Effects of CDTI Alerting System Properties on Pilot Multi-Task Performance

Christopher Wickens

Angela Colcombe

Follow this and additional works at: https://corescholar.libraries.wright.edu/isap_2007

Part of the Other Psychiatry and Psychology Commons

Repository Citation

This Article is brought to you for free and open access by the International Symposium on Aviation Psychology at CORE Scholar. It has been accepted for inclusion in International Symposium on Aviation Psychology - 2007 by an authorized administrator of CORE Scholar. For more information, please contact corescholar@www.libraries.wright.edu, library-corescholar@wright.edu.
EFFECTS OF CDTI ALERTING SYSTEM PROPERTIES ON PILOT MULTI-TASK PERFORMANCE

Christopher Wickens:
MA&D Division of Alion Science and Technology, Boulder, Colo.

Angela Colcombe:
The Boeing Company. Seattle Washington

In a dual task context, pilots monitored a simulated cockpit display of traffic information (CDTI) supported by an imperfect auditory or visual alerting system. Across four experiments:

1. Varying the alerting threshold, to reduce the false alarm rate (but increase the miss rate) degraded both the concurrent flight control task, and conflict detection accuracy, with only a slight improvement in detection speed.

2. Changing the dual task context from flight control, to ATC communications greatly degraded performance on both the concurrent task and the conflict detection task, revealing the high cognitive demands of traffic monitoring.

3. Changing ATC communications to a visual data link display greatly improved dual task performance.

4. Auditory conflict alerts were generally more effective than visual, but sometimes degraded the concurrent task more than visual.

Introduction

Designers of aircraft conflict alerting systems are faced with a dilemma. The desire to provide pilots with enough advanced warning (conflict “look-ahead-time” LAT) to avail ample time for planning, selecting and executing avoidance maneuvers trades off against the inevitable loss in automation accuracy in predicting future conflicts, as LAT increases (as in any forecast). This reliability decrement is the result of unpredictable trajectories of the aircraft involved (Kuchar & Yang, 2000, Rantanen & Thomas, 2006; Rantanen et al, 2004).

To this inevitable imperfect reliability is added four other elements characteristic of airborne alerts, the focus of our experiments:

1. As in most imperfect alerting systems, the threshold of traffic conflict alerts is typically set low, in order to avoid conflict misses (or late alerts), but with a consequently high false alarm rate (Bliss, 2003). These false alarms at best, can be annoying or disruptive of ongoing tasks, and at worst can lead to the so-called “cry wolf” syndrome, in which all alerts (including true ones) are delayed in their response, or even ignored altogether (Dixon & Wickens, 2006).

2. Alerts provide their greatest benefit in a multi-task context, because they relieve the pilot of the need for continuous monitoring of traffic (or terrain or weather) data; yet either alert misses or alert false alarms will disrupt concurrent task performance, in potentially different ways, contingent upon the differing cognitive impacts of these two types of “automation failures”. (Meyer, 2004, Dixon and Wickens, 2006). In particular, a miss-prone alerting system, will force the pilot to spend more time monitoring the raw data on a traffic display, to catch the conflicts that the system may miss. This attention diversion can disrupt concurrent task performance. The false alarm prone system will increase the frequency with which pilots need to deal with alerts, interrupting the concurrent tasks, unless those alerts are ignored altogether (the cry wolf syndrome).

3. Alerts are typically auditory, but this modality may be “pre-emptive” and disruptive of those ongoing tasks in a way that visual alerts are not (Iani and Wickens, 2007, Latorella, 1996).

4. Different concurrent tasks may be disrupted differently by the tasks of airborne traffic monitoring and alert responding.

Thus in four experiments we investigate these issues as 48 pilots monitored a simulated cockpit display of traffic information (CDTI), coupled with an imperfect alerting system for traffic conflicts, while carrying out different concurrent tasks. This paper focuses primarily on between-experiment comparisons, in which we examine (a) the effect of shifting the alert threshold (experiment 1 vs experiment 2); (b) the effect of changing the qualitative nature of the concurrent task from flight control, to ATC communications (experiment 1 vs experiment 3); (c) Changing the nature of ATC communications from traditional voice communications to data link-mediated visual (text) communications.
Methods

In all experiments, pilots monitored a screen presenting a simple 2D Cockpit Display of Traffic Information (CDTI), presenting own ship in the lower center, with traffic aircraft continuously entering (approximately every 10 seconds) and leaving the display on quasi random paths, so that at any given time, there were typically four aircraft present. All were flying at the same simulated altitude of own ship. Sometimes a traffic aircraft would enter on a trajectory that would penetrate the 3 mile conflict radius around own ship (a conflict traffic). Non conflict aircraft would fly on a range of trajectories from those just over 3 mile separation, to those several miles over. Upon detecting a conflict, the pilot was asked to click on the relevant aircraft with a mouse, and then press one of two buttons to indicate which way the conflict aircraft should maneuver to best avoid the conflict. (While this response task was not entirely natural, it was designed to force pilots to pay close attention to the traffic geometry, and hence impose a substantial load on traffic monitoring).

Traffic conflicts were coupled with a conflict alert that was either auditory (“traffic traffic”) or visual (color flashing). However this alert was imperfectly reliable, as would characterize such systems with a long tactical look-ahead time (e.g., 2-5 minutes). While overall reliability was constant at 0.75, its qualitative nature varied between experiments 1, 3 and 4 (which all had a typically high (4/1) ratio of false alarms to misses) and experiment 2, which had a 50/50 ratio of false alarms to misses. This variation captures the effect of a designer’s decision to raise the alert threshold in order to lower false alarm rate.

While monitoring traffic on the CDTI, pilots performed a concurrent tracking task, simulating flight control, in experiments 1 and 2. This was first order (e.g., simulating control of pitch and bank, in turbulence), and located adjacent to the CDTI. Visual separation was such that it was easy to notice the flashing color in the visual conflict alert condition, even as the eyes were fixated on the center of the tracking display. Hence scanning was not necessary in order to notice the alert, in the visual condition.

In experiment 3, the concurrent task was changed to one which simulated the demands of air traffic control, the pilot hearing a string of heading airspeed and altitude commands, and being required to compare these with another set of actual values, and compute the difference between them (i.e., the difference between command and actual for each of the three variables). The same task was used in experiment 4, except that the information was presented visually on a display that occupied the same location as the tracking display from experiments 1 and 2.

A between-subjects design was used between experiments, while modality (auditory versus visual) was varied within experiments. The effects of two additional variables manipulated within experiments – tracking demand and likelihood vs. binary alert – are not discussed here, but see Colcombe & Wickens, 2006, Wickens and Colcombe, in press). All subjects were student pilots at the University of Illinois, possessing at least a private pilots’ ratings. Pilots were paid for their participation.

Results

We present below only the most important statistically significant effects, focusing primarily on the between experiment manipulations. Further details can be found in Colcombe & Wickens (2006), or, regarding experiments 1 and 2, Wickens and Colcombe (in press). Across all experiments we focus on three dependent variables: RT and accuracy of discriminating conflicts from non-conflicts, and concurrent task performance.

Experiment 1 vs experiment 2: reducing false alarms

This contrast examined the effect of reducing the false alarm rate, but at the necessary expense of increasing miss rate, given an alert system with constant sensitivity. Thus such a shift might be observed if the alert threshold were increased to reduce annoying false alarms. The consequence of this shift was a slightly (1 sec) shorter response time to failures (observed only when the concurrent tracking task was easy; F 2,22 = 4.37; p<.05). This RT shift, suggested that the “cry wolf” effect was diminished when false alarm rate was decreased. However this faster response was accompanied by a marked 40% decrease in the accuracy of discriminating conflicts from non-conflicts (F 2,22 = 55.7; p<.01), as if suggesting a speed accuracy tradeoff.

Equally prominent was the marked decrease in concurrent task accuracy that resulted from miss-prone automation (12% increase in error: F2,22 = 3.74; p<.05). This deterioration can be directly attributed to the increased need to monitor the raw traffic data, in the face of more automation misses, an effect that Meyer (2004), and Dixon and Wickens...
(2006) have labeled the loss of reliance upon the automated alerting system.

**Experiment 1 vs Experiment 3: Changing the concurrent task**

The change of the concurrent task from the perceptual motor demands of flight control, to the cognitive voice demands of ATC communications, in some respects mimics the kinds of changes that might result from IFR to VFR, coupled with an autopilot for inner loop flight control. This change greatly degraded the accuracy of detecting conflicts (a 30% reduction, $F_2, 24 = 10.28; p<.01$), although it did not affect response time. Such a change clearly reflects the high cognitive demands of the ATC communications competing for resources with those high cognitive demands of traffic trajectory extrapolation, a competition that was greatly reduced with the perceptual motor tracking task, whose demands are more resource-related.

**Experiment 3 vs Experiment 4: Implementing data link**

The procedures for experiment 4 were identical to those of experiment 3 except that the synthetic controller voice was replaced by a visual text display, in which information was conveyed at the same rate as the voice. As such, it mimics the behavior of uplinked data link information. While the predictions of multiple resource theory (Wickens et al, 2003) would suggest that this added form of visual resource competition should degrade performance of one or both tasks, just the opposite effect was observed here. Conflict detection accuracy was nearly 50% higher in experiment 4 with visual data link ($F_2,24=19.06$), and accuracy for the ATC communications task was also improved, from 86% to 96% ($F_2,24 = 24.98, p<.01$).

While this finding is incompatible with the visual competition predicted by multiple resource theory (Wickens, 2002), it is quite consistent with the strategic auditory preemption effect, predicted by preemption theory (Latorella, 1996; Wickens and Liu, 1988; Wickens Dixon and Seppelt, 2005). According to this view, when auditory material is presented of sufficient length so as to require substantial working memory demands, operators must switch attention to deal with the material before it is forgotten. When equivalent information is presented on a permanent visual display (e.g., the data link screen), such switching is no longer urgent, and can be done during more optimal times, thereby better preserving the ongoing visual task. Indeed it is this flexibility of use of data link, for lower priority communications tasks, that serves as one of its main advantages (Navarro and Sikorski, 2001; Helleberg and Wickens, 2003).

**Within experiment comparisons: auditory versus visual alerts**

As was explained with regard to ATC task modality (auditory versus visual data link), so with alert modality, there is again a contrast between multiple resource predictions, that both CDTI and concurrent tasks will be better performed with an auditory alert, and those predictions of pre-emption theory, that The CDTI task will improve, but visual concurrent tasks will degrade, when an auditory (rather than visual) CDTI alert is presented. Across the three experiments with visual concurrent tasks, the results bearing on these theory contrasts were ambiguous. On the one hand, the two experiments with flight control as a concurrent task did indeed show an advantage for auditory alerts (Experiment 1, for both speed and accuracy; experiment 2, for accuracy only), a finding consistent with both theories. On the other hand, the pattern of effects for the concurrent task were entirely ambiguous: experiment 1 (false alarm prone automation), an advantage for the visual alert: experiment 2 (greater automation miss-rate), an advantage for the auditory alert, and experiment 4 (visual data link) no difference.

Finally when we examined the data for experiment 3, with an auditory concurrent task, they revealed that the CDTI conflict was better performed (faster, and more accurate) with visual alerts (consistent with multiple resource theory), but the concurrent auditory task was slightly better supported by auditory alerts on the traffic task, although this effect was small (only a 2% accuracy difference) and actually inconsistent across other variables manipulated in the experiment.

**Discussion**

In a series of four experiments, we have varied three critical factors that could influence the effectiveness and workload of a CDTI alerting system in a dual task context. First, our variation of the threshold revealed that a reduction in alert false alarm rate produces only a small gain in the speed of response, and is more than offset by a loss of conflict detection accuracy, as well as a loss in concurrent task (simulated flight path tracking) performance.

Second, variation of alerting modality revealed the general advantage for auditory over visual alerts, even in these circumstances in which the visual alerts could be easily seen, while attention was focused on the
concurrent task (i.e., no scanning). This effect is predicted by both multiple resource and pre-emption theory. However the benefit of auditory alert presentation to the concurrent task was not always clear.

Third, the concurrent task context in which the alert system is embedded is quite important, and can be extrapolated to be so in high workload single pilot operations: concurrent flight path control is not greatly disruptive compared to concurrent ATC voice interaction. But the latter problems can be considerably offset by the use of visual data link, which allows more flexible scheduling of responding.

In conclusion, the full generality of these results to the flight deck can be questioned, because of the relatively low fidelity of the simulation. It is important that research on these topics be continued in higher fidelity environments.

Acknowledgments

This research was supported by a grant NASA NAG 2-1535 from NASA Ames Research Laboratory. Dr. David Foyle was the scientific technical monitor. The authors acknowledge the valuable assistance in software development from Sharon Yeakel and Ron Carbonari.

References


Wickens, C. D., & Colcombe, A (in press). Performance consequences of imperfect alerting automation associated with a cockpit display of traffic information *Human Factors*
