Eye-tracking and Pupilometry to Assess Flight Simulator Training

Abstract. Airline pilots are assumed to be capable of performing their work in both normal and difficult circumstances. However, as aviation safety has improved, pilots have been increasingly implicated in aviation accidents with pilot error being identified as causal more frequently. Flying an aircraft within an airline is a complex task. A significant amount of research has been published over the past 4 decades concerning the mechanisms that underlie the expert performance of complex tasks. Affect has been identified as is a key determinant of performance. Affect has been shown to influence the development of higher-order mental abilities such as situation assessment and decision-making Simulators play an invaluable role in training and assessment. They have been used to train pilots how to control and use their emotions to facilitate optimal performance. In aviation, pilots have long been taught how to stay calm under difficult circumstances and they have been able to do this through the opportunity simulators provide to practice building confidence. The underlying premise is the pilot's confidence can be strengthened through training which provides the opportunity to experience mastery under high stress conditions. Therefore, affective state is a key variable for investigation in the development and evaluation of aviation simulation training. This project used eye tracking and pupilometry to assess pilot's affective state while undertaking normal and difficult tasks in a flight simulator. Fixation duration and saccade rate corresponded reliably to pilot self-reports of affective state, while pupil size and saccade amplitude did not show a strong comparison to changes in affective state. The implications from the data collected are discussed in terms of using eye-tracking technology to objectively measure pilots' affective levels during simulation training.

1. Introduction - Aviation context/background

+++TIM – please add here – as discussed this will be dependent on the journal you feel we should target..

Despite the aviation industry's success in achieving continual improvements in safety, the identification of Non-Technical skills (NTS) as a root cause in many accidents has remained a problem. It is clear that the industry is making attempts to address this problematic situation, as seen by the push to implement NTS training and assessment to counter the unfortunate effects of

NTS deficits. However, the assessment of NTS is still seen to be challenging. No model or framework has yet been devised that is able to be effectively used in practice.

2. Simulation Training Assessment

The aviation industry has used flight simulation one of its cornerstones of training for many decades. Despite the popularity of simulators to deliver training in NTS methods for evaluating their use are largely limited to subjective self-reports or final action analysis. It is widely accepted that, to address remaining issues of human error, there is more to be done in the area of assessment. Research has identified that current simulator training assessment techniques need to be expanded to incorporate more complex analyses of operator error (Lenne, Regan, Triggs & Haworth, 2004).

Performance measures currently provide information about the end result of training, such as the number of correct actions and response times. The continuing lack of research into new methods of simulation assessment still leaves it open to criticism. The Australian and International Pilots Association have highlighted the lack of a "scientific basis" and recommend that "at the present time this new synthetic training is not proven and it needs to be done carefully and cautiously"(ABC, 2007). The development of objective measures of cognitive processes will significantly impact the quality and effectiveness of simulator-based training.

This paper introduces a new approach to assessment of trainee performance in simulation. Flight simulators have been used to train pilots how to control and use their emotions to facilitate optimal performance for many years. Yet this use of simulation has not been formally recognized or measured. If training programs are going to maximize the strengths of human affective states to provide more effective training and increase the likelihood of success in professional settings characterized by high task demands, then methods of assessment which are objective yet go beyond the limits of technical skills must be further developed. Simulation training works on the premise that a professional's **confidence** in his/her affective skills can be strengthened through training whereby they have been given the opportunity to experience mastery while in a highly aroused state, for example they have been able to work through fear and successfully translate it into high goal performance.

Mavin (2010) through an investigation of the criteria used to assess pilots' performance examined individual check captain's personal evaluation criteria. Significantly, check captains identified that the first element of the decision making process was having the **confidence** to be able to make decisions. A candidate could fail if they were viewed as being unable to take the final step in taking responsibility or demonstrated uncertainty in decision-making. The ability to take charge of the situation was the first important step for command. The pivotal and essential

role affective skills play in this ability is demonstrated in the following quote: *I've always believed that flying aero planes is nothing more than an exercise in self-confidence* (Mavin, 2010 p100). Affective states such as fear, hesitancy, agitation, impatience and nervousness have no role to play.

3. Affective Skills

The work of Damasio (1995) in neuroscience has convincingly demonstrated the interdependence between emotions and activities previously considered to require only rational thought, such as problem-solving and decision-making. In Lazarus's (1966) transactional model of stress, people appraise task demands in terms of their own resources to cope. A professional needs to perceive they have the individual affective resources to cope with their workload. This perspective support the conceptualization of affect as a further resource that can act as an input to support cognitive processing. While the role of affect has been increasingly recognized in psychology, few theoretical models investigate how affect contributes to whether a task can be successfully completed under high stress conditions. Primarily affect has been conceptualized as an output of too many task demands such as feelings of anxiety being the result of work overload.

Emotion however is more than a response to job demands. It prepares individuals to respond to eliciting stimuli by coordinating a system for responses: anger prepares body to fight, and fear prepares it for flight (Matsumoto & Wilson, 2008). Affective skills are a concept that has been recognized in the field of professional sport for some time. Affective Intelligence is a subscale of a tool developed to measure 'mental toughness' in athletes. Affective intelligence recognizes that optimal performance is dependent on an athlete being able to remain in control of their emotions, no matter what obstacles they encounter and be able to be actively able to bring their emotions into play to facilitate optimal performance (Gordon & Gucciardi, 2010; Gucciardi & Gordon, 2009; Connaughton et al., 2008).

Some research has reported that motivation to achieve can only be positively influenced by emotions that are perceived to be 'positive'. Positive affective experiences facilitate the retrieval of positive self- and task-related information, whereas negative affective experiences facilitate the retrieval of negative self- and task-related information (Linnenbrink & Pintrich, 2002; Daniels et al, 2009). However, unpleasant activated states have also been shown to be motivating during goal striving because they lead individuals to mobilize effort (Ballard, 2009). There are likely to be situations in which too much unpleasant activation is detrimental to motivation. There will be individual differences in how much unpleasant activation is too much before individuals start to lose motivation. It has always been a goal of simulation training to gradually expose trainees to increasing levels of stress until they are able to tolerate a significantly higher stressful state before they start to withdraw effort. The goal of training is to provide the

opportunity for trainees, through repetition, to prove to themselves that they can maintain optimum performance under highly stressful conditions and that they do not need to withdraw but they can use their emotions to support their own performance.

If anxiety and fear are not controlled and used to support performance they inevitably degrade and undermine performance both at a technical and cognitive skills level. In research which evaluates simulator training of cognitive skills such as situation assessment and decision-making, affective state has therefore become a key variable of interest (Tichon, 2010).

4. Training Affective Skills

Objective judgment and decision-making are crucial pilot skills which rely on higher-order cognitive pathways to process. Emotions associated with stress such as fear and anger unless they are controlled can overwhelm prefrontal cognitive processes (Russo et al., 2005). In aviation, pilots have long been taught to practice staying calm (Lehrer, 2009). The ability to control emotion varies across pilots and often relates to experience. An expert, in the domain of affective skills, can be viewed as having developed over time and through experience the level of skill required to control his/her emotions during high stress operations.

Simulation training for both pilots and other professionals working in high stress environments has always, in addition to training technical skills, been teaching individuals how to control and use their emotions to facilitate optimal performance. Simulation training works on the premise that a professional's confidence in his/her affective skills can be strengthened through training whereby they have been given the opportunity to experience mastery while in a highly aroused state, for example they have been able to work through fear and successfully translate it into high goal performance. Additionally, emotions help individuals identify associations between event-based triggers and behavioral consequences so as to be better equipped in similar situations in the future (Matsumoto & Wilson, 2008). As a resource that supports identification of prior experiences, emotions are clearly a resource that can be brought to bear by individuals either in the training or operational context.

5. Measuring Affect

Research is currently constrained by the necessity of relying on subjective self-report measures to assess a trainee's affective state. Subjective measures of stress include interviews, open-ended and scaled response questionnaires. "Subjective measures are particularly useful as a means to assess person-environment 'fit' based on direct subjective report of stressful experience."(reference). However, great caution should be exercised in using subjective self-report measures to assess stress in other domains and purposes. For example, if a researcher is

interested in evaluating stress produced by a particular situation, subjective measures may be misleading. An environment perceived as stressful by one individual may not be perceived as stressful to another," (Redden et al, 2004; p554).

While technical performance can be measured objectively, to date affect cannot. The aim of the current project was to identify a reliable set of objectively measurable features and responses that highly correlate to target emotions via eye-tracking during flight simulation.

Simulators provide the advantage of being able to replicate high stress operational environments safely and realistically. Within these replicated environments the opportunity arises to test and develop more advanced measures. Increasingly it is being recognized that capturing behavioral data from participants such as facial expressions or head movements may be a more accurate representation of how and what they feel and a better alternative to self-report questionnaires that are not only subject to bias but also interrupt participant's affective-cognitive processes (Reynolds & Picard, 2004; Ahn et al 2010).

Feature extraction via pupilometry as an objective measurement tool for affect recognition has been garnering a significant increase in interest (Liao etal, 2006; Bailenson et al., 2008). Stress has been a key variable of interest with investigations revealing features that are potentially sensitive and robust to stress (Liao et al.,2006). Research has identified a number of parameters relating to the eyes and their movement which are influenced by affective state and specifically state anxiety. Blinks, saccades and pupil dilation have all been reported as varying systematically with manipulations to stress or measured anxiety levels. Chapman et al. (1999) conditioned subjects to expect an electrical shock to their finger-tip, producing raised levels of anxiety and stress. During periods shortly before a shock, the team recorded increases in the cycling of pupil size (i.e. variability in pupil size over time) which they attributed to a rise in anxiety levels. Partla and Surakka (2003) exposed subjects to images designed to produce positive or negative arousal in subjects and reported changes in the maximum, short-term pupil dilatory response which they termed PSV (Pupil size variation). Some tentative links between eye-movement and affective state have also been reported for subjects observing static faces portraying a range of emotional expressions. (Susskind et al. 2008). In particular, the authors report increases in peak saccade velocity in response to fearful expressions. Perhaps one of the more widely investigated measures is blink rate. The general consensus is that blinking increases as anxiety levels increase (e.g. Harrigan & O'Connell, 1996), however the opposite result has also been reported (Liao et al. 2006).

The research reported here makes up the second step in a two part study. Two pilot studies where designed to commence an investigation of commercially available monitoring equipment to provide a reliable set of affect data. The initial pilot was conducted in a lower-anxiety driving simulation. The expected physiological responses to increasing anxiety where significant enough to be clearly captured by the initial pieces of objective monitoring equipment trialled in Pilot

One (Tichon et al 2011). The study reported here is Pilot Two which was expected to generate higher levels of stress and anxiety as the simulator exposure is a real-world training scenario.

6. Method

Pilot Study Two was conducted in January 2011. This pilot was undertaken at a flight training facility located at Aviation High, Hendra, Queensland. Eye responses were recorded while participants undertook a flight scenario designed to become increasingly difficult throughout the 30 minute flight. Before commencing and after completing the training scenario, the participant was required to fill in a questionnaire assessing their affective state.

6.1 Participants

Volunteers were recruited from among currently enrolled undergraduate students of the Griffith University School of Aviation. All had prior flying experience. Prerequisites for participation were normal or corrected-to-normal vision. Twelve subjects took part in the study ranging in age from 26 to 51 years.

6.2 Apparatus

6.2.1 Simulator

The experiments were conducted in a GeoSim Cockpit style Fixed Wing Synthetic Trainer (see Figure 1).

The simulator is able of replicating, to a varying degree of fidelity, a number of aircraft, including single and dual engine piston aircraft, and larger turbo prop. The simulator has an enclosed fiberglass shell replicating a small aircraft cockpit. Main controls include dual yoke controls, rudder and throttle quadrant. Other secondary controls and systems are made available, including trim wheel, radio's and a selection of switches for operating lights, beacons, and other associated systems. More sophisticated controls include magneto switch, starter button, and cowl flap controls.

The simulator allows for the adjustable seating for two pilots.

6.2.2 Virtual Scene

Pilot instruments were confined within the main fiber glass shell, and located in front of the pilot. The computer utilizes the 'Flight Simulator X' program, and for this study, was replicating analogue instruments consistent with a Cessna 172 aircraft. The simulator allows for an outside visual system. This system consisted of an over head projector, projecting an image on a large 120" screen using a high quality LCD laser projector.

Once all measuring equipment had been safely attached, a 30 minute simulated flight was conducted. All participants were given 5 minutes to familiarize themselves with the aircraft controls and systems during taxiing undertaken but the trainer. Students were then required to take over the controls and conduct the following maneuvers:

1. Manual takeoff

2. Level off at 1000 feet

3. Thirty degree turns, in both directions

4. Climbing and descending turns and 30 degrees, at 80 knots, at 500 feet per minute (for the level of flight experience all participants found this maneuver extremely difficult to complete)

5. Landing on the runway they departed

6.2.4 Feature Extraction via Eye tracking and Pupilometry

Eye movement and pupil size data were be recorded using an SR Research EyeLink II head-mounted tracker, performing binocular 500Hz sampling. The eye tracker is connected to a dedicated PC (the EyeLink 'host PC') that has custom hardware, on a PCI card, to communicate with the eye tracker and collect its data. The EyeLink II system also requires a second "display PC" which manages interaction with the subject using the eye tracker. The host PC effectively acts as a peripheral to the display PC.

Data collected by the eye-tracker includes relative pupil size, eye position and movement velocity. The final data includes some derived "events": blinks, fixations and saccadic movements that are calculated by the eye-tracker host before it sends the data file to the display PC. This data file is accumulated on the host PC and transmitted at the end of a session.

The eye tracker requires an initial calibration procedure for each participant. During this the subject is fitted with the head harness, and then responds to fixed stimuli, enabling the eye

tracker software to calculate how to extrapolate from eye movement and position to gaze direction for the individual subject.

Feature measures which were derived from the eye-tracking data were: Blinking Frequency (BF), Average Eye Closure Speed (AECS), Percentage of Saccadic Eye Movement (PerSac), Gaze Spatial Distribution (GazeDis), Percentage of Large Pupil Dilation (PerLPD),

6.2.5 Survey

A self-report survey was used in this initial stage to gain an indication from participants of those emotions that were engendered by the test simulation. A suitable general measure of emotional states was ascertained to be the Multiple Affect Adjective Checklist (MAACL-R) which has been extensively used in the investigation of the impact of stress on psychological functioning (Hunsley, 1990) and is currently in use in simulator training evaluations. In studies of acute stress, the Army Research Laboratory (ARL) has found that temporary stress effects such as anxiety, depression and hostility are revealed by the Multiple Affect Adjective Checklist-Revised (MACCL-R). The checklist consists of 132 adjectives that comprise five primary subscales (anxiety, depression, hostility, positive affect, and sensation seeking). The checklist can be completed in approximately five minutes (Redden et al, 2004). The drawback of the scale is that it more easily achieves independence between positive and negative affect rather than between two dimensions of negative affect, however the goal is to use the items as a starting point from which to commence the research enabling later theoretically derived dimensions of affect to be identified enabling reliable differentiation.

6. Results

The categories investigated were:

- **#blinks**: total number of blinks during trials in which anxiety was reported.
- **%sacc**: percentage of time spent making saccadic eye movements during trials in which anxiety was reported.
- **sacc amplitude**: average amplitude of saccades during trials in which anxiety was reported.
- **sacc vel**: average velocity of saccades during trials in which anxiety was reported.

- **pupil size variation**: maximum pupil size change over a 4 second window relative to average pupil size over the past ten seconds. By using the running average of pupil size we could discount the effects of pupil dilation in response to changing light levels and other extraneous factors.

7. Discussion

8. Conclusion.

9. Future Research: Linking Affect to Perceptual Processing

People can learn to distinguish between task-relevant information and task-irrelevant information (Sohn, Douglass, Chen & Anderson, 2000). Through eye-tracking data, learning should be reflected in the pattern of attention distribution or eye fixation. In simulation training we want to know what is happening with participant's perceptual processes at specific decision points within the scene and the subsequent impact of evoked emotion. In successful training participants should learn to pay more attention to on-task regions relative to off-task regions (Sohn et al, 2000). If the training is properly designed, eye-movement data should show that users look at irrelevant regions less and less as they practice more.

In future research we propose to monitor affective state while simultaneously tracking people's perceptual processing in order to examine the impact of stress at different decision points. The eye tracking system used in the current study can also record the scene observed by the participant via a head-mounted camera. Eye tracking data is rendered onto the camera's video stream as a small fixation cross and this video is automatically synchronised with the eye movement data. This allows offline matching of salient visual events (e.g. obstacle appears, target acquired) with the physiological data.

 It is expected that high affect levels of negative may negatively influence visual processing of cues. For example when angry, a person may focus for less time on an important cue they had

spent far longer visually processing during an earlier training session. The ability to synthetically evoke intense affective states during a complex cognitive task allows accurately timed measurement of user responses and will help to answer such questions as is there an optimal degree of affective intensity which supports the laying down of a long term memory trace but over which intensity level cognition degrades? Such insights will assist the development of a model to guide use of eye movement monitoring to evaluate high affect training in virtual environments.