Proceeding of the 30th European Association for Aviation Psychology Conference, Sardinia, Italy, 24-28 September 2012

The Evaluation of Pilots Performance and Mental Workload by Eye Movement

*Wen-Chin Li[1], Fa-Chung Chiu[2], Ka-Jay Wu[3]

[1, 2 &3]Psychology Department, National Defense University,
Beitou District, Taipei City 112, Taiwan
Tel: +886-2-28962979 ext. 11, Fax: +886-2-28964426
w.li.2002@cranfield.ac.uk

E-mail: w.li.2002@cranfield.ac.uk

ABSTRACT

Pilots make important decisions often using ambiguous information, while under stresses and with very little time. During flight operations detecting the warning light of system failure is a task with real-world application relates to measurement of pilot's performance and eye movement. The demand for a pilot’s visual and situational awareness in multiple tasks can be detrimental during pilots’ mental overload conditions. The purpose of this research is to evaluate the relationship between pilot’s mental workload and operational performance by eye tracking. Collecting eye movement data during flight operations in a virtual reality of flight simulator provided useful information to analysis participants’ cognitive processes. There were 36 pilots participated in this research, the experience of flight hours between 320 and 2,920, the range of age between 26 and 51 years old. The apparatus included Applied Science Laboratories (ASL) eye tracking, IDF flight simulator and NASA_TLX for data collection. The results show that pilots with high SA detecting hydraulic malfunction have shorter total fixation duration on Air Speed Indicator and longer total fixation duration on Altitude Indicator, Vertical Speed Indicator, Right multi-display and Left multi-display compared with pilots without detecting the signal of hydraulic malfunction. Pilots’ total fixation time on Integration Control Panel, Altitude Indicator, Attitude Indicator and Right Multi-display, and pilots’ subjective rating on NASA-TLX effort dimension for the mission of close pattern have significant relationship with pilots’ performance on the operational time for completing the tactic mission. Experienced pilots operate aircraft familiar with monitoring Airspeed Indicator and kinetic maneuvering result in less fuel consumption. This study could provide guidelines for future training design to reduce pilots mental workload and improve situational awareness for enhancing flight safety.

Keywords: Eye Movements, Human Errors, Mental Workload, Situational Awareness

INTRODUCTION

The human is an essential factor for the success or failure of flight operations. Military pilot needs to be able to perform competing tasks in a highly dynamical, threatening environment and, at the same time, have to maintain a high level of attention to different system instruments and environmental conditions (Wickens, 2002). The natural limitation of cognitive processes and the vast number of (often parallel) tasks are reasons for increasing critical stress for pilots. Under high pressure of flight mission and dynamic aircraft maneuvers in the tactic missions, the pilot faces additional difficulties and increased workload during multi-missions and adverse environmental conditions (Ahlstrom, 2003). Because workload can negatively affect operator performance and increase the probability of operating hazards, researchers have spent a great deal of effort developing measures of operator workload (Averty et al., 2004).

Aside from using self-reported subjective workload ratings as a gauge for evaluating operator's workload, pupillary response has also been proposed as an index of the amount of cognitive processing (Beatty, 1982). Eye movement measurement offers deep insights into human- machine interaction and the mental processes of pilots. Measurements based on different aspects of ocular behavior, such as the number of fixations, dwell time, and the dilation of pupil, have been used to reveal the status of mental workload. There was evidence that increasing in workload could increases dwell time and the frequency of long fixations (Van Orden, Limbert, Makeig, & Jung, 2001). Athe´nes et al. (2002) suggested that workload could increase the error of operation ,and the decreased fixation duration appeared to predict upcoming errors in the auditory task. Peter, Jennifer and Joey (2001) found that experts had significantly shorter dwells, more total fixations, more aim point and airspeed fixations and fewer altimeter fixations than novices. Experts were also found to have better defined eye-scanning patterns. The most importance conclusion reached by Bellenkes et al.(1997) is that scanning difference between novice and experts are correlated with better performance by experts. Experts should have shorter dwell and more fixations than novices on all the instruments. Fox, Merwin, Marsh, McConkie, and Kramer (1996) proposed that experts are more likely to use peripheral vision to process a broader range of visual cues than novices do. Furthermore, there is a close connection between which display item a person is looking at and which item individual is thinking about, as well as between fixation duration and amount of processing (Rayner, 1998). The pattern of acquisition of cue-based information during the performance of a task provides an opportunity to assess the application of distinct cognitive skills. In summarize, the eye movements are useful to reveal the diagnostic information that enables the development of appropriate strategies which efficiently target a particular feature of the performance of a task.

There has been much speculation and analysis undertaken to describe the impact of distributed air and ground operations on the controller and flight crew engaged in flight operations (Endsley, 1997). The pilots have to make decisions and share decisions not only about the management of the airspace, but also about the operating state (the mode of control) of that airspace, the workload may be increased dramatically during abnormal situations and system failures (Weiner, 1989). Workload can negatively affect pilots' performance and increase the error of operation (Athe´nes et al., 2002). Wickens (2002) define workload as the load imposed on the limited information processing resources of the unaided (without assistance of automation) human operator described as the “baseline” or “manual” condition. This load can be imposed from two qualitatively distinct sources, the single task difficulty of the task that might otherwise be automated, and the multitask load in which the baseline (vs. automated) task is performed. Controlling aircraft is also a stressful operation which needs high situation awareness to make risk assessment, might increasing pilot's mental workload. Some studies suggest that workload increases dwell time (Bellenkes, Wickens, & Kramer, 1997). Pilots make important decisions often using ambiguous information, while under stresses and with very little time. Workload has impact to cognitive processes, therefore eye movements may serve as a window into the operations of the situational awareness and reflect the mental state of pilots. This research applies the eye-tracking technology to investigate cognitive effort involved in detecting information presented in stressful situation for measuring pilots' situational awareness. By examining pilots' eye movement's patterns and performance compared with pilots' subjective stress levels by NASA-TLX, it is hope to discover the role of cognitive effort in flight operations for improving the training effectiveness of situational awareness and aviation safety.

METHOD

1. Participants

Participants consisted of 36 pilots recruited from R.O.C. air force pilots. The flying hours is between 320 and 2920 hours, the rank is between first lieutenant and general, the age is between 26 and 51 years old (Table 1).

Table1 Demographical Variables of Participants *(N=36)*

Variables	Group	Frequencies (%)

Age	25－30	29(80.6%)
	31－35	2(5.6%)
	36－40	1(2.8%)
	41－45	2(5.6%)
	Above 46	2(5.6%)
Rank	First lieutenant	9(25.0%)
	Captain	19(52.8%)
	Major	1(2.8%)
	Lieutenant Colonel	5(13.9%)
	Above colonel	2(5.6%)
Qualification	Not combat ready	12(33.3%)
	Combat ready	12(33.3%)
	Two aircraft team leader	4(11.1%)
	Four aircraft team leader	2(5.6%)
	Daytime back seat instructor	1(2.8%)
	Night back seat instructor	3(8.3%)
	Training instructors	2(5.6%)
Flying hours	Under 500	16(44.4%)
	501-1000	11(30.6%)
	1001-1500	3(8.3%)
	1501-2000	3(8.3%)
	Above 2001	3(8.3%)

2. Apparatus

2.1 Eye-tracker

Eye-tracking data were collected using an ASL (Applied Science Laboratory) Mobile Eye that is head-mounted and 76 gram in weight. When combined with an optional head-tracking device and eye/head integration software, the eye tracker can also measure a pilot’s eye line of gaze with respect to stationary surfaces in the environment. The Mobile Eye interfaces either after or during recording operations with a laptop for data processing. It is designed to be durable under a variety of active applications and its light weight is suitable to detect the eye movement when pilot operates the air vehicle under active and dynamic flight task.

2.2 Flight Simulators

This study used a fighter simulator, a dynamic high fidelity trainer that replicates actual aircraft performance, navigation and weapon systems. This simulator provides a realistic representation of the flight management system. The instructors can supervise the participated pilot’s performance and the instrument data from three screens. In instrument fight task, the integrate control panel (ICP)、speed、altitude、attitude、head-up display (HUD) , Left and Right Multi-display provide critical information. Pilots cross check those instrument to maintain the speed、altitude 、heading and position. Therefore, this study set those five gauges as the area of interests (AOIs) to analyze the eye movement data.

2.3 NASA-TLX

The NASA Task Load Index (TLX) is a popular technique for measuring subjective mental workload. It relies on a multidimensional construct to derive an overall workload score based on a weighted average of ratings on six subscales: (1) Mental demand: How much mental demand and perceptual activities you would use; (2) Physical demand: How much is the degree of physical demand; (3) Temporal demand: How much is the degree of time pressure; (4) Performance: How do you feel about the flying time and the performance in flight? (5) Effort: How much difficult do you think? (6) Frustration : How much frustration and disappointment do you feel (Hart & Staveland,1988).

3. Scenario for Flight Simulator

Close Pattern: The initial setting is 7-mile visibility with calm wind and R/W 36L. The aircraft was heading 360° on the center-line of R/W, ALT 2000ft and SPD 300KIAS. At the strip of R/W, maintain Throttle 80%, extended S/B, and set 60~70° bank with 2~2.5 G. Make level at ALT 1500ft on downwind. At this phase, keep parallel with center-line of R/W and cross-check indicator of S/B, heading, ALT and SPD. Follow by the L/G down and maintains SPD at 200 KIAS above. Establish attitude for turning and descending at included angle 45° with R/W at SPD 200 KIAS above and minimum throttle at 75%. Adjust bank to intercept the Final's descending course. Operate flare tenderly is quite important and maintain 10° pitch referring to GUN CROSS after touching for reducing SPD.

4. Procedure

Participants were asked to fly instrument scenario using the simulator, the procedure included: (1) an orientation to the experiment (10 minutes); (2) eye-tracker calibration in the cockpit of flight simulator (5-10 minutes); (3) operate instrument task on flight simulator (10-15 minutes); (4) rate NASA TLX (10-15minutes). Each session was conducted by an eye-tracker operator and a flight instructor. The instructor evaluates pilot's performance base on the flight simulator, and the simulator control panels records the fuel consumption and flight time. Eye-scan patterns, video, verbal protocol data were collected. The amount of fuel consumption reflects the pilot's performance on flight task. There is an active warning light of hydraulic malfunction at the stage of final Approach to evaluate pilots' situation awareness.

RESULT AND DISCUSSION

1. Situational Awareness on total fixation duration and NASA-TLX dimensions

Pilot's situational awareness is the key factor to identify the potential risk during flight operations. The results of t-test show the differences of total fixation duration of AOIs and NASA-TLX dimensions between pilots' with situational awareness and without situational awareness (table 2). There is a significance of Left multi-display (p=.01) and margin significances of Air speed indicator (p=.07), Altitude Indicator (p=.08), Vertical Speed Indicator (p=.06), and Right multi-display (p=.06) at total fixation duration on different instruments (AOIs) of IDF flight simulator. There are no significant differences on the NASA-TLX dimensions for pilots' situational awareness. Table 2 shows that pilots with high SA detecting hydraulic malfunction with longer total fixation duration on Left Multi-display for seeking information to bank the aircraft for landing compared with pilots without detecting the signal of hydraulic malfunction.

Table 2 The differences of Situational Awareness on Total Fixation Duration and NASA-TLX

Dimensions	*M*(*SD*)		*T*	Effect size D	*p*
	Y	N			
Integration Control Panel	.59(.51)	.76(1.04)	-.65	.21	.52
Air Speed Indicator	75.42(30.93)	90.46(25.85)	-1.51	.52	.07
Altitude Indicator	.17(.43)	.03(.09)	1.47	.41	.08
Attitude Indicator	.01(.06)	.02(.06)	-.13	.17	.90
Vertical Speed indicator	.17(.49)	.01(.01)	1.61	.42	.06
Right Multi-display	.64(.85)	.30(.41)	1.62	.48	.06
Left Multi-display	1.42(1.92)	.25(.54)	2.71	.76	.01
Mental Demand	53.64(15.37)	58.21(15.76)	-.86	.29	.39
Physical Demand	37.50(18.95)	39.64(15.87)	-.35	.12	.73
Temporal Demand	49.77(13.49)	56.43(15.50)	-1.36	.47	.18
Performance	57.73(21.20)	52.86(13.40)	.77	.26	.40
Effort	58.64(15.52)	57.14(12.51)	.30	.10	.75
Frustration	60.91(16.67)	61.07(17.45)	-.03	.01	.98

Parenthesis show as Standard Deviations

During flight operations detecting the warning light of system failure is a task with real-world application relates to measurement of pilot's performance and eye movement. The demand for a pilot’s visual and situational awareness in multiple tasks can be detrimental during pilots’ mental overload conditions. While much is known about the outcomes in aeronautical decision-making by pilots' performance (Li & Harris, 2008), less is known about the processes that underlie these differences. There is no significant difference between pilots detecting and non-detecting hydraulic malfunction by NASA-TLX 6 dimensions. The information collected by eye tracking device are objective cognitive processes, on the other side, the information collected by NASA-TLX are subjective results of pilots’ feelings of stress. How pilots’ subjective feeling of stress effect pilots operational performance and eye movement pattern among AOIs are complicated cognitive processes, and need further researches to improve pilot’s situational awareness and quality of decision-making.

2. Total fixation duration and NASA-TLX dimensions to predict the fuel consumption

The indication of fuel consumption was one of criteria for measuring pilot’s performance because the capacity of fuel tank is limited for fighter pilots to complete their tactic missions. Pilot’s total fixation duration on the AOIs and NASA-TLX dimensions are predictors, and the fuel consumption as criteria for conducting multiple regression analysis. The results show as table 3. The percentage of accountable variation R^2 equal to .88, and *Adj* R^2 equal to .77; $F(13, 22) = 5.76$, $p < .001$. There is less fuel consumption while pilots have longer total fixation duration on the Integration Control Panel (p = .001), Air Speed Indicator (p =.000) and Altitude Indicator (p = .003) during close pattern. However, there is more fuel consumption while pilots have longer total fixation duration on the Altitude Indicator, Vertical Speed Indicator, and Right Multi-display. The more fuel consumption, the poorer performance of flight operations. There is no predict effect between total fixation duration and Left Multi-display. There are significances on the NASA-TLX dimensions for predict pilots performance with fuel consumption. Pilots with more mental demand (p = .002) during close pattern consume more fuel. However, pilots feel more efforts (p = .001) and better performance (p = .024) on the dimensions of NASA-TLX consuming less fuel (table 3). Pilots’ total fixation time on Integration Control Panel, Altitude Indicator, Attitude Indicator and Right Multi-display, and pilots’ subjective rating on NASA-TLX effort dimension for the mission of close pattern have significant relationship with pilots’ performance on the operational time for completing the tactic mission. Experienced pilots operate aircraft familiar with monitoring Airspeed Indicator and kinetic maneuvering result in less fuel consumption.

Table 3 Total fixation duration and NASA-TLX dimensions to predict the fuel consumption

	Fuel Consumption (β)	*t*	*p*
Integration Control Panel	-.79	-3.74	.001
Air Speed Indicator	-.63	-4.70	.000
Altitude Indicator	-.85	-3.38	.003
Attitude Indicator	.84	2.71	.013
Vertical Speed indicator	.78	4.78	.000
Right Multi-display	.50	3.04	.006
Left Multi-display	-.07	-.58	.569
Mental Demand	.53	3.62	.002
Physical Demand	-.08	-.42	.676
Temporal Demand	-.18	-1.22	.237
Performance	-.34	-2.43	.024
Effort	-.56	-5.03	.001
Frustration	.29	1.52	.143

R^2 = 88; Adj R^2 = .77;. R^2 =; $F(13, 22) = 5.76$, $p < .001$

CONCLUSION

The present study is using eye-tracking to understand the pilots' decision processes in a simulated flight of close pattern. Eye movements provide numerous clues to underlying cognitive processes pilots encode information, and what information they use or ignore related to the performance of fuel consumption under stress situations. The military aviation has several challenges for pilot's situational awareness training while stressing the military's training resources. Military conducts ab-initio training, it transforms complete novices into minimally proficient and safe pilots. This basic flight training is conducted in formal aviation training division; the qualified pilots are then transferred to operational field units for tactic mission training. This study is necessary for pilots to establish what areas of interests (instruments) appears essential to situational awareness, a dynamic mental workload that incorporates fast-time scan for collecting information from different gauges to deal with changing situations. There is a raising need for conducting further research regarding eye movement pattern and mental workload in the real-time flight operations.

REFERENCES

Ahlstrom, U. (2003). Current trends in the display of aviation weather. *Journal of Air Traffic Control, 45*(3), 14-21.

Athènes, S., Averty, P., Puechmorel, S., Delahaye, D., & Collet, C. (2002). *ATC complexity and controller workload: Trying to bridge the gap.*

Averty, P., Collet, C., Dittmar, A., Athe´nes, S.,& Vernet-Maury, E.(2004). Mental workload in air traffic control: an index constructed from field tests. *Aviation, Space, and Environmental Medicine, 75,* 333–341.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychological Bulletin, 91,276-292.

Bellenkes, A. H., Wickens, C.D. & Kramer, A.F. (1997). Visual scanning and pilot expertise: The role of attentional flexibility and mental model development. *Aviation Space and Environmental Medicine, 68,* 569-579.

Endsley, M. R. (1997). *Level of automation: Integrating humans and automated systems.*

Fox, J., Merwin, D., Marsh, R., McConkie, G., & Kramer, A. (1996). Information extraction during instrument flight: An evaluation of the validity of the eye-mind hypothesis. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting.*

Hart, S. G. and Staveland, L.E. (1988). Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.), *Human Mental Workload.*

Li, W-C., & Harris, D. (2008). The evaluation of the effect of a short aeronautical decision-making training program for military pilots. *International Journal of Aviation Psychology, 18*(2), 135-152.

Peter, K., Jennifer, S., & Joey, H. (2001). Comparison of expert and novice scan behaviors during VFR flight. *The 11th International Symposium on Aviation Psychology.* Columbus: The Ohio University.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin, 124*(3), 372.

Van Orden, K. F., Limbert, W., Makeig, S., & Jung, T. P. (2001). Eye activity correlates of workload during a visuospatial memory task. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 43*(1), 111-121.

Wickens, C. D. (2002).Situation Awareness and Workload in Aviation. *Psychological Science, 11*(4), 128–133.

Wiener, E. L (1989). *Human Factors of Advanced Technology ("Glass Cockpit") Transport Aircraft.* (NASA Contractor Report 177528). Moffett Field, CA: NASA Ames Research Center.