Fatigue Effects in Fluorescent Penetrant Inspection

Colin G. Drury, Brian Green, John Schultz and Elizabeth L. Henry
University at Buffalo, Department of Industrial Engineering
438 Bell Hall, Buffalo, NY 14260
drury@buffalo.edu

Abstract

An experiment was performed to test the effects of fatigue factors on performance and stress in a high fidelity simulation of a fluorescent penetrant inspection of aircraft turbine engine blades. Five factors found in year 1 to be potentially related to inspection fatigue were tested in a mixed experimental design using 80 participants recruited from the local community. Many main effects and interactions were significant in performance analyses, although the vigilance decrement was not a strong effect in this task.

Introduction

This report follows the 2004 report on the potential for fatigue in repetitive inspection tasks in aviation. The motivation for the work remain unchanged: 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. A common thread in all these 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. Our earlier paper (Saran, Schultz and Drury, 2004) examined 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. The analysis presented in that paper formed the basis for the design of the current experimental study of temporal factors in aircraft inspection. The first experiment, reported here, was designed as a factorial experiment to find the significant interactions among key variables, so that subsequent experiments could explore these in a more parametric manner. Only performance results are presented here as they are of the most immediate impact in aviation inspection and in the design of parametric experiments for the final project years.

Methodology

Participants: A total of eighty participants completed this study. Several participants were removed from the analysis for the following reasons: 1) inadequate computer skills, 2) not completing all three sessions, 3) not following inspection procedures as outlined in the training. Participants were selected based on calls from people in the local community who responded to a newspaper advertisement. All participants: 1) had previous industrial experience or were currently university students, 2) had sufficient computer skills to use a mouse and a keyboard, 3) had sufficient color vision, 4) were available for testing at sessions both at 9am and 3am, and 5) could complete one training session and 2 experimental sessions within about one weeks time. Participants were paid $15 per hour for their participation as well as a $20 bonus after completing the third session.

Materials: Two Dell Pentium 4 computers were used to run a simulation created in Visual Basic. Pictures of jet turbine blades with spots of fluorescent penetrant super-imposed on them were presented to the participants. Sixty-three blades were photographed from 6 views so that the inspector could look at all sides of the blade (Figure 1). The simulation allowed participants to view each of the six

possible views of the blade by rotating 90 degrees in any direction. Visits to aviation FPI facilities were used to help develop the simulation, and to ensure that FPI inspectors found it valid. The simulation included recordings of hangar noise (80db) to recreate the ambient sounds realistic to this process.

In the simulation, targets were considered cracks that were hidden under spots of fluorescent penetrant. Participants were able to use the computer mouse to "swab" the spots of fluorescent penetrate. A defect was considered a spot of fluorescent penetrate that could not be removed by swabbing it with the swab tool. A magnification tool was included in the simulation that allowed inspectors to magnify areas of the blade at 2 times the regular size. This tool was used at the inspector's discretion. To report a crack, inspectors clicked a mouse button that opened a dialog box to write a brief description of the crack, e.g. "front view at upper right corner". The program recorded a Notepad text file which kept a time stamp of each inspector action including: blade numbers, start and stop points of swabbing, blade view changes, and reports. This information made it possible to classify any errors as: 1) if inspectors looked at a blade view where a crack was present, 2) if inspectors swabbed the area over a crack and failed to report it (or reported it accurately), and 3) if an inspector swabbed a crack but failed to report it.

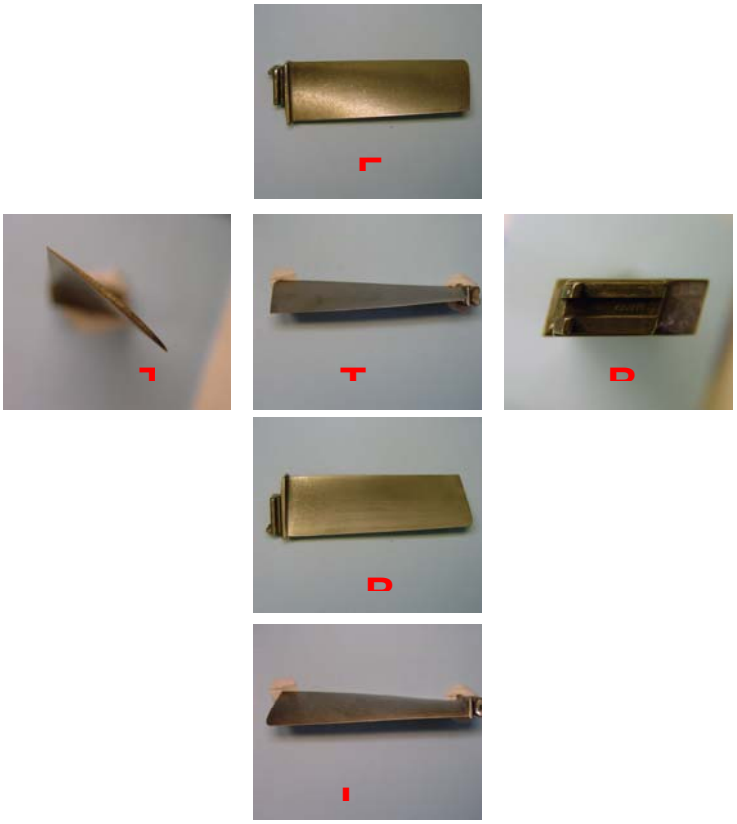

Figure 1: Six views of each blade in the simulation

Design and Procedure: Participants were randomly assigned to code numbers after passing the pre-screening process described above. Four between participants factors were tested: 1) illumination (light/dark), 2) time on task (1 hour or 2 hours), 3) breaks vs. no breaks (break condition consisted of a 3 minute break for every 20 minutes on task), and 4) defect rate (7% or 15%). In the analysis, defect rate was treated as a covariate (mean rate = 0.11) since the number of cracks seen by any participant varied in any 20-min interval depending upon their rate of working. Within-participant factors were of time of day (day/night) and 20-minute period within each session.

All participants completed a training session that included a series of paper and pencil measures of demographics and an informed consent form. Scales from sleep research were used at the beginning and end of the session: the SOFI and SSS sleep scales. Tests of visual acuity, color vision testing, an adaptation of the Folkhard scale, and the Group Embedded Figures Test (given with no time limit) were collected for possible use as covariates.

Training began with a self-paced PowerPoint presentation about the fluorescent penetrant and visual inspection processes. Comprehension questions were included to insure that participants read and understood the material. Next, participants were given a training exercise guided by the experimenters. A tutorial script was read aloud to explain how to use the tools in the simulator (i.e. swab tool, magnify tool, etc). Participants were allowed to practice with, and ask questions about, these tools. The tutorial emphasized that cracks are most likely to occur around the edges of the blades and that the majority of the time spent inspecting should be on that area of each blade. The tutorial included a nine blade practice set with feedback. The experimenters remained present to assure that every participant saw each of the three defects and reported them properly.

After completion of the tutorial, participants completed a computerized mental workload assessment (NASA-TLX) and then completed the SOFI/SSS tests again (creating documentation of pre/post levels of sleepiness). The participants were then instructed about how to use the Actigraph sleep watches and how to complete their written sleep logs. The second day of experimentation (either at 9 am or 3 am) consisted of the SOFI/SSS scales followed by the predetermined experimental condition (between subject factors). Prior to beginning the experiment, participants listened to a brief audio recording of the instruction and were given an optional 3 blade practice set. Upon completion of the simulation program participants completed the NASA TLX workload inventory as well as completing a second SOFI/SSS inventory. The third day of the experiment resembled the second day with the following exceptions. If the second day was conducted at 9am then the third was the

3am condition (and vice versa), and the order of defects was different to prevent some participants from seeing several obvious defects early in the experiment and others being presented with less obvious defects which are more likely to be missed.

Data Handling: The data was coded as a *hit*, *miss*, *false alarm*, or a *correct rejection*. Probability of Detection (PoD) was calculated as the total number of true positives (hits) divided by the number of true positives (hits) plus the total number of false negatives (misses). Likewise, Probability of False Alarm (PoFA) was calculated as total false positives (false alarms) divided by true negatives (correct rejections) plus false positives (false alarms). Speed data came from total blades inspected, mean time to accept a blade and mean time to reject a blade. Each measure was calculated for each 20 minute period to allow for comparisons between participants in the break condition versus those in the no break condition.

Results

Two General Linear Models (GLM) analyses of variance were performed because of the experimental design that included half the participants with a one-hour task and half with a two-hour task. The first ANOVA was of just the two-hour participants. The GLM ANOVA could not calculate all terms because of some missing cells, but the summary of only the significant effects is given in Table 1. The second used all conditions, but separated out the three time blocks: One hour participants in their only hour, two-hour participants in each of their first and second hours. In these analyses, actual defect rate and Run (first, second) were used as covariates, but only Run was significant for the three time blocks ANOVA. Table 2 shows significant results. Note that Run was only significant for speed measures, with a 21% improvement in throughput from the first to the second run.

Over both analyses there were fewer effects on accuracy measures than on time measures. The most consistent effect was of Time on Task, shown in Figure 2 for both speed and accuracy measures. For PoD there was an initial increase followed by level performance in Day conditions, but high initial performance followed by a slow drop and a final end surge at Night. For PoFA there was just a gradual decrease in false alarms over the whole two hours. Finally the speed data showed a steady performance improvement over the two hours.

Both analyses showed an interaction between Light/Dark and Day/Night for Total Blades, with the data for the three time block analysis plotted here as Figure 3. Higher throughput was achieved when the internal lighting matched the external conditions. Another example of a significant interaction is the Breaks/No Breaks X Day/Night for two speed measures in one analysis and PoFA in the other. The Total Blades measure is plotted for the two-hour data in Figure 4. Breaks give higher throughput in Day conditions, but No Breaks is faster at Night. The PoFA data show the same effect with less false alarms where there was better throughput.

Obviously not all interactions can be presented in five pages, and indeed the analysis to TLX and sleep variables will be needed to provide a full picture of the experiment, but there are interesting significant interactions on which to base further experiments.

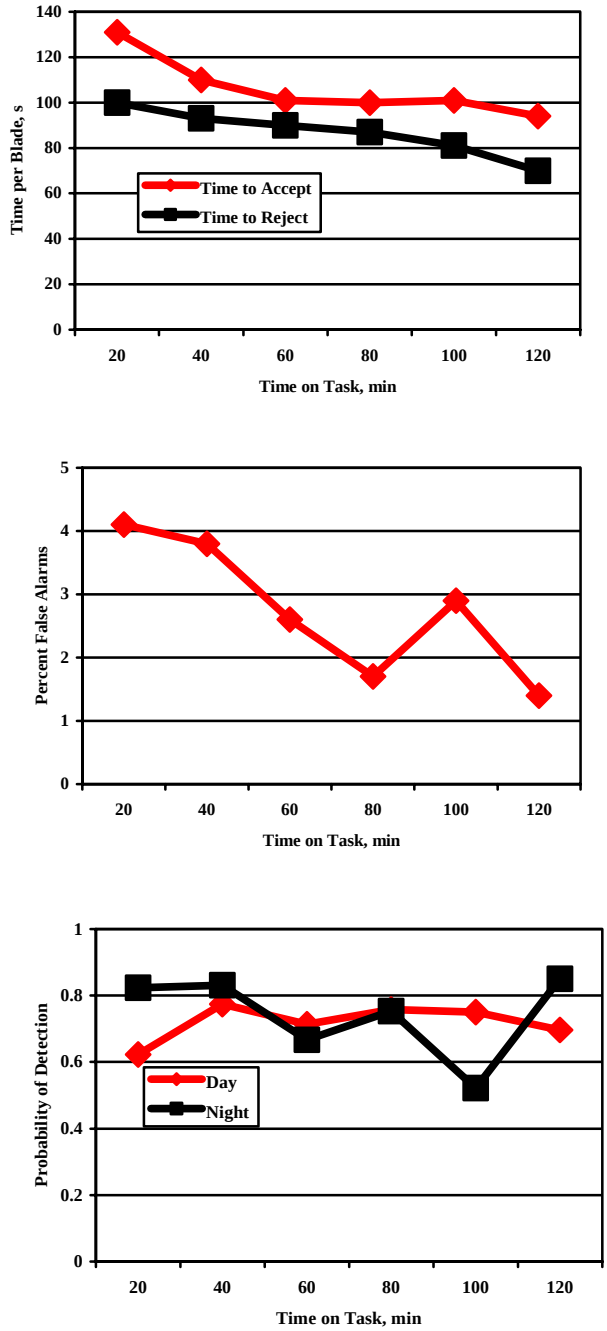

Figure 2: Time on Task effects for 2-hour participants

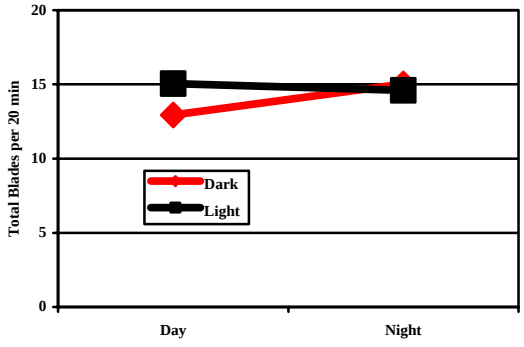

Figure 3: Day/Night X Dark/Light interaction

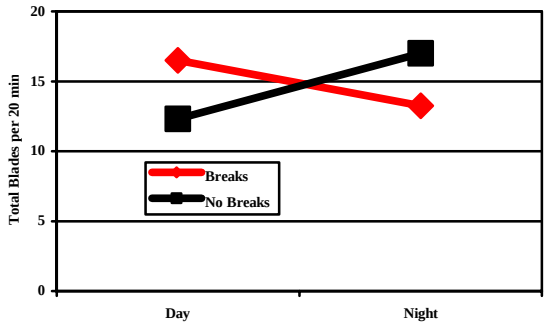

Figure 4: Day/Night X Breaks/No Breaks interaction

Discussion and Conclusions

The experiment reported here was a simulation of one type of repetitive inspection in aviation. It was checked with FPI inspectors who assured us it was at least face valid. The design chosen for this first experiment was a between-participants design with 5 industrially-experienced participants in each of the 16 ($=2^4$) cells of the design. With the addition of a single within-participants factor of Day/Night, this gave a 2^5 design so that many two-way interactions could be measured. The design ensured that there were no unwanted carry-over effects between conditions (except perhaps Day/Night which was given in a random order) and so was safe if less powerful than a within-participants design. The intention is to use the results of this experiment to design a more focused set of parametric experiments to measure more explicitly the effects of significant main effects and interactions across more levels of these variables. For example, now that we have established that Day/Night interacts with both Breaks/No Breaks and Light/Dark, we can include more Break durations and Light levels in future experiments provided we also perform the tests at Day and Night.

On a practical level, we concluded that there were the expected significant individual differences and that these can interact with some variables, e.g. Day/Night. Thus we can expect night working to affect inspectors differentially, so that not all may be suitable for Night work. We also showed that across a long period of continuous inspection (up to two hours), performance measured by PoD may change differentially with Day/Night. The other dependant variables, PoFA and speed measures, all appear to improve with Time on Task. Any vigilance decrement may well be limited to Night conditions. In this combined search and decision task (Drury, 2001) with performance times measured in minutes rather than milliseconds per blade, vigilance decrement does not appear at the same magnitude as in typical laboratory vigilance tasks (e.g. Parasuraman and Davis, 1977). Horowitz, Cade, Wolfe and Cziesler (2003) have already reported that the search function may not show the classic vigilance decrement phenomenon shown by primarily decision tasks.

Finally, there appeared to be interesting effects of working in light conditions vs. dark conditions, where the best condition for throughput was the one with the best match to the outside light levels, i.e. Day vs. Night. This may have practical implications for setting light levels in FPI inspection, which is always carried out at low levels to illumination so as to be able to view the fluorescence under UV lighting.

Acknowledgement

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

References

Drury, C. G. (2001). Human Factors and Automation in Test and Inspection, In G. Salvendy, *Handbook of Industrial Engineering, Third Edition*, Chapter 71, John Wiley & Sons, New York, 1887-1920.

Horowitz, T. S., Cade, B., Wolfe, J. M. and Cziesler, 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.

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.

Saran, M., Schultz, J. and Drury, C. G. (2004). Temporal effects in aircraft inspection: what price vigilance research. *Human Factors and Ergonomics Society 48th Annual Meeting*, New Orleans, LA, September 20-24. 2004.

Source	PoD	PoFA	Total Blades	Accept Time	Reject Time
Time (6 x 20 min intervals)		P = 0.019	P = 0.004	P < 0.001	P < 0.001
Subjects	P < 0.001		P < 0.001		
Light/Dark*Day/Night			P = 0.034		
Breaks/No Breaks*Day/Night			P = 0.001		P < 0.001
Day/Night*Time	P = 0.008				
Day/Night*Subjects		P < 0.001	P < 0.001	P < 0.001	P < 0.001
Breaks/No Breaks*Day/Night*Time				P = 0.048	

Table 1: Significant ANOVA results for the two-hour participants only

Source	PoD	PoFA	Total Blades	Accept Time	Reject Time
Run			P < 0.001	P = 0.003	P < 0.001
Light/Dark*Breaks/No Breaks				P = 0.036	P = 0.006
Light/Dark*Breaks/No Breaks*Time Block		P = 0.013			P = 0.025
Subjects	P < 0.001	P < 0.001	P < 0.001	P < 0.0	P < 0.001
Day/Night			P = 0.030		
Time			P = 0.001	P = 0.001	
Light/Dark*Day/Night			P < 0.001		
Breaks/No Breaks*Day/Night		P = 0.025			
Time Block*Day/Night		P = 0.001			
Time Block*Day/Night*Time	P = 0.014				

Table 2: Significant ANOVA results for the three one-hour time blocks