The Complexity of Signal Detection in Air Traffic Control Alert Situations

Kenneth R. Allendoerfer
Federal Aviation Administration
Atlantic City Int'l Airport, NJ

Shantanu Pai
Engility Corporation
Egg Harbor Township, NJ

Ferne J. Friedman-Berg
Federal Aviation Administration
Atlantic City Int'l Airport, NJ

Air traffic controllers continually monitor the traffic situation in their sectors and take action when they detect potentially hazardous situations. Automation systems simultaneously and independently monitor the situation and provide alerts when the situation meets defined criteria. The decisions made by the controllers and the automation systems may agree or disagree. Signal Detection Theory (SDT) provides a theoretical framework for understanding how controllers and automation systems make these decisions. However, traditional SDT provides an incomplete explanation of decision-making in the real-world ATC situations. In this paper, we examine instances where controllers take actions independently of the alert and where controllers take actions in response to an alert, but delay their actions until more information is available. Results from this study are applicable to other domains where operators are tasked to monitor situations while simultaneously monitoring the output of an alerting system.

SIGNAL DETECTION THEORY

Signal Detection Theory (SDT) is a theoretical framework for understanding how people make decisions under uncertain and noisy conditions (Green & Swets, 1988). In the SDT framework, a monitor looks for specific information known as the signal. Information that is not part of the signal is considered noise. Alerting systems have been considered in SDT terms in a variety of domains (Sorkin & Woods, 1985).

According to SDT, the monitor sets an internal criterion for how much evidence must be accumulated before it decides that a signal is present. The criterion affects the number of hits, misses, correct rejections, and false alarms. If the signal and noise are hard to distinguish, and the monitor sets a low criterion, there will be a high rate of hits and a high rate of false alarms. If the monitor sets a high criterion, there will be a low rate of false alarms but also a low rate of hits. When the monitor is a human, the criterion is based on the perceived costs and benefits of the four decision categories, and the criterion can shift dynamically as costs and benefits change. When the monitor is an automation system, the criterion is set by the system designers and can be fixed or dynamic depending on the capabilities of the system.

Federal Aviation Administration (FAA) air traffic controllers receive alerts that warn of potentially hazardous situations. In air traffic control (ATC), the cost of a miss (i.e., the controller concludes that no hazard exists when one actually does) could be extremely high. Failing to take necessary action or taking an action late could lead to a loss of separation or even an accident. The cost of a false alarm (i.e., the controller concludes that a hazard exists when one actually does not) is not zero, but is much lower than a miss. Taking an unnecessary control action wastes fuel, causes delays, and creates workload for the controller and pilot.

A model ATC monitoring system

ATC is a complex decision-making environment requiring continuous monitoring for potential hazards. Controllers and automation systems independently monitor the traffic. The system generates an alert when it determines that a potential hazard exists according to its internal criteria. The alert contributes to the controller's overall assessment of the situation. In this way, controllers monitor both the traffic and the automation simultaneously. Figure 1 depicts a combined human-alert system where three separate monitoring activities occur simultaneously.

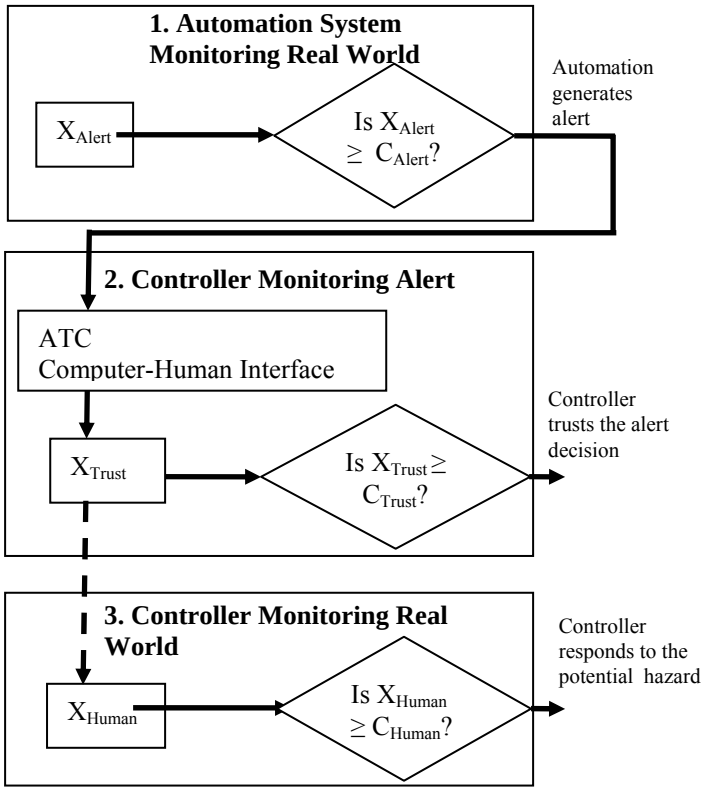

Figure 1. Three monitor signal detection model applied to ATC (adapted from Sorkin & Woods, 1985).

10.1518/107118108X348954

1. The ATC automation system continuously monitors parameters such as aircraft speed, heading, altitude, and current distance from terrain and other aircraft. In SDT terms, these parameters contribute to an overall statistic, X_{Alert}. The automation system activates an alert when X_{Alert} is greater than a criterion, C_{Alert}.

2. Controllers monitor for alerts generated by the automation system. When controllers see or hear an alert, they decide whether they trust the alert (and therefore should respond) or distrust the alert (and therefore should wait for more evidence before responding). In Figure 1, the controller's trust of the alert is denoted by X_{Trust} and the criterion is denoted by C_{Trust}.

3. Controllers actively monitor the traffic independent of the automation system. Controllers have knowledge and experience regarding the traffic that may not be represented in the alert algorithms. Controllers also have information, such as pilot intent, that the automation does not have. For this reason, controller decisions about whether a potential hazard exists will sometimes disagree with the automation system. When controllers disagree with the automation, they may take action without being prompted by an alert or may not take action even though an alert has activated. In Figure 1, the evidence accumulated by the controller is denoted by X_{Human}, and the controller's internal criterion is C_{Human}. The dotted arrow in Figure 1 indicates that the presence of an alert contributes towards the controller's assessment of the signal X_{Human}. However, controllers may respond to a situation based on their own assessment of the situation, even when no alerts are active.

ATC automation systems activate Conflict Alerts (CAs) when the flight paths of two aircraft are projected to lose separation within a short time. CAs appear on the radar display with special codes (e.g., text, flashing, color) and, in some environments, with accompanying audible alerts. Minimum Safe Altitude Warnings (MSAWs) activate when the altitude of one aircraft is projected to be lower than a designated safe minimum for a geographic area. ATC procedures do not require controllers to issue control actions in response to CAs or MSAWs. Instead, procedures require controllers to assess the situation indicated by the alert and determine if any action is required. In many cases, no response is the right response (Allendoerfer, Friedman-Berg, & Pai, 2007).

To illustrate our model using actual CA and MSAW data, we used data collected for a larger human factors study of ATC alerts (Allendoerfer, Friedman-Berg, & Pai, 2007). In particular, we asked the following questions.

1. Do controllers take immediate action when an alert activates or do they wait for more information? If controllers wait, it would suggest that controllers and automation disagree in some situations, and that controllers sometimes distrust alerts that have activated.

2. If controllers do not respond to alerts immediately, what are some possible explanations for this delay?

3. Do controllers respond to situations without being prompted by the alert? If so, this would suggest that controllers monitor the traffic situation independently and make their own decisions about whether a potentially hazardous situation exists. It would also suggest that controllers sometimes distrust alerts that have *not* activated.

METHOD

We analyzed recordings provided by the automation and communication systems at 5 Air Route Traffic Control Centers (ARTCCs) and 17 Terminal Radar Approach Control (TRACON) facilities. The automation data included the activation time of the alert, the call signs and locations of the aircraft involved, and derived values, such as the projected closest point of approach and the closing speed. The communication data included the radio transmissions between controllers and pilots and the phone or intercom calls between controllers.

For each alert identified in the automation data, we listened to the voice recordings 5 minutes before and 5 minutes after the alert activation and transcribed all communications related to the affected aircraft. In most cases, the controllers did not respond to the alert at all or responded to the situation before the alert activated. From the original dataset, we identified CAs and MSAWs for which positive control instructions were issued in response to the alert; these were considered as hits in SDT terms. A control instruction is a change in speed, altitude, heading, or route. For control instructions that were issued after the alert activation, we calculated a response time.

We also determined how many of the aircraft involved in a CA or MSAW situation received control instructions *before* the alert activated. This helps us identify controller actions that were not prompted by the alert, but instead resulted from controllers' own assessments of the traffic situation.

RESULTS

The median response time (i.e., the time after the alert activated until the controller began to issue a control instruction), following a CA was 88 seconds. This delay in responding suggests that controllers often wait to see how a situation develops before taking action in response to CAs. This is not an indication that such alerts are unnecessary or nuisances, because the alerts eventually did lead to controller action. However, such delays suggest that controllers do not consider many CAs to be situations requiring immediate action.

Response times

The median response time following an MSAW activation was 38 seconds. Like CAs, this indicates that controllers wait for low altitude situations to develop further before responding. However, controllers do not wait as long for MSAWs as they do for CAs. These results suggest that controllers assessing MSAW situations find them to be more urgent than many CAs. This could be because MSAWs have a greater degree of certainty than CAs (i.e., the ground will not unexpectedly turn) or because the controller has fewer

options for addressing MSAW situations (e.g., the only option may be to issue a climb).

The response times showed a wide range. The shortest response time to a CA was 3 seconds, the longest was 294 seconds, with an interquartile range of 111 seconds. The shortest response time to an MSAW was 3 seconds, the longest was 339 seconds, with an interquartile range of 27 seconds. The wide ranges suggest that controller responses to alerts depend on the individual situation–controllers trust the alert in some situations and respond to it immediately. In other situations, they do not trust the alert completely and wait for additional evidence that a response is necessary.

Controller responses before alert activation

Of the 394 aircraft involved in a CA that received a controller response, 67% received the response before the automation system activated the alert. Of the 56 aircraft that were involved in a MSAW that received controller response, 68% received the response before the alert activated. This result indicates that controllers appear to continually search for potential hazards on their own and take actions proactively rather than waiting for the alert system to tell them when action is necessary.

DISCUSSION

Explanations for controller response times

SDT provides a reasonable explanation for controller response times. Figure 2, adapted from Wickens (1992), depicts the automated systems probability of observing a specific value of signal X_{Alert} in the presence of noise or signal. The curve to the left represents the noise distribution and that on the right represents signal distribution. The system criterion is indicated by the vertical line C_{Alert}. The automation system generates an alert for all values of X_{Alert} to the right of C_a and rejects all signal values to the left of C_{Alert}. The criterion is selected so the system minimizes misses at the cost of increasing the number of false alarms.

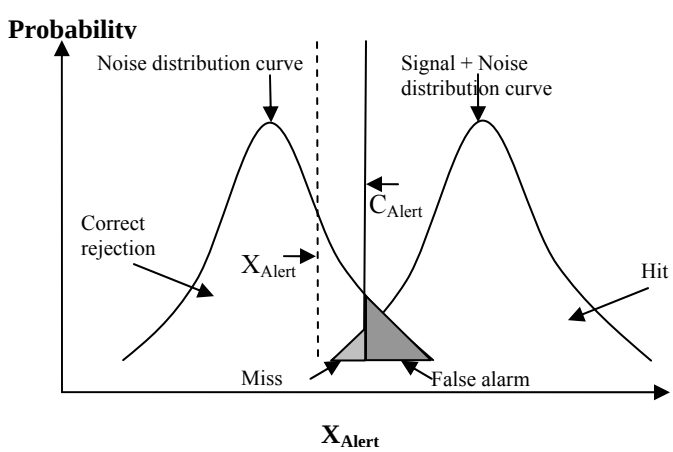

Figure 2. Plot of probability of signal detection in a signal-noise environment.

Several factors can affect X_{Alert} and push it toward or away from the criterion C_{Alert}, as shown in Figure 3. These factors include error in the surveillance data, aircraft and airspace configurations, and maneuvers. Surveillance data contains inherent error that can make it appear that a potential hazard exists when it does not or can hide a real potential hazard. Some aircraft configurations (e.g., two aircraft 20 nm apart, opposite heading) and maneuvers (e.g., one aircraft turning toward another) are more indicative of potential hazards than others. Note that the direction of the factors affecting X in the various figures are drawn arbitrarily. For example, aircraft configurations could push X_{Alert} toward C_{Alert} in some situations and away from C_{Alert} in others.

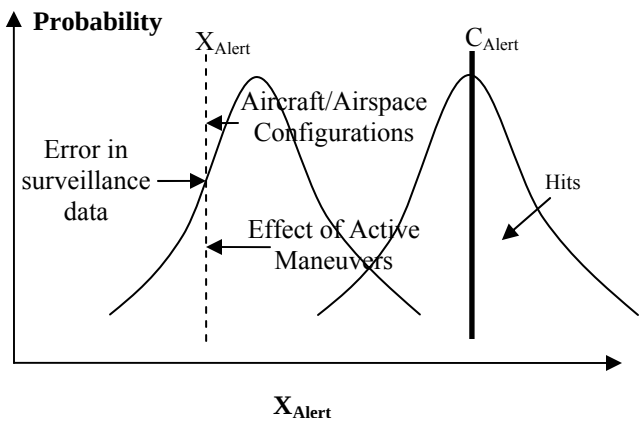

Figure 3. Factors influencing the alert signal in the automation system

These factors affect if and when the signal will reach the criterion and whether an alert will activate. When an alert activates, and the controller sees or hears it, the controller must decide whether to trust it. Because ATC safety alerts are designed to activate before a loss of separation occurs and are designed to provide controllers with enough time to prevent an operational error (OE), controllers often have time to allow a situation to develop before they absolutely must take action.

Figure 4 shows how different factors affect controllers' decision whether to trust the alert. Note that controller's perception of the alert, $X_{Trust,}$ is subject to similar forces as the signal calculated by the automation system, $X_{Alert.}$ If a human believes that the alerting system is generally accurate for this type of situation, users will naturally come to rely on it (Parasuraman & Riley, 1997). On the other hand, a history of false alarms in similar situations can increase desensitization towards the alerts. Increased trust in the alert will push X_{Trust} toward C_{Trust}.

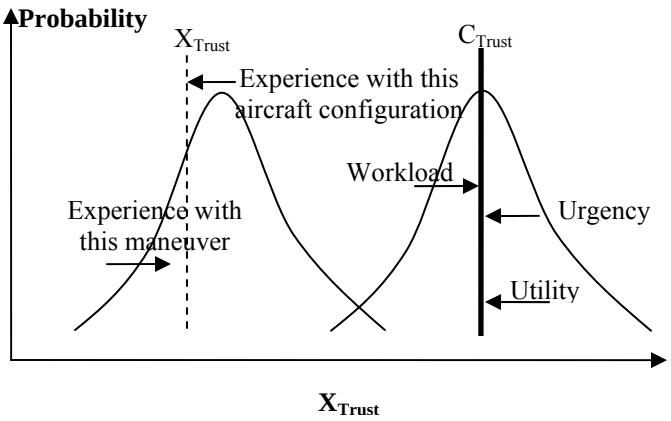

Figure 4. Factors influencing controller trust of an active alert.

The controller's decision-making criterion regarding the alert, C_{Trust}, can be affected by factors that the controller experiences but the automation system does not.

Workload. Controllers are normally responsible for multiple aircraft at the same time. If the controllers' workload is very high, they may not respond immediately to an alert because they have higher priority tasks elsewhere. Increased workload will lead the controller to adopt a higher criterion C_{Trust} for responding. However, controllers might respond immediately to an alert in an equivalent situation when their workload is otherwise low. Thus, increased workload may lead to increased response time and vice versa.

Urgency. Some alerts indicate situations that will become hazardous in a minute or two, others indicate situations that already have or that will become hazardous in a few seconds. High urgency will push C_{Trust} to the left; less evidence is needed to respond to very urgent situations. High urgency should lead to faster response times, even when workload is high and trust is low. When the urgency of the situation is low, the controller may decide to not respond immediately and revisit the situation later when more information is available or when workload has changed.

Outside utility of action. Controller actions may have costs and benefits beyond addressing the potential hazard. For example, controllers typically address a conflict situation by instructing pilots to change their heading, speed, or altitude. Each action may have a ripple effect on other aircraft in the vicinity that are not involved with the conflict. This effect can be beneficial, resulting in better routes or shorter delays for other aircraft, or it can be costly, resulting in unnecessary route changes or workload. This is especially true in the terminal ATC environment where controllers have a smaller amount of airspace. If responding to an alert has outside benefits, controllers may shift C_{Trust} to the left, requiring less evidence before responding. If responding to the alert would create numerous other problems, controllers may shift C_{Trust} to the right, requiring more evidence before responding.

The factors that introduce variability into the level of trust and the factors that introduce variability into the controller trust criterion may help explain the high variability in response times. As a consequence of lowering C_{Trust} in response to outside factors, a controller may respond to a CA more quickly than a controller with a higher $C_{Trust.}$ For example, we would predict that workload, urgency, and outside utility would correlate with response time.

Explanations for controller responses before alert activation

Figure 5 depicts how controllers monitor the situation independently from the alert system. The signal, as perceived by the controller independently of the alert, is denoted by X_{Human}. Controllers' internal criterion for response is denoted by C_{Human}. X_{Human} and C_{Human} are both subject to forces similar to those depicted in Figures 3 and 4. However, in this case, the output of the controller's decision to trust an active alert also the alert influences X_{Human}. That is, if the alert activates and the controller trusts it ($X_{Trust} > C_{Trust}$), the alert serves as additional evidence that a potential hazard exists, and shifts X_{Human} to the right. However, we are skeptical that the reverse is true–it seems unlikely that controllers treat the absence of an alert as evidence of the absence of a hazard. This force is represented by a dotted arrow in Figure 5 as the influence of $X_{Trust..}$

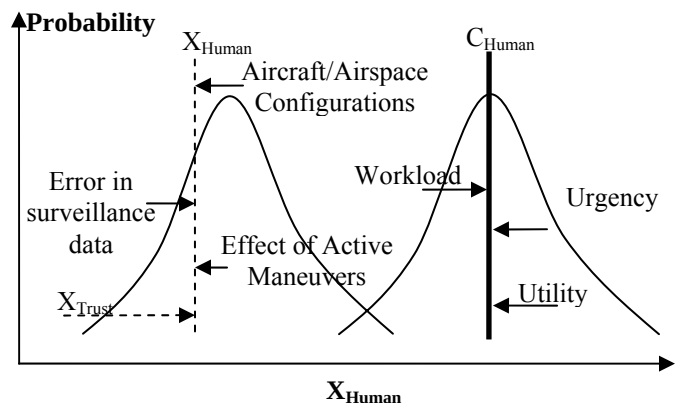

Figure 5. Factors influencing controller assessment of the signal and criterion selection.

Other response categories

The traditional SDT categories do not always apply cleanly to real world applications. In particular, ATC alerts are intended to provide controllers with time to prevent the hazardous situation from occurring. Therefore, the timeliness of the alert is critical and must be considered when evaluating its effectiveness. All ATC alerts incorporate a look-ahead time parameter that is intended to activate an alert *before* a loss of separation occurs with enough time for controllers to respond and the results of their actions to be realized.

Incorporating a look-ahead time further complicates the SDT model. For example, a potential loss of separation might exist and CA might activate (i.e., an SDT hit). However, due to surveillance error effects on X_{Alert} the alert may activate too late to prevent the loss. How should we judge the performance of the alert? It is successful in that it detected a

hazardous situation, but it is unsuccessful in that it did not help controllers prevent an OE. Even so, the alert did activate in time to prevent a collision and contributed positively to safety. We call this category Hit But Late (HBL). A large number of HBLs is undesirable, despite the high hit rate, and would indicate that the alerting system needs improvement. In the SDT model laid out in this paper, a HBL would occur when $X_{Alert} \geq C_{Alert}$ (the alert activates) and $X_{Human} < C_{Human}$ (the controller has not taken action independently). It is unclear how X_{Trust} and C_{Trust} affect HBLs because the alert activated late. It is possible that the controller distrusts the alert and would not have acted even if it had activated on time. It is also possible that the controller trusts the alert too much, and chose to not respond because the alert had not activated. To guard against HBLs, the alert algorithms incorporate long look-ahead times.

However, long look-ahead times affect the other decision categories. For example, suppose two aircraft are heading toward the same fix in opposite directions. The algorithm projects their positions and determines that they will lose separation in 75 seconds. However, unknown to the system, one aircraft is planning to turn 20 degrees to the left in 12 seconds. Because the algorithm does not have access to this intent information, a CA activates. In this case, the false alarm occurred only because the time parameter forced the algorithm to make a decision before all the information was known. If the system were able to wait 12 seconds longer, the false alarm would not have occurred. We call these situations False Alarms But Early (FABEs). Controllers' knowledge of the traffic may cause them to distrust alerts that fall in this category and wait before taking action. In the example, the controller expects the turn to occur and waits to respond to the alert until after the turn is complete and no potential hazard exists. From the SDT standpoint, FABEs result from situations where $X_{Alert} \geq C_{Alert}$ (the alert activates), $X_{Human} < C_{Human}$ (the controller has not taken action independently), and $X_{Trust} < C_{Trust}$ (the controller does not trust the alert because the controller has information that the alert does not have).

Like traditional SDT hits and false alarms, HBLs and FABEs trade off each other. A longer look-ahead time will yield fewer HBLs but also more FABEs. A shorter look-ahead time will yield fewer FABEs but also more HBLs. Choosing the right look-ahead time is essential to the effectiveness of the alert and FABEs may play a significant a role in controller trust and response times.

CONCLUSION AND RECOMMENDATIONS

We examined how an alerting system and a human monitor interact and how the complexities in the overall human-automation system can lead to different decisions. We discussed how traditional SDT can be expanded to include multiple monitors operating simultaneously, controller trust in automation, and the effect of look-ahead time. A truly comprehensive analysis of ATC alerts must consider all these factors in addition to the usual SDT considerations of sensitivity and criterion placement.

In particular, future research should examine how three separate monitoring systems interact. Researchers should explore how different levels of urgency, workload, and outside utility affect response criteria placement and controllers' trust in the alert system. For example, which of these factors trumps the others? How high must workload become before a controller does not respond to an urgent alert?

Researchers also should examine how different alert presentation techniques affect criterion placement and trust. For example, the ATC alerts could be enhanced to track variables that correlate with controller workload, such as number of aircraft being handled. When the system detects a high workload situation, it could delay presentation of the alert until workload improves or the situation becomes more urgent.

Alternately, Sorkin and Woods (1985) suggest using an alert that is coded to indicate the conservatism of the criterion. For example, an alert triggered by a conservative criterion (i.e., one that requires more evidence) might be coded in flashing red to indicate that the alert is "very certain" that a potential hazard exists. Controllers might trust such alerts more because the alert would provide information about its own trustworthiness.

The data used for analyses comes from the ATC environment. However, we believe that environments that employ alerting systems face issues similar to those discussed in this paper. In particular, our approach may be useful to environments where human operators maintain some autonomy to respond independently of the alert.

REFERENCES

Allendoerfer, K., Friedman-Berg, F., & Pai, S. (2007). *Human factors analysis of safety alerts in air traffic control* (DOT/FAA/TC-07/22).

Green, D.M., & Swets J.A. (1988). *Signal Detection Theory and Psychophysics*. Los Altos, CA: Peninsula.

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

Sorkin, R. D., & Woods, D. D. (1985). Systems with human monitors: A signal detection analysis. *Human-Computer Interaction*, *1*, 49-75.

Wickens, C. D. (1992). *Engineering psychology and human performance*. New York: HarperCollins.