ATC Signaling Systems:
A Review of the Literature on Alarms, Alerts, and Warnings

Keith J Ruskin, MD
Professor of Anesthesia and Critical Care
University of Chicago
5841 S Maryland Avenue, MC4028
Chicago, IL 60637
773-702-6700
ruskin@uchicago.edu

Chase Corvin, MD, MBA
Department of Surgery
University of Chicago

Stephen Rice, PhD
Professor of Human Factors
Embry-Riddle Aeronautical University

FAA Program Manager
Dan Herschler
NextGen Human Factors Division (ANG-C1)

TECHNICAL REPORT DOCUMENTATION PAGE

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
4. TITLE AND SUBTITLE ATC Signaling Systems: A Review of the Literature on Alarms, Alerts, and Warnings		5. REPORT DATE Sep 2020 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Keith J Ruskin, MD; Stephen Rice, PhD; Chase Corvin, MD		8. PERFORMING ORGANIZATION REPORT
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Chicago Department of Anesthesia and Critical Care 5841 S Maryland Avenue, MC4028 Chicago, IL 60637		10. WORK UNIT NO. 11. CONTRACT OR GRANT NO. 692M151940006
12. SPONSORING AGENCY NAME AND ADDRESS Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591		13. TYPE OF REPORT AND PERIOD COVERED Final Report for Phase 1: Sep 2019 - Sep 2020 14. SPONSORING AGENCY CODE Human Factors Division, ANG-C1

15. SUPPLEMENTARY NOTES

ATO Research Requirements Roundtable item AJM-FY19-5. The FAA does not have a Human Factors handbook or guide to assist in the development of system design or associated training requirements for integration of alarms and alerts into air traffic systems. While the FAA has a few standards associated with AT system alarms and alerts, there is no relevant, guiding documentation for how to develop or integrate alarms and alerts or how to properly train system users of alarms and alerts. The FAA should determine a standardized methodology to determine how and when to use various kinds of alarms and alerts during system design and development in order to help reduce the number of each and their potential for repetitiveness, redundancy, conflict or confusion, and nuisance caused by any of these conditions. Similarly, end-user training standards should be developed to help reduce confusion, misunderstanding, misinterpretation, of alarms and alerts due to absent or inadequate understanding of each

16. ABSTRACT

The Federal Aviation Administration's air traffic control organization encompasses a variety of facilities that include towers, terminal radar approach control facilities (TRACONs), and air route traffic control centers (ARTCCs). ATO facilities are dynamic environments that require rapid decision making to manage rapidly evolving situations. Controllers routinely interact with pilots of varying skill levels who are flying aircraft with different capabilities at different speeds, altitudes and trajectories. Controllers rely upon signals to indicate that an abnormal event may occur in the future or when an occurrance requires immediate attention to avoid an accident or incident that may result in harm or loss of life. The most critical goals for developing signals are to provide well-formatted information to the controller, clearly and accurately indicating potential hazards, and minimizing unnecessary alarms. Signals should be easy for controllers to identify and should be consistent across all equipment with which a controller is required to interact. This comprehensive literature review describes the work done as part of the first phase of a five-year project in which we document the current understanding of signals and identify potential knowledge gaps. In the remaining phases we will develop a signaling philosophy for air traffic control, create a handbook that will offer advice on development of signals that will be used by air traffic controllers, validate the handbook with stakeholders, and train controllers on the optimal use of signals.

17. KEY WORDS Signal; alarm; alert; warning; air traffic control; aviation; automation; safety	18. DISTRIBUTION STATEMENT		
19. SECURITY CLASSIF. (OF THIS REPORT) Unclassified	20. SECURITY CLASSIF. (OF THIS PAGE) Unclassified	21. NO. OF PAGES 78	22. PRICE N/A

Form DOT F 1700.7 (08/72)

ATC Signaling Systems:
A Review of the Literature on Alarms, Alerts, and Warnings

Introduction

The Federal Aviation Administration's air traffic control organization (ATO) encompasses a variety of Air Traffic Control (ATC) facilities that include towers, terminal radar approach control facilities (TRACONs), and air route traffic control centers (ARTCCs). ATC facilities are often high-stress environments that require rapid decision making in the setting of dynamic situations. Controllers routinely interact with pilots of varying skill levels who are flying aircraft with different capabilities at different speeds, altitudes and trajectories. ATC is a so-called "High-3" industry: high-technology, high-intensity, and high-reliability, in which errors can be life-threatening and are unacceptable. Organizational and technological aspects of the work environment in these organizations are also highly complex. In this respect, ATC shares many similarities with other transportation industries (e.g., rail and shipping), and with finance and medicine, among others. In fact, Owens (2017) ATC has been described as an integral part of the global "knowledge economy."

Air traffic controllers continually evaluate the impact of such factors as weather, converging traffic, and emergency situations, using this information to prioritize tasks and solve problems before they affect flight safety. Controllers generally rely on prospective memory, which involves remembering things that must be done in the future (Dismukes, 2012) to perform their jobs. Some situations require controllers to rely only upon established procedures, understanding of their airspace, and prospective memory to help aircraft avoid some problems (*e.g.,* severe weather) for which there is no alarm or alert depicted on their display. They must also constantly anticipate events and work proactively to avoid loss of separation or other situations that would require an alarm, alert or warning (collectively called *signals*). Excessive false, misleading, and nuisance signals can impair a controller's ability to function effectively, especially during periods of high workload, while missed signals can lead to adverse events such as loss of separation.

The purpose of a signal is to attract the attention of a human operator when there is a possibility of an unwanted or undesirable event occurring in the future (*e.g.*, a warning), or when an unwanted event is currently taking place (*e.g.*, an alert or an alarm), so that the operator can intervene before that event causes harm or loss of life.[1] Many signals (*e.g.*, alarms) are therefore designed to be intrusive and distracting. In fact, some signals, such as fire alarms, are specifically designed to be so loud and distracting that they cannot be ignored. Air traffic controllers rely upon accurate, timely signals to maintain safety within the National Airspace System, but their performance is contingent upon reliable automation. In a study of the effects of imperfect automation on air traffic controllers, for example, Rovira and Parasuraman (2010) found that both false alarms and misses had serious negative effects on their performance.

Frequent interruptions from nonactionable alarms can disrupt the controller's prospective memory, causing him or her to forget a possibly essential action that was planned for some point in the future. Improved alarm design may, however, alleviate this deleterious effect on prospective memory (Loft, 2013). Auditory distractions may also impair prospective memory by using cognitive resources even if the controller is not paying attention to the sound. (Banbury *et*

[1] The definitions used for certain signals in this report may differ from those used in other government publications. For example, the RC-12P and RC-12Q aircraft flight manual describes a *warning* as "An operating procedure, practice, etc., which, if not correctly followed, could result in personal injury or loss of life."

al., 2001) There are additional important reasons to improve the sensitivity (correct activation of an alarm) and specificity (lack of activation when the monitored condition does not exist) of alarm s and to harmonize alarms across equipment and the ATO. *Alarm fatigue* is defined as either an increase in an operator's response time or a decrease in his or her response rate to an alarm that is caused by exposure to excessive numbers of alarms (Ruskin & Hueske-Kraus, 2015). This phenomenon has been observed and implicated in errors within multiple industries other than aviation. The role of false alarms is also being explored in healthcare, where organizational and technological aspects of the hospital environment are highly complex. The Joint Commission, the body responsible for accrediting most hospitals, recognized the clinical significance of alarm fatigue in 2019 and made clinical alarm management a National Patient Safety Goal. Alarm fatigue has also been described in passenger security screening and shipping (Dillon, 2018). The Deepwater Horizon catastrophe, for example, was caused in part because a combustible gas alarm system had been inhibited for a year before the event to prevent false alarms from disturbing the crew in the middle of the night (Hilzenrath 2010).

Alarm systems that operators perceive to be too unreliable are likely to provoke the so-called "cry-wolf effect" (Breznitz, 1984) in which the operator either disables or ignores the alarm. This can be especially problematic during periods of high workload when the operator does not have time to assess the reliability of the aid and chooses instead to simply abandon it (Bliss, 2000). This behavioral outcome—turning off the alarms—has been noted in the ATO before and raises concerns about the effectiveness of alarms with poor reliability (Wickens, Rice, Keller, Hutchins, Hughes, & Clayton, 2009). The actual function of the alarm may not be the same as the perception of that function, which may also degrade the operator's trust. For example, if a smoke detector sounds an alarm because a toaster burns a piece of bread, it has functioned correctly. People who hear the alarm often perceive it to be a false alarm because there was no fire because they incorrectly assess the result as a 'fire alarm' and not a 'smoke alarm'. The operator must therefore understand what the alarm is supposed to do and what its thresholds are.

Of course, unreliable alarms do not always cause air traffic controllers to mistrust automation. Wickens et al. (2009) conducted an archival study of the effect of false alarms associated with conflict avoidance on the Cry Wolf Effect, and found no strong evidence that there was a lack of response or a delay in responding to a signal. To the contrary, Wickens et al. found that controllers frequently responded to a potential threat before a signal was activated and hypothesized that false alarms may in some cases actually enhance controllers' situation awareness. This was also hypothesized by Breznitz (1984) in a discussion of the positive features of false alarms. This result was likely due to the professionalism and continually high task engagement displayed by air traffic controllers. (The context of an alarm affects whether or not it is considered to be *false*. If, for example, a controller recognizes a potential problem and addresses it before the associated alarm goes off, this would be considered a "hit" by the controller and the alarm would be false. If, however, the controller did not recognize the problem before the alarm called his or her attention to it, the alarm would be valid and considered to be a hit.)

Even when they are perfectly reliable, the intrusive nature of auditory and visual alarms can increase an operator's stress level during an abnormal event (Peryer, Noyes, Pleydell-Pearce, & Lieven, 2005), particularly when the operator has already noticed the problem and is actively managing it (*i.e.*, a nuisance alarm). At this point, the alarm is no longer beneficial and has become a source of unhelpful annoyance. For this reason, it is essential to design alarms that are reliable and salient enough to be acknowledged by controllers and allow for actionable responses to avert undesired events. Alarms must also minimize unnessary disruptions in workflow while attracting a controller's attention to a potential hazard and improving his or her ability to successfully complete the duties of controlling multiple aircraft.

Purpose

The Federal Aviation Administration has tasked our group with the development of a comprehensive handbook for the design of signaling systems in ATC. In addition, we will develop training materials that teach controllers how best to manage them. The design handbook and training materials will provide information on how controllers might adjust the system to improve their ability to manage nuisance alerts without inadvertently altering response bias in ways that would degrade safety. The first two stages of this comprehensive five-year project are to: 1) summarize the current literature on signaling in air traffic control and other domains; and 2) develop a signaling philosophy that will be used to guide development of the handbook. This comprehensive review summarizes the current literature on alarms, alerts and warnings in multiple domains, including aviation, ground transportation, healthcare, shipping and nuclear power. We focus primarily on auditory and visual signals, but also evaluate the role of tactile signals. We then explain how this information creates a foundation for the next phases of this project.

Automation

Automation is defined as the use of a machine to replace or augment tasks that are typically performed by humans (Wickens 1992). Automation can be deployed at varying levels, from low levels of automated assistance (e.g., painting projected trajectories on a radar display) to fully autonomous operation (Parasuraman & Riley, 2016). Automated systems offer significant benefits such as reduced workload, increased efficiency of operations (Dixon, Wickens & Chang, 2005; Rice & Keller, 2009; Rice & Trafimow, 2012), and improved safety and accuracy through warnings, alerts and alarms (Casner & Schooler, 2014; Hancock et al., 2013; McFadden et al., 2004; Meyer, 2001; Mouloua & Koonce, 1997; Parasuraman & Hancock, 2001; Parasuraman & Mouloua, 1996; Scerbo & Mouloua, 1999; Sheridan, 1998; Wiener, 1988). Automation is therefore widely integrated into complex work environments, particularly those with little room for error, such as aviation (Endsley, 2018; Hancock, 2019; Parasuraman et al., 2000; Sheridan, 2002; Vagia et al., 2016).

Automation can, however, have a negative impact on workload, situation awareness, and human-system interaction. Automation may negatively affect users' performance by causing complacency (Bailey & Scerbo, 2007; De Boer & Dekker, 2017; Landman et al., 2017; Parasuraman & Manzey, 2010; Parasuraman et al., 1993), loss of situational awareness (Cummings et al., 2016; Endsley & Kaber, 1999; Endsley & Kiris, 1995; Parasuraman et al., 2008; Wickens, 2008), and/or degradation of manual skills (Casner et al., 2014; Haslbeck, & Hoermann, 2016; Parasuraman & Riley, 1997). It is difficult for human operators to monitor highly automated systems for extended periods of time (Dekker & Woods, 2002; Mouloua & Koonce, 1997; Mouloua & Parasuraman, 1994; Parasuraman & Mouloua, 1996; Rice, 2009; Rice & Keller, 2009; Rice, Trafimow & Hunt, 2010; Scerbo & Mouloua, 1998; Vincenzi, Mouloua, & Hancock, 2004; Wiener, 1988). Completely removing a human operator from automated tasks can impair recovery from a system failure, but allowing both the automated system and the human operator to develop options for resolution of a problem may also reduce performance

(Endsley, 1999). These factors may directly affect an operator's ability to safely recover control of a system and revert to "manual" operations during a system failure.

Furthermore, if the automation is imperfect (*i.e.*, prone to failure or a significant degradation in performance), or its users are under time pressure (Rice, Trafimow, Keller, Hunt & Geels, 2011), an operator may lose trust in the system and fail to use it appropriately or miss crucial information (Dixon & Wickens, 2006; Dixon, Wickens & McCarley, 2007; Endsley & Kiris, 1995; Lee & See, 2004; Rice, 2009; Rice & McCarley, 2010; Rice, Trafimow & Hunt, 2010; Wickens & Dixon, 2007). Understanding the impact of automated systems on users' performance and error-management techniques may provide critical information for the successful development of advanced systems in the aviation field (Endsley, 2017; Gawron, 2019; Kontogiannis & Malakis, 2009; McBride et al., 2014; Sheridan, 2002; Sheridan et al., 1983; Wickens et al., 2009; Wickens et al., 2015; Wickens et al., 1998), including signaling systems for air traffic controllers.

Signals

Signals are one component of automation that are commonly used in air traffic control and many other domains that attract a human operator's attention to a situation that requires their attention. The term *signal* describes all stimuli that serve the general function of notifying a human operator of a situation that might require their intervention (*i.e.* alarms, alerts and warnings) and may include auditory, visual, and other sensory stimuli. Researchers have historically used various terms interchangeably, making it difficult to communicate about the problems associated with excessive signals and potential solutions. The taxonomy proposed by Bliss and Gilson (1995) is based upon the timing between a signal and its associated hazard. We use this taxonomy because it accurately classifies the types of signals used by air traffic control.

According to Bliss and Gilson, an *alarm* is defined as a transient sensory signal (usually auditory or visual) that indicates the presence of an ongoing danger that requires immediate corrective action. An *alert* indicates that an adverse event may occur sometime in the near future, usually soon enough for the operator to remember the alert. For example, a Conflict Alert has two components: An *Imminent* alarm that indicates an ongoing loss of separation that requires immediate action necessary and a *Predicted* alert that indicates a potential loss of separation that is predicted within 120-180 seconds. The Predicted conflict alert requires the controller's attention, but may not require any additional action. While alarms and alerts are temporary dynamic signals that are triggered by a changing situation, a *warning* is usually a permanent written indication of a static and unchanging hazard.

The healthcare industry provides useful examples that help to distinguish between signal types. For example, a warning may indicate that a patient has an allergy to a medication; this is an ongoing and permanent state. An alert might occur some time before a patient is expected to deteriorate, notifying the physician or nurse of this unexpected change in events and allowing them to react before the patient becomes critically ill. (Bartkowiak *et al*., 2019) Lastly, an alarm would notify the practitioner to an urgent danger that must be addressed immediately, such as a life-threatening heart rhythm. Within the ATC environment, a warning might be an indication that an aircraft is not equipped for RNAV/RNP, an alert may notify the controller that a conflict may occur between two aircraft that are on crossing routes, and an alarm would sound when an aircraft is transmitting a Mode C transponder code of 7500, indicating a hijacking in progress. In both settings, alarms demand immediate action, alerts indicate that an operator has time to take corrective action, and warnings provide an ongoing notice of a potential hazard.

Mistrust in Unreliable Signals

We anticipate any strategy that increases the accuracy of signals and controllers's trust in them will include improving the reliability of the underlying automation. Much of the current understanding about mistrust comes from research in the field of social psychology, where definitions of mistrust typically focus on human-to-human trust and behavior (Lee & See, 2004; Spector, 2004, Levine, 2018). Although trust and mistrust are sometimes portrayed on a continuum, many researchers feel that the two terms are not necessarily opposites and may arise from activation of different regions of the brain (Dimoka, 2010). Part of this confusion occurs because there are many definitions of mistrust that depend on who or what is the target of trust and how mistrust affects behavior. Mistrust may impede the adoption and use of automated systems such as those that generate signals used by ATC, albeit in a slightly different way than that of human-human interactions (Lee & See, 2004). Behavioral outcomes associated with mistrust of an automated system vary based upon the circumstances of the failure. If, for example, an automated aid tends to generate false alarms, the operator's mistrust would likely cause him or her to ignore and ultimately stop responding to those alarms. However, an operator who mistrusts an automated warning system because it tends to miss events (e.g. a luggage screening system that misses weapons) is likely to begin checking other data in order to assess the validity of the alarm. Both behaviors were affected by the level of workload imposed by other concurrent tasks (Manzey, 2014). One is a result of false alarms while the other is a result of the system failing to alert the operator to a potential hazard. This section will now discuss the understanding of mistrust and provide a definition of mistrust that can be applied to this project.

Parasuraman and Riley (1997) have described the ways in which operators interact with automated systems and how the reliability of these systems can change operator behavior. Operators can often engage or disengage with automation at any time, and do so based on how reliable they view the automation to be. This use is affected by various factors such as mental workload, risk and trust. *Misuse* is defined as over-reliance on automation, which can lead to monitoring or decision failures. *Disuse* occurs when the automation is neglected or underused, and commonly occurs in the setting of frequent false alarms (false alarms or misses). Lastly, *abuse* occurs when the operator uses the automation in a manner that has negative consequences for human performance. For example, Rice, Trafimow and Hunt (2010) found that operators tend to second-guess automated systems even when it is highly reliable, causing combined human-automation performance to be worse than the automation itself.

Meyer (2001; 2004) proposed a different kind of taxonomy of how automation errors affect human performance. In this taxonomy, *compliance* with the automation occurs when the operator follows instructions provided by the automated system and *reliance* upon the system occurs when the operator takes no action unless the system instructs him or her to do so. For example, if a fire alarm sounds, the intended response is for occupants to vacate the building. People who exhibit a high level of compliance will leave the building immediately. If a system tends to produce an inordinate number of false alarms, the level of compliance decreases because people mistrust the system and ultimately ignore the alarms. Conversely, the absence of a fire alarm (the sound and not the product) implies that there is no fire. People with a high level of *reliance* will trust the silence of the alarm to indicate that all is safe. If the alarm fails to notify building occupants of smoke or a fire, this automation "miss" will decrease reliance. Rice (2009), Rice and McCarley (2010), and Geels-Blair, Rice and Schwark (2013) tested the effects of automation false alarms and misses on both compliance and reliance behavior in various single- and multiple-task scenarios. All three studies concluded that high rates of false alarms dramatically decreased compliance and also had a lesser, but significant, detrimental effect on reliance. In contrast, high rates of misses can significantly decrease operator reliance and also decrease compliance to a lesser extent. Thus, either type of automation error will cause mistrust for both compliance and reliance behaviors.

Keller and Rice (2010) developed the concept of *System-Wide Trust Theory* in order to explain how the reliability of one aid can affect trust in other related aids. For example, an automated system might be designed with multiple aids that sound alarms when particular events take place. System-wide trust theory predicts that failure of one of these aids would decrease the operator's trust in the related aids, even when he or she is told *a priori* that the aids are unconnected. For example, failure of the low fuel light in a car may cause the driver to lose trust in the rest of the gauges, even though he or she knows that the other gauges are not connected to the fuel indicator. Rice and Geels (2010) subsequently presented operators with four diagnostic aids and found similar effects even when only one of the aids produced false alarms or missed actual events. Geels-Blair, Rice and Schwark (2013) then expanded the study to eight aids, finding that while the effect still existed, there was some dissipation of the mistrust in the other aids. This implies that the number of aids has an effect on operator system wide trust.

Signals in Healthcare, Nuclear Power, and Ground Transportation

As part of this research, the role of signals was reviewed in domains that share similar characteristics to air traffic control, such as use of automated technology, reliance on prospective memory, and dependency upon signals to attract the operators' attention to potentially hazardous events. Human performance experts have explored the problems associated with signals in healthcare, nuclear power, shipping, and railroads and have proposed solutions that offer guidance for our handbook. Experience with signaling systems in these domains can potentially be applied to the development of guidance and training in the use of signals for ATC.

Healthcare

Electronic medical devices are an integral part of patient care, and provide vital life-support and physiological monitoring that can improve safety in the operating room and intensive care unit. The signals generated by these devices are intended to warn clinicians about any deviation of physiological parameters from their normal value, preferably before a patient has been harmed. Life support devices (*e.g.,* ventilators and cardiopulmonary bypass machines) also employ alarms to alert healthcare providers to potentially life-threatening failures. In one study, 8975 alarms occurred during 25 consecutive cardiac procedures. A total of 359 alarms were recorded during the average procedure—approximately 1.2 alarms per minute (Schmid, 2011). Medical equipment manufacturers deliberately set alarm defaults to high sensitivity to avoid missing true events; this is identitical to the approach used in designing equipment for ATC.

According to the principles of signal detection theory, the alarm criteria have been set to a very liberal sensitivity because false alarms are presumed to be less damaging than misses. This assumption should be considered with caution because many studies have demonstrated that false alarms actually degrade trust more than misses (Dixon, Wickens & McCarley, 2007; Rice, 2009; Rice & McCarley, 2010). As a result, most alarms have low specificity and low positive predictive value and are therefore often ignored (Cvach, 2012). This problem is compounded when alarms are implemented across multiple parameters, leading to a cascade of alarms that create a noisy, distracting environment while doing little to improve patient care—a situation called *alarm flood* (Ruskin, 2015). System wide trust failure may then cause the clinician to mistrust and ignore even alarms that may indicate a potentially life-threatening event.

As mentioned previously, researchers often use signaling terms interchangeably, which can complicate attempts to understand and address the problems created by excessive alarms. The current standard for medical alarms is International Electrotechnical Commission (IEC) 60601 1-8, which specifies basic safety and performance requirements, including alarm categories that are prioritized by degree of urgency and consistency of alarm signals (Xiao & Seagull, 1999). The IEC standard does not, however, address the problems associated with the high sensitivity of sensors and low specificity of alarm conditions. Conversely, a valid alarm may give the health care provider very little time to react to a life-threatening event. Signals should therefore be designed to give the healthcare provider enough time to take action to prevent an adverse outcome. The duration of an appropriate time delay is, however, contingent upon operational parameters, most notably the rate at which the situation deteriorates. This also applies to our philosophy of ATC alarms: equipment designers and possibly controllers can identify the operational parameters that are relevant to a given signal and then adjust the alarm thresholds to provide the controller with enough time to react to a hazard while minimizing the rate of nonactionable alarms. Ultimately, there is a tradeoff between the accuracy of the alarm and the amount of time given to the operator after the alarm sounds.

Medical signals can be further subdivided according to the underlying condition that results in their activation. *Clinical alarms* indicate that the patient requires immediate attention, while technical alarms indicate that the biomedical equipment requires attention. For example, ventricular fibrillation (a heart rhythm that is fatal if not immediately treated) triggers a clinical alarm while a disconnected sensor or a poor-quality blood pressure tracing might cause a technical alarm. Xiao and Seagull (1999) have proposed a taxonomy that distinguishes between signals based on their usefulness for medical personnel who monitor medical processes. Xiao et al.'s taxonomy includes the following distinctions:

- *False alarms* in medicine occur when no danger exists, often because sensor thresholds are set to maximize sensitivity. Although false alarms are common in healthcare, the context in which they occur is different in the ATC environment. Although signals are generated only when a defined parameter threshold is exceeded, these signals may not be useful to the controller if, for example, a loss of separation alarm sounds even though the controller intends to issue (or has already issued) an instruction that will ensure separation. Even though actual false alarms may not occur, the perception of false alarms can still have detrimental effects on trust. In other words, if the operator perceives an alarm to be false, trust is still degraded even though the system is not actually producing false alarms.
- *Nuisance alarms* indicate a problem in a specific context but the alarm has been activated in a different context that will not lead to harm. For example, a low blood pressure alarm will usually activate when a blood pressure cuff is inflated on the same arm in which an intra-arterial catheter is directly measuring blood pressure. This occludes blood flow through the intra-arterial catheter, causing the system to artificially interpret a low pressure. The resulting alarm is the same as if the patient's blood pressure was actually low. Nuisance alarms can occur in the ATC environment if a controller is already aware of the problem and taking action to address the situation (Wickens, et al., 2009). For example, two airplanes may be on a converging course, but the controller knows that one will be turning onto a final approach course well before a loss of separation occurs.

- *Inopportune alarms* occur at the wrong time, such as when an alert signals a condition too far in the future for the operator to take immediate action. In this case, the operator might not remember that the alarm activated when it becomes necessary to address the problem. This problem may be affected by an operator's prospective memory, which is influenced by age, substance use, and certain genetic diseases (Einstein, 1990; Nilsson, 2002).
- *Actionable alarms* indicate a physiologically abnormal state which requires a clinician to intervene in order to avoid patient harm. A mild deviation might require only assessment of the patient and heightened alertness for further change, while others might indicate an urgent, life-threatening problem (Karnik & Bonafide, 2015).
- *Nonactionable alarms* can be caused by monitoring artifact (e.g., the electrical interference of a cautery device in the operating room causing a "ventricular fibrillation" arrhythmia alarm for the patient), or a true deviation from the alarm limits that represents a clinically insignificant abnormality (e.g., a ventilator's apnea alarm activating during induction of general anesthesia when the patient is expected to not be breathing on his or her own).

Nuclear Power

The nuclear power industry has developed strategies for signal design and management that may apply to some aspects of air traffic control. Operators of advanced nuclear power plants are responsible for monitoring a large number of alarms in a highly complex system. The cognitive resources of an operator who is presented with a large number of alarms in a short time (*e.g.,* during an *alarm flood*) can be quickly exhausted, impairing their ability to identify the most critical failures and take corrective action (Ahmed, 2013). This decreases the operator's ability to detect faults and decreases the ability to appropriately respond to faults (Niwa & Hollnagel, 2001).

The impact of workload on alarm and task performance has been studied through the use of dual-task exercises (Bliss & Dunn, 2000). Participants were asked to perform a primary and secondary task, which consisted of tracking and monitoring information. They were also asked to react to various auditory and visual alarms while completing the psychomotor tasks. An analysis of response data such as frequencies, accuracy, and tracking error was performed. The results of the study indicated a negative correlation, that is, as the primary task workload increased alarm performance decreased (Bliss & Dunn, 2000). The data aligns prior studies conducted by Lysaght (1989), showing that as mental workload of one task increases, performance on all other tasks will eventually become degraded. Similar studies by Dixon and Wickens (2006), Keller and Rice (2010), and Rice, Trafimow, Keller and Bean (2011) also support these findings.

As part of a strategy to manage alarm flood in the nuclear power industry, Lin *et al*. (2017) demonstrated the utility of control charts to help power plant operators monitor a variety of alarms. In this study, operators and engineers were asked to manipulate a simulated nuclear power plant system using a *monitoring aid system* that consisted of either textual information or a graphical display, while performing a secondary task. The authors found that the *monitoring aid system* (MAS) significantly improved alarm detection rate, secondary task performance, and subjective mental workload, all of which generated significant improvements in performance. The authors further concluded that their MAS might reduce mental workload as well.

Wu and Li (2018) reviewed the problems associated with alarms in nuclear power and developed several recommendations to improve their usability. For an alarm system to be usable, it should result in alarm floods less than 1% of the time. They suggest that *pre-alarms* might help operators to take preventative action before a failure occurs or to intervene in the early stages of a failure. Such a pre-alarm would draw the operator's attention to an abnormal state before it becomes critical, but the challenge is to develop pre-alarms that provide meaningful information without increasing the overall volume of signals that the operator must process. *Group alarms* combine several lower-level alarms into a single, high-level alarm that alerts the operator to the most critical underlying alarm state. Group alarms can reduce the number of alarms that an operator would receive during a critical event while allowing the operator to see all of the underlying conditions by clicking on the master notification (Wu & Li, 2018). Some of these recommendations may be applicable to air traffic control. For example, a controller might receive an auditory or visual alert that would serve as a pre-alarm when two airplanes are on converging paths but are not in imminent danger of loss of separation (*e.g.*, a tactical conflict alert or a strategic conflict probe).

Aviation

Much of the current understanding about warnings, alerts and alarms comes from the aviation industry. This section will discuss past research and improvements that have been made in the development of signaling systems on the flight deck and in air traffic control, and then describe knowledge gaps that require additional study.

Signals

Signals are common in aircraft and ground crew transportation technology. Several features of the signal's design can impact its success, including the frequency of warnings, visual layout, auditory pitch and tone, saliency and limitations of the technology used to activate the signal. These features are highlighted through some examples within aviation including the

Terrain Avoidance and Warning System (TAWS), the Enhanced Ground Proximity Warning System (EGPWS), and the Final Approach Runway Occupancy Signal.

Haberkorn et al. (2014) examined how pilots interact with conventional traffic advisory systems to avoid collision during visual flight. A total of 21 pilots were observed during simulated flight while using four displays that combined various predictive features. They were presented with single and multiple potential conflicts that were visible on both a moving map and on the simulator display. Pilots also answered a survey asking about mental workload demands and display usability. Participants preferred using a display that offered a combination of aircraft symbols, path trajectory, and a priority cue, but initiated avoidance maneuvers that did not always conform to regulations. When presented with multiple potential conflicts, the pilots reacted more slowly when confronted with the traffic warning. The authors conclude that improvements are needed in order to improve prediction of traffic situations and reduce display scanning time. This suggests that it is necessary to consider the number of concurrent potential conflicts when designing signals for air traffic controllers and to understand that operators may react more slowly when faced with multiple conflicts.

Alerts and Alarms

As mentioned previously, alerts and alarms are dynamic and time sensitive, while warnings are static and usually permanent. The purpose of an alert is to notify the operator of an event that may require attention in the future, but does not require immediate action. Alerts therefore give the operator more time to address the hazard than do alarms.

Early studies of aircraft alerting systems focused on implementing an increasing number of systems into the cockpit to help pilots prevent accidents (Veitengruber et al., 1977). These studies were a result of numerous aviation accidents between the early 1960s and late 1970s. Following the accident investigations, the Federal Aviation Administration provided a detailed summary stating that the majority of the accidents could have been avoided (Hanson et al., 1982). The increased implementation of alerting systems have been described in the Boeing 707 and the Boeing 747 airplanes. The Boeing 707 had approximately 188 possible alerts while the upgraded 747 employed approximately 455 alerts (Veitengruber et al., 1977). This increase in alerts can also be examined in Douglas DC-8 and Douglas DC-10 airplanes as well. DC-8 models contained 172 alerting signals, while DC-10 models incorporated 418 total alerting signals. Since that time, researchers have evaluated aspects of alerts that include physiological (i.e., visual or auditory) and psychological (i.e., emotions or mental state) reactions.

Jurgensohn *et al*. (2001) investigated methods to improve air traffic controllers' ability to detect possible collisions between aircraft using an early warning system. Participants monitored a screen that simulated air traffic converging from different locations into a single stream for a final approach to an airport while a few aircraft flew in random directions. Participants received either an early alert (15s to collision), which prompted them to monitor the situation or an immediate alert (5s to collision), which prompted them to click an "Infringement" menu item. Additionally, the reliability of these alerts ranged between 100% reliability, 90% reliability, 70% reliability, and no warning. As expected, participants' performance decreased in the 70% reliability and the no warning condition, and best performance occurred when they received an immediate alert five seconds before the collision. The average response time for participants receiving the five-second alert differed very little between the various reliabilities, suggesting that participants were prepared to take the appropriate action regardless of the accuracy of the alert. On the other hand, the average response time of the 15 second condition was significantly different between the reliabilities. The authors concluded that a warning that occurred five seconds before loss of separation occurred seemed elicit the most reliable response because the participant could take immediate “action” while the 15 second warning simply directed the participant’s attention to the potential “conflict,” with the level of attention decreasing as the reliability of the warning was degraded. This study seems to suggest that there is an optimal amount of time between activating a signal and the potential hazard, although a five-second warning is not sufficient to ensure safety in an operational environment. The authors conclude that understanding how controllers direct their attention in response to alerts will result in improved signals that can better support their performance.

Sensory Design

Despite the abundance of alerting signals, pilots still rely heavily on visual information provided by the instrument panel. Numerous studies have investigated the reliability of visual displays within the cockpit (Karmakar et al., 2019; Kim et al., 2019; Moacdieh et al., 2013). Wickens and Colcombe (2007) evaluated multi-task performance when using cockpit displays of traffic information and found three major components to the efficiency of the cockpit display of traffic information alerts: 1) threshold; 2) modality; and 3) the total number of alerts. They argued that adjusting these components within the system can improve pilot performance and decrease false alarms.

The amount of information available to pilots has increased over the years since there are more daily aircraft operations. One of the more modern technologies includes primary flight displays (PFD). This display provides a variety of important information such as navigation data, terrain information, and weather data. Moacdieh et al. (2013) examined participants’ performance when using the modern PFD. The results of the study indicate that performance decreased due to the amount of clutter on the display. In addition, the authors used eye-tracking data and found that as clutter increased, so did spatial density, which caused participants to become more distracted. The study concludes that some changes need to be addressed in the modern PFDs to reduce clutter and increase performance.

In the air traffic control environment, researchers have compared ATCs' performance between different types of visual alerts while completing an auditory task (Giraudet et al., 2015). Participants' reactions to the auditory task were more accurate when they saw the more salient alarm, which included a text labeled "ALRT" and four moving yellow chevrons. Researchers believe that the "enhanced visual design freed-up attentional resources which, in turn, improved the cerebral processing of the auditory stimuli" (p. 246). Multiple studies have explored the design of accurate and reliable signals that help prevent prevent potentially fatal incidents that (Long & McGarry, 2009; Nikolic et al., 2004; Santel, 2016; Thomas, & Rantanen, 2006; Tippey et al., 2017). Visual features are a key component in signal design. Nikolic et al. (2004) used a physiological approach to examine how multi-colored and high definition displays affect a pilot's attention. The results suggested that abrupt visual changes on a solid background resulted in an almost perfect performance and the fastest reaction times. The researchers recommend higher fidelity display interfaces and the introduction of visual onsets to increase performance and decrease reaction time. This may be applicable to the design of air traffic control signals and will be addressed during the second year of this project.

Because flying an aircraft requires a high level of visual activity, designers began using auditory alerts to help reduce pilots' visual workload (Wiener, 1977). The effectiveness of auditory alerts was measured in a simulated cockpit and shown to increase reaction times more than visual alerts (Reinecke, 1976). Subsequently, it was noticed that pilots began turning off the auditory alerts (King & Corso, 1993; Wiener, 1977; Yeh & Wickens & Yeh, 1988). These studies show that pilots generally disable auditory systems because of the high noise level in the cockpit and increased level of subjective workload. In another study of auditory alerts, King and Corso (1993) studies subjects who were asked to perform a visual search task in a laboratory setting. Although the intensity of the sound was varied, no distinction was made between alarms and alerts in this study. Subjects disabled auditory signals in order to decrease their subjective workload, which suggests that auditory displays require additional processing that increases cognitive workload. Although studies suggest that auditory alerts may degrade performance if they are overused, several recent studies have shown that the utility of auditory alerts depends upon their informational content. For example, a survey of pilots conducted by Peryer *et al* revealed that loud, continuous alerts are distracting to pilots and may cause a startle response. Pilots in this study preferred an auditory alert that contained a nonspeech component to attract their attention followed by a vocal alert that described the problem or commanded an immediate action. (Peryer, Noyes , Pleydell-Pearce, & Lieven, 2005) The strategies suggested by Peryer *et al* can guide future development to ensure that auditory signals used by ATC are designed to maximize informational content while minimizing additional workload.

Signals may also produce *habituation*, in which operators filter out a signal that is repetitive but has no apparent consequence, causing people to become less responsive to a sound that is heard frequently. Ironically, individuals with high working memory capacity appear to habituate to repetitive signals more quickly. (Sörqvist, Nöstl & Halin, 2012) Varying auditory alerts may therefore help prevent pilot habituation. In one study of signals indicating an unsafe landing gear condition (Fasano, 2012), 10 participants were recruited to perform non-precision instrument approaches while they were exposed to a variety of different landing gear aural alerts during this task. Alerts included a consistent alert over a long period of time, another alert that changed in pitch and loudness, and an alert that changed in duration. Pilots' performance during the task were recorded and pilots answered a questionnaire to gain insight into their perceptions about which alert they found more useful. The authors concluded that an alert that continuously changed in pitch had a resulted in the best performance. The qualitative data from the study also aligned with initial results as almost all the pilots in the study preferred the alert that changed in pitch. This suggests that time-varying components of auditory cues (i.e., tempo, intensity, or type of signal) should be considered as part of guidance for a new ATC signaling strategy.

Prior studies have suggested that using a synthesized voice alert to convey flight-critical information improves response times because it is unique and attracts attention. As described above, the study by Peryer *et al* (2005) concluded that pilots prefer a non-verbal signal followed by a voice command. The reason for this was that a signal composed only of speech may be difficult to detect in a noisy environment with multiple conversations. However, there has been conflicting evidence as to whether or not the synthesized voice should be preceded by an auditory alert to gain the pilots' attention before activating the spoken signal. In a study by Simpson and Williams (1980), four Boeing 727 pilots in a flight simulator were asked to respond to different combinations of alerts, including an auditory tone and a voice recording. The authors found that that pilots took less time to respond to the alert and take appropriate action when they did not receive an initial alert tone. The authors also noted that additional information can be conveyed to pilots with no increase in comprehension time.

One study later evaluated the impact of controllers' experience on their response to various auditory alerts. Air traffic controllers responded to Short Term Conflict Alerts MSAWs, and Area Proximity Warning. Participants heard either a traditional auditory alert or an enhanced auditory alert, which consisted of a “beep” followed by a female voice speaking the name of the alert followed by another beep. The duration of the two signals was identical, and the time that was required for a response was measured from the onset of the sound. The enhanced auditory signal produced improved controllers’ performance and efficiency across all three critical incident situations, regardless of controller experience (Kearney *et al.*, 2016). The authors concluded that the observed improvement in performance resulted from improved situational awareness and a decreased startle response to a series of beeps. This allowed the controllers to identify the problem and begin formulating a solution more quickly.

Specific features of auditory signals can have a significant effect on their effectiveness. One example is the impact of gender and speaking tone used in verbal cockpit warnings (Arrabito, 2009). Participants identified verbal warnings through an auditory channel while performing a tracking task. Subjects heard either a male or female communicate specific warning codes in three voice styles: 1) whisper; 2) urgent; and 3) monotone. In the first experiment, there was little or no background noise, while the second experiment included noise that resembled that of typical radio communications. The results of the first experiment showed no significant differences between talker or listener gender, although urgent and monotone voices provided the fastest response times. The second experiment found that the male voice in either monotone or urgent communication produced higher accuracy and quicker response times. This study may be applicable if spoken alarms are used to indicate that a hazard requires urgent attention. The use of spoken signals may require additional considerations in the aviation environment, however, because of the focus on speech as a primary means of communication. Controllers must differentiate between conversations with pilots, conversations with other controllers, and conversations with supervisors. Possible solutions include spoken signals that are acoustically distinct from normal human speech or are preceeded by a tone or some other auditory display as occurs in airplane cockpits and as suggested by Kearney (2016).

Finally, the limitations of the technology underlying the warning system can also impact their effectiveness. *Controlled flight into terrain* (CFIT) occurs when an otherwise airworthy aircraft is flown into the ground or into an obstacle due to pilot error. According to the FAA (n.d.), CFIT was responsible for approximately 50% of aviation-related crashes between 1979 and 1989. TAWS alerts pilots when they are potentially flying too close to terrain (Federal Aviation Administration, 2000). Although the implementation of TAWS was beneficial to pilots, the reliability of the system was not perfect. TAWS was solely dependent upon a radio altimeter, which did not give pilots adequate time to make a change in their trajectory (Ziółkowski & Skłodowski, 2018). In contrast, EGPWS uses GPS to locate hazardous terrain in relation to the aircraft, improving the reliability of the system (He, et al., 2007).

Researchers have acknowledged that collision avoidance technology offers opportunities for improvement (Thomas & Rantanen, 2006). Additional work is necessary to "identify factors that affect operators' trust, workload, and situation awareness when interacting with collision avoidance systems and aid in the development of guidelines for mitigation of the adverse effects of imperfect technology" (p. 502). The researchers identify several human factors issues with current collision detection systems and recommend additional research focused on these specific areas that may offer guidance for ATC signal design.

1. How the detrimental effects of nuisance alerts and alarms can be mitigated, possibly by depicting uncertainty, providing the pilot with better situation awareness and/or improved training strategies;

2. Determination of the optimal settings for when alerts and alarms should occur, including threshold, look-ahead time and other factors related to uncertainty, as well as multi-level signals;
3. Whether automated conflict resolution algorithms significantly improve conflict resolutions (*vs* unaided decisions); and
4. How best to share responsibility between pilots and air traffic control in free flight conditions, including normal operations and conflict situations.

Thomas & Rantanen, 2006

Earcons, or auditory icons, are another signal that may provide more information about the nature of a specific hazard. One common example of an earcon is the sound of crumpling paper that most computers currently use to indicate that a file is being deleted. Earcons were first described approximately 25 years ago and were shown at that time to improve an operator's ability to navigate a menu. (Brewster *et al*., 1993; Brewster *et al*., 1996) Earcons have been studied in medicine, and have been shown to facilitate monitoring of vital signs in multiple patients. (Hickling *et al*., 2016) This suggests that earcons may also be useful for air traffic control signaling, in which controllers are required to monitor multiple aircraft at once. Perry *et al*. (2007) asked participants to complete a computer-based training program while responding to iconic auditory, visual or basic auditory alerts. The results of the study indicated that participants had significantly less training time when responding to the earcons as well as a higher recognition of the alerts compared to the visual and basic auditory sounds. The application of the study can be used to improve alerts that inform air traffic controllers about events that do not require significant time pressure.

One possible method for using voice alerts may be the use of *spearcons*, which are earcons that consist of a spoken phrase that has been electronically sped up, sometimes to the point where it is no longer recognized as speech. Spearcons are easily distinguished from normal speech, and have been shown to improve navigation through menu systems (Walker BN *et al*, 2013), and facilitate the monitoring of multiple patients' vital signs (Li SYW *et al*, 2019). This may address the problem posed by a synthetic speech signal being less noticeable than it is in a flight deck situation, although further research would be required to establish the utility of this technique in air traffic control.

Tactile Signals

The most common signals used in aviation are visual and auditory, but tactile signals may improve an operator's response to critical events. Some studies have examined the effectiveness of other sensory modes such as tactile alerts to reduce the mental workload pilots endure during flight (Eriksson et al., 2006; Jennings et al., 2004; Raj & Braithwaite, 1999). The Tactile Situational Awareness System (TSAS) developed by the United States (U.S.) Naval Aerospace Medical Research Laboratory (Jennings et al., 2004) is one of the current tactile systems used. TSAS was originally created to aid helicopter or fixed-wing pilots during landing at low visibility conditions (Myers, 2008). The system operates by providing the pilot with situational awareness (SA) tactile cues that alert the pilot when they are moving into a potentially hazardous situation. Tactile alerting systems have also been shown to shift drivers' attention to other vehicles and to improve their ability to avoid forward and rear-end collisions (Meng, 2015).

One high-fidelity study flight-tested TSAS using deck landing maneuvers similar to motions experienced when attempting to land on ships at sea (Jennings et al., 2004). The participants included 11 Bell 205 helicopter pilots who performed this dynamic task in various low visibility environments. The results of the study suggested that TSAS improved pilots' SA in both ideal and degraded visibility conditions. Although there was significant evidence showing the benefit of TSAS to improve SA and performance, there was little to no evidence showing a decrease in the mental workload associated with landing during high sea state environments.

The placement of vibrotactile alerting systems has also been tested (Craig & Sherrick, 1982; Salzer et al., 2011). Most tactile displays are located in the torso to allow pilots to have their hands free. Salzer et al. (2011) investigated whether a vibrotactile thigh placement would have equal reliability. They found that vertical orientation could be accomplished by using tactile feedback located on the thigh. The application of this study can introduce another alternative to the vibrotactile alerting system in the torso to meet the individual preferences of the pilot. In another study, Ngo *et al.* (2012) used a low-fidelity air traffic simulation to measure the effectiveness of auditory and tactile cuing. Whenever an aircraft violated separation requirements, the aircraft was highlighted in red and a 500-Hz tone sounded. A 200-Hz vibrotactile cue was randomly activated whenever separation was lost. Participants responded more quickly to a conflict when both an visual cue and auditory cue were present. Interestingly, participants who experienced a vibrotactile cue alone did not detect conflicts more rapidly, but pairing the auditory and tactile cues consistently improved performance. This is partially consistent with studies in other domains. Wrist-worn tactile alerts have, for example, been found to improve a physician's ability to detect changes in a patient's vital signs (Ng, 2007). This suggests that critically important ATC signals (*e.g.*, imminent loss of separation) might include a multisensory component such as a simultaneous auditory alarm and tactile stimulation.

Reliability and Trust

It is currently impossible to produce perfectly-reliable anticollision technology because prediction algorithms must work in a probabilistic environment. It is therefore necessary to identify factors that affect pilots' trust in collision avoidance systems and develop procedures that mitigate the effects of imperfect technology. In a review of advanced aviation technology, Thomas and Rantanen (2006) examined in-cockpit traffic displays and false conflict alerts. They recommend additional research to guide the development of effective alerting algorithms. Moreover, variables such as alert threshold and look-ahead time can impact the rate of false alarms as they do in ATC. Indicating the level of uncertainty around predicted conflicts may also improve trust in the system. Lastly, Thomas suggests that future research consider how best to divide responsibility between ATC and pilots for avoiding conflicts.

Xu *et al*. (2007) examined the effects of conflict alerting system reliability and task difficulty on pilots. The participants included 24 pilots who were tasked with observing the development of conflicts between themselves and an intruder on a 2-D cockpit display of traffic information. The results of the study indicated that nearly half of the pilots relied on the automated system to improve miss estimations. Interestingly, the pilots who relied on the automation showed improved performance on difficult traffic trials by 83%. Overall, the automated alerts presented to pilots caused them to review the traffic data more clearly.

There is also evidence, however, that suggests that the pilot-automation coordination decreases awareness and may result in errors (Sarter *et al*., 2007; Sarter & Woods, 1994). One experiment used a scenario that consisted of an hour-long flight with 12 challenging automation-related events. The participants included ten captains and ten first officers to ensure the pilots were experts in using the standard autopilot system within the cockpit. The authors collected data that included participants' behavior, mental load, and eye-tracking. The results showed that pilots failed to acknowledge manual mode selections and did not notice automatic mode changes. The authors highlight that pilot monitoring strategies are not effective and, therefore, must promote new training programs using automated interfaces to support effective system monitoring.

Controllers may be required to defer non-critical tasks during periods of high workload. A forgotten task may, however, impact flight safety (Wilson et al., 2018). Several studies have therefore explored the "cognitive and situational factors that support the execution of deferred actions after a period of interposed activity" (p. 360). Interestingly, the interruptions did not appear to negatively affect controllers' ability to resume a task at a later time, which is consistent with near infrared spectroscopy studies that show that experts have higher oxygenation and better performance at high taskloads. (Bunce *et al*, 2011) While air traffic controllers may be able to safely defer alerts, alarms require an immediate response as they indicate the potential for a significantly more dangerous scenario. Several studies have explored ways to augment prospective memory in air traffic controllers. In one such study, controllers operating a simulator were required to perform several nonroutine actions when accepting handoff of an aircraft. Causing the target aircraft to flash improved the controllers' prospective memory and enhanced conflict detection (Loft, 2011). New signalling systems should therefore be designed to minimize the amount of ambient noise a controller is exposed to since extraneous noise impairs prospective memory. (Banbury *et al*., 2001) Using visual signals may, however, be more complicated in the ATC environment because flashing symbols are already used to indicate events like an IDENT transponder code.

In a study of how expertise and level of responsibility in the cockpit affects pilots' reactions to alerts (Zheng *et al*., 2014), the expertise of the pilots varied from less expert to more expert while the level of responsibility alternated between pilot flying and pilot monitoring. The authors concluced that expert pilots were more sensitive to alerts while monitoring the aircraft. On the other hand, expert pilots who were flying were more likely to disable the alerts. Less experienced pilots expressed being more annoyed from the same alerts but displayed more tolerance to the noise. The study showed considerable differences in how pilots react to alerts based on responsibility and expertise. Although no similar studies have been performed for air traffic controllers, the results obtained by Zheng *et al* support the possibility of including information about adjusting signal parameters in training materials in order to accomodate different controller skill levels.

<u>Managing Nuisance Alerts</u>

Integrating an alert cue with current systems does not always ensure appropriate action, particularly if the cue is not necessary or it is considered to be a nuisance (Friedman-Berg et al., 2008; Mumaw, 2017). Throughout their daily operations, controllers may experience nuisance alerts that occur when an auditory or visual warning is activated but fails to provide useful or novel information (Allendoerfer et al., 2008; Friedman-Berg et al., 2008; Maltz & Shinar, 2003). There are several different types of nuisance alerts that often provide controllers with information that is either redundant or not pertinent to their current situation. This may negatively impact controllers' performance despite their tendency to aptly handle these alerts. Air traffic controllers are proactive, actively detecting and resolving conflicts, but also rely on automated systems to alert them to possible collisions. Everson and Fieldsend (2006) developed a multiobjective evolution strategy to show a region of optimal receiver operating characteristics that minimize false positives when tuning systems. Unfortunately, optimal settings that can be implemented safely on a wide scale remain elusive. One solution may be to include training for individual controllers that will allow them to tune signals to accommodate operational factors (such as aircraft that are established on RNP procedures) and individual experience while also updating the automated systems to incorporate new procedures (*e.g.*, RNP).

To investigate the impact of nuisances on performance, one study collected data from en route and terminal facilities, including over 36,000 Conflict Alerts (CA) and over 11,000 Minimum Safe Altitude Warnings (MSAWs; Friedman-Berg et al., 2008). "In en route, controllers responded to CAs 38% of the time; in terminal, 56% of the time. En route controllers responded to 9% of the MSAWs, whereas terminal controllers responded to 39%," (page 106) which means in most cases, controllers did not respond to the majority of these signals. Interestingly, the data indicated that when controllers responded to potentially hazardous situations, they typically responded before the CA or MSAW was activated. This suggests that air traffic controllers already use a proactive approach to preventing mishaps.

Signals may actually cause distractions or increase air traffic controllers' cognitive workload (Friedman-Berg et al., 2008), negatively affecting future trust in the system and its performance. At the time this study was published, the FAA also conducted a systematic study observing safety alerts during live operations and examining automation data and voice recordings related to alerts (Allendoerfer et al., 2007). The FAA study produced results similar to those of Friedman-Berg, indicated that controllers ignored the majority of CAs and MSAWs but that no operational errors or deviations occurred. When controllers did respond to potential conflict situations, the majority did so before receiving the alert. Researchers concluded that 81-87% of CAs and 87-97% of MSAWs are nuisance alerts or unnecessary, which is consistent with prior research. Ultimately, Friedman-Berg et al. recommended that the Federal Aviation Administration "develop prototype alarms that incorporate gradations of urgency or likelihood and evaluate these prototypes through human factors testing" (p. 108). There is, of course, a tradeoff between minimizing the number of signals that controllers perceive as nuisance alerts and allowing the controller to set extremely liberal limits that might not provide the controller with enough time to respond to an unexpected, potentially hazardous situation. Developing appropriate “hard” limits for signal activation and training materials that will allow controllers to tune signals to accommodate airspace complexity and their own levels of expertise will be addressed in future phases of this project.

Another group of researchers investigated the role of signal detection theory (SDT) and how it relates to controllers’ performance (Allendoerfer et al., 2008; Paielli et al., 2009). Automation and voice recording was examined during the time surrounding an alert at different facilities. The majority of controllers did not respond to the alert or respond to the situation before the alert activated. For CAs and MSAWs, controllers waited an average of 38 to 88 seconds, respectively, before responding. SDT provides a reasonable explanation for controllers’ behavior as potentially critical situations meet the criterion for alerting controllers. The researchers recommended implementing look-ahead times to help prevent unnecessary alerts, although this approach may introduce a few trade-offs.

To help provide more reliable automation and reduce the rate of operational errors, the National Aeronautics and Space Administration (NASA) Ames Research Center developed the Tactical Separation-Assisted Flight Environment (TSAFE) and compared its performance to CAs (Paielli et al., 2009). Unlike CAs, TSAFE uses multiple predicted trajectories for each flight to account for the unknown factor of pilot intent. Although this technique increases the likelihood of producing a more realistic scenario, it is also important to not increase the false alarm rate. The researchers offer several different techniques for minimizing the false alarm rates with the TSAFE system. The first technique considers the prediction time horizons for horizontal trajectory and vertical trajectory. Through a series of complex algorithms measuring the degree of flight-plan conformance, the horizontal trajectory computes a shorter Dead Reckoning (DR; the path the airplane would take if it continued on its current speed and heading) prediction, which reduces the rate of false alerts. The vertical trajectory minimizes the DR altitude projection while a pilot is executing a climb or descent. After this period, the DR projection is increased and the flight plan-based trajectory is reduced “to minimize false alerts due to inaccurate climb or descent rate predictions” (p. 187).

Secondly, the authors suggest incorporating the “second-alert” rule which modifies the current CA requirement of needing loss of separation (LoS) for two of three consecutive radar samples. For TSAFE, the rule only applies when LoS is predicted to be more than one minute away and it only requires the second alert to occur within one minute of the first. The last recommendation modifies the standard altitude rounding rule used in the host computer. When a new altitude amendment is issued, false alerts occur during the delay in initiation of the required action while the aircraft is still flying near its previous cleared altitude. However, delaying the rounding to the new altitude for one minute would help minimize these false alerts. Using the measures mentioned above, TSAFE produced 40% fewer false alerts than the CA during a 2-hr sample of traffic data (Paielli et al., 2009).

Although TSAFE produced fewer false alerts, there are still problems associated with issuing these alerts. In the past, air traffic control organizations created different auditory alerts that conveyed the varying levels of urgency, which were designed as a series of auditory pulses (Cabrera & Ferguson, 2006). The purpose of these alerts is to minimize stress; however, these alerts may actually induce stress and produce a startle effect in controllers. (Cabrera & Ferguson, 2006, Kearney *et al.*, 2016) Controllers may become annoyed and desensitized to signals that are frequently activated but do not provide useful information; this may decrease their effectiveness. The authors recommend that new auditory alerts should be designed while considering "perceptual prominence, the meaning communicated by the sound, and the ability to distinguish alert sounds" (p. 245). Unfortunately, designing a new automated alert does not guarantee that air traffic controllers will find them to be useful.

Distinguishing Alerts and Alarms

In 2015, Hah et al. (2015) developed recommendations for NextGen alarm and alert presentation of safety-critical messages to ATC. The researchers created a taxonomy of current alarms and alerts and conducted a cognitive walkthrough evaluation of six supervisory controllers' responses to alarms or alerts. Overall, the researchers developed five recommendations they believed would improve NextGen alarms, alerts, indicators and notifications (AAIN):

1. Air traffic control alarms and alerts must be defined clearly as either alarms or alerts.
2. The NextGen system must use a consistent coding scheme (*e.g.,* display color). For instance, red be used for alarms only, not for alerts.
3. Aural annunciations that are used for alarms or alerts should be distinct from each other.
4. The salient features of alarms and alerts should correspond to the severity of the hazard.
5. Alarms and alerts should not hinder other elements on the display, which will interfere with controllers performing other tasks.

Hah et al., 2015

In addition, Hah *et al*. (2015) recommended that there be a display in the Service Operation Centers that shows all the systems on one monitor display. Although this does not impact the controller's workflow, routing maintenance-related alarms might allow managers to spot problems earlier and facilitate troubleshooting. As also suggested by prior studies, Hah *et al.* also recommended reducing the number of nuisance alarms and alerts. A high number of nuisance alerts can increase controller workload and decrease overall performance, desensitizing them and reducing trust in the automation. (Hah, 2017) Recommendations included the development of more sophisticated alert suppression functions in order to reduce nuisance alerts. For example, alert parameters could be adjusted to account for human factors data such as *look-ahead time*, which estimates controllers' response time to identifying and resolving hazardous situations (Allendoerfer et al., 2007). Furthermore, alerts could be graded so that they become more noticeable as the situation becomes more critical. Finally, the authors recommend that any future systems should be created with a user-centered design and evaluation with human-in-the-loop simulations. After a list of alarms and alerts has been created, studies should be conducted in a high-fidelity simulator in order to evaluate the presentation methods and effectiveness of the alarms and alerts. As this project nears completion, it may be possible to conduct these studies at the FAA's William J Hughes Technical Center.

Continuing the line of research dedicated to a clear distinction between alarms and alerts, Hah et al. (2017) created a categorization scheme to distinguish between the two and applied it in NextGen ATC systems. Using a cognitive walkthrough method, data was gathered from six controllers and the following recommendations were proposed:

1. The visual and auditory features of the alarms and alerts shall be decided in the order of red and yellow—with flashing if needed. The size of the alarm or alert

element must also be considered in addition to changing the limited block to the full data block if necessary, such as HIJK in the limited data block in ERAM. (Dual coding, such as color and shape, must be applied because controllers who are deficient in color vision are permitted to work with a waiver.)

2. Consistent naming conventions should be used to facilitate both standardization and communication between controllers in different facilities. Where different codes are used on controller displays (*e.g.*, HIJK in ERAM *vs* HJ in STARS), the full English word should be used for communication between different facilities.

3. If an alarm or alert has an aural announcement, it should be different from the others regardless of how often it occurs. For instance, in STARS, EM, HJ and RF issue the same warning tone.

4. The visual presentation of an alarm and alert must be clearly distinguishable from the background, and the contrast level must be high. For instance, the Automatic Terminal Proximity Warning Cone in STARS was presented over other displayed information.

5. Use a more pronounced feature in displaying any altitude-related alarms and alerts. For example, they could be displayed in a different color.

6. Instead of issuing a general status warning such as OUTAGE in ERAM, the controller should be provided with a detailed list or a link to more information in order to minimize the extra steps a controller must take.

Hah et al., 2017

Alarms

Alarms notify an operator of imminent danger that requires immediate in order to prevent serious harm. Bliss et al (1999) reviewed incidents gathered from the National Aeronautics and Space Administration's Aviation Safety Reporting System, categorizing the data by alarm issue (i.e., false or missed alarm), aircraft type, and flight phase. They found that the rate of alarm-related accidents was increasing, with incidents mostly occurring during the approach phase of flight. The majority of incidents involved GPWS and TCAS. The data supported the results of prior studies that concluded that these two systems were historically associated with high false-alarm rates and therefore a primary focus of aviation experts (Bliss & Gilson, 1998). The authors also found that the most reported false-alarms occurred in medium-large transport aircraft during approach and departures. Software designers were encouraged to redesign TCAS and GPWS algorithms in order to ensure that aircrew made the correct response to alarms.

The overall effectiveness of alarms has also been researched in the domain of unmanned aerial systems (UAS). Dixon, Wickens and Chang (2005) and Dixon and Wickens (2006) conducted multiple UAS simulations in order to study how pilots monitor system alarms. In their simulations, the pilots flew a series of missions while completing a goal-oriented task of finding waypoints. In addition, they were asked to search for targets of opportunity and to monitor a dynamic interface that occasionally revealed a system failure. These system failures were announced by auditory and visual alarms that notified the pilot of the failure. The reliability of the alarm aids was manipulated to test the effects of false alarms and misses on pilot performance. Pilots experienced difficulty finding all of the targets of opportunity and detecting system failures, especially when the alarm aid was unreliable. The authors concluded that automated aids must be reliable in order to provide a benefit to pilots. Wickens and Dixon (2007) concurred with this, and after conducting a meta-analysis of the literature, concluded that aids should be at least 70% reliable to have any hope of improving human-automation performance.

Rovira and Parasuraman (2010) studied how imperfect automation (a miss *vs* a false alarm) affected air traffic controllers' ability to detect potential conflicts. In this study, controllers used a conflict probe automation, which highlights aircraft that are projected to conflict. When the automation was perfectly reliable, controllers were able to detect conflicts more rapidly and more accurately than without the automation. When the accuracy of the automation was degraded to less than 100% reliability, however, controllers detected significantly fewer conflicts. In particular, they performed worse during a miss (only 25% of conflicts detected) compared to the false alarm (50% of conflicts detected). The authors suggest that although this automation may help controllers, additional research is necessary to understand factors that affect human-automation interaction. (p. 423)

Air traffic flows are typically organized along varying altitudes to ensure aircraft are provided with safe levels of lateral and vertical separation. If the procedures that maintain safe separation fail, the TCAS system is one of the final safeguards to prevent a collision. TCAS provides pilots with auditory and visual warnings and a "resolution advisory" that pilots are required to follow. Resolution advisories (RA) are negotiated between the TCAS systems in the potentially conflicting aircraft, and typically command one pilot to climb while the system in the other aircraft will command its pilot to descend. A typical TCAS aural alert might be, "Traffic, traffic. Climb, climb." Immediate compliance is critical for safe operation, and pilots are required to follow a TCAS RA even if doing so results in violating an ATC clearance. It is, however, important to minimize the number of false alarms, which causes pilots to distrust the system. Although multiple studies have evaluated the effectiveness of TCAS, it is unclear whether a similar system would be beneficial for ATC signaling systems and controller training.

Additional alarm systems have been added to the cockpit to help prevent errors that may impact flight safety. Runway incursions are a common error that can result in collisions (Schönefeld & Möller, 2012) and are especially hazardous because there may be very little time or space to prevent loss of separation when it occurs on or near a runway. Sammut and Mangion (2014) proposed and tested the effectiveness of a directive cockpit alarm that indicates a potential conflict and also gives guidance on how resolve that conflict. During simulator trials, the researchers found that directive alarms increased crew performance accuracy immediately after the alarm was activated. The authors note that the use of this alarm system (i.e., a system that provides corrective guidance instead of simply notifying the pilot) may improve safety.

Conveying Urgency in Alarms

Arrabito et al. (2004) conducted a case study to investigate the urgency of non-verbal auditory alarms using pilots operating CH-146 Griffon helicopters. The five alarms tested were Crypto, Emergency Locator Transmitter, Low Rotor, Radalt, and Selcal. Following the initial experiment, the pilots were asked to fill out a questionnaire on their perception of each alarm. The results suggested that the pilots rated the Emergency Locator Transmitter as the most urgent. These alarms sound at multiple frequency levels as sirens. Radalt (i.e., high pitch with decaying offset) and Crypto (i.e., continuous high pitched) were rated as least urgent. The authors note that the use of specific alarms to represent an immediate danger (i.e., Emergency Locator Transmitter alarm) can help reduce pilot mistrust as the auditory systems represent the urgency of specific events.

Several studies have investigated alarms and the various components that could influence how an individual reacts to alarm systems (e.g. Bliss, 1997; Bliss, Gilson, & Deaton, 1995; Burt et al., 1995; Stankovic et al., 2014). One study conducted a psychophysiological evaluation of how individuals perceive the urgency of aural signals (Burt et al., 1995). The authors explain that the initial development of auditory cues was implemented because cockpit lighting signals were becoming ineffective (Wheale, 1981). Participants were provided with three different types of auditory signals and were asked to rate the urgency of each signal. Electroencephalogram (EEG) was also recorded in order to directly measure the brain's response when various auditory signals played. The results of the study suggest that reaction times differ between the three auditory signals. Similarly, the EEG results indicated that as the urgency of an event increased, participants became more sensitive to, and aroused by, the signal. The authors concluded that the use of psychophysiological measures to evaluate perceived urgency of alarms can help improve pilot behavior when a dangerous event is about to occur.

Bliss et al. (1995) investigated influences of alarm urgency and the "cry-wolf" effect. The participants were presented with a cognitively demanding task while simultaneously attending to various alarm urgencies. Visual alarms such as red, yellow and green were shown to illustrate urgency. The authors found that approximately 90% of participants did not respond to every alarms that was presented. The remaining 10% of participants responded to all alarms. The results also showed that use of both visual and aural alarms was effective in gaining attention. The results of this study are consistent with prior publications that concluded that there is a tendency for humans to disregard alarms based on urgency. If this type of alarm is used in future applications, the urgency of the alarm should correspond with the urgency of the needed response. Otherwise, users will become frustrated by unnecessarily urgent alarms (Rice, et al. 2010).

A *binary alarm* emits display messages to system operators when specific thresholds are exceeded (Sorkin & Woods, 1985). *Likelihood alarm* technology uses probability matching to fundamentally predict the likelihood of an event to occur (Bliss et al., 1995). Clark et al. (2009) conducted a study on binary and likelihood alarms in a decision-making task to detect which system accrued more false alarms or misses. The results of this study suggest that a system equipped with likelihood alarm technology improves decision-making accuracy by 20% (Clark et al., 2009). The authors note that the implementation of this technology can expand to various roles not only aviation but in medical diagnostics, power plants, and military operations.

The Impact of Workload on Alarm Responsiveness

The impact of cognitive workload on alarm responses and decision-making can be evaluated using the *information-processing model* (Wickens & Liu, 1988), in which the thinking process is compared to the function of a computer: The human mind takes in information, organizes it, and stores it for future use. In a series of two experiments, examining likelihood alarm technology, workload and task-critical information impact decision-making and bias (Bustamante, 2008), workload and task-critical information was found to directly influence response bias in decision-making. In other words, participants did not search for information regarding alarms, even though some alarms were accurate. The findings from the study add to prior research on design changes within aviation displays and decision-making support tools.

Singh et al (2009) observed the effects of automation training, reliability and workload on automated-induced complacency in 120 non-pilots performing simulated flight and who were asked to target automation malfunctions. The study revealed that automation-induced complacency was present during multi-task and high static conditions. In other words, as workload and automation reliability increased, automation-induced complacency increased as well. The researchers argue that the impact of the debate between human-centered and technology-centered approaches can be evaluated within this study. Thus, human monitoring automated systems do not necessarily guarantee a reliable system.

Responding to alarms often requires the performance of additional tasks, which ultimately impacts responsiveness. Bliss (1997) examined the differences between manual and vocal reactions to alarms of moderate reliability. Twenty-four pilots were recruited to perform tracking and monitoring tasks. Simultaneously, the pilots were presented with 75% true alarms and were asked to either press a key to respond or respond vocally, depending on the condition. The results showed that pilot reaction times were higher when they were asked to respond manually. Pilots responded to a higher number of true alarms when they responded vocally. The authors note that this result could be due to the hands-free nature of a vocal response as pilots are predominantly occupied with other manual tasks during flight.

Followup studies have focused on the influences of decision-making, stress and emotional state on participants' performance (Bechara & Damasio, 2005; Starcke et al., 2008). One study investigated decision-making under stress using eye-tracking software (Stankovic et al., 2014). The study utilized a validated measure called the Matching Familiar Figures Test (MFFT) to rate cognitive impulsivity, and pupillometry measures to record psychophysiological data. While participants completed the MFFT, a short-term aversive noise was introduced to represent an acute stressor. Eye-tracking data showed that as stress increased, the accuracy of the MFFT decreased. The authors highlight that participants who were in the stress condition made decisions faster than those who were in the control group. In addition, the the pupillometry measurements demonstrated higher levels of arousal during the stress portion of the task, suggesting that acute stress plays an important role in decision-making.

One example of how increased cognitive workload from multiple conflicting alarms can lead to a crash is the Air France 447 mishap. While flying through an area of thunderstorms, the airplane's pitot tube became covered with ice, causing the airspeed indicator to become unreliable. The autopilot unexpectedly handed control back to the flight crew while the flight management system simultaneously reverted to "alternate law." In this mode, many of the airplane's flight envelope protections are disabled. The flight crew was unprepared to manually fly the airplane and had lost situation awareness, especially as to which functions were automated and which required manual control. The crew had not been trained to manage a partial automation failure combined with an airspeed near the stalling speed. Confusing messages on the electronic centralized aircraft monitor further impaired the flight crew's ability to regain control of the aircraft, as did confusion as to which crew member was actually flying the airplane. The plethora of sometimes-contradictory visual and auditory alarms, combined with confusion, loss of situation awareness,and stress-induced cognitive impulsivity, ultimately caused the airplane to crash in the Atlantic Ocean.

Effects of Inattentional Blindness and Deafness on Alarm Response

Inattentional deafness is a phenomenon in which, under certain circumstances, an auditory stimulus is unnoticed by a human operator. *Inattentional blindness* refers to a human operator who does not notice a visual stimulus. Both of these phenomena have been implicated in accidents in transportation. Auditory alarms in the cockpit are typically used to attract a pilot's attention to a potential hazard while decreasing visual workload (Edworthy et al., 1991).

However, a review of aviation safety reports has shown a consistent pattern of accidents caused by a pilot's inability to react appropriately to alarms (Bliss, 2003; Edworthy et al., 1991; Wickens et al., 2009). Dehais et al. (2012) sought to understand the cognitive biases associated with missed alarms. Pilots were asked to perform landing tasks using a 3-axis motion flight simulator. During landing, pilots given a continuous alert for an unsafe landing gear at the same time they faced wind shear. More than half of the pilots did not report hearing the auditory landing alarm during debriefing and the authors concluded that inattentional deafness was responsible for the pilots' lack of response to the auditory cues. One possible explanation for this effect is that management of the windshear condition overhwelmed the attentional resources of some participants, but the small sample size in this study did not reach statistical significance necessary to support this hypothesis.

Multiple additional studies have similarly shown that the inability to detect alarms may occur when attentional and perceptive resources are overwhelmed by increased cognitive workload (Dehais et al., 2014; Santangelo et al., 2007; Scannella et al., 2013; Wickens et al., 2009). The goal of one such study was to determine whether inattentional deafness to critical alarms is observed during flight simulations. (Dehais et al., 2014) Participants were asked to perform landings under two conditions: During one landing, an auditory alarm sounded; during the other landing, auditory alarm sounded while the pilot experienced wind shear. (Dehais et al., 2014). The wind shear condition was associated with the highest cognitive workload and also produced measurable psychophysiological stress. Almost 40% of the participants did not report hearing or react to the alarm during the critical wind shear condition while all of the participants perceived the alarm in the neutral condition (*i.e.*, no wind shear). The authors concluded that the sudden, dramatic increase in cognitive workload temporarily impaired the subjects' ability to perceive the alarm and attributed this to central inattentional deafness. Interestingly, these results suggest that when confronted with a combination of visual tasks and auditory tasks (*e.g.*, alarm identification), visual tasks take precedence and may increase the risk of inattentional deafness.

To further explain this phenomenon, researchers investigated P300 event-related potential as an indicator of inattentional deafness (Giraudet et al., 2015). Participants were asked to complete a landing task and an auditory task (oddball task). EEG electrode caps and Electro-OculoGraphic (EOG) electrodes were placed on the participant to record data while completing the dual-task. Next, they were given a subjective workload rating scale called the NASA TLX to identify individual workloads. As expected, results showed that the highest mental workload was exerted in the dual-task. The P300 amplitude diminished significantly in the landing task with high mental load. The use of P300 event-related potentials might be used to determine the ability of pilots or controllers to detect alarms that would potentially help to mishaps.

A more recent study proposed to predict inattentional deafness by assessing pre-stimulus EEG signals (Duprès et al., 2018). Participants were asked to perform a critical landing task in a flight simulator and were instructed to press a trigger when they heard a sound that occurred only rarely. The authors concluded that inattentional deafness to a single alarm is possible and propose the use of adaptive alarms as a possible countermeasure. The authors propose that classification performances above a chance level can potentially predict the occurrence using EEG engagement ratios. Therefore, the authors highlight that the engagement ratio is a vital component to assess a pilot's aural performance and predict inattentional deafness.

Inattentional deafness has also been demonstrated in air traffic controllers. Air traffic controllers are constantly tasked with responding accurately to multiple events. Giraudet et al. (2015) investigated inattentional deafness to an auditory alert on a simulated control task.

Controllers were asked to guide either one plane (low cognitive workload) or two planes (high cognitive workload), while also responding to visual notifications related to peripheral aircraft. Throughout the task, they heard either standard tones, which they were instructed to ignore, or alarm tones, which they were instructed to report. Participants failed to report 28.8% of alarms in the low cognitive workload condition and 46.2% of alarms in the high cognitive workload condition. Furthermore, participants in the high cognitive workload condition detected fewer visual notifications; however, their performance in guiding the aircraft was not altered. When controllers are experiencing a high cognitive workload (which is a fairly common occurrence in the field), current auditory alarms may not be salient enough to attract the controllers' attention. Although few if any countermeasures for inattentional deafness have been proposed, pulsing yellow chevrons have been shown to work as a countermeasure to inattentional blindness (Imbert *et al*., 2014), as has brief, selective information removal (Dehais *et al*., 2012). Tactile signals (discussed above) may also help. Additional research is still needed to develop measures that can help to prevent failure to perceive an alarm due to the effects of increased workload.

Increasing workload may also degrade an operator's ability to localize alarms. Edworthy et al. (2018) exposed a group of physicians to eight auditory clinical alarms already known to have relatively high localizability in six conditions—while varying workload. Participants were asked to read or carry out mental arithmetic (secondary tasks) while in quiet conditions or exposed to typical ICU sounds. Performance in the localizability task was best in the control condition (without a secondary task) and worst in those tasks which involved both a secondary task and noise. They also found that participants in a quiet environment without a secondary task missed one alarm in ten occurrences, while participants in the highest workload conditions missed an alarm on every fourth occurrence. Possible explanations included inattentional deafness and alarm fatigue.

Attempting to investigate other methods of preventing unintentional alarm omission, Kearney *et al*. (2019) explored the impact of alert designs on air traffic controllers' eye movement patterns and situational awareness. The controllers' number of fixations, fixation duration, and saccade velocity were captured while controllers responded to acoustic alerts and semantic alerts. The researchers argued that "semantically designed verbal warnings tailored to specific hazard situations may improve hazard-matching capabilities without a trade-off in annoyance" (p. 308). Overall, the researchers concluded that the semantic design increases controllers' situational awareness "by increasing fixation numbers," which allows the controller to gather more situation-specific critical information and develop a problem-solving approach (p. 313).

To help mitigate the possibility of an in-air collision, researchers have explored different methods of using fuzzy signal detection theory (SDT), TSAFE, and Traffic Collision Avoidance System (Dennis, 2003; Kuchar et al., 2004; Masalonis & Parasuraman, 2003). Fuzzy SDT allows for a more realistic detection of possible conflicts and decreases the probability of a false alarm when detecting potential conflicts. In addition, TSAFE helps transfer the burden of detecting conflicts from controllers to computers; however, this algorithm needs further improvement and testing before implementation (Dennis, 2003).

Management of Signaling Changes within the Air Traffic Control Culture

Although the guidance on alarms, alerts and warnings that will be developed during this program will improve the efficiency and safety of the system, changes to ATC signals may require personnel within the ATO to receive additional training. Management of these changes and inclusion of all stakeholders will allow new systems to be successfully adopted and implemented to achieve the desired safety and efficiency outcomes. Careful consideration must be given to the current culture within ATC, and a well-designed implementation plan must be devised that accounts for this culture. A robust implementation plan will ensure that controllers, mid-level managers, and higher leadership are all satisfied with the outcomes of the improvements that the handbook that is the ultimate goal this project will propose. This section will therefore address key factors affecting organizational change and innovation, organizational learning, managing and measuring change, and pitfalls to avoid with organizational change. Examples are included from the aviation industry as well as other high reliability organizations where a culture of safety and quality are paramount.

Building and improving a culture of safety requires that an organization enables its members to focus on safety, provides the necessary tools and environment to enact changes that improve safety, and supports an iterative process of continuous improvement. A strong safety culture is believed to be linked to fewer safety incidents (Tear, 2020). The aviation industry has a long-standing culture of safety with constant improvements that been developed proactively or in response to accidents or near misses. There are many examples of how international air traffic control organizations have developed this safety culture. For example, Callari et al. (2019) discuss the Maastricht Upper Area Control Centre (MUAC) operated by EUROCONTROL as an example of a “mindful organization,” with a laudable safety record. MUAC has created a system that prioritizes constant safety evaluation and learning within its air traffic organization.

Perceptions about an air traffic organization’s safety culture may, however, vary based upon an individual’s position within that organization. Perception towards a safety culture is also impacted by individual societal views on power-hierarchies and challenges to authority (Tear, 2020). An assessment of Swedish air traffic control centers found that managers consistently perceived the organization's safety culture more positively than did non-managers (Ek 2007). This highlights the importance of periodic, multi-level assessment of how all members of the organization perceive the culture in order to demonstrate areas for improvement.

Conclusions and Phase 2

The most critical goals for developing ATO alerts and alarms are to provide clearly formatted information to the controller (that is not available from other sources), to clearly and accurately indicate potential hazards, and to minimize false alarms. At a minimum, signals should be consistent across all equipment with which controllers are required to interact. Alarms should also be unique and easily recognizable. An MSAW alert, for example, is distinct from a conflict alert and should include visual and auditory components, and signals with high levels of urgency may include other components such as tactile stimulation. Both auditory and visual alarms should indicate the level of hazard (e.g., an aircraft that is 100 feet below decision altitude on a precision instrument approach is much more likely to encounter an obstacle than is an aircraft that is 100 feet below the required altitude at the initial approach fix).

Multiple factors, including noise, lighting and inattentional blindness or deafness can affect a controller's ability to detect a signal. Air traffic control operating irregularities seem to occur more frequently during periods of moderate to low workload and normal operational complexity (Stager & Hameluck, 2007). Overuse of automated systems may also exacerbate the effects of complacency. Automation must decrease the controller's workload and not unnecessarily detract from the controller's ability to complete other time-critical tasks. Any proposed new alarm model will include recommendations for increasing the rate of alarm perception and the ability to draw a controller's attention to a potential hazard.

The purpose of Phase 2 will be the development of a signaling philosophy for the product. Our work will be guided by the understanding, described in this literature review, of how signals have been developed and implemented in aviation and other industries. We intend to analyze and integrate research concepts and findings in studies of safety-critical systems across application domains. We will augment this information with interviews, observation of air traffic controllers, and consultation with FAA's NextGen researchers to ensure that our signaling philosophy will help controllers throughout the ATO in their vital mission to maintain safe operations. This philosophy will include the provision of theoretical frameworks that address central cognitive constructs and drivers. It will also include a clear and detailed description of the signaling process in ATC facilities that focuses on principal stakeholders' information exchange, priority determination, human-automation interactions, problem solving, and action execution. In addition, we will include a section that provides well-informed, actionable recommendations for signal design and implementation. The latter will necessarily concentrate on important aspects such as striving for optimal signaling redundancy, maintaining operator situation awareness, trust, minimizing cognitive stress, workload, and complacency, and facilitating clear and uncongested communication within each ATC facility and between all vested parties and platforms.

This comprehensive literature review summarizes the information necessary to develop strategies that will increase the reliability of signals, make signals more attention-grabbing, and provide possible approaches to ensure timely, accurate reactions to signals by controllers. As the project moves into its later phases, we hope to interview air traffic controllers, who will help us to understand the chronology of controller activities, provide a breakdown of the importance of each activity, and offer examples of situations where signals have been shown to be important, redundant or distracting. From these insights, we will develop a signaling philosophy that mirrors the typical chronology of controller actions as aircraft approach and depart airspace sectors.

This philosophy will then be used to develop a handbook during Phase 3 that will be organized by common task behaviors. This handbook will provide bulleted recommendation items and explanations that relate to the perception of individual and multiple signals. Examples will take the form of "best practices" that should accompany each signal set and "actions to avoid." Where possible, algorithmic tables will be presented, similar to troubleshooting guides commonly found in technical manuals. All topic sections will be cross-referenced to an index for rapid, intuitive search. Each of these phases is a critical component of the project's ultimate goal, which is to provide accurate, reliable signals that help air traffic controllers to maintain the highest level of safety in the world.

References

Allendoerfer, K., Friedman-Berg, F., & Pai, S. (2007). *Human factors analysis of safety alerts in air traffic control* (Tech. Rep. No. DOT/FAA/TC-07-22). Atlantic City, NJ: Federal Aviation Administration.

Allendoerfer, K. R., Pai, S., & Friedman-Berg, F. J. (2008). The complexity of signal detection in air traffic control alert situations. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *52*(1), 54-58. doi:10.1177/154193120805200113

Amaldi, P., & Khalil, S. (2013). Designing alarm device for socio-technical systems: co-active or coercive design? *In Proceedings of the 3rd International Conference on Application and Theory of Automation in Command and Control Systems*, 132-135.

Arrabito, G. R. (2009). Effects of talker sex and voice style of verbal cockpit warnings on performance. *Human factors*, *51*(1), 3-20.

Arrabito, G. R., Mondor, T. A., & Kent, K. J. (2004). Judging the urgency of non-verbal auditory alarms: a case study. *Journal of Ergonomics*, *47*(8), 821-840. https://doi-org.ezproxy.libproxy.db.erau.edu/10.1080/0014013042000193282

Bailey, N. R., & Scerbo, M. W. (2007). Automation-induced complacency for monitoring highly reliable systems: The role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science*, *8*(4), 321-348. doi:10.1080/14639220500535301

Bakare, A. K., & Junaidu, S. B. (2013). Integration of radar system with GPS-based traffic alert and collision avoidance system (TCAS) for approach control separation. *Journal of Aviation Technology and Engineering*, *2*(2), 6.

Banbury SP, Macken WJ, Tremblay S, Jones DM. Auditory distraction and short-term memory: phenomena and practical implications. *Hum Factors*. 2001;43(1):12-29. doi:10.1518/001872001775992462

Bartkowiak B, Snyder AM, Benjamin A, et al. Validating the Electronic Cardiac Arrest Risk Triage (eCART) Score for Risk Stratification of Surgical Inpatients in the Postoperative Setting: Retrospective Cohort Study. *Ann Surg*. 2019;269(6):1059-1063. doi:10.1097/SLA.0000000000002665

Bechara, A., & Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and Economic Behavior*, *52*(2), 336-372. doi:10.1016/j.geb.2004.06.010

Bliss JP, Gilson RD, Deaton JE. Human probability matching behaviour in response to alarms of varying reliability. Ergonomics 1995;38:2300-12.

Bliss, J. P. (1997). Alarm reaction patterns by pilots as a function of reaction modality. *The International Journal of Aviation Psychology*, *7*(1), 1-14. doi:10.1207/s15327108ijap0701_1

Bliss, J. P. (2003). Investigation of alarm-related accidents and incidents in aviation. *The International Journal of Aviation Psychology*, *13*(3), 249-268. doi:10.1207/S15327108IJAP1303_04

Bliss, J. P., & Dunn, M. C. (2000). Behavioural implications of alarm mistrust as a function of task workload. *Journal of Ergonomics, 43*(9), 1283–1300. https://doi.org/10.1080/001401300421743

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 Human Factors and Ergonomics Society Annual Meeting, 1*, 6.

Bliss, J. P., & Gilson, R. D. (1998). Emergency signal failure: Implications and recommendations. *Ergonomics, 41*(1), 57-72. doi:10.1080/001401398187323

Bliss, J. P., Gilson, R. D., & Deaton, J. E. (1995). Human probability matching behaviour in response to alarms of varying reliability. *Ergonomics, 38*(11), 2300-2312. doi:10.1080/00140139508925269

Brewster SA, Wright PC, Edwards ADN. 1993. An evaluation of earcons for use in auditory human-computer interfaces. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (CHI '93)*. Association for Computing Machinery, New York, NY, USA, 222–227. DOI:https://doi.org/10.1145/169059.169179

Brewster S., Raty VP., Kortekangas A. (1996) Earcons as a Method of Providing Navigational Cues in a Menu Hierarchy. In: Sasse M.A., Cunningham R.J., Winder R.L. (eds) *People and Computers XI*. Springer, London. https://doi.org/10.1007/978-1-4471-3588-3_12

Breznitz, S. (2013). *Cry wolf: The psychology of false alarms*. Psychology Press. Hillsdale, NJ: Lawrence Erlbaum

Bunce S.C. et al. (2011) Implementation of fNIRS for Monitoring Levels of Expertise and Mental Workload. In: Schmorrow D.D., Fidopiastis C.M. (eds) *Foundations of Augmented Cognition. Directing the Future of Adaptive Systems. FAC 2011. Lecture Notes in Computer Science, vol 6780*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-21852-1_2

Burt, J. L., Bartolome, D. S., Burdette, D. W., & Comstock, J. R. (1995). A psychophysiological evaluation of the perceived urgency of auditory warning signals. *Ergonomics*, *38*(11), 2327-2340. doi:10.1080/00140139508925271

Bustamante, E. A. (2007). Using Likelihood alert technology in cockpit displays of traffic information to support free flight. *In 2007 International Symposium on Aviation Psychology* (p. 100).

Bustamante, E. A. (2008). Implementing Likelihood Alarm Technology in Integrated Aviation Displays for Enhancing Decision-Making: A Two-Stage Signal Detection Modeling Approach. *International Journal of Applied Aviation Studies*, *8*(2), 241–261. https://search-ebscohost

Cabrera, D., & Ferguson, S. (2006). Considerations arising from the development of auditory alerts for air traffic control consoles. *Proceedings of the 12th International Conference on Auditory Display* (pp. 242-245). London, UK.

Callari TC, McDonald N, Kirwan B, Cartmale K (2019) Investigating and operationalising the mindful organising construct in an Air Traffic Control organisation, *Safety Science* 120, n838-849. doi:10.1016/j.ssci.2019.08.027

Casner, S. M., Geven, R. W., Recker, M. P., & Schooler, J. W. (2014). The retention of manual flying skills in the automated cockpit. *Human Factors*, *5*6(8), 1506-1516. doi: 10.1177/0018720814535628

Casner, S. M., & Schooler, J. W. (2014). Thoughts in flight: Automation use and pilots' task-related and task-unrelated thought. *The Journal of Human Factors and Ergonomics Society*, *56*(3), 433-442. doi:10.1177/0018720813501550

Clark, R. M., Peyton, G. G., & Bustamante, E. A. (2009). Differential effects of likelihood alarm technology and false-alarm vs. miss prone automation on decision making. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *53*(4), 349-353. doi:10.1518/107118109X12524441081028

Craig, J. C., & Sherrick, C. E. (1982). Dynamic tactile displays. *Tactual perception: A sourcebook*, 209-233.

Cummings, M. L., Gao, F., & Thornburg, K. M. (2016). Boredom in the workplace: A new look at an old problem. *The Journal of Human Factors and Ergonomics Society*, *58*(2), 279-300. doi:10.1177/0018720815609503

Cvach M. Monitor alarm fatigue: an integrative review. Biomed Instrum Technol 2012;46:268-77.

De Boer, R., & Dekker, S. (2017). Models of automation surprise: Results of a field survey in aviation. *Safety*, *3*(3), 20. doi:10.3390/safety3030020

Dehais, F., Causse, M., Régis, N., Menant, E., Labedan, P., Vachon, F., & Tremblay, S. (2012). Missing critical auditory alarms in aeronautics: Evidence for inattentional deafness? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 1639-1643. doi:10.1177/1071181312561328

Dehais, F., Causse, M., Vachon, F., and Tremblay, S. (2012). Cognitive conflict in human–automation interactions: a psychophysiological study. *Appl. Ergon*. 43, 588–595. doi: 10.1016/j.apergo.2011.09.004

Dehais, F., Causse, M., Vachon, F., Régis, N., Menant, E., & Tremblay, S. (2014). Failure to detect critical auditory alerts in the cockpit: evidence for inattentional deafness. *Human Factors*, *56*(4), 631-644.

Dehais F, Roy RN, Scannella S. Inattentional deafness to auditory alarms: Inter-individual differences, electrophysiological signature and single trial classification. *Behav Brain Res*. 2019;360:51–59. doi:10.1016/j.bbr.2018.11.045

Dennis, G. D. (2003). *TSAFE: Building a trusted computing base for air traffic control software* (Doctoral dissertation, Massachusetts Institute of Technology).

Dillon FL, Burns WJ, John RS. Insights for Critical Alarm-Based Warning Systems. *Decision Analysis*, 2018; 15(3); 154–173. Doi: 10.1287/deca.2018.0369

Dimoka, A. (2010). What Does the Brain Tell Us About Trust and Distrust? Evidence from a Functional Neuroimaging Study. MIS Quarterly, 34(2), 373-396. doi:10.2307/20721433

Dismukes, R. K. (2012). Prospective Memory in Workplace and Everyday Situations. *Current Directions in Psychological Science*, 21(4), 215–220. https://doi.org/10.1177/0963721412447621

Dixon, S. R., & Wickens, C. D. (2006). Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence in high workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *48*(3), 474-486. doi:10.1518/001872006778606822

Dixon, S. R., Wickens, C. D., & Chang, D. (2005). Mission control of multiple unmanned aerial vehicles: A workload analysis. *Human Factors*, *47*(3), 479-487.

Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2006). How do automation false alarms and misses affect operator compliance and reliance. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *50*(1), 25-29. doi:10.1177/154193120605000106

Duprès, A., Roy, R. N., Scannella, S., & Dehais, F. (2018). Pre-stimulus EEG engagement ratio predicts inattentional deafness to auditory alarms in realistic flight simulator. *In 3rd International Mobile Brain/Body Imaging Conference* (MOBI 2018), 1-3.

Edworthy, J., Loxley, S., & Dennis, I. (1991). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *The Journal of the Human Factors and Ergonomics Society*, *33*(2), 205-231. doi:10.1177/001872089103300206

Edworthy J, Reid S, Peel K, et al. The impact of workload on the ability to localize audible alarms. *Appl Ergon*. 2018;72:88–93.

Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(4), 717–726. doi:10.1037/0278-7393.16.4.717

Endsley, M. R. (2017). From here to autonomy: Lessons learned from human-automation research. *Human Factors*, *59*(1), 5-27. doi: 10.1177/0018720816681350

Endsley, M. R. (2018). Level of automation forms a key aspect of autonomy design. *Journal of Cognitive Engineering and Decision Making*, *12*(1), 29-34. doi:10.1177/1555343417723432

Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness, and workload in a dynamic control task. *Ergonomics*, *42*(3), 462-492. doi: 10.1080/001401399185595

Endsley, M. R., & Kiris, E.O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors, 37*(2), 381-394. doi: 10.1518/001872095779064555

Eriksson, L., van Erp, J., Carlander, O., Levin, B., van Veen, H., & Veltman, H. (2006). Vibrotactile and visual threat cueing with high G threat intercept in dynamic flight simulation. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *50*(16), 1547-1551. doi:10.1177/154193120605001607

Everson, R. M., & Fieldsend, J. E. (2006). Multiobjective optimization of safety related systems: An application to short-term conflict alert. *IEEE Transactions on Evolutionary Computation*, *10*(2), 187-198.

Fasano, V. (2012). *Evaluating dynamic 'landing gear unsafe auditory alerts as a defense against habituation* (Master's Thesis). http://commons.erau.edu/

Federal Aviation Administration. (n.d). *Lessons learned from civil aviation accidents: prevailing cultural/ organizational factors*. https://lessonslearned.faa.gov/ll_main.cfm?TabID=2&LLID=57&LLTypeID=6

Federal Aviation Administration. (2000). *Advisory circular: airworthiness criteria for the installation approval of a terrain awareness and warning system (taws) for part 25 airplanes.* (AC 25-23). https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC25-23.pdf

Federal Aviation Administration. (2007). *Flight procedure standards branch: instrument procedures handbook.* Washington, D.C.: U.S. Department of Transportation, Federal Aviation Administration, Flight Procedure Standards Branch.

Folk, C. L., Remington, R.W., & Wright, J. H. (1998). The structure of attentional control: Contingent attentional capture by apparent motion, abrupt onset, and color. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 317–329.

Friedman-Berg, F., Allendoerfer, K., & Pai, S. (2008). Nuisance alerts in operational ATC environments: Classification and frequencies. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *52*(1), 104-108. doi:10.1177/154193120805200123

Gawron, V. (2019). *Automation in aviation - Guidelines*. (Mitre Publication No. MTR190018). https://www.mitre.org/sites/default/files/pdf/pr-16-3426-lessons-lost-automation-in-aviation-guidelines.pdf

Geels-Blair, K., Rice, S., & Schwark, J. (2013). Using system-wide trust theory to reveal the contagion effects of automation false alarms and misses on compliance and reliance in a simulated aviation task. *International Journal of Aviation Psychology*, *23*(3), 245-266.

Giraudet, L., Imbert, J. P., Bérenger, M., Tremblay, S., & Causse, M. (2015). The neuroergonomic evaluation of human machine interface design in air traffic control using behavioral and EEG/ERP measures. *Behavioural Brain Research*, *294*, 246-253.

Giraudet, L., Imbert, J. P., Tremblay, S., & Causse, M. (2015). High rate of inattentional deafness in simulated air traffic control tasks. *Procedia Manufacturing*, *3*, 5169-5175.

Giraudet, L., St-Louis, M. E., Scannella, S., & Causse, M. (2015). P300 event-related potential as an indicator of inattentional deafness? *PLoS one*, *10*(2), e0118556.

Haberkorn, T., Koglbauer, I., & Braunstingl, R. (2014). Traffic displays for visual flight indicating track and priority cues. *IEEE Transactions on Human-Machine Systems*, *44*(6), 755-766. doi:10.1109/THMS.2014.2352496

Hah, S., Hallman, K., & Williams, B. (2015). NextGen Air Traffic Control and Technical Operations Alarms and Alerts Evaluation. U.S. Department of Transportation Federal Aviation Administration

Hah, S., Hallman, K., Williams, B., & Heinz, K. (2017). Evaluation of NextGen air traffic control alarms and alerts using cognitive walkthrough method. In Proceedings of the *Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 252-256).

Hancock, P. (2019). Some pitfalls in the promises of automated and autonomous vehicles. *Ergonomics*, *62*(4), 479-495. doi: 10.1080/00140139.2018.1498136

Hanson, D. C., Boucek, G. P., Smith, W. D., Hendrickson, J. F., Chikos, S. F., Howison, W. W., & Berson, B. L. (1982). Flight Operations Safety Monitoring Effects on the Crew Alerting System. *Proceedings of the Human Factors Society Annual Meeting*, *26*(9), 783–787. https://doi.org/10.1177/154193128202600907

Hancock, P. A., Jagacinski, R. J., Parasuraman, R., Wickens, C. D., Wilson, G. F., & Kaber, D. B. (2013). Human-automation interaction: Research and practice. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *21*(2), 9-14. doi: 10.1177/1064804613477099

Haslbeck, A., & Hoermann, H. J. (2016). Flying the needles: Flight deck automation erodes fine-motor flying skills among airline pilots. *Human Factors*, *58*(4), 533-545. doi: 10.1177/0018720816640394

Hickling, A., Brecknell, B., Loeb, R. G., & Sanderson, P. (2017). Using a Sequence of Earcons to Monitor Multiple Simulated Patients. *Human Factors*, 59(2), 268–288. https://doi.org/10.1177/0018720816670986

Hilzenrath D (2010) Technician: Deepwater Horizon warning system disabled. Washington Post (July 23), http://www .washingtonpost.com/wp-dyn/content/article/2010/07/23/ AR2010072305419.html

Hitchcock, E. M., Warm, J. S., Matthews, G., Dember, W. N., Shear, P. K., Tripp, L. D., . . . Parasuraman, R. (2003). Automation cueing modulates cerebral blood flow and vigilance in a simulated air traffic control task. *Theoretical Issues in Ergonomics Science*, *4*(1-2), 89-112. doi:10.1080/14639220210159726

Imbert, J. P., Hodgetts, H. M., Parise, R., Vachon, F., Dehais, F., and Tremblay, S. (2014). Attentional costs and failures in air traffic control notifications. *Ergonomics* 57, 1817–1832. doi: 10.1080/00140139.2014.952680

Jennings, S., Cheung, B., Rupert, A., Schultz, K., & Craig, G. (2004). Flight-Test of a Tactile Situational Awareness System in a Land-based Deck Landing Task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(1), 142–146. doi: 10.1177/154193120404800131

Jurgensohn, T., Park, I., Sheridan, T. B., & Meyer, J. (2001). Conflict warnings and the search and detection of collisions in a simulated atc-like task. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 45*(4), 458-462. doi:10.1177/154193120104500441

Karmakar, A., Bhushan, B., Santel, C., Sameer, A., & Bhattacharya, B. (2019). Interface Design of Low-Cost Collision Alerting System for Sports Aviation. *Research into Design for a Connected World*, 407-415.

Karnik A, Bonafide CP. A framework for reducing alarm fatigue on pediatric inpatient units. *Hosp Pediatr* 2015;5:160-3.

Kearney, P., Li, W. C., & Lin, J. J. (2016). The impact of alerting design on air traffic controllers' response to conflict detection and resolution. *International Journal of Industrial Ergonomics, 56,* 51-58.

Kearney, P., Li, W., Yu, C., & Braithwaite, G. (2019). The impact of alerting designs on air traffic controller's eye movement patterns and situation awareness. *Ergonomics*, *62*(2), 305-318. https://10.1080/00140139.2018.1493151

Keller, D. & Rice, S. (2010). System-wide versus component-specific trust using multiple aids. *Journal of General Psychology*, *137*(1), 114–128.

Kim, S. H., Lee, J. H., Park, D., Lee, Y., & Yun, M. H. (2019). Effects of visual complexity levels and information decluttering methods for cockpit displays on human search performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 96-100. doi:10.1177/1071181319631057

King, R. A., & Corso, G. M. (1993). Auditory displays: If they are so useful, why are they turned off. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *37*(9), 549-553. doi:10.1177/154193129303700907

Kontogiannis, T., & Malakis, S. (2009). A proactive approach to human error detection and identification in aviation and air traffic control. *Safety Science*, *47*(5), 693-706. doi:10.1016/j.ssci.2008.09.007

Kuchar, J., Andrews, J., Drumm, A., Hall, T., Heinz, V., Thompson, S., & Welch, J. (2004, September). A safety analysis process for the traffic alert and collision avoidance system (TCAS) and see-and-avoid systems on remotely piloted vehicles. *In AIAA 3rd" Unmanned Unlimited" Technical Conference, Workshop and Exhibit* (p. 6423).

Kuchar, J. E., & Drumm, A. C. (2007). The traffic alert and collision avoidance system. *Lincoln Laboratory Journal*, *16*(2), 277.

Landman, A., Groen, E., Paassen, M., Bronkhurst, A., & Mulder, M. (2017). Dealing with unexpected events on the flight deck: A conceptual model of startle and surprise. *Human Factors, 59*(1), 1-12. doi: 10.1177/0018720817723428

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50-80. doi: 10.1518/hfes.46.1.50_30392

Levine EE, Bitterly TB, Cohen TR, Schweitzer ME. Who is trustworthy? Predicting trustworthy intentions and behavior. *J Pers Soc Psychol*. 2018;115(3):468-494. doi:10.1037/pspi0000136

Lewis, M. F. (1973). *Frequency of anti-collision observing responses by solo pilots as a function of traffic density, ATC traffic warnings, and competing behavior* (No. FAA-AM-73-6). Civil Aerospace Medical Institute.

Li SYW, Tse MK, Brecknell B, Sanderson PM. Spearcon Sequences for Monitoring Multiple Patients: Laboratory Investigation Comparing Two Auditory Display Designs. *Hum Factors*. 2019;61(2):288-304. doi:10.1177/0018720818797502

Lin JT, Chen YC, Wu SC, Hwang SL. Development and evaluation of a monitoring-aid system for a nuclear power plant in control room system manipulation. *Work*. 2017;57(4):611–625. doi:10.3233/WOR-172587

Loft, S., Smith, R. E., & Remington, R. W. (2013). Minimizing the disruptive effects of prospective memory in simulated air traffic control. *Journal of Experimental Psychology: Applied*, 19(3), 254–265. https://doi.org/10.1037/a0034141

Loft, S., Smith, R. E., & Bhaskara, A. (2011). Prospective memory in an air traffic control simulation: External aids that signal when to act. *Journal of Experimental Psychology: Applied*, 17(1), 60–70. https://doi.org/10.1037/a0022845

Long, K. M., & McGarry, K. A. (2009). Why are those lights flashing? direct-to-pilot warnings for preventing runway incursions. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *53*(1), 61-65. doi:10.1518/107118109X12524441078987

Lysaght, R. J., Hill, S. G., Dick, A. O., Plamondon, B. D., & Linton, P. M. (1989). *Operator workload: Comprehensive review and evaluation of operator workload methodologies* (No. TR-2075-3). Analytics Inc Willow Grove Pa.

Maltz, M., & Shinar, D. (2003). New alternative methods of analyzing human behavior in cued target acquisition. *The Journal of the Human Factors and Ergonomics Society*, *45*(2), 281-295. doi:10.1518/hfes.45.2.281.27239

Masalonis, A., & Parasuraman, R. (2003). Fuzzy signal detection theory: Analysis of human and machine performance in air traffic control, and analytic considerations. *Ergonomics*, *46*(11), 1045-1074. doi:10.1080/0014013031000121986

Manzey D, Gérard N, Wiczorek R. Decision-making and response strategies in interaction with alarms: the impact of alarm reliability, availability of alarm validity information and workload. *Ergonomics* 2014;57:1833-55.

McBride, S. E., Rogers, W. A., & Fisk, A. D. (2014). Understanding human management of automation errors. *Theoretical Issues in Ergonomics Science*, *15*(6), 545-577. doi:10.1080/1463922X.2013.817625

McFadden, S. M., Vimalachandran, A., & Blackmore, E. (2004). Factors affecting performance on a target monitoring task employing an automatic tracker. *Ergonomics*, *47*(3), 257-280. doi:10.1080/00140130310001629748

Meng F, Spence C. Tactile warning signals for in-vehicle systems. *Accid Anal Prev*. 2015;75:333-346. doi:10.1016/j.aap.2014.12.013

Meyer, J. (2001). Effects of Warning Validity and Proximity on Responses to Warnings. *Human Factors*, *43*(4), 563–572. https://doi.org/10.1518/001872001775870395

Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *The Journal of the Human Factors and Ergonomics Society, 46*(2), 196-204. doi:10.1518/hfes.46.2.196.37335

Moacdieh, N. M., Prinet, J. C., & Sarter, N. B. (2013). Effects of modern primary flight display clutter: Evidence from performance and eye tracking data. *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1) 11-15.

Molloy, R., & Parasuraman, R.(1996). Monitoring an automated system for a single failure: Vigilance and task complexity. *Human Factors, 38*(2), 311-322. doi: 10.1177/001872089606380211

Mouloua, M. E., & Koonce, J. M. (1997). *Human–automation interaction: Research and practice*. Lawrence Erlbaum Associates, Inc.

Mumaw, R. J. (2017). Analysis of alerting system failures in commercial aviation accidents. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 110-114. doi:10.1177/1541931213601493

Myers, D. (2008). *Tactile situational awareness system (tsas)*. (Report No. N68335-09-C-0025). Department of Defense. https://www.sbir.gov/content/tactile-situational-awareness-system-tsas-4

Ng G, Barralon P, Dumont G, Schwarz SK, Ansermino JM. Optimizing the tactile display of physiological information: vibro-tactile vs. electro-tactile stimulation, and forearm or wrist location. *Conf Proc IEEE Eng Med Biol Soc*. 2007;2007:4202-4205. doi:10.1109/IEMBS.2007.4353263

Ngo, M. K., Pierce, R. S., & Spence, C. (2012). Using multisensory cues to facilitate air traffic management. *The Journal of Human Factors and Ergonomics Society*, *54*(6), 1093-1103. doi:10.1177/0018720812446623

Nikolic, M. I., Orr, J. M., & Sarter, N. B. (2004). Why pilots miss the green box: How display context undermines attention capture. *The International Journal of Aviation Psychology*, *14*(1), 39-52. doi:10.1207/s15327108ijap1401_3

Nilsson LG, Nyberg L, Bäckman L, Genetic variation in memory functioning, *Neuroscience & Biobehavioral Reviews*, 26 (7): 841-8, 2002, doi:10.1016/S0149-7634(02)00070-2.

Paielli, R. A., Erzberger, H., Chiu, D., & Heere, K. R. (2009). Tactical conflict alerting aid for air traffic controllers. *Journal of Guidance, Control, and Dynamics*, *32*(1), 184-193.

Parasuraman, R. (1987). Human-computer monitoring. *Human Factors*, *29*(6), 695-706. doi: 10.1177/001872088702900609

Parasuraman, R., & Hancock, P. A. (2001). Adaptive control of mental workload. In P. A. Hancock & P. A. Desmond (Eds.), *Human factors in transportation. Stress, workload, and fatigue* (p. 305–320). Lawrence Erlbaum Associates Publishers.

Parasuraman, R., & Manzey, D. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, *52*(3), 381-410. doi: 10.1177/0018720810376055.

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced 'complacency.' *The International Journal of Aviation Psychology*, *3*(1), 1-23. doi:10.1207/s15327108ijap0301_1

Parasuraman, R., & Mouloua, M. (1996). *Automation and human performance: Theory and applications* (1st ed.). Mahwah, N.J: Lawrence Erlbaum Associates. doi:10.1201/9781315137957

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, and abuse. *Human Factors, 39*(2), 230-253. doi: 10.1518/001872097778543886

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 30*(3), 286-297. doi:10.1109/3468.844354

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness: Mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making*, *2*(2), 140-160. doi: 10.1518/155534308X284417

Perry, N. C., Stevens, C. J., Wiggins, M. W., & Howell, C. E. (2007). Cough Once for Danger: Icons Versus Abstract Warnings as Informative Alerts in Civil Aviation. *Human Factors*, *49*(6), 1061–1071. https://doi.org/10.1518/001872007X249929

Peryer G, Noyes J, Pleydell-Pearce K, Lieven L (2005) Auditory Alert Characteristics: A Survey of Pilot Views, *The International Journal of Aviation Psychology*, 15:3, 233-250, DOI: 10.1207/s15327108ijap1503_2

Pritchett, A., Haga, R., & Thakkar, D. (2014). Pilot responses to traffic events during NextGen high traffic density terminal operations. *In Digital Avionics Systems Conference (DASC), 2014 IEEE/AIAA 33rd* (pp. 2C4-1).

Raj, A. K., & Braithwaite, G. (1999). The tactile situation awareness system in rotary-wing aircraft: Flight test results. *Current Aeromedical Issues in Rotary Wing Operations*, *5*.

Reinecke, M. (1976). Experimental cognition for qualification of voice warning systems in aircraft. *Deutsche Gesellschaft fuer Luft- und Raumfahrt, Jahrestagung, 9 th, Munich, West Germany*.

Rice, S. & Geels, K. (2010). Using system-wide trust theory to make predictions about dependence on four diagnostic aids. *Journal of General Psychology, 137*(4), 362-375.

Rice, S. & Keller, D. (2009). Automation reliance under time pressure. *Cognitive Technology, 14*(1), 36-44.

Rice, S. & McCarley, J. (2011). Effects of response bias and judgment framing on operator use of an automated aid in a target detection task. *Journal of Experimental Psychology: Applied, 17*(4), 320-331.

Rice, S. & Trafimow, D. (2012). Time pressure heuristics can improve performance due to increased consistency: A PPT methodology. *Journal of General Psychology, 139*(4), 273-288.

Rice, S., Trafimow, D. & Hunt, G. (2010). Using PPT to analyze sub-optimal human-automation performance. *Journal of General Psychology, 137*(3), pp. 310 - 329.

Rice, S., Trafimow, D., Keller, D., Hunt, G. & Geels, K. (2011). Using PPT to correct for inconsistency in a speeded task. *Journal of General Psychology, 138*(01), pp. 12-34.

Rice, S., Trafimow, D., Rahmania, A., Keller, D., Hunt, G., Taruc, K., Dirhamsyah, M. & Ridha, M. (2010). Sound the alarm--death approaches (suarakan genderang--maut menyerang). The Tsunami Early Warning System in Indonesia. *International Journal of Technology, Knowledge and Society 6*(4), pp. 13-26.

Rovira, E., & Parasuraman, R. (2010). Transitioning to future air traffic management: Effects of imperfect automation on controller attention and performance. *The Journal of Human Factors and Ergonomics Society*, *52*(3), 411-425. doi:10.1177/0018720810375692

Santangelo, V., Belardinelli, M. O., & Spence, C. (2007). The suppression of reflexive visual and auditory orienting when attention is otherwise engaged. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 137-148. doi:10.1037/0096-1523.33.1.137

Salzer, Y., Oron-Gilad, T., Ronen, A., & Parmet, Y. (2011). Vibrotactile "on-thigh" alerting system in the cockpit. *Human Factors*, *53*(2), 118. doi:10.1177/0018720811403139

Sammut, A., & Mangion, D. Z. (2014). Runway conflict alerting: advisory or directive? *The Aeronautical Journal*, *118*(1205), 713-741.

Santel, C. G. (2016). *Factors influencing the usability of collision alerting systems in gliding* (Doctoral dissertation, Technische Universität Darmstadt).

Sarter, N. B., Mumaw, R. J., & Wickens, C. D. (2007). Pilots' monitoring strategies and performance on automated flight decks: An empirical study combining behavioral and eye-tracking data. *The Journal of the Human Factors and Ergonomics Society*, *49*(3), 347-357. doi:10.1518/001872007X196685

Sarter, N. B., & Woods, D. D. (1994). Pilot interaction with cockpit automation II: An experimental study of pilots' model and awareness of the flight management system. *The International Journal of Aviation Psychology*, *4*(1), 1-28. doi:10.1207/s15327108ijap0401_1

Scannella, S., Causse, M., Chauveau, N., Pastor, J., & Dehais, F. (2013). Effects of the audiovisual conflict on auditory early processes. *International Journal of Psychophysiology*, *89*(1), 115-122. doi:10.1016/j.ijpsycho.2013.06.009

Scerbo, M. W., & Mouloua, M. (1999). *Automation technology and human performance: Current research and trends.* Mahwah: NJ: Lawrence Erlbaum.

Schmid F, Goepfert MS, Kuhnt D, et al. The wolf is crying in the operating room: patient monitor and anesthesia workstation alarming patterns during cardiac surgery. Anesth Analg 2011;112:78-83.

Schönefeld, J., & Möller, D. P. F. (2012). Runway incursion prevention systems: A review of runway incursion avoidance and alerting system approaches. *Progress in Aerospace Sciences*, *51*, 31-49. doi:10.1016/j.paerosci.2012.02.002

Sheridan, T. B. (1998). Allocating functions rationally between humans and machines. Ergonomics in Design: *The Quarterly of Human Factors Applications, 6*(3), 20-25. doi:10.1177/106480469800600305

Sheridan, T. B. (2002). Humans and automation: System design and research issues. Santa Monica, CA, USA: *Human Factors and Ergonomics Society*.

Sheridan, T. B., Vámos, T., & Aida, S. (1983). Adapting automation to man, culture and society. *Automatica*, *19*(6), 605-612. doi:10.1016/0005-1098(83)90024-9

Simpson, C. A., & Williams, D. H. (1980). Response time effects of alerting tone and semantic context for synthesized voice cockpit warnings. *The Journal of the Human Factors and Ergonomics Society, 22*(3), 319-330. doi:10.1177/001872088002200306

Singh, A. L., Tiwari, T., & Singh, I. L. (2009). Effects of automation reliability and training on automation-induced complacency and perceived mental workload. *Journal of the Indian Academy of Applied Psychology*, *35*(2009), 9-22.

Sorkin, R. D., & Woods, D. D. (1985). Systems with human monitors: A signal detection analysis. *Human-Computer Interaction*, *1*(1), 49-75. doi:10.1207/s15327051hci0101_2

Sörqvist, P., Nöstl, A. & Halin, N. Working memory capacity modulates habituation rate: Evidence from a cross-modal auditory distraction paradigm. *Psychon Bull Rev* 19, 245–250 (2012). https://doi.org/10.3758/s13423-011-0203-9

Spector MD, Jones GE (2004) Trust in the Workplace: Factors Affecting Trust Formation Between Team Members, *The Journal of Social Psychology*, 144:3, 311-321, DOI: 10.3200/SOCP.144.3.311-321

Stankovic, A., Aitken, M. R., & Clark, L. (2014). An eye-tracking study of information sampling and decision-making under stress: implications for alarms in aviation emergencies. *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*, 125-129).

Starcke, K., Wolf, O. T., Markowitsch, H. J., & Brand, M. (2008). Anticipatory stress influences decision making under explicit risk conditions. *Behavioral Neuroscience*, *122*(6), 1352-1360. doi:10.1037/a0013281

Stager P, Hameluck D (1990) Ergonomics in air traffic control, *Ergonomics*, 33:4, 493-499. DOI: 10.1080/00140139008927156

Tear MJ, Reader TW, Shorrock S, Kirwan B. (2020) Safety culture and power: Interactions between perceptions of safety culture, organisational hierarchy, and national culture, *Safety Science*, 121, 550-561. DOI: 10.1016/j.ssci.2018.10.014.

Thomas, L. C., & Rantanen, E. M. (2006). Human factors issues in implementation of advanced aviation technologies: A case of false alerts and cockpit displays of traffic information. *Theoretical Issues in Ergonomics Science*, *7*(5), 501-523.

Tippey, K. G., Roady, T., Rodriguez-Paras, C., Brown, L. J., Rantz, W. G., & Ferris, T. K. (2017). General aviation weather alerting: the effectiveness of different visual and tactile display characteristics in supporting weather-related decision making. *The International Journal of Aerospace Psychology*, *27*(3-4), 121-136.

Vagia, M., Transeth, A. A., Fjerdingen, S. A. (2016). A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? *Applied Ergonomics, 53*(Part A), 190-202. doi: 10.1016/j.apergo.2015.09.013

Veitengruber, J. E., Boucek, G. P., & Smith, W. D. (1977). *Aircraft alerting systems criteria study* (Vol. 1). Department of Transportation, Federal Aviation Administration, Systems Research and Development Service.

Walker BN, Lindsay J, Nance A, et al. Spearcons (speech-based earcons) improve navigation performance in advanced auditory menus. *Hum Factors*. 2013;55(1):157-182. doi:10.1177/0018720812450587

Wheale, J. L. (1981). The speed of response to synthesized voice messages. *British Journal of Audiology*, *15*(3), 205-212. doi:10.3109/03005368109081439

Wickens, C.D. (1992). *Engineering psychology and human performance*. New York, NY: Harper-Collins.

Wickens, C. D. (2008). Situation awareness: Review of Mica Endsley’s 1995 articles on situation awareness theory and management. *Human Factors, 50*(3), 397-403. doi: 10.1518/001872008X288420

Wickens, C. D., Clegg, B. A., Vieane, A. Z., & Sebok, A. L. (2015). Complacency and automation bias in the use of imperfect automation. *The Journal of Human Factors and Ergonomics Society*, *57*(5), 728-739. doi:10.1177/0018720815581940

Wickens, C. D., & Colcombe, A. (2007). Dual-task performance consequences of imperfect alerting associated with a cockpit display of traffic information. *Human Factors*, *49*(5), 839-850.

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. doi:10.1080/14639220500370105

Wickens, C., Dixon, S., Goh, J., & Hammer, B. (2005). Pilot dependence on imperfect diagnostic automation in simulated UAV flights: An attentional visual scanning analysis. *13th International Symposium on Aviation Psychology, Dayton, OH, 2005*

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2015). *Engineering psychology and human performance*. Psychology Press.

Wickens, C. D., & Kessel, C. (19801980). Processing resource demands of failure detection in dynamic systems. *Journal of Experimental Psychology: Human Perception and Performance, 6*(3), 564.

Wickens, C. D., & Liu, Y. (1988). Codes and modalities in multiple resources: A success and a qualification. *The Journal of the Human Factors and Ergonomics Society*, *30*(5), 599-616. doi:10.1177/001872088803000505

Wickens, C. D., Mavor, A. S., Parasuraman, R., & McGee, J. P. (1998). *The future of air traffic control: human operators and automation.* Washington, DC: National Academy Press.

Wickens, C. D., Rice, S., Keller, D., Hutchins, S., Hughes, J., & Clayton, K. (2009). False alerts in air traffic control conflict alerting system: Is there a "cry wolf" effect? *The Journal of the Human Factors and Ergonomics Society*, *51*(4), 446-462. doi:10.1177/0018720809344720

Wiener, E. L. (1977). Controlled flight into terrain accidents: System-induced errors. *The Journal of the Human Factors and Ergonomics Society*, *19*(2), 171-181. doi:10.1177/001872087701900207

Wiener, E. L. (1988). *Human factors in aviation*. In E. L. Wiener and D. C. Nagel (Eds.), Cognition and Perception (pp. 433-461): Elsevier

Wilson, M. D., Farrell, S., Visser, T. A. W., & Loft, S. (2018). Remembering to execute deferred tasks in simulated air traffic control: The impact of interruptions. *Journal of Experimental Psychology*. *Applied*, *24*(3), 360-379. doi:10.1037/xap0000171

Xiaojun Wu & Zhizhong Li (2018) A Review of Alarm System Design for Advanced Control Rooms of Nuclear Power Plants, International Journal of Human–Computer Interaction, 34:6, 477-490, DOI: 10.1080/10447318.2017.1371950

Xiao Y, Seagull FJ. An Analysis of Problems with Auditory Alarms: Defining the Roles of Alarms in Process Monitoring Tasks. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 1999;43:256-60.

Xu, X., Wickens, C. D., & Rantanen, E. M. (2007). Effects of conflict alerting system reliability and task difficulty on pilots' conflict detection with cockpit display of traffic information. *Ergonomics*, *50*(1), 112-130. doi:10.1080/00140130601002658

Yeh, Y.-Y., & Wickens, C. D. (1988). Dissociation of Performance and Subjective Measures of Workload. *Human Factors*, 30(1), 111–120. https://doi.org/10.1177/001872088803000110

Zheng, Y., Lu, Y., Yang, Z., & Fu, S. (2014). Expertise and responsibility effects on pilots' reactions to flight deck alerts in a simulator. *Aviation, Space, and Environmental Medicine*, *85*(11), 1100-1105.

Ziółkowski, A., & Skłodowski, M. (2018). On value and using of TAWS/FMS alert data in examination of air accidents, the case of warsaw-smolensk flight on 10 april 2010. *The Aeronautical Journal*, *122*(1255), 1301-1329. doi:10.1017/aer.2018.59