreevy, & Hitchcock, 1987; Gempler, & Wickens, 1998; Helleberg, Wickens, & Xu, 2000; Hoekstra & Bussink, 2003; Johnson, Battiste, & Bochow, 1999; Johnson, Battiste, Delzell, Holland, Belcher, & Jordan, 1997; May, Campbell, & Wickens, 1995; Merwin, & Wickens, 1996; Merwin, Wickens, & O'Brien, 1998; O'Brien, & Wickens, 1997; Palmer, 1983; Scallen, Smith, & Hancock, 1997; Smith, & Lee, 1984; Thomas, & Johnson, 2001; van Westrenen, & Groeneberg, 2003; Wickens, & Helleberg, 1999; Wickens, & Morpew, 1997; Wickens, Helleberg, & Xu, 2002; Wing, Barmore, & Krishnamurthy, 2002; v Doorn, Bakker, & Meckiff, 2001), True conflicts per hour (Magill, 1997), Number of free flight cancellations (DiMeo, Kopardekar, Ashford, Lozito, & Mackintosh, 2001).
2.1.2. Efficiency
Number of maneuvers (Scallen, Smith, & Hancock, 1996)
2.1.3. Economy
2.2. Human
2.2.1. Performance
Falsely predicted conflicts (Magill, 1997), Time in predicted conflict (Wickens, Helleberg, & Xu, 2002), Time until the first maneuver was initiated (Wickens, Gempler, & Morpew, 2000)
2.2.2. Workload
NASA-TLX (Alexander & Wickens, 2001, 2002; Merwin & Wickens, 1996; Wickens, Gempler, & Morpew, 2000)
2.2.3. Situation Awareness
2.2.4. Trust
Estimate of the criterion Beta (Merwin & Wickens, 1996)

7.3.2. Measurement Scales

Measurement involves the assignment of a number system to represent the values of the variables of interest. There exist four distinct measurement scales—the nominal, ordinal, interval, and ratio scales—and the measured variables must be explicitly associated with the appropriate scale and its corresponding mathematical properties (Krantz, Luce, Suppes, & Tversky, 1971; Ghiselli, Campbell, & Zedeck, 1981). The measurement scale is also a possible

taxonomic criterion. Many of the direct measures are ratio measures. However, the measures derived from these are only ordinal at best. Hence, it is only possible to state that one system imposes a higher workload on a pilot than another, but not how much higher. The identification of measurement scales is also inexorably linked to development of criteria, as described above. This task is crucial in order to ensure correct interpretation and proper statistical treatment of the data.

7.4. Summary

In Table 7.1., it is noteworthy that the measures of the most critical aspects of human interaction with cockpit automation (e.g., CDTI) are largely missing, namely workload and situation awareness. There are two reasons for this: (1) This review focused on direct and primary measures; establishing the relationships between overt, directly measurable variables and covert cognitive functions would necessitate a comprehensive review of psychological literature, which was clearly outside the scope of this project; (2) the inferences made about covert variables in the literature vary widely and are based on a wide range of direct measures. There is also little data available to establish firm relationships between measurable variables and covert cognitive constructs in the context of CDTIs.

Workload and situation awareness are high-level constructs. Although they have been defined widely and sufficiently accurately in the literature, their measurement continues to pose significant problems. Consequently—despite the fact that automation has the most significant impact on these very aspects of performance—their use as dependent variables in experimental settings or in evaluation of operational systems is often limited by the lack of available measures, or lack of well-established relationships between them and measurable variables.

Human performance in interacting with complex and dynamic systems is the result of the integration of multiple subtasks into a whole. This whole ultimately greatly influences reliability and workload of the operator. Hence, it is important to determine at the subtask level what type—or types—of perceptual and cognitive processes pilots and controllers use in their tasks, what factors in terms of concurrent tasks and demands as well as display design affect their performance, and how their performance further may affect their situation awareness. To answer these questions, a thorough and comprehensive task analysis must be performed.

7.5. Recommendations

Despite the rather cautionary tone of the above conclusions and the undeniable difficulties measurement issues pose to systems' evaluation, a number of recommendations can be nevertheless made to ensure the reliability and validity of measures employed:

1. All measurement should be preceded by a thorough task analysis. Task analysis can also be used to establish criteria for measures.
2. Calibration of all measurement instruments, including human raters, is another critical prerequisite for measurement. Some subjective measurement systems, such as the NASA-TLX workload index, contain standard procedures for calibration (Hart & Staveland, 1988).
3. Explicit determination of the scales of measures must be done to ensure proper statistical treatment of the data and apposite conclusions drawn from them.

4. Thorough review of the research literature establishing relationships between directly measurable task variables and covert cognitive constructs of interest (e.g., workload) must be done before making inferences about the latter on the basis of the former. It is important to bear in mind, however, that such relationships are often indeterminate in the literature, and that research literature often contains many controversies and conflicting results.
5. Explicit listing of assumptions, and careful checking for violations of them, is yet another prerequisite for validity of measures. Many simplifying assumptions are often necessary in complex task and they must be considered in making inferences about the results so that biases can be identified and conclusions moderated accordingly.
6. The taxonomic structure presented in Table 7.1. should be helpful in checking assumptions and identifying the respective strengths and weaknesses of measures employed.

References

- Abbott, T.S., Moen, G.C., Person, L.H., Keyser, G.L., Yenni, K.R., & Garren, J.F. (1980). *Early flight test experience with cockpit displayed traffic information (CDTI)* (NASA Technical Memorandum 80221/AVRADCOM Technical Report 80-B-2). Hampton, VA: NASA-Langley.
- Alexander, A. L., & Wickens, C. D. (2001). Cockpit display of traffic information: The effects of traffic load, dimensionality, and vertical profile orientation. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*.
- Alexander, A. L., & Wickens, C. D. (2002). Does traffic load modulate the effects of traffic display formats on performance? (No. ARL-02-9/NASA-02-7). Savoy, IL: University of Illinois, Aviation Research Laboratory.
- Alexander, A. L., Wickens, C.D., & Merwin, D. (2004, in press). Perspective and Co-planar cockpit display of traffic information: implications for maneuver choice, flight safety and mental workload. *International Journal of Aviation Psychology*.
- Aviation Week & Space Technology* (1993, Nov. 22). United to Test TCAS Use for Altitude Changes, pp. 63.
- Battiste, V., & Johnson, W. W. (1998). Development of a cockpit situation display for free-flight. *1998 World Aviation Conference*. Anaheim, CA.
- Bliss, J. O. (2003). An investigation of alarm related incidents in aviation. *International Journal of Aviation Psychology* 13(3), 249-268.
- Bliss, J.P., Freeland, M.J., & Millard, J.C. (1999). Alarm related incidents in aviation: A survey of the aviation safety reporting system database. *Proceedings of the 43rd Annual Meeting of the Human Factors & Ergonomics Society* (pp. 6-10). Santa Monica, CA: HFES.
- Chandra (year?)
- Chapanis, A. (1959). *Research techniques in human factors*. Baltimore, MD: Johns Hopkins University Press.
- Chappell, S. L. (1990). Human factors research for TCAS. *Proceedings of the Royal Aeronautical Society Conference on the Traffic Alert and Collision Avoidance System (TCAS)—Its development, testing, Installation and Operational Use* (6-1/6-19).
- Chappell, S.L. & Palmer, E.A. (1983). Primary separation between three aircraft using traffic displays. *Proceedings of the Human Factors society 27th Annual Meeting* (pp. 767-771). Santa Monica, CA: Human Factors Society.
- Chemiavsky, E. A. (1997, May). *Simulation-based assessment of operational performance of traffic alert and collision avoidance system (TCAS) II logic version 7.0*. (MTR 97W0000033). MITRE: McLean, VA.

- Ciemier, C. E., Gambarani, G. P., Sheridan, E. H., Stangle, S. A., Stead, R. P., & Tillotson, D. H. (1993, Apr.). *The results of the TCAS II transition program (TTP)*. Annapolis, MD: ARINC Research Corporation.
- DiMeo, K., Kopardekar, P., Ashford, R., Lozito, S., & Macintosh, M.-A. (2001). *Sahred-separation: Empirical results and theoretical implications* (No. ATM 2001-10-24b). Moffett Field, CA: NASA Ames Research Center.
- Dixon, S., & Wickens, C. D. (2003). *Imperfect automation in unmanned aerial vehicle flight control* (AHFD-03-17/ MAAD-03-2). Savoy, IL: University of Illinois, Aviation Human Factors Division.
- Ellis, S.R., McGreevy, M.W., & Hitchcock, R.J. (1987). Perspective traffic display format and airline pilot traffic avoidance. *Human Factors*, 29, 371-382.
- Federal Aviation Administration (2001). *Air Carrier Operational Approval and Use of TCAS II* (AC120-55B). Washington, DC: Author
- Federal Aviation Administration (2000). *Air Traffic Control* (7110.65M). Washington, DC: Author.
- Federal Aviation Administration (1998). *TSO-C147, Traffic Advisory System (TAS) Airborne Equipment* (TSO-C147). Washington, DC: Department of Transportation, Federal Aviation Administration.
- Federal Aviation Administration (1996). *TCAS transition program (TTP)*. Newsletter, Issue #27, Washington, DC.
- Federal Aviation Administration (1993). *Air Carrier Operational Approval and Use of TCAS II* (AC120-55A). Washington, DC: Author
- Federal Aviation Administration (1993). *Airworthiness approval of traffic alert and collision avoidance systems (TCAS II) and mode S transponders*. (AC 20-131A). Washington, DC: US Department of Transportation, Federal Aviation Administration.
- Federal Aviation Administration (1990). *Introduction to TCAS II*. Washington, DC: U.S. Government Printing Office.
- Feldman, E., & Weitzman, D. O. (1999). The false alert problem in automated conflict detection: Crying wolf too often. *Journal of ATC*, 41(2), 51-55.
- Gempler, K. S., & Wickens, C. D. (1998). *Display of predictor reliability on a cockpit display of traffic information*. University of Illinois Institute of Aviation Final Technical Report (ARL-98-6/ROCKWELL-98-1). Savoy, IL: Aviation Res. Lab.
- Ghiselli, E.E., Campbell, J.P., & Zedeck, S. (1981). *Measurement theory for the behavioral sciences*. San Francisco, CA: W. H. Freeman & Co.

- Green, D.M., & Swets, J.A. (1988). *Signal detection theory and psychophysics*. New York: Wiley.
- Hart, S.G., & Loomis, L.L. (1980). Evaluation of the potential format and content of a cockpit display of traffic information. *Human Factors*, 22, 591-604.
- Hart, S.G., & Staveland, L. E. (1988). Development of NASA TLX (Task Load Index): Results of experimental and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183).
- Helleberg, J., Wickens, C. D., & Xu, X. (2000). *Pilot maneuver choice and safety in a simulated free flight scenario* (Technical Report ARL-00-1/FAA-00-1). Savoy, IL: University of Illinois, Aviation Res. Lab.
- Hennessy, R. T. (1990). Practical human performance testing and evaluation. In H. R. Booher (Ed.), *MANPRINT: An approach to systems integration* (pp. 433-479). New York: Van Nostrand Reinhold.
- Hoekstra, J., & Bussink, F. (2003) *Free flight: How low can you go?* DASC Report.
- Hopkin, V.D. (1980). The measurement of the air traffic controller. *Human Factors*, 22(5), 547-560.
- Hopkin, V. D. (1995). *Human factor in air traffic control*. Bristol, PA: Taylor & Francis.
- Irvine, R. (2001) A geometric approach to conflict probability estimation. *Proceedings of the 4th US/Europe Air Traffic Management R&D Seminar*, Santa Fe, December 3-7, 2001.
- Irvine, R. (1998). *The GEARS conflict resolution algorithm* (No. AIAA-98-4236): AIAA.
- Jago, S., & Palmer, E. (1980). Separation monitoring with four types of predictors on a cockpit display of traffic information.(pp. 439-447).
- Johnson, W. W., Battiste, V., & Bochow, S. H. (1999). A cockpit display designed to enable limited flight deck separation responsibility. Paper presented at the *World Aviation Conference*.
- 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., Bilimoria, K.D., Thomas, L.C., Lee, H.Q., & Battiste, V. (2003). Comparison of human and automation generated conflict resolutions. *AIAA Guidance, Navigation, & Control Conference, August 2003*.
- 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*.

- Johnson, W. W., Liao, M., & Tse, S. (1999). The effect of symbology location and format on attentional deployment within a cockpit display of traffic information. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting*.
- Klass, P.J. (1994). New TCAS software cuts false alerts. *Aviation Week & Space Technology*, 135-136.
- Kopardekar, P., Sacco, N., & Mogford, R. (2001). Comparison of air and ground conflict detection and resolution algorithms and their implications. *NASA-Ames Technical Report*.
- Kosso, P. (1989). Science and objectivity. *Journal of Philosophy*, 86, 245-257.
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement, volume 1: Additive and polynomial representations*. New York: Academic Press.
- Kreifeldt, J.G. (1980). Cockpit displayed traffic information and distributed management in air traffic control. *Human Factors*, 22(6), 671-691).
- Kremer, H., Bakker, B., & Blom, H. (1997) Probabilistic versus geometric conflict probing. *Proceedings of the 1st US/Europe Air Traffic Management R&D Seminar*, Saclay, France, June 17-20, 1997.
- Krishnan, K., Kertesz, J. S., & Wise, J. A. (2000). Putting four dimensions in "perspective" for the pilot. Paper presented at the Proceedings of the IEA 2000/HFES 2000 Congress.
- Krois, P. (1999a). *Alerting systems and how to address the lack of base rate information*. (No. Unpublished manuscript.). Washington, DC: FAA.
- Krois, P. (1999b, July 25). *White Paper: Human factors assessment of the URET conflict probe false alert rate*. Washington, DC: Federal Aviation Administration.
- Krozel, J. (1999). On Conflict Detection. *Paper presented at the joint EUROCONTROL/ RTCA CD&R Working Group Meeting*, NASA/Ames Research Center, CA, January 26-28, 1999
- Krozel, J. & Peters, M. (1997). Conflict detection and resolution for free flight. *The Air Traffic Control Quarterly Journal Special Issue on Free Flight*, 5(3).
- Kuchar, J. K. (2001, March 27-29). *Managing uncertainty in decision-aiding and alerting system design*. Paper presented at the 6th CNS/ATM Conference, Taipei, Taiwan.
- Kuchar, J.K, & Yang, L.C. (2000) A review of conflict detection and resolution modeling methods. *IEEE Transactions on Intelligent Transportation Systems*, 1(4), 179-189.
- Kulikowski, A. J. (1967). Enroute control and CAS. *Journal of ATC*, 9, 19-21.
- Larsen, R. (1967). The case for CAS. *Journal of ATC*, 9, 15-17.
- Layton, C., Smith, P.J., & McCoy, C.E. (1994). Design of a cooperative problem-solving system for en-route flight planning: An empirical evaluation. *Human Factors* 36(4), 94-119.

- Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10), 1243-1270.
- Lee, J., & Moray, N. (1994). Trust, self-confidence and operator's adaptation to automation. *International Journal of Human Computer Studies*, 40, 153-184.
- Lee, J.D., & See, K.A. (in press). Trust in technology: Designing for appropriate reliance. *Human Factors*.
- Livack, G. S., McDaniel, J. I., Battiste, V., & Johnson, W. W. (1999, April 12-15). Human factors considerations in automatic dependent surveillance-broadcast (ADS-B) flight operations. The ICAO Fourth Global Flight Safety and Human Factors Symposium, Santiago, Chile.
- Love, W. D. (1998). Preview of TCAS II version 7. *Air Traffic Control Quarterly*, 6(4), 231-247.
- Lubkowski, D. J. (1993). *TCAS logic improvements*. McLean, VA: MITRE Center for Advanced Aviation System Development.
- Magill, S. A. N. (1997). Trajectory predictability and frequency of conflict avoiding action. *Proc. CEAS Free Flight Conference*. DERA.
- Maltz, M., & Shinar, D. (2003). New alternative methods of analyzing human behavior in cued target acquisition, *Human Factors*, 45(2), 281-295.
- May, P.A., Campbell, M., & Wickens, C.D. (1995). Perspective displays for air traffic control: Display of terrain and weather. *Air Traffic Control Quarterly*, 3(1), 1-17.
- McLaughlin, M.P. (1997, June). *Safety study of the traffic alert and collision avoidance system (TCAS II)* (Tech. Report MTR 97W0000032). MITRE: McLean, VA.
- Mecham, M. (1991). Human factors play large role in Fokker TCAS cockpit adaptation. *Aviation Week & Space Technology*, 165-166.
- Meister, D. (1989). *Conceptual aspects of human factors*. Baltimore, MD: The Johns Hopkins University Press.
- Merwin, D.H. & Wickens, C.D. (1996). *Evaluation of perspective and coplanar cockpit displays of traffic information to support hazard awareness in free flight*. University of Illinois Institute of Aviation Technical Report (ARL-96-5/NASA-96-1). Savoy, IL: Aviation Res. Lab.
- Merwin, D. H., Wickens, C. D., & O'Brien, J. V. (1998). Display format-induced biases in air traffic avoidance behavior. Paper presented at the *World Aviation Congress (98WAC-71)*.
- Metzger, U., & Parasuraman, R. (2001). Conflict detection aids for air traffic controllers in free flight: Effects of reliable and failure modes on performance and eye movements. **(Source?)**

- Miller, C.A., Williamson, T., Walsh, J.A., Nivert, L.J., Anderson, J.L., & Zeitlin, A.D. (1994, September). Initiatives to Improve TCAS-ATC Compatibility. *Journal of ATC*, 6-7, 10-12.
- MITRE. (2002). *Traffic alert and collision avoidance system*. Retrieved June 30, 2002, from <http://www.caasd.org/proj/tcas/>
- Mondolini, S., & Conway, S. (2001). An airborne conflict resolution approach using a genetic algorithm (No. AIAA 2001-4054): AIAA.Mouloua, M., & Koonce, J. M. (Eds.). (1997). *Human-automation interaction : research and practice*. Mahwah, NJ: Lawrence Erlbaum.
- Moray, N. (2003). Monitoring, complacency, skepticism and eutectic behavior. *International Journal of Industrial Ergonomics*, 31, 175-178.
- O'Brien, J.V. & Wickens, C.D. (1997). Free flight cockpit displays of traffic and weather: Effects of dimensionality and data base integration. *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting* (pp. 18-22). Santa Monica, CA: Human Factors and Ergonomics Society.
- Owen (1993).
- Paielli, R. A., & Erzberger, H. (1997). Conflict probability estimation for free flight. *Journal of Guidance, Control, and Dynamics*, 20(3), 588-596.
- Palmer, E.A., Jago, S.J., Baty, D.L., & O'Connor, S. L. (1980). Perception of horizontal aircraft separation on a cockpit display of traffic information. *Human Factors*, 22(5), 605-620.
- Palmer, E.A. (1983). Conflict resolution maneuvers during near miss encounters with cockpit traffic displays. *Proceedings of the Human Factors Society 27th Annual Meeting* (pp. 757-761). Santa Monica, CA: HFES.
- Parasuraman, R. (1987). Human-computer monitoring. *Human Factors*, 29(6), 695-706.
- Parasuraman, R., Hancock P. A., & Olofinboba, O. (1997). Alarm effectiveness in driver-centered collision-warning systems. *Ergonomics*, 40(3), 390-399.
- Parasuraman, R., Masalonis, A.J., & Hancock, P.A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors*, 42(4), 636-659.
- Parasuraman, R., Molloy, R., & Singh, I.L. (1993). Performance consequences of automation-induced "complacency." *The International Journal of Aviation Psychology* 3, 1-23.
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230-253.
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model of types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics*, 30A(3), 286-297.

- Prinzo, W.V. (2003). Pilot's visual acquisition of traffic: Operational communication from an in-flight evaluation of a cockpit display of traffic information. *The International Journal of Aviation Psychology*, 13(3), 211-232.
- Pritchett, A. (2001) Reviewing the role of cockpit alerting systems. *Human Factors and Aerospace Safety*, 1(1), 5-38.
- Rantanen, E.M., & Nunes, A. (2003, Apr. 14-17). Taxonomies of measures in air traffic control research. *Proceedings of the 12th International Symposium on Aviation Psychology* (pp. 977-980). Dayton, OH.
- Rantanen, E. M., Wickens, C. D., Xu, X., & Thomas, L. C. (2003). Developing and validating human factors certification criteria for cockpit displays of traffic information avionics. *Paper presented at the FAA General Aviation, Aviation Maintenance, and Vertical Flight Human Factors Research Program Review Conference*, University of Nevada Reno, September 10-11, 2003.
- RTCA. (2000). Application descriptions for initial cockpit display of traffic information (CDTI) applications (No. RTCA/DO-259). Washington, DC: RTCA, Inc.
- RTCA (1997). *Minimum operational performance standards for traffic alert and collision avoidance system II (TCAS II) airborne equipment* (RTCA DO-185A). Washington, DC: RTCA, Inc.
- RTCA. (1998). *Guidance for implementation of cockpit display of traffic information* (RTCA/DO-243). Washington, DC: RTCA, Inc.
- RTCA (1997) *Minimum operational performance standards for traffic information service (TIS) data link communications* (RTCA DO-239). Washington, DC: RTCA, Inc.
- SAE G-10 Cockpit Display of Traffic Information Subcommittee. (2003, Nov. 25). *Human interface criteria for cockpit display of traffic information technology* (ARP 5365). Society of Automotive Engineers.
- Sarter, N. B., & Schroeder, B. (2001). Supporting decision making and action selection under time pressure and uncertainty: The case of in-flight icing. *Human Factors*, 43(4), 573-583.
- Scallen, S. F., Smith, K., & Hancock, P. A. (1996). Pilot actions during traffic situations in a free-flight airspace structure. *Proceedings of the 40th Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica, CA: HFES.
- Scallen, S. F. Smith, K., & Hancock, P.A. (1997). Influence of color cockpit displays of traffic information on pilot decision-making in free flight. *Proceedings of the 41st Annual Meeting of the Human Factors Society*. Santa Monica, CA: HFES.
- Scheffler, I. (1967). *Science and subjectivity*. Indianapolis, IN: Bobbs-Merrill.
- Sheridan, T. (1997). Supervisory control. In G. Salvendy (Ed.), *Handbook of human factors*. New York: Wiley.

- Sheridan, T.B. (2002). *Humans and automation: System design and research issues*. New York: Wiley.
- Smith, J., Ellis, S., & Lee, E. (1984) Perceived threat and avoidance maneuvers in response to cockpit traffic displays. *Human Factors*, 26(1), 33-48.
- Smith & Lee (1984).
- 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.
- St. John, M., & Manes, D.I. (2002). Making unreliable automation useful. *Proceedings of the 46th Annual HFES Meeting*.
- Thomas, L.C., & Johnson, W.W. (2001). Evaluation of CDTI dynamic predictor display technology. Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting.
- Thomas, L.C., Wickens, C.D., & Rantanen, E. (2003). Imperfect automation in aviation traffic alerts: A review of conflict detection algorithms and their implications for human factors research. *Proceedings of the 47th Annual HFES Conference*. Santa Monica, CA: HFES.
- Tillotson, D., & Gambarani, G. (1993). Lessons learned during the traffic alert and collision avoidance system (TCAS) transition program (No. DOT/FAA/RD-93/32). Washington, DC: FAA.
- Uhlarik, J., & Comerford, D.A. (2003, March). *Information requirements for traffic awareness in a free-flight environment: An application of the FAIT analysis* (DOT/FAA/AM-03/5). Washington, DC: Office of Aerospace Medicine, Federal Aviation Administration.
- van Doorn, B. A., Bakker, G. J., & Meckiff, C. (2001) Evaluation of advanced conflict modelling in the Highly Interactive Problem Solver. *Proceedings of the Fourth USA/Europe Air Traffic Management R&D Seminar*, Santa Fe, 3–7 December 2001.
- van Westrenen, F, & Groeneberg, J. (2003). Co-operative conflict resolution: An experimental validation of aircraft self-separation in a classroom environment. *International Journal of Aviation Psychology*, 13(3), 233-248.
- Vickers, T.K. (1992). FAA's International TCAS Symposium. *Journal of ATC*, 14.
- Wickens, C.D. (1986). The effects of control dynamics on performance. In K.R Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of perception and performance, Vol 2: Cognitive processes and performance*. pp 39-1/39-60. New York: Wiley & Sons.

- Wickens, C.D., Gordon, S.E., & Liu, Y. (1998). *An introduction to human factors engineering*. New York: Addison Wesley Longman.
- Wickens, C.D., Gempler, K., & Morphey, M.E. (2000). Workload and reliability of predictor displays in aircraft traffic avoidance. *Transportation Human Factors*, 2(2), 99-126.
- Wickens, C.D., & Helleberg, J. (1999). *Interactive perspective displays for airborne hazard awareness*. (No. ARL-99-1/ROCKWELL-99-1). Savoy, IL: University of Illinois Aviation Res. Lab.
- Wickens, C.D., Helleberg, J., & Xu, X. (2000). Decision and workload implications of free flight and the cockpit display of traffic information (CDTI). Paper presented at the IEA 2000/HFES 2000 Congress.
- Wickens, C. D., Helleberg, J., & Xu, X. (2002). Pilot maneuver choice and workload in free flight. *Human Factors*, 44(2), 171-188.
- Wickens, C.D., Lee, J., Liu, Y., & Gordon-Becker, S.E., (2004). *An introduction to human factors engineering*, 2nd Ed. New York: Pearson Education.
- Wickens, C. D., Mavor, M., & McGee, J. (1997). *Flight to the future: Human factors of air traffic control*. Washington, DC: National Academy of Science.
- Wickens, C. D., Mavor, M., Parasuraman, R., & McGee, J. (1998). *The future of air traffic control*. Washington, DC: National Academy of Science.
- Wickens, C.D., Mavor, A.S., Parasuraman, R., & McGee, P. (1998). Airspace system integration: The concept of free flight. In *The future of air traffic control*. Washington DC: National Academy Press.
- Wickens, C.D., Miller, S., & Tham, M. (1996). The implications of data-link for representing pilot request information on 2D and 3D air traffic control displays. *International Journal of Industrial Ergonomics*, 18, 283-293.
- Wickens, C.D., & Morphey, E. (1997). *Predictive features of a cockpit traffic display: A workload assessment*. University of Illinois Institute of Aviation Technical Report (ARL-97-6/NASA-97-3). Savoy, IL: Aviation Res. Lab.
- Wiener, E. L. (1989). Human factors in advanced technology ("glass cockpit") transport aircraft (No. NASA-TR-177528). Moffett Field, CA: NASA Ames Research Center.
- Wing, D. J., Adams, R. J., Barmore, B. E., & Moses, D. (2001). *Airborne use of traffic information in a distributed air-ground traffic management concept: Experiment design and preliminary results*. Paper presented at the 4th USA/Europe Air Traffic Management R&D Seminar, Santa Fe, NM.

- Wing, D. J., Barmore, B. E., & Krishnamurthy, K. (2002, August 5-8, 2002). Use of traffic intent information by autonomous aircraft in constrained operations. Paper presented at the AIAA Guidance, Navigation, and Control Conference and Exhibit, Monterey, CA.
- Winterberg, R, Mulder, M., & Van Paassen, M. M. (2003) Direct Manipulation In Aircraft Four-Dimensional Trajectory Planning Interfaces. *Proceedings of the 2003 ISAP Conference*, Dayton, OH.
- Woods, D.D. (1995). The alarm problem and directed attention in dynamic fault management. *Ergonomics*, 38(11), 2371-2394.
- Wright (year?)
- Xu, X., Wickens, C. D., & Rantanen, E. M. (in preparation). *Effects of air traffic geometry and conflict warning System reliability on pilot's conflict detection with cockpit display of traffic information (CDTI)*.
- Yang, L. C., & Kuchar, J. K. (1997). Prototype conflict alerting system for free flight. *Journal of Guidance, Control, and Dynamics*, 20(4), 768-773.
- Yeh, M., & Wickens, C.D. (2001). Display signaling in augmented reality: The effects of cue reliability and image realism on attention allocation and trust calibration. *Human Factors*, 43(3), 355-365.
- Yeh, M., Merlo, J.L., Wickens, C.D., & Brandenburg, D.L. (2003). Head up vs. head down: The costs of imprecision, unreliability, and visual clutter on cue effectiveness for display signaling. *Human Factors*, 45(3), 390-407.

APPENDIX

SUMMARY OF PILOT-IN-THE-LOOP STUDIES THAT INCORPORATED DIFFERENT CONFLICT DETECTION ALGORITHMS INTO CDTIS FOR USE IN FREE FLIGHT SIMULATIONS

Reference	Domain	Type	N	Task	Independent Variable + Levels	Dependent Variable + Measure	Results	Comment	CDTI/Algorithm Specifications
DiMeo et al. (2002)	Aviation	PIL	3 x 4 ATC, 3 x 2 crew	Shared flight plan changes between pilots and ATC	4 Scenarios: (1) URET, no CDTI, (2) URET + CDTI, Shared-Separation Level 1 and 2	Subjective safety, workload, and SA rating			
DiMeo et al. (2002)	Aviation	PIL	3 x 4 ATC, 3 x 2 crew	Shared flight plan changes between pilots and ATC	4 Scenarios: (1) URET, no CDTI, (2) URET + CDTI, Shared-Separation Level 1 and 2	Conflict resolution strategies	Pilots sometimes used more complex strategies than controllers (combined speed and heading changes); controllers resolved conflict earlier than pilots		
DiMeo et al. (2002)	Aviation	PIL	3 x 4 ATC, 3 x 2 crew	Shared flight plan changes between pilots and ATC	4 Scenarios: (1) URET, no CDTI, (2) URET + CDTI, Shared-Separation Level 1 and 2	Lateral separation in conflict resolution	Pilots accomplished lower mean lateral separation than controllers		
DiMeo et al. (2002)	Aviation	PIL	3 x 4 ATC, 3 x 2 crew	Shared flight plan changes between pilots and ATC	4 Scenarios: (1) URET, no CDTI, (2) URET + CDTI, Shared-Separation Level 1 and 2	Number of free flight cancellations	Pilots did not cancel FF, controllers did		
DiMeo et al. (2002)	Aviation	PIL	3 x 4 ATC, 3 x 2 crew	Shared flight plan changes between pilots and ATC	4 Scenarios: (1) URET, no CDTI, (2) URET + CDTI, Shared-Separation Level 1 and 2	Air-Ground and Air-Air communications	Shared-Separation Level 1 requirement for pilots to inform ATC of maneuvers increased comm. Wklod	No details on data or analysis	
Magill (1997)	Aviation	Sim.			Along-track and vertical prediction errors				
Magill (1997)	Aviation	Sim.			Separation minima (5, 10, 15 miles)	Falsely predicted conflicts (FAR)	FAR increased as separation minima increased		

Reference	Domain	Type	N	Task	Independent Variable + Levels	Dependent Variable + Measure	Results	Comment	CDTI/Algorithm Specifications
Magill (1997)	Aviation	Sim.			Phases of flight (combinations of climb, level, and descent)	True conflicts per hour	True conflicts by phases of flight: climb/desc = 34%, level/desc = 22%, climb/level = 18%, level/level = 17%, climb/climb = 5%, desc/desc = 4%		
Magill (1997)	Aviation	Sim.			LAT (5, 10, 15, 20 min)	Falsely predicted conflicts (FAR)	FAR increased as LAT increased		
Magill (1997)	Aviation	Sim.			Separation minima (5, 10, 15 miles)	True conflicts per hour	True conflicts increased as separation minima increased		
Dowek et al. (2001)	Aviation	sim		Monte Carlo Sim			Validation of conflict portion of algorithm not clarified	Intent of paper is to highlight resolution part of algorithm	3-D position, ground track, vertical speed, ground speed; "tactical", short lookahead times, state information only; produces solution maneuvers; PZ = +/- 1000ft vert, 5 nm rad; at LAT, DCP is measured every 10 sec
Prandini et al. (2000)	Aviation	sim		Monte Carlo Sim		Monte Carlo sims; "For each pair of simulated trajectories, execute the conflict detection algorithm at every radar measurement time (radar measurement time set equal to 12 seconds), each time using the updated flight plans and time horizon."		discussion of false alerts, which are calculated by SOC curves derived in Monte Carlo simulations	lat, long only; scaled Brownian motion perturbation which varies linearly with time; or a multivariate Gaussian random variable; course drift = 15 knot standard dev speed (along track) fluctuation; 1 nm standard dev crosstrack error, PZ = 5 nm

Reference	Domain	Type	N	Task	Independent Variable + Levels	Dependent Variable + Measure	Results	Comment	CDTI/Algorithm Specifications
Irvine (2001)	Aviation	sim		Monte Carlo Sim		Graphs of conflict probability were produced using finite set of conflict configurations	graphs of conflict probability illustrate the limits on performance of medium-term conflict detection	Discussion of false alerts focuses on setting appropriate probability of conflict, improving predicting techniques to minimize uncertainty parameters; refers to Paielli (1998)	lat, long; PZ = 5 nm rad; DCP, time to DCP, crossing angle, speed of each aircraft (set to 480 kn, not varied); along-track (15 knot [0.25 nm/min] or 0.22 nm/min) standard dev and cross-track error (given by equation)
Paielli & Erzberger (1997)	Aviation	sim		Monte Carlo Sim		Gaussian statistical model (of rms prediction errors)	Gaussian model validated empirically using actual air traffic data; Monte Carlo used for algorithm itself		lat, long, and vert; PZ = +- 2000 ft, 5 nm rad; path crossing angles, DCP's, and times to DCP's; RMS prediction errors (for lat, long, and vert) are treated as normally distr. (Gaussian)
Yang & Kuchar (1997)	Aviation	sim		Monte Carlo Sim	intruder aircraft's heading and altitude changes; host aircraft's avoidance response latency	Monte Carlo sims run on state information, deriving the probability of conflict	preliminary free-flight sims conducted at Ames in 1996, on 747-400 simulator - conflicts were scripted to examine pilot response, also to test alerting logic with 4 stages of alerting (inc TCAS alerts); no reference given for this PIL study	discussion of false alerts, which are calculated by SOC curves derived in Monte Carlo simulations; 3-stage alert: 1) potential threat >10 min or up to 200 nm away, 2) continued threat = display change, 3) continued threat = aural alert requesting action	lat, long, and vert; PZ = +- 1000 ft, 5 nm rad; 2 intruder velocities, 9 intruder hdg angles, initial positions varied over grid of 200 nm^2 with 1 nm incr.; turns: no turn, left or right 10 or 20 deg, speed incr. and decr. of 10 & 50 kts; Gaussian, normally distr. errors of 50 m lat, 30 m vert; course drift = 15 kt SD speed (along track) fluctuation; 1 nm dtandard dev crosstrack error; tracking errors from Paielli & Erzberger's work

Reference	Domain	Type	N	Task	Independent Variable + Levels	Dependent Variable + Measure	Results	Comment	CDTI/Algorithm Specifications
Johnson et al. (1999)	Aviation	PIL	8 crews	avoid conflicts	intent info was either shared or unshared, conflicts were either true or "apparent", initial altitudes were either separated or not separated; high vs. low flight-deck responsibility	success in avoiding conflicts	Positive subj responses to availability of info, colors, display options; no performance data		lat, long, & vert; PZ = 5 nm, +-1000 ft vert
Johnson et al. (1997)	Aviation	PIL	10 crews	avoid conflicts	enhanced vs. basic CDTI conditions	success in avoiding conflicts	Each crew flew all 8 scenarios w/ at least one conflict each; results showed min separation was never lost, enhanced display led to proactive resolution	authors state explicitly that no false alarms were programmed into the scenarios; 2 no alert/info only levels (low and moderate threat), 3 alert levels (high threat, very high threat + Traffic advisory, and very high threat + Resolution advisory)	lat, long, & vert; PZ = 5 nm, +- 1000 ft vert; along-track error StDev of 15 kn, cross track error StDev of 1 nm (based on Paielli & Erzberger); probabilities of turns or altitude changes included (4 turns/hr, 4 alt changes/hr), with magnitudes randomly drawn from given distributions;
Wing et al. (2002); Wing et al. (2001)	Aviation	PIL	16 pilots	avoid conflicts	tactical vs. strategic CDTI operation; low vs. high operational complexity	success in avoiding conflicts	Pilots flew 3 en-route trials for each of the 4 IV scenarios; authors conclude that pilot subjective responses indicated overall acceptance of display type and info provided; prelim results showed generally successful conflict resolution	3 stage alert: 1) inform pilot of potential conflict, no action required, 2) conflict detected, action by ownship required, 3) LOS has occurred, immediate action by ownship required; intruder aircraft generated by NASA software, Traffic and Events Manager (TMX)	lat, long, & vert; PZ = 5 nm, +-1000 ft vert; lookahead was 5 min for state info, 8 min for intent info; three categories of conflicts could occur (state-only, intent-only, and blunder); uncertainty parameters not specified (assumed to be same as in Yang & Kuchar);

Reference	Domain	Type	N	Task	Independent Variable + Levels	Dependent Variable + Measure	Results	Comment	CDTI/Algorithm Specifications
Wing et al. (2002)	Aviation	PIL	16 pilots	avoid conflicts	Time pressure (5 and 8 min vs. 3.5 min)	success in avoiding conflicts	Lateral maneuvers chosen most often (probably due to use of 2-D single-view CDTI)		
v Doorn et al. (2001)	Aviation	PIL	ATC's	detect conflicts	conflict geometry (crossing, in-trail, and head-on); collision risk thresholds of 10-4, 10-6, and 10-8	success in detecting conflicts	single ATC monitored simple conflicts in horiz plane only, with various collision risks, to establish meaningful display parameters		PZ = 5 nm rad; real-time scenarios (lat, long only)
Hoekstra & Bussink (2003)	Aviation	PIL	8 crews	avoid conflicts	low vs. high traffic; managed vs. free flight terminal approach	success in avoiding conflicts	8 sets of pilots flew enroute descent and terminal approach trials; pilots reported higher workload ("some effort" vs. "little effort") for free flight arrival compared to free flight cruise/descent, but thought the free flight descent was easy, and all three segments rated very high on "acceptability"	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; false alarms are assumed to be minimized by setting look-ahead time to 5 min (as established by prior NLR research, specifically Magill, 1997)	lat, long, & vert; PZ = +- 1000 ft vert, 5 nm en-route, 3 nm near terminal; uncertainty parameters not specified - traffic generated by NLR software "Traffic Manager"
Palmer (1983)	Aviation	PIL	16 pilots	Resolve conflicts	Time pressure (80 sec vs. 45, 20 sec)	success in avoiding conflicts	More vertical maneuvers as time pressure increased		2-D single-view CDTI
Alexander & Wickens (2002)	Aviation	PIL	18 pilots	Resolve conflicts	Display type (45° 3-D vs. 2-D coplanar)	success in avoiding conflicts	More vertical maneuvers overall; tendency increased for 2-D coplanar		
Alexander & Wickens (2001)	Aviation	PIL	18 pilots	Resolve conflicts	Display type (45° 3-D vs. 2-D side-view coplanar vs. 2-D rear-view coplanar)	success in avoiding conflicts	Vert maneuvers most common overall - climbs for 2-D, descents for 3-D; Rear-view coplanar and 3-D were "safer" in terms of avoiding predicted conflicts and also implementing climbs than side-view coplanar	Altitude maintenance of combo (lat/vert) maneuvers best with side-view, worst with 3-D; rear-view coplanar concluded to be best overall display	

Reference	Domain	Type	N	Task	Independent Variable + Levels	Dependent Variable + Measure	Results	Comment	CDTI/Algorithm Specifications
Merwin & Wickens (1996)	Aviation	PIL	n/a	Detect and resolve conflicts	Display type, (60° 3-D vs. 30° 3-D vs. 2-D coplanar)	success in avoiding conflicts	Best detection with 2-D coplanar, (2-D > 30° 3-D > 60° 3-D); More vertical maneuvers (vs. other types of maneuvers) in all three display types; also more vertical maneuvers in 2-D coplanar than either 3-D display	No apparent time cost of integration across the two views in the 2-D coplanar display; A descent bias was noted for the 30° 3-D display; an ascent bias was seen in the 60° 3-D display; both biases are attributed to the foreshortening effects of the viewpoint location	
Merwin & Wickens (1996)	Aviation	PIL	n/a	Resolve conflicts	Conflict geometry (crossing, in-trail, and head-on; ascending, descending, level)	success in avoiding conflicts	2-D display produced more maneuvers in opposite direction of intruder than 3-D (which encouraged same-direction vert maneuvers); Head-on (135°) conflicts were most difficult to avoid; left-approaching traffic resulted in more conflicts than right; Ascending and descending conflicts more difficult than level conflicts		
Wickens & Helleberg (1999)	Aviation	PIL		Resolve conflicts	Display type (2-D coplanar vs. 3-D variable-viewpoint display)	success in avoiding conflicts	Vertical maneuvers preferred over lateral; 3-D showed greater number of vertical maneuvers than 2-D; 3-D display resulted in more time spent in conflict than 2-D, but less maneuver deviations*	3-D variable-viewpoint was used, indicating that pilots were not put off by information access cost of interaction; * The decreased path deviation (increased efficiency) of maneuvers created using the 3-D display did not result in safer maneuvers	
Reference	Domain	Type	N	Task	Independent Variable + Levels	Dependent Variable + Measure	Results	Comment	CDTI/Algorithm Specifications

					Variable + Levels	Variable + Measure			Specifications
Wickens & Helleberg (1999)	Aviation	PIL		Resolve conflicts	Conflict geometry (crossing, in-trail, and head-on; ascending, descending, level)	success in avoiding conflicts	Fewer vertical maneuvers when traffic was ascending into conflict (vs. descending into conflict or same altitude as ownship)		
Thomas & Johnson (2001)	Aviation	PIL	8 pilots	Resolve conflicts	Conflict geometry (360° of lateral conflict angles in 20° increments)	success in avoiding conflicts	Overtake (0°-40°) conflicts resulted in least safe (least successful in avoiding conflict) resolutions; Overtake conflicts were least efficient (in terms of number of vectors and distance deviations/time increases from original flight path)	Avoidance maneuvers limited to lateral only	2-D single-view CDTI
Helleberg et al. (2000)	Aviation	PIL	12 pilots	avoid conflicts	Conflict geometry (crossing, in-trail, and head-on; ascending, descending, level)	success in avoiding conflicts	Vertical maneuvers preferred over lateral; vertical maneuvers tended to be opposite the vertical direction of intruder; lateral maneuvers also opposite the lateral direction of intruder	Climbs were chosen more often than descents - climbs occurred when intruder was level with ownship as well as descending into conflict	2-D coplanar CDTI
Merwin et al. (1998)	Aviation	PIL	30 pilots	avoid conflicts	Display type (2-D coplanar vs. 30° 3-D vs. 60° 3-D CDTIs)	success in avoiding conflicts	2-D coplanar produced more vertical maneuvers; both 3-D displays showed more lat+vert maneuvers than 2-D); 2-D coplanar was generally better for predicted conflicts, but not actual conflicts		

Reference	Domain	Type	N	Task	Independent Variable + Levels	Dependent Variable + Measure	Results	Comment	CDTI/Algorithm Specifications
Merwin et al. (1998)	Aviation	PIL	30 pilots	avoid conflicts	Conflict geometry (crossing, in-trail, and head-on; ascending, descending, level)	success in avoiding conflicts	When chosen, lateral maneuvers showed tendency to turn towards intruder; Ascending/descending conflicts were most difficult across all display types (slight benefit of 2-D coplanar over both 3-D displays)	As noted in Merwin & Wickens (1996) integration task did not produce costs for 2-D coplanar display; however, as clutter and/or effort increases, benefits of coplanar may be attenuated by the increased integration costs	2-D coplanar CDTI
Smith et al. (1984)	Aviation	PIL		avoid conflicts	Conflict geometry (crossing, in-trail, and head-on; ascending, descending, level)	success in avoiding conflicts	Lateral maneuvers most common, tendency to turn towards intruder		display was 2-D single view CDTI
Ellis et al. (1987)	Aviation	PIL		detect and avoid conflicts	Display Type (2-D single-view vs. 3-D 30° CDTI)	success in avoiding conflicts	3-D CDTI supported faster conflict detection, except for head-on conflicts; 3-D CDTI resulted in more "appropriate" maneuver tendencies, as rated by pilots after the fact; 3-D showed tendency for more vertical maneuvers than other types		
Gempler & Wickens (1998)	Aviation	PIL		avoid conflicts	Conflict geometry (crossing, in-trail, and head-on; ascending, descending, level)	success in avoiding conflicts	Vertical maneuvers favored over airspeed, then lateral; Least successful conflict avoidance when intruder was descending into conflict		2-D coplanar CDTI

Reference	Domain	Type	N	Task	Independent Variable + Levels	Dependent Variable + Measure	Results	Comment	CDTI/Algorithm Specifications
Wickens & Morphew (1997)	Aviation	PIL		avoid conflicts	Conflict geometry (crossing, in-trail, and head-on; ascending, descending, level)	success in avoiding conflicts	Pilots avoided conflicts more when intruder was descending (vs. ascending or level); pilots avoided more overtake conflicts than head-on or crossing		2-D coplanar CDTI
O'Brien & Wickens (1997)	Aviation	PIL		avoid conflicts	Display type (2-D coplanar vs. 30° 3-D vs. 60° 3-D CDTIs)	success in avoiding conflicts	Vertical maneuvers favored over lateral; 2-D showed preferences for more combo (lat+vert) maneuvers than 3-D		
van Westrenen & Groeneberg (2003)	Aviation	PIL	12 and 24 pilots	avoid conflicts	Conflict geometry (crossing, in-trail, and head-on; ascending, descending, level)	success in avoiding conflicts	conflicts appeared to be detected fairly well, although a number of conflicts did occur; CDTI was equipped with ASAS conflict prevention module, restricting conflict-producing maneuvering; Conflicts occurred most often with descending (or ascending) aircraft ("bots", which did not have conflict prevention module)	Deliberately trained pilots to resolve conflicts vertically; flew through airspace w/ other pilots, "bots"; flew 12, 14, and 24 craft "superconflicts" successfully	2-D coplanar CDTI
Scallan et al. (1997)	Aviation	PIL		avoid conflicts	Conflict geometry (crossing, in-trail, and head-on; ascending, descending, level); use of color symbology (proximity indications)	success in avoiding conflicts	use of color coding produced earlier maneuvers (presumably due to faster detection of threat); preference for vertical maneuvers w/ no color; use of color encouraged airspeed changes; use of color produced earlier maneuvering but no benefit to safety	the tendency for vertical maneuvers could be due to the 2-D coplanar display or the short-term conflicts	2-D coplanar CDTI

Reference	Domain	Type	N	Task	Independent Variable + Levels	Dependent Variable + Measure	Results	Comment	CDTI/Algorithm Specifications
Wickens et al. (2002)	Aviation	PIL		avoid conflicts	Conflict geometry (crossing, in-trail, and head-on; ascending, descending, level)	success in avoiding conflicts	use of CDTI aided detection of non-CDTI aircraft (seen out the window) vs. a no-CDTI condition; vertical maneuvers were most common, especially in opposite direction of traffic; climbs were more common than descents, and were also chosen when traffic was level with ownship and esp. in head-on conflicts; head-on conflicts were resolved more safely in terms of less time in conflict; BUT vertical maneuvers resulted in more time in conflict than lateral (??)		2-D coplanar CDTI