

Temporal Effects in Aircraft Inspection: What Price Vigilance Research?

Colin G. Drury, Monique Saran and John L. Schultz
 University at Buffalo, Department of Industrial Engineering
 438 Bell Hall, Amherst, NY 14260
 drury@buffalo.edu

Abstract

This paper examines issues of fatigue in inspection by using an established function analysis of inspection to show its characteristics, and then proposing a four-level classification of temporal effects to help future applications. This classification divides the temporal effects into four components: weekly, daily, hourly, and minute time scales. The analysis presented here will form the basis for the design of future experimental studies of temporal factors in aircraft inspection.

Introduction

Failures of both airframe inspection and engine inspection have highlighted the potential impact of human limitations on inspection system performance. Accidents that have occurred due to engine inspection failure include the Sioux City and Pensacola accidents. The 1989 Sioux City crash was the result of inspection not finding a crack in an engine disk. Remnants of fluorescent penetrant were found in the crack after the crash. These remnants helped to determine that the crack was large enough to be seen when the inspection occurred but why it was missed is not known. The 1996 Pensacola crash was due to a fan hub in the left engine having an undetected crack. Both of these crashes could have been prevented if the cracks had been located during inspection. In a 1998 incident to an Aloha Boeing 737 aircraft, evidence was found of multiple site fatigue damage leading to structural failure. The resulting National Transportation Safety Board investigation report issued in 1989 attributed the incident to the failure of the operators' maintenance program to detect corrosion damage. A common thread in all three incidents was that inspection failure occurred during inspection tasks of normal working duration, i.e. a working shift with typical breaks. A number of visual and Non-Destructive Inspection (NDI) techniques require the inspector to work continuously on repetitive tasks for extended periods. Examples are fluorescent penetrant inspection of engine rotor blades, eddy current inspection of large batches of wheel bolts, and magnetic particle inspection of landing gear components. Such tasks typically occur on all shifts and can involve inspecting at low periods of the human circadian rhythm. Inspectors may be subject to the effects of cumulative fatigue from overtime and shift work.

In all of these inspection tasks, the *a priori* similarity to classical vigilance tasks suggests that performance (defect detection) may decrease with time spent inspecting. However, much skepticism exists regarding the relevance

of vigilance studies to the operational environment. In the case of aircraft inspection tasks, there is the added complication of the relevance of shift-work and circadian rhythm studies to these particular tasks. Thus, we have two issues:

1. Can we expect the findings from the vigilance literature to apply to aircraft inspection?
2. How well might the studies of circadian rhythms and cumulative fatigue from shift working apply to vigilance, and then to aircraft inspection?

Note that both of these issues concern the temporal effects of inspection work. This paper examines these issues by using an established function analysis of inspection to show its characteristics, and then proposing a four-level classification of temporal effects to guide future applications. Indeed, the analysis presented here will form the basis for the design of future experimental studies of temporal factors in aircraft inspection.

Analysis of Inspection Tasks in Aviation

To understand inspection, and to provide a link between inspection and the psychology / human factors literature, we use the generic functions which comprise all inspection tasks whether manual, automated or hybrid. We have recently undertaken a systematic analysis of all of the inspection techniques involved in NDI of aircraft (Drury, 2003), so far covering Fluorescent Penetrant Inspection (FPI), Visual inspection, Borescopes, Eddy Current and Ultrasonics. All were studied in aircraft maintenance settings to perform Hierarchical Task Analyses and thus derive a set of Good Practices related to human and system functioning. Each of these NDI techniques exhibited all of the generic functions, although some required much preparation prior to the actual inspection. Table 1 shows these functions, with the specific application to NDI in aviation. We can go further by taking each function and listing its correct outcome, from which we can logically

derive the possible errors, and even analyze level of functioning using Rasmussen’s Skill / Rule / Knowledge hierarchy (Drury and Prabhu, 1994).

The functions of search and decision are the most error-prone, although for much of inspection, especially NDI, setup can cause its own unique errors (Murgatroyd, Worrall and Waites, 1994). Search and decision have been the subjects of considerable mathematical modeling in the human factors community, with direct relevance to visual inspection.

In the visual aspects of inspection tasks, the inspector must move his/her eyes around the item to be inspected to ensure that any defect will eventually appear within an area around the line of sight in which it is possible to achieve detection. This area, called the visual lobe, varies in size depending upon target and background characteristics, illumination and the individual inspector’s peripheral visual acuity. As successive fixations of the visual lobe on different points occur at about three per second, it is possible to determine how many fixations are required for complete coverage of the area to be searched. We have useful models of visual search applicable to inspection (Wolfe 1994; Drury and Hong 2000), but the point made here is that all inspection tasks in aviation do involve some search, in contrast to many laboratory vigilance tasks.

Decision-making is the second key function in inspection. This is where each indication is judged as being a defect or not a defect. An inspection decision can have four outcomes (Table 2). These outcomes have associated probabilities, for example, the probability of detection is the fraction of all defective items rejected by the inspector shown as p_2 in Table 2.

Decision of Inspector	True State of Indication	
	Non-defect	Defect
Accept, i.e. Call non-defect	Correct accept, p_1	Miss, $(1 - p_2)$
Reject, i.e. Call defect	False alarm, $(1 - p_1)$	Hit, p_2

Table 2. Four outcomes of inspection decisions

At this point, the obvious rational decision making models such as Signal Detection Theory are usually invoked to equate inspection to simple decisions. From the analysis in Table 1, it is clear that inspection is not merely the decision function. The use of models such as signal detection theory to apply to the *whole* inspection process is misleading in that it ignores the search function. For example if the search is poor, then many defects will not be located. At the overall level of the inspection task, this means that probability of detection (PoD) decreases, but this decrease has nothing to do with setting the wrong decision criteria. Even such devices as ROC curves should

be applied only to the decision function of inspection, not the overall process, unless search failure can be ruled out on logical grounds.

Temporal Aspects of Inspection

Temporal effects in the literature occur over four times scales:

1. Weeks, where the issues are shift work and cumulative fatigue from hours of work, sleep loss, days worked, overtime and shift work.
2. Days, where circadian rhythms are predominant, so that time of day is the main driver.
3. Hours, where the issues are times spent continuously on tasks, and the timing, nature and duration of rest periods
4. Minute, where the concern is sequential effects in repetitive tasks: does the detection of a defect on one item inspected affect the behavior or performance on subsequent items?

Each of these is reviewed in turn before examining in more detail their relevance to aircraft inspection. To help obtain background data on the hours of work and shift work patterns of NDI inspectors, a survey “Aircraft Maintenance Personnel Survey of Work Hours” was given to samples of NDI inspectors at several airlines. The survey, Folkard (2002), asks about hours of work, shift systems, breaks, vacation days and some symptoms of stress. Here we present simple summary statistics, from our first group of 40 NDI inspectors at two airlines. The sample was older and more experienced than typically found for AMTs. Comparing the age and experience distributions to the population demographics of Aviation Maintenance Technicians found in a national sample compiled by the Bureau of Labor Statistics (BLS, Washington, 1991), our sample was significantly older with a median age of 46.5 year versus a BLS median age of 36.2 years (Wilcoxon test, $t = 645$, $p < 0.001$). Our sample was also more experienced with a median of 24.0 years as an Aviation Maintenance Technician versus a BLS median of 9.4 years (Wilcoxon test, $t = 780$, $p < 0.001$). Selected questions on hours of work and rest are given in Table 3.

	Median	Minimum	Maximum
Hours of work per week	40	30	56
How long before a work break?	2.0	1.0	4.0
How many minutes does break last?	12.5	0	45
How many days annual leave?	31	11	40

Table 3. Sample work characteristics of NDI Inspectors

The temporal work characteristics appear about what would be expected, with 40-hour weeks, 2 hours between

breaks and 10-minute breaks. The relatively long vacation periods presumably arise from the high seniority typical of NDI inspectors, and confirmed here by the high age and experience statistics.

Weeks

The cumulative fatigue effects of shifts may span a period of a week or more. Fletcher and Dawson (2001) showed how fatigue builds up over the course of a week and its interactions with circadian variations. Their model was validated with a field study using OSPAT (Occupational Safety Performance Assessment Technology) performance tests and a VAS (Visual Analog Scale) measurement of alertness. French and Morris (2003) developed the FADE model that was validated using results from pattern recognition test from the NASA Space Cognitive Assessment Test (SCAT) battery and a divided attention version of the Maniken Task. Both models show the cumulative effects of shift work over a week and show circadian lows that occur daily.

Days

The daily variations in performance that an individual goes through are cyclic and predictable. The circadian rhythms or internal biological clock combined with environmental cues (*zeitgebers*) make people diurnal or active during the day. In general, humans show the same type of phasic behavior in performance as these biological rhythms, but there are individual differences in the timing of the onset of phases. Typically, people experience a circadian low, in measures such as body temperature, at approximately 0400 each day. Other variables relating to human bodily functions have been found to have lower values at night including heart rate, blood pressure and urinary excretion (Folkard 2002; Fletcher and Dawson, 2001). Studies of shift work contain strong evidence for circadian rhythm influence on performance decrements and contain recommendations for ameliorating performance decrements associated with circadian variations (Della Rocco, Comperatore, Caldwell, Cruz 2000; Fletcher and Dawson 2001; Folkard 2002). Vigilance effects (see **Hours**) appear quite sensitive to diurnal effects.

Hours

The vigilance decrement is a decline in performance that occurs along the hourly time scale. Typically, performance drops during the first 15 minutes on task and continues to decline until about 30 minutes into a task (Teichner, 1974).

Parasurman and Davies (1977) discussed vigilance in depth from a decision theory (SDT) approach and stated the decline in performance was based on the task characteristics of successive vs. simultaneous and the event rate or the numbers of stimuli over time. Their taxonomy

of vigilance showed that sensitivity decrement was related to these two factors. More recently, See, Howe, Warm and Dember (1995) conducted a meta analysis of the sensitivity decrement in vigilance and determined that these task characteristics are a large component of the vigilance decrement but that the sensory-cognitive component must be investigated as well. For aircraft inspection work this last distinction is not relevant, no targets are uniformly “sensory” in See et al, terminology.

Vigilance shares many characteristics of the inspection task such as rare signals, time on task, high memory load, and spatial and temporal uncertainty, but is different in other ways, as detailed later.

Minutes

Sequential effects are those found on time scales of seconds or minutes, and represent the influence of recent prior targets on subsequent performance. Tsao (1984) found that “following the detection of a faulty item, stopping time decreases for the second and third items, increases for the sixth and seventh items, and then levels off.” This was true with different target difficulty levels and for different informed or feed-forward defect rates. A re-analysis of the Panjwani and Drury (2003) data on rare-event inspection found a negligible sequential effect. There may be small sequential effects, but they are unlikely to influence the aircraft inspection task significantly due to the very low event rate for this task, and to their small absolute magnitude.

Relevance to Aircraft Inspection

From the site visits, the hours of work survey and Folkard’s study in the aviation maintenance industry, it does appear that temporal effects are likely in aircraft inspection tasks. Shift working is common, although most inspection in component shops is still on day shift. Both night shifts and changing shift schedules have been shown to reduce performance on tasks similar to inspection, e.g. vigilance tasks. While it is still not clear how closely vigilance mimics aviation inspection tasks, it is quite clear that vigilance tasks are particularly sensitive to the effects of circadian lows and cumulative fatigue from shift working. Thus, inspection tasks with vigilance-like characteristics are performed at times when decrements would be expected. The integrative models of Folkard (2002), Fletcher and Dawson (1998) and French and Morris (2003) all give sound advice on avoiding cumulative fatigue states. The typical work/rest schedule is 2 hours work followed by 10 minutes rest, which would again give cause for concern if vigilance tasks were indeed close mimics of inspection. The vigilance decrement literature shows performance declines over periods of less than one hour for some types of vigilance task. Tasks particularly susceptible to decrements are those where there is no

constantly available comparison standard, and where signals are rare, both characteristics of aircraft inspection. Other factors causing a vigilance decrement are less relevant: untrained personnel and symbolic stimuli. Overall, we can compare the attributes of classical vigilance tasks with those of aircraft inspection, as shown in Table 4.

The task for subsequent years of this project is to use the classification scheme and literature comparison to produce guidelines and industry best practices for alleviating problems of fatigue in aircraft inspection. As part of the production of such guidelines, we will need to validate them using a simulation of particular inspection tasks. That simulation will need to be realistic with respect to both the task itself and the participants performing the task.

Conclusions

It is evident that we must be careful of assuming that vigilance and inspection are identical, and thus applying vigilance findings to inspection tasks blindly. Inspection, especially aircraft NDI, has many complex subtasks, only a few of which are likely to have the characteristics of vigilance tasks.

As we move to the broader fields of temporal effects, such as circadian rhythms or shift work, we must not assume that vigilance findings hold. Indeed, a recent paper on time of day effects (Horowitz, Cade, Wolfe, and Czesler, 2003) found the usual effect of peaks and troughs of circadian rhythm on a vigilance task, but none on a simple search tasks performed at similar times.

Acknowledgement

This work was financed by the Federal Aviation Administration, Grant No. 03-G-012; Contract monitor, William K. Krebs.

References

- Bureau of Labor Statistics (1991). BLS, Washington.
- Della Rocco, P. S., Comperatore, C., Caldwell, L. and Cruz, C. E. (2000). *The Effects of Napping on Night Shift Performance*. FAA Tech Report.
- Drury, C. G. (2003). Human Factors and Inspection: Best Practices. *Proceedings of the 93rd Congress of the International Ergonomics Association Annual Conference*, Seoul, Korea.
- Drury, C. G. and Hong, S.-K. (2000). Generalizing from single target search to multiple target search. *Theoretical Issues in Ergonomics Science*. **1**(4), 303-314.
- Drury, C. G. and Prabhu, P. V. (1994). Human factors in test and inspection. In G. Salvendy and W. Karwowski (eds.), *Design of Work and Development of Personnel in Advanced Manufacturing*, J. Wiley, New York, 355-402.
- Fletcher, A. and Dawson, D. (1998). A Work-Related Fatigue Model Based on Hours-of-Work. *Managing fatigue in transportation: proceedings of the 3rd Fatigue in Transportation Conference*, Fremantle, Western Australia, 189-208.
- Fletcher, A. and Dawson, D. (2001). Field-based validations of a work-related fatigue model based on hours of work. *Transportation Research Part F* **4**, 75-88.
- Folkard, S. (2002). *Work Hours of Aircraft Maintenance Personnel*. Civil Aviation Authority, CAA Paper 2002/06.
- Folkard, S. and Monk, T. (1985). *Hours of Work, Temporal Factors in Work-Scheduling*, John Wiley and Sons, Chichester.
- French, J. and Morris, C. S. (2003). Modeling Fatigue Degraded Performance in Artificial Agents. *Proceeding of the Human Factors and Ergonomics Society 47th Annual Meeting*, 307-310.
- Horowitz, T. S., Cade, B. ., Wolfe, J. M. and Czesler, C. A. (2003). Searching night and day: a dissociation of effects of circadian phase and time awake on visual selective attention and vigilance. *Psychological Science*, **14**(6), 549-557.
- Huey, B. M. and Wickens, C. D. (1993). *Workload Transition: Implications for Individual and Team Performance*, The National Academy of Sciences.
- Murgatroyd, R. A., Worrall, G. M. and Waites, C. (1994). A Study of the Human Factors Influencing the Reliability of Aircraft Inspection, AEA/TSD/0173. Rislely, AEA Technology.
- Panjawani, G. and Drury, C. G. (2003). Effective Interventions in Rare Event Inspection. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Denver, CO*, 41-45.
- Parasuraman, R. and Davies, D.R. (1977). A taxonomic Analysis of Vigilance Performance. In R. R. Mackie (ed), *Vigilance, Theory, Operational Performance, and Physiological Correlates*, Plenum Press, New York.
- Teichner, W. H. (1974). The Detection of a Simple Visual Signal as a Function of Time of Watch. *Human Factors*, **16**(4), 339-353.
- See, J., Howe, S. R., Warm, J. S. and Dember, W. N. (1995). Meta-Analysis of the Sensitivity Decrement in Vigilance. *Psychological Bulletin*, **117**(2), 230-249.
- Tsao, Y.-C. and Wang, T.-G. (1984). *Inspectors' Stopping Policy After Fault Detection*. *Human Factors*, **26**(6), 649-657.
- Wolfe J. M. (1994). Guided Search 2.0: A Revised Model of Visual Search. *Psychonomic Bulletin and Review*, **1**(2), 202-238.

Function	Inspection Description
1. Initiate	All processes up to accessing the component. Get and read workcard. Assemble and calibrate required equipment. For FPI this includes part preparation steps.
2. Access	Locate and access inspection area. Be able to see the area to be inspected at a close enough level to ensure reliable detection. For component inspection, the parts are typically brought to the inspector rather than the inspector going to the airframe.
3. Search	Move field of view across component to ensure adequate coverage. Carefully scan field of view using a good strategy. Stop search if an indication is found.
4. Decision	Identify indication type. Compare indication to standards for that indication type.
5. Response	If indication confirmed, then record location and details. Complete paperwork procedures. Remove equipment and other job aids from work area and return to storage. If indication not confirmed, continue search (3).

Table 1. Generic function description and application to Non-Destructive Inspection

VIGILANCE TASK ATTRIBUTE	INSPECTION TASK ATTRIBUTE
Important Signals	Cracks or other defects that can have direct safety consequences.
Rare Signals	Defects can range from quite common, e.g. corrosive areas on older aircraft, to extremely rare (e.g. cracks in jet engine titanium hubs). Under most circumstances far less than 1 out of 10 inspected components will contain a reportable defect.
Low Signal Strength	Most defects are perceptually difficult to detect, often occurring within a background of non-defects, e.g. cracks among dirt marks and scratches.
Long Time on Task	Time on task can vary from a few minutes to about 2 hours without a break. Scheduled breaks are typically four 15-min breaks per shift, but many tasks are self-paced so that inspectors can break early or continue beyond scheduled time to complete an area or component.
High Memory Load	Prototypical defects are usually stored in the inspector's memory, rather than being presented as part of the task. Sometimes typical defects are illustrated on workcards, but workcards are often poorly integrated into the inspection task.
Low Observer Practice	Inspectors are highly skilled and practiced, after 3-10 years as an AMT before becoming an inspector. However, for some rare defects, even experienced inspectors may literally never have seen one in their working lifetime.
Sustained Attention on One Task	Inspectors may have some tasks where just one defect type is the target, but these are often interspersed with other tasks (e.g. different components) where different defects, often less rare defects, are the target.
Time Uncertainty	Defect occurrence is rarely predictable although inspectors often return to the same area of the same aircraft or engine and attempt to predict when defects are likely.
Spatial Uncertainty	While the actual occurrence of defects at specific places on specific components may be unpredictable, the inspector can have much useful information to guide the inspection process. Training, service bulletins and shared experiences can help point inspectors to specific locations where defects are more likely.
Low Feedback	Aircraft inspectors do not get good feedback, mainly because there is no easy way to find what truly is a signal, especially a missed signal. Feedback on missed defects only comes when one is found at a subsequent inspection, or when an operational incident occurs. Even feedback on false alarms is sporadic. Feedback of both Misses and False Alarms is at best severely delayed and therefore of little use to the inspector.
Unrealistic Expectations	For more common defects, expectations from training can translate relatively faithfully into practice. However, for very rare defects, expectation may still be unrealistically high after considerable practice.
Isolated Inspection Environment	The hangar and even the shop inspection environment are typically noisy, social and distracting. Both noise and social interaction and even some forms of distraction have been found to <u>improve</u> vigilance performance in laboratory tasks.

Table 4. Comparison between attributes of vigilance tasks and aircraft inspection tasks