EFFECTS OF RADAR SIDE CONFLICT PROBE FUNCTIONALITY ON ATC CONFLICT DETECTION PERFORMANCE AND EFFICIENCY

Anthony J. Masalonis
Spectrum Software Technology
Atlantic City International Airport, NJ
Ben Willems
Federal Aviation Administration
Atlantic City International Airport, NJ
Thomas Fincannon
Applied Research Associates
Atlantic City International Airport, NJ

We investigated integrating Conflict Probe (CP) on air traffic controllers' Radar Side (R-Side) displays. Eight controllers worked realistic, high-traffic simulation scenarios alone, using both R-Side and Radar Associate Side (RA-Side) displays. We manipulated CP presence on the R-Side—like today, it always appeared on the RA-Side—and the presence of yellow alerts for near-conflicts. We used established controller performance and workload metrics, plus novel operational analyses not used in past studies. R-Side CP had few workload effects, but increased voice communications when we included yellow alerts. It improved the efficiency of correcting conflict-inducing clearances, and seemed to facilitate proactive control to avert more urgent alerts. Though our simulated CP was less reliable than the current operational version, it showed evidence of benefiting performance and acceptance. Participants commented that R-Side yellow alerts were desirable in moderation. Future research should assess the appropriate alerting criterion.

In Air Traffic Control (ATC), a conflict occurs when two aircraft are closer than the minimum separation standard or an aircraft violates an unauthorized airspace volume. For many years, ATC workstations have included a short-term, tactical Conflict Alert (CA), which alerts controllers to conflicts predicted within about the next two minutes. Each conflicting aircraft's datablock—textual information near the airplane's location symbol—flashes, and the callsigns of the pair appear in a list. The Conflict Probe (CP) helps controllers detect and resolve aircraft conflicts earlier, to enhance safety and efficiency and help controllers manage their own workload. The CP grew out of research and development activities by the Federal Aviation Administration (FAA) and its contactors, who developed the Automated En Route ATC (AERA) concept (Goldmuntz et al., 1981). MITRE's User Request Evaluation Tool (URET, Brudnicki & McFarland, 1997), and the National Aeronautics and Space Administration (NASA)'s Conflict Prediction and Trial Planning (CPTP) tool (Paielli & Erzberger, 1997), extended AERA's capabilities. CP is now an integral part of the operational en route ATC system. It alerts sooner than CA—up to 20 minutes for aircraft conflicts and 40 minutes for airspace—and uses flight plan information, rather than current speed and heading alone. CP and CA are each useful in different settings and are complementary.

A typical en route ATC workstation includes two displays: the Radar Side (R-Side), and the Radar Associate Side (RA-Side), also known as the Data Side (D-side). The R-Side shows information essential to keeping aircraft separated, including each controlled aircraft's position as derived from radar, textual information about each flight, and the aforementioned CA alerts. The RA-Side includes a list of current and future controlled aircraft, with further information about each flight, and its own trajectory-based position display of aircraft.

Currently CP appears only on the RA-Side display. The RA-Side is sometimes operated by a separate controller in a two-person sector team, necessitating communication between the controllers when dealing with conflicts. At other times, a controller works a sector alone and uses both the R- and RA-Side displays, but must attend primarily to the R-Side to support the primary task, separating aircraft. In these settings, controllers often cannot make productive use of the RA-Side, including the CP. Therefore, integrating the CP onto the R-Side should enhance controllers' conflict resolution ability.

The present study was part of a succession of Human-in the-Loop (HITL) experiments for the FAA's Separation Management/Modern Procedures (SepMan) program. Under this program, Zingale, Willems, Schulz, and Higgins (2012) tested the implementation of CP on the R-Side, comparing various display methods and alerting criteria in a HITL simulation using current and retired controllers. They found evidence that controllers working alone and using both displays chose to view the details of the alert for more of the CP alerts when CP was on the R-

Side than when it was only on the RA-Side. Their data also suggested that a rule presenting only conflicts predicted within six minutes on the R-Side resulted in more frequent viewing of notifications by the R-Side controller in a two-person team. Their questionnaire data showed a preference for displaying CP information on the R-Side.

Zingale et al. (2012) analyzed objective safety and efficiency measures, such as losses of separation where controllers did not prevent a pair of aircraft from violating the prescribed separation minima, and efficiency of working aircraft through the sector in terms of time and distance traveled. None of these measures showed significant benefits of R-Side CP, so they recommended further CP research using more varied traffic scenarios. Therefore, we further investigated the CP location question in a HITL study analyzing a wider variety of objective performance data, including controller behaviors not assessed in previous studies of this research program. We also introduced more variety between the scenarios run in each condition to lessen predictability.

Another question regarding CP that had not been previously explored was the degree of reliability required for this automation to support performance and to be accepted by controllers when integrated on the R-Side (Masalonis, Rein, Messina, & Willems, 2013; Rein, Masalonis, Messina, & Willems, 2013). We defined reliability as a combination of Hit and False Alert rates, after Wickens and Dixon (2007). We set out to study whether improvements over the reliability of the currently-fielded algorithms would make R-Side CP functionality operationally acceptable. Here, we will focus mostly on how integration of CP information on the R-side affected operational performance, perceptions, and behaviors. We reported other analyses and results for the same simulation in Willems, Masalonis, Fincannon, Puzen, & Bastholm (2016). We predicted that CP information on the R-Side would improve controllers' conflict resolution performance and reduce workload. We also hypothesized that controllers would perceive locating CP on the R-Side as beneficial.

Method

Participants

Eight current, en route, Certified Professional Controllers—seven male, one female; mean age 50.13 years (SD = 3.94), ranging in age from 42 to 54 years—from the En Route Automation Modernization (ERAM) National User Team participated in a HITL experiment at the Research Development and Human Factors Laboratory at the FAA's William J. Hughes Technical Center (WJHTC). All reported having worked traffic at their facility in the preceding 12 months. Their mean ATC experience was 26.04 years (SD = 2.88), ranging from 21.83 to 30.50 years.

Apparatus

We conducted a high-fidelity ATC simulation using tools developed at WJHTC and/or regularly used there for research and testing: the Distributed Environment for Simulation, Rapid Engineering, and Experimentation (DESIREE), the Simulation Driver and Radar Recorder (SDRR), the Target Generation Facility (TGF), and the ERAM Evaluation System (EES). This was the first HITL simulation to interface EES to DESIREE.

Each controller used an R- and RA-Side workstation. The R-side workstations included a radar display (BARCO 29″ LCD, resolution 2048 x 2048); and a Cortron keyboard, Keypad Selection Device (KSD), and trackball. The RA-side workstations contained an EIZO 30″ LCD monitor, resolution 2560 x 1600, showing the Aircraft List [ACL] View) and Graphic Plan Display (GPD) depicting trajectory-based data, and a Cortron keyboard and trackball. The controllers and the simulation pilots used push-to-talk (PTT) communications through a voice switching and control system that simulated the current operational system. The simulation pilots made requests and responded to clearances as real-life pilots would. They worked in rooms separated from the experiment rooms. DESIREE collected real-time subjective workload assessments every two minutes via the custom-made Workload Assessment Keypad (WAK), based on the Air Traffic Workload Input Technique (ATWIT; Stein, 1985). The WAK has 10 buttons labeled 1 through 10. Participants may press the 1 to indicate very low workload, the 10 for very high workload, or any button in between. We provided a handout detailing the anchors for the scale.

We used a variety of questionnaires to collect participants' demographic information, and their subjective opinions about the CP functionality and concept and about other aspects of the simulation. Over-the-shoulder observers, who were also experienced controllers, used forms to rate participants' performance during each scenario.

In all conditions, CP information appeared on the RA-Side. The system identified aircraft involved in one or more conflicts with a color-coded square(s) depicting the number and type(s) of alert (aircraft or airspace). The color code for airspace alerts was orange. Aircraft alerts were red, meaning that CP predicts the centerlines of the two trajectories to violate separation minima, or yellow, meaning that CP predicts the adherence bounds—buffer zones surrounding each centerline to allow for prediction uncertainty—to get closer than the separation minima. In the conditions where CP was also available on the R-Side, a color-coded square in each conflicting aircraft's data block showed the total number of alerts and the color of the most severe alert; controllers could click it to show the trajectories. CP also provided a list of all alerted aircraft, with the same color-coded square for each flight and an indication of the number of minutes until the conflict.

We simulated a high-altitude sector in Indianapolis Center (ZID). Traffic ranged from about 7 to 26 aircraft over the course of each scenario. ZID personnel consider the real-life sector's capacity to be 19 aircraft.

Design and Procedure

In this paper, we mainly discuss the manipulation of two variables: CP Location, with two levels, Present and Absent (on R-Side); and Algorithm, with two levels (Improved and Legacy). The Legacy condition used the fielded algorithms; the Improved condition introduced features that engineering studies had shown to improve reliability (Crowell, Fabian, Young, Musialek, & Paglione, 2011, 2012). However, most participants reported not noticing or being sensitive to these differences, and objective data showed little evidence that the enhancements affected controller preferences or performance (Willems et al., 2016). The main difference between Legacy and Improved Algorithms of interest to this paper is that in the CP Present, Legacy Algorithm condition, the R-Side displayed CP alerts not only for red and orange alerts, but also for yellow alerts.

Depending on the analysis, we either employed a 2 x 2 design or focused on one of the independent variables. The overall experiment contained other manipulations, in particular a CP Reliability variable which we were not able to manipulate correctly due to simulation errors. The result of this error was that all the conditions had a lower CP reliability than that available in the currently fielded system. We took advantage of this fact to see if R-Side CP would provide benefit at the lower reliability presented to our participants; if so, it should also do so with the current or an improved algorithm.

Four controllers participated at a time. In each run, two participants worked side by side in each of two experimental rooms. The simulations were independent; controllers did not interact with each other. We did not present identical scenarios simultaneously to multiple participants. Each participant worked alone, operating both the R- and RA-Sides. Each group of four spent four days at WJHTC, a full day of training followed by about two and one-half days of testing. The testing sessions comprised twelve 50-minute scenarios. We conducted a short debrief session at the end of most experimental days, and a longer one at experiment end on the fourth day.

Results

Workload

We conducted multiple regression on the individual WAK ratings, with aircraft count and time-on-task as covariates, and unique run nested within Participant as a random effect. The ratings did not vary according to Location, $F(1, 28.1) = 0.55$, $p < .47$; Algorithm, $F(1, 28.02) = 0.069$, $p < .80$; or their interaction, $F(1, 28.02) = 0.092$, $p < .77$. We conducted a similar multiple regression on the communication workload data, except that the only random effect was Participant. In preliminary analysis, this model was more reliable for this model than the model used for WAK. The Location x Algorithm interaction significantly affected the number of voice transmissions, $F(1, 536.1) = 4.83$, $p < .03$. A post-hoc Tukey test showed that the significance resulted from 19 more transmissions per hour in the R-Side Present, Legacy condition than the R-Side Absent, Legacy condition. Similar regression analyses on the number of various types of commands issued showed no meaningful effects of Location or Algorithm.

Subjective Assessments

The results of the subjective assessments are further covered in Willems et al. (2016). To summarize a few results relevant to the present paper, CP on the R-Side increased controller ratings of CP usefulness, and increased over-the-shoulder observer ratings of participants' ability to use the CP in appropriate situations and in a timely and effective manner. Controllers also said that they were more likely to believe and respond to CP alerts when the information was on the R-Side. Observer ratings did, however, suggest reduced Situation Awareness (SA) with R-Side CP due to controllers reacting to the CP information instead of detecting conflicts on their own.

Proactive Altitude Clearances

We identified all cases where the controller instructed a pilot to "expedite" a climb or descent. Controllers do not use "expedite" often, reserving it for urgent situations such as impending conflicts. For this analysis, we determined whether each expedited clearance occurred after a CA activation on the given aircraft. This sequence of events would indicate that the highly tactical CA automation detected a conflict event before the participant, and the expedite clearance was a purely reactive response to something the controller had not known about. We classified each event according to whether a CP alert had activated for that particular aircraft. Therefore, if an event happened during an R-Side CP Present condition but there was no CP alert for the aircraft in question, it was classed as "No Alert." In four of the 15 cases where the expedited aircraft had not received a CP Alert on the R-Side, there was a CA associated with the aircraft (see Figure 1). In the seven situations where the aircraft did receive a CP alert, the controller always proactively gave an altitude clearance before an imminent conflict triggering a CA could develop.

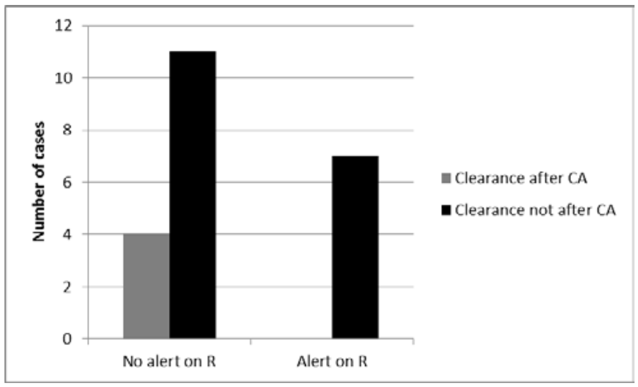

Figure 1. Number of Expedite clearances occurring after/not after CA, by R-Side CP Presence.

We did not have enough observations for statistical significance; with Yates' correction for low expected frequencies (Yates, 1934), $\chi^2(1, N = 22) = 0.85$, $p < .36$. However, it is operationally notable that in the entire experiment, it was never the case that a situation where the controller received a CP alert on the R-Side escalated to the point where it seemed necessary to expedite an altitude change due to a CA. This finding suggests that CP on the R-Side facilitated proactive conflict resolution behavior.

Modified Clearances

Clearances requiring an amendment sometimes indicate that the initial clearance was inappropriate for some reason, such as being the cause of a potential conflict. However, quickly amending these indicates good performance: the controller has corrected a less-than-ideal decision before a conflict or other problem resulted.

For this analysis we focused on cases where the participant gave and then later revised an instruction. We identified 22 relevant cases for analysis, all of which involved altitude clearances. Of these, 10 had an R-Side CP alert on the given aircraft, and 12 did not—either by virtue of being a CP Absent condition or because there was no alert for that particular aircraft.

We compared the time elapsed between the initial and corrected/rescinded clearances. Participants took 37 seconds longer to correct altitude clearances in R-Side Absent than in R-Side Present events (see Figure 2). A within-subjects *t* test showed the difference to be statistically significant, $t(19) = 2.50$, $p < .03$.

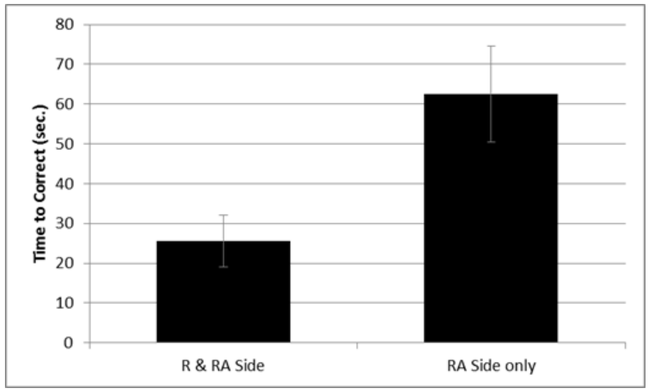

Figure 2. Time to correct altitude clearances, by R-Side CP Presence. Error bars represent standard error of the mean.

Behavioral Observations and Participant Comments

Observers and experimenters witnessed productive use of the R-Side display during the simulations, with controllers issuing clearances to aircraft with yellow CP alerts on the RA-Side in R-Side CP Absent runs. Conversely, situations occurred where the lack of a yellow alert on the R-Side resulted in a delay in acting, and aircraft nearing the separation minima—sometimes resulting in a CA—followed by the controller expressing discomfort with the lack of an R-Side alert. Cases such as these serve as anecdotal evidence that some yellow alerting may be beneficial on the R-Side. Therefore, during the debriefing sessions, experimenters raised the topic of presenting yellow CP alerts on the R-Side. Some participants held that R-Side yellow alerts could represent too much information: two of the eight controllers explicitly articulated a benefit to only showing red alerts, and a third participant called running with yellow alerts on the R-Side a "waste."

Comments in the debriefing in favor of R-Side yellow alerts focused on the notion that these alerts might be acceptable if the reliability were higher. Five of the eight controllers mentioned the potential benefit of presenting yellow alerts defined according to a more intuitive criterion. As mentioned earlier, CP's current definition of a yellow alert is when the conformance boxes surrounding the trajectory line, rather than the centerlines themselves, are predicted to violate separation standards. This rule sometimes misleads controllers as to why a given situation resulted in a yellow vs. red vs. no alert. Therefore, in the debriefing we discussed the concept of coloring an alert yellow based on the predicted distance between the centerlines: if this distance was greater than 5 nautical miles for aircraft at the same altitude, it would result in a yellow alert, regardless of the proximity between the conformance boxes. The question then arises as to what the largest separation should be that would still generate a yellow alert. Not all participants commented on this topic, but those who did exhibited a range on the exact distance they would prefer, with three explicitly stating preferred distances: 5.5, 6, and 8 nautical miles.

Discussion

Analyzing additional types of operational performance not addressed previously in this research program or, in some cases, any previous ATC HITLs, allowed the derivation of a fuller picture of CP's effects on controller performance. The CP display integration reduced the time to correct suboptimal altitude clearances by more than 50%, a result both statistically and operationally significant. Analysis of expedited clearances, while not powerful enough for statistical significance, showed that R-Side CP alerts always prevented controllers from having to issue an expedited clearance in response to a CA. We recommend using these performance measurements in future work.

Integrating CP on the R-Side might cause controllers to react too often to alerts, especially those that do not necessarily require action, resulting in higher workload and lower performance. The present results provide a mixed answer to this concern. The lack of increase in subjective workload or most objective workload measures with CP on the R-Side suggests that this is not an issue. However, R-Side CP increased communications workload, and hindered performance on tasks like timely verbal handoffs to the next frequency. Our operationally-experienced observers also indicated that R-Side CP might compromise SA. This latter finding corroborates previous research, such as Endsley and Kaber (1999), who stated that depending on automation can reduce SA.

The debrief discussions of yellow alerts show mixed opinions about whether these should appear on the R-Side, and how the answer to this question is affected by reliability. The consensus among participants appeared to be that with an acceptable reliability level, yellow alerts would be beneficial especially if defined in an intuitive manner, such as the number of miles of predicted separation between trajectory centerlines. The present study set out partly to establish the overall level of reliability needed for CP acceptance on the R-Side, but did not systematically investigate how red versus yellow alerts should be defined, whether yellows should appear on the R-Side, and how reliability affects these answers. These are all important future research areas.

The experimental design introduced a potential confound. The R-Side yellow alerts appeared only in the Legacy Algorithm condition, where CP had a higher false alert rate, so that the effects of yellow alerts cannot be totally removed from the lower reliability experienced in this condition. However, as mentioned earlier, several participants commented that they did not notice the reliability differences introduced by the algorithmic manipulations, and this factor should not seriously compromise the conclusions. Furthermore, the fact that the CP reliability was lower in the same conditions where R-Side yellow alerts were present served as a catalyst for the debriefing discussion and revealed useful insights.

CP's reliability in the runs analyzed for this paper was lower than in the fielded system, enabling us to conclude that the observed benefits of CP, and the fact that controllers were generally willing to accept the R-Side integration as evaluated in this experiment, would extend to the field, even without improving the fielded algorithm.

References

Brudnicki, D. J., & McFarland, A. L. (1997). *User Request Evaluation Tool (URET) conflict probe performance and benefits assessment*. McLean, VA: The MITRE Corporation.

Crowell, A., Fabian, A., Young, C., Musialek, B., & Paglione, M. (2011). *Evaluation of prototype enhancements to the En Route Automation Modernization's conflict probe* (DOT/FAA/TC-TN-12/3). Atlantic City International Airport, NJ: Federal Aviation Administration William J. Hughes Technical Center.

Crowell, A., Fabian, A., Young, C., Musialek, B., & Paglione, M. (2012). *Evaluation of prototype enhancements to the En Route Automation Modernization's conflict probe with an updated traffic sample from Chicago* (DOT/FAA/TC-TN-12/56). Atlantic City International Airport, NJ: Federal Aviation Administration William J. Hughes Technical Center.

Endsley, M., & Kaber, D. (1999). Level of automation effects on performance, situation awareness, and workload in a dynamic control task. *Ergonomics, 42*(3), 462–492.

Goldmuntz, L., Kefaliotis, J. T., Weathers, D., Kleiman, L. A., Rucker, R. A., & Schuchman, L. (1981). *The AERA Concept* (FAA-EM-81-3). Washington, DC: Federal Aviation Administration.

Masalonis, A. J., Rein, J. R., Messina, J., & Willems, B. (2013). Accuracy assessment of air traffic conflict probe prototype for operational evaluations. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57*(1), 66–70.

Paielli, R. A., & Erzberger, H. (1997). Conflict probability estimation for Free Flight. *Journal of Guidance, Control, and Dynamics, 20,* 588–596.

Rein, J. R., Masalonis, A. J., Messina, J., & Willems, B. (2013). Meta-analysis of the effect of imperfect alert automation on system performance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57*(1), 280–284.

Stein, E. (1985). *Air traffic controller workload: An examination of workload probe* (DOT/FAA/CT-TN84/24). Atlantic City International Airport, NJ: Federal Aviation Administration William J. Hughes Technical Center.

Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science, 8*(3), 201–212.

Willems, B., Masalonis, A. J., Fincannon, T., Puzen, G. M., & Bastholm, R. A. (2016). *Separation management human-in-the-loop simulation: Evaluations of conflict probe alert reliability requirements and conflict probe algorithmic improvements* (DOT/FAA/TC-16/38). Atlantic City International Airport, NJ: Federal Aviation Administration William J. Hughes Technical Center.

Yates, F. (1934). Contingency table involving small numbers and the χ^2 test. Supplement to the *Journal of the Royal Statistical Society, 1*(2), 217–235.

Zingale, C. M., Willems, B., Schulz, K., & Higgins, J. S. (2012). *Separation management (SepMan 2): Human factors simulation: Evaluation of conflict probe location and format, display monitors, and pointing devices* (DOT/FAA/TC-12/62). Atlantic City International Airport, NJ: Federal Aviation Administration William J. Hughes Technical Center.