Detecting Fatigue in Commercial Flight Operations Using Physiological Measures

Lisa C. Thomas

Kimberly Craig

Christopher Gast

Robert Grube

Mike Muhm

See next page for additional authors

Authors
Lisa C. Thomas, Kimberly Craig, Christopher Gast, Robert Grube, Mike Muhm, and Emma Romig

Detecting Fatigue in Commercial Flight Operations using Physiological Measures

Lisa C Thomas
Kimberly Craig
Christopher Gast
Robert Grube
Mike Muhm
Emma Romig
Boeing Commercial Airplanes
Seattle, WA

The purpose of this study is to determine whether any technology exists to unobtrusively, reliably, and accurately detect symptoms of fatigue in real time before fatigue affects performance. Airline pilots are fitted with a variety of physiological measurement devices (e.g. EEG, blink rate, etc.) that have been demonstrated in the literature to be related to fatigue. Each crew of two pilots performs simulated gate-to-gate flight operations under rested and fatigued conditions, during which physiologic and performance parameters are continuously monitored. In addition, audio, video, and simulator data are recorded for post-session evaluation. Ultimately, if one or more technologies proves effective, we can incorporate it into the flight deck for real-time fatigue detection capability as part of a larger fatigue risk management system. The usefulness of this type of approach extends beyond the commercial flight deck to any work environment that requires multi-shift or other non-traditional scheduling.

"In the operational environment, sleepiness-related performance deficits are typically the result of interactions involving multiple factors such as recent and long-term sleep-wake history, the circadian rhythm of alertness, effects of time on task, cognitive workload, and individual differences in resilience and sensitivity to sleep loss. Such factors should be taken into consideration when devising operational work-rest schedules and when assessing individual[s] who complain of deficits in alertness and performance" (Balkin, 2011).

For any fatigue risk management program to be effective, it is important to gain better understanding of all of the various factors that contribute to fatigue and, more importantly, when the effects of fatigue may negatively affect the pilots' abilities to cope with those situations. A tired pilot may not always make mistakes and in fact may rarely make mistakes, but he is at greater risk in all situations than an alert pilot. If it is possible to quantify how each factor contributes to the experience of fatigue, insight may be gained into how best to mitigate those factors and/or the resulting fatigue.

We are particularly interested in how the differences in the types of commercial flight operations may affect fatigue in different ways. Long-haul pilots must deal with circadian disturbances with time zone transitions, which result in disrupted sleep schedules and ineffective recovery periods (Bourgeois-Bougrine et al, 2003). Even when they have opportunities to sleep on the aircraft during longer flights, it is not always restful given the noise, temperature, lighting, and other factors which reduce the quality of sleep (Rosekind, Miller, Gregory, & Dinges, 2000). For short-haul pilots, relatively high workload contributes to fatigue (Bourgeois-Bougrine et al, 2003); "because takeoffs and landings are extremely task-intensive, it logically follows that a flightcrew member who has performed six sets of takeoffs and landings will be more fatigued than the flightcrew member who has performed only one takeoff and landing" (FAA, 2011). If fatigue has different causal factors and different manifestations, our ability to measure and even predict the effects for various types of operations may aid the design of more effective mitigation strategies.

Several studies have examined the physiological symptoms as well as performance of fatigued and rested pilots during real operations (Thomas et al, 2006; Samel, Wegmann, & Vejvoda, 1997; Wright & McGown, 2001; Signal, Gale, & Gander, 2005). Thomas and his colleagues collected data from crews who provided sleep/wake diary, actigraphy, psychomotor vigilance task (PVT) performance, and subjective ratings during four international flights. They found that the amount of sleep obtained by pilots in the 24 hours preceding the flight had significant effects on both self-reported fatigue levels as well as PVT performance; specifically, that less sleep resulted in higher fatigue ratings and slower response times in the PVT task.

A review by Mathis & Hess (2009) recommends that researchers interested in measuring fatigue and sleepiness create “a battery of subjective and objective tests to answer a specific question in order to achieve the most appropriate description,” since each test measures a particular aspect of a particular factor and any one test run in isolation may misrepresent (or miss) what is happening physically, physiologically, or cognitively on a larger scale.

Physiological and performance measures of fatigue and sleepiness

There are several strategies for detecting fatigue and/or sleepiness, and determining fatigue effects on performance of tasks of varying cognitive or physical effort. Objective physiological measures of physiological responses related to fatigue include brain activity, eye movements and eye-closure related measures, heart rate related measures, actigraphy, postural measures, and voice/speech analysis. Some of these measures are more intrusive than others (devices that require electrodes or are worn on the body), which limit their usefulness in an applied setting. A number of companies are marketing devices that are far less obtrusive to track and detect fatigue in situations where fatigue effects could be potentially harmful, such as driving, long-haul trucking, shipping, and aviation, with varying levels of success and acceptance by the targeted user community.

Objective performance measures of fatigue effects include performance on PVT and other stimulus-response tasks, tasks along the cognitive effort continuum (e.g. simple tasks such as basic addition or word games through complex problem solving), pattern-matching and monitoring tasks, and decision-making. In some cases, the performance measured is the subject’s primary task (e.g. flight path deviations made by a fatigued pilot) and as such is operationally relevant; however it can be difficult to determine when fatigue is having an effect if the task does not lend itself to purely objective standards (e.g. was the landing really well done, good, or just acceptable) or there are other factors contributing to a change in performance (e.g. is a pilot’s delay in setting the flaps due to fatigue or a competing task). In other cases, the tasks are somewhat arbitrary and are exclusively used as a fatigue detection metric and serve no operational purpose (e.g. the PVT), but they tend to have clearer, more explicitly defined parameters of good/fatigued performance.

Subjective measures of fatigue effects include a variety of self-report rating scales and sleep diaries, where people indicate the frequency and duration of their sleep periods and assign ratings to the quality of sleep obtained. Subjective measures are likely to be less reliable compared to objective measures, since they are more heavily influenced by those factors that mask fatigue or sleepiness, as noted in studies that show low subjective ratings of fatigue (subject reports feeling sufficiently alert) but clear fatigue effects in objective measures (subject’s PVT scores are slower than their typical “alert” scores).

Current Study

The current study is intended to evaluate a variety of physiological measurement devices and other sensors embedded into a flight deck and used in the context of commercial flight deck operations. Data from these tools and from flight performance will be analyzed to determine whether one or more of these devices can be used to effectively and reliably detect fatigue, and possibly even predict performance decrements due to fatigue effects.

We have selected a wide range of embedded physiological devices that have been validated by prior research to effectively detect symptoms of fatigue and/or sleep deprivation. Data from these devices will be collected and evaluated to determine whether fatigue effects and symptoms can be effectively detected in the context of long and short haul flight operations. We will also be collecting performance data based on pilot interaction with the flight deck, flight performance as evaluated using a Flight Operations Quality Assessment (FOQA) analysis, and decision-making as rated by a panel of subject matter experts who observe the pilots and rate their actions and decisions against a set of criteria or expectations. FOQA algorithms and criteria can be applied to evaluate flight performance (e.g. deviations from the planned flight path, quality of landings, etc).

We will compare all our physiological measures against subjective ratings (KSS, SP) and objective (PVT) measures that are well-established measures of fatigue. In addition, we will compare the subjective/objective measures of fatigue with the measures of performance.

Expected Results

- We expect to see an association between pilot errors and physiological evidence of fatigue effects (e.g. long blinks, head-nodding, low scores on PVT, KSS, SP, EEG evidence of micro-sleep events, etc).
  - The time series of physiological measures will be compared to the time series of pilot performance measures (incorporating adjustment for relative fatigue state) to determine what predictive ability the physiological measures may hold. Both cumulative and instantaneous measurements of each variable will be examined for their ability to predict pilot error with high selectivity.
- We expect to see associations between evidence of fatigue and/or sleepiness in the various embedded devices. For example, if the EEG shows a change in brain activity concordant with increased drowsiness or sleepiness, there will be a concurrent increase in percent eye closure and in posture modification during that time period. Similarly if the EOG shows slow eye movement, the eyelid position tracking data should indicate long fixations or slow blink rates.
  - The time series of each physiological measurement will be compared to one another to determine the level of agreement. For EEG measurements, the PSG method will be considered the standard to which other EEG measurements are compared.
- We expect to see evidence of deteriorations in crew resource management (CRM) as quantified by LOSA-style evaluation during time periods when fatigue symptoms are detected by the various devices.
- We expect to observe pilot error rates that are dependent on fatigue state (rested/fatigued) and cumulative workload (early/late) and their interaction.

Method

Subjects. Thirty two crews consisting of a Captain and First Officer from the same airline company will be recruited for inclusion in the study. In addition, up to 10 additional pilots will be recruited to allow final improvements to elements of the study. All participants must be qualified in

Boeing 777 aircraft. Because of the paucity of female commercial airline pilots, the selection of participants will be limited to males. Additional demographic information, such as age, flight hours, and fatigue-relevant habits (e.g., smoking, normal caffeine intake, known sleep abnormalities) will be collected.

Apparati. The study will be conducted in the 777 Cab, a high fidelity flight simulator. The 777 Cab is equipped to record data from pilot interactions with the controls and displays and this data will be time-synched to the physiological device data outputs. Physiological measures include EEG (electroencephalograph), EOG (electrooculograph), ECG (electrocardiograph), actigraphy, blink rate, head posture, body posture, and voice samples. Additional measures include psychomotor vigilance test (PVT) scores, subjective fatigue and sleepiness scores, and behavioral observations via audio/video recordings.

Scenarios. Half of the total participants will be assigned to long-haul flight scenarios, and the other half to a series of four shorter flights that will be conducted in succession to simulate short-haul operations. The sessions will be matched in terms of total flight duty time (eight hours), and involve gate to gate operations, from simulated dispatch interactions, through the flight itself, and ending with a taxi to the arrival gate. Further, at the end of the long-haul flights and in the last flight of each short-haul series, the crew will encounter either a medical emergency or a holding pattern scenario.

All crews will be scheduled to arrive at the lab following a minimum of 72 hours off-duty. This is expected to reduce any accumulated fatigue that may have been acquired during their preceding flight schedules and produce a set of participants who are all within a narrow range of restedness relative to their own baseline at the start of the simulator-based data collection phase. Within each group (long-haul and short-haul), pilots will fly two sessions, one each under Rested and Fatigued conditions. The Rested session will be scheduled when the pilots' predicted alertness is at its peak (i.e. normal daytime schedule after a sufficient rest opportunity). The Fatigued sessions will be scheduled when predicted alertness is at a minimum (i.e. overnight). Prior to the Fatigued sessions, the crew will be given a normal sleep opportunity with a required wakeup time, then remain awake for an extended period of time and participate in scheduled activities during the day to eliminate rest opportunities. For half of the crews, the Rested session will be first, followed by the Fatigued session; the other half of crews will experience the Fatigued session first, followed by the Rested (see Table 1 for full breakdown of conditions).

Table 1.
Outline of each unique session type by flight condition, relative fatigue level, and order of fatigue condition.

Crew	**Session 1**	**Session 2**
Crew A	Rested; Long Haul	Fatigued; Long Haul
Crew B	Fatigued; Long Haul	Rested; Long Haul
Crew C	Rested: Short Haul	Fatigued; Short Haul
Crew D	Fatigued; Short Haul	Rested: Short Haul

Procedure. Prior to participation in the simulator-based data collection sessions, the pilots will be asked to wear an actigraph watch and use the Jeppesen CrewAlert app to enter subjective fatigue ratings and perform a psychomotor vigilance task (PVT) periodically throughout each day for 2 weeks.

On the first day of the simulator-based sessions, both pilots will be brought to the flight simulator. They will be outfitted with the selected instruments and run through any calibration steps that are required. Once the devices are in place and working, the pilots will be briefed as though they were going to fly an actual flight. They will interact with (simulated) dispatch, tower personnel, cabin crew, and ground crew while preparing the plane for takeoff, and perform all duties associated with entering and executing a flight plan. During the gate-to-gate operations, they will encounter a number of glitches, anomalies, and other decision points, which are designed to gather information about the pilots' awareness, vigilance, and alertness as they perform these tasks.

At four times during each session, pilots will be asked to perform extra tasks targeted at acquiring subjective and objective data on fatigue and workload. These Fatigue Assessment Breaks consist of performing the PVT, responding to the KSS and SP subjective ratings scales, reading sentences aloud from a provided sheet, looking for changes in event related potential, and three minutes of rest with eyes closed.

They will be served a standard meal (sandwich, chips, cookie, and water) at approximately the half-way point in the session, and will be allowed to leave the cab when necessary for short periods of time. They will also be offered coffee, water, and granola bars periodically throughout the sessions. No formal napping periods will be allowed, however if one or both pilots falls asleep during the session, the experimenters will not intervene.

After the flight session is concluded, the pilots will be returned to their hotel and will be asked to follow the predetermined procedures for either rest or extended wakefulness. The second session will proceed just like the first, with the crew performing the other of the two flight scenarios under similarly challenging conditions.

Results and Discussion

Data collection on this study began in August 2012 and will continue through early 2014. At the time of this paper's acceptance for publication, we have not acquired sufficient data for preliminary analyses. However, initial results will be reported at the time of presentation.

Conclusions

The results of this study are expected to contribute to our understanding of fatigue in the commercial flight deck; how different factors may differentially affect pilots' physiological symptoms and flight operations performance. With a more thorough understanding of what fatigue looks like in the flight deck, we may be able to more effectively design fatigue risk management programs for a variety of risk-inducing situations.

References

Balkin, TJ. (2011). Performance deficits during sleep loss: Effects of time awake, time of day, and time on task. In MH Kryger, T Roth, & WC Demont (Eds.), *Principles and Practice of Sleep Medicine, 5th edition* (Chapter 65). St Louis, MO: Saunders, Elsevier.

Basner, M, & Dinges, DF. (2011). Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. *Sleep, 34*(5), 581-591.

Bourgeois-Bougrine, S, Cabon, P, Gounelle, C, Mollard, R, & Coblentz, A. (2003). Perceived fatigue for short- and long-haul flights: A survey of 739 airline pilots. *Aviation, Space, & Environmental Medicine, 74*, 1072–7.

FAA (2011a). FAA and Industry are taking action to address pilot fatigue, but more information on pilot commuting is needed. *Office of Inspector General Audit Report No. AV-2011-176*, September 2011.

FAA (2011b). Flightcrew member duty and rest requirements. *US DOT FAA Docket No. FAA-2009-1093; Amendment Nos. 117-1, 119-16, 121-357,* December 2011.

FAA Fact Sheet (2010). *Pilot Fatigue.* Retrieved from http://www.faa.gov/news/fact_sheets/news_story.cfm?newsId=11857

Fogt, DL, Cooke, WH, Kalns, JE, & Michael, DJ. (2011). Linear mixed-effects modeling of the relationship between heart rate variability and fatigue arising from sleep deprivation. *Aviation, Space, & Environmental Medicine, 82*, 1104-9.

Hursh, SR, Raslear, TG, Kaye, AS, & Fanzone, JF. (2006). Validation and calibration of a fatigue assessment tool for railroad work schedules, summary report. Washington DC: Federal Railroad Administration.

Hursh, SR, & Van Dongen, HPA. (2011). Fatigue and performance modeling. In MH Kryger, T Roth, & WC Demont (Eds.), *Principles and Practice of Sleep Medicine, 5th edition* (Chapter 66). St Louis, MO: Saunders, Elsevier.

Kaida K, Akerstedt, T, Kecklund, G, Nilsson, JP, & Axelsson, J. (2007). Use of subjective and physiological indicators of sleepiness to predict performance during a vigilance task. *Industrial Health*, 45, 520-526.

Reifman, J, & Gander, P. (2004). Commentary on the three-process model of alertness and broader modeling issues. *Aviation and Space Environmental Medicine, 75*(3, Supplemental), A84-88.

Samel, A, Wegmann, HM, & Vejvoda, M. (1997). Aircrew fatigue in long-haul operations. *Accident Analysis and Prevention, 29*(4), 439-452.

Thomas, MJW, Petrilli, RM, Lamond, N, Dawson, D, & Roach, GD. (2006). Australian long haul fatigue study. *Proceedings of the 59th Annual International Air Safety Seminar (IASS).*

Van Dongen, HPA, & Hursh, SR. (2011). Fatigue, performance, errors, and accidents. In MH Kryger, T Roth, & WC Demont (Eds.), *Principles and Practice of Sleep Medicine, 5th edition* (Chapter 67). St Louis, MO: Saunders, Elsevier.

Van Dongen, HPA, Mott, CG, Huang, JK, et al. (2007). Optimization of biomathematical model predictions for cognitive performance impairment in individuals: accounting for unknown traits and uncertain states in homeostatic and circadian processes. *Sleep, 30,* 1129-1143.

Vogel, AP, Fletcher, J, & Maruff, P. (2010). Acoustic analysis of the effects of sustained wakefulness on speech. *Journal of the Acoustical Society of America, 128*(6), 3747-56.