Journal of Human Performance in Extreme Environments

Volume 7 | Issue 2 Article 9

Published online: 10-1-2003

Assessment of Real Data and Theoretical Issues in Extreme Aviation Environments

Peter I. Terrence
University of Central Florida

Richard D. Gilson
University of Central Florida

Peter A. Hancock
University of Central Florida

Recommended Citation

Terrence, Peter I.; Gilson, Richard D.; and Hancock, Peter A. (2003) "Assessment of Real Data and Theoretical Issues in Extreme Aviation Environments," *Journal of Human Performance in Extreme Environments*: Vol. 7 : Iss. 2 , Article 9.
DOI: 10.7771/2327-2937.1034
Available at: http://docs.lib.purdue.edu/jhpee/vol7/iss2/9

In complex tasks such as aviation, overall response time possesses an inherent decision-making component. In contrast, pure RT measurements focus on highly automated, almost reflexive, reactions to stimuli.

Assessment of Real Data and Theoretical Issues in Extreme Aviation Environments

Peter I. Terrence, Richard D. Gilson, *and* ***Peter A. Hancock,*** *University of Central Florida, Orlando, Florida*

We investigated performance in extreme aviation situations using data recorded from actual in-flight emergencies that resulted in an accident. Response times to such sounds as alarms, auditory cues, and critical verbal statements were extracted from cockpit voice recorder (CVR) transcripts. Preliminary screening identified 14 CVR transcripts which permitted response time evaluation. Results from these selections showed crewmember response times ranging from 1 second to 41 seconds, with a mean and standard deviation of 9.57 s and 10.56 s, respectively. Despite the evident problems of sample size resulting in a positively skewed distribution and the limitations on the inferences drawn from these results, we contend that these data render insights into actual emergency response performance and point to valuable avenues for future exploration.

Introduction

This paper examines performance in real-world emergencies which resulted in aviation accidents. Cockpit voice recorders (CVR) offer an objective performance measurement in these problematic circumstances. Response time measurements gleaned from these recordings differ markedly from laboratory-based reaction time (RT) measurements. In complex tasks such as aviation, overall response time possesses an inherent decision-making component. In contrast, pure RT measurements focus on highly automated, almost reflexive, reactions to stimuli. Comparative control circumstances for emergencies are impossible and inferences from normative behavior are rarely applicable (see Hancock & Scallen, 1998). In order to better capture such crucial situations, the important first step is the formation of a descriptive database.

Real-World Response: The Existing Evidence

Some limited insights can be gained from studies relating to the reaction time (RT) measurements of driver braking behavior, but we are not aware of response time investigations into actual aviation emergencies. Johansson and Rumar (1971) found a 0.66 s median driver brake RT for anticipated accident situations with no RT longer than 2 s. For unanticipated versus anticipated accident situations, the average median RT measurements across all participants were 0.73 s and 0.54 s, respectively. Graham's (1999) work on auditory icons as emergency warnings for drivers showed no mean brake RT's greater than 1 s across all reported conditions. Even extended-duration, emergency driver response data yield mean RT's no longer than 1.74 s (Muto, 1982). Other research into the effects of stimulus change and laboratory surprise on RT measures rarely show latencies above 1.5 s (see Niepel, 2001, Meyer et al., 1991, Niepel et al., 1994). Sanders and McCormick (1993) discuss real surprise RT, with the complexity of both the stimulus and the subsequent response mitigating the increase in response latency. Response times in actual driving situations range from 1.5 s to 2 s. However, response times in actual aviation environments are greater, though the degree of the increased latency has not been documented (Gilson, 2002). Accordingly, since much of the research into surprise stimuli is based on reaction, not response time, and also

http://dx.doi.org/10.7771/2327-2937.1034

focuses on driving studies, generalization to extreme aviation incidents is problematic.

Information Source

The first step in analyzing real-world data was to ascertain the location of timed-response recordings to emergency situations. Response time data is an important sources of objective information in uncontrolled environments, despite its limitations on generalizability. The primary warehouse of this type of information was concluded to be the National Transportation and Safety Board's (NTSB) aviation incident database, specifically the transcripts of the cockpit voice recorders (CVR) typically installed in commercial and many privately owned aircraft.

The CVR maintains a four-channel, auditory record of events in the cockpit prior to the unit's deactivation following an accident. Although these recording devices are intended to record verbal communications, other auditory stimuli such as engine noise, warning systems, and landing gear movement may be discernible on the recording. Most CVR models record continuously for 30 min, while newer models have a recording time of 2 hr. Sophisticated audio equipment extracts the information from the CVR, following an incident. A panel of NTSB officials, FAA members, the aircraft operator, representatives from aircraft's manufacturer and engine manufacturer, and the pilot's union, then transcribes the recording for its inclusion in the final report (National Transportation and Safety Board [NTSB], 2003).

We hypothesized that responses to auditory signals in an emergency situation in the cockpit would be accompanied by "long" response latencies from the crew. Pilots and other crewmembers are well trained to react quickly to emergency situations, therefore a delay in responding should be readily apparent and identifiable. However, it should be noted that the delayed response time may not necessarily be a problem at all. Perhaps the best course of action during an emergency is for the crew to take the time to assess the situation and respond accordingly (Green et al., 1996). There is no incentive to react blindly to an emergency with a potentially fatal response. If the correct response to a situation were always to initiate a set of commands as quickly as possible to avoid exacerbating the problem, then there is little need for human intervention. Automation of the response would serve the situation better in these instances. Sheridan and Parasuraman (2000) outlined the case for when and when not to automate a given procedure. The equations contained in their work paint a concise picture of the automation question using the probabilities of a system failure with the benefits and costs of correctly and incorrectly perceiving the status of the situation. The authors state the difficulties inherent in determining the actual costs and benefits in a particular set of circumstances for automation, which are magnified in extreme aviation environments. The application of the aforementioned equation becomes a challenging undertaking given the unquantifiable nature of some of these costs and benefits.

The basis of this paper is not the importance of speed or a delay in response time as the sole bearer of responsibility for aviation incidents and catastrophes. It is possible that a failure to execute a command within a particular time frame would have no bearing on whether or not a crash occurs. Consequently, performing a function quickly is not necessarily a life-saving act. Rather, a response time analysis affords a window into human performance during extreme aviation emergencies. An analysis of events as they unfold in time, instead of a summary of the final outcome, will help to identify links between theories and expand current human performance models of real-world situations. The CVR transcripts allow researchers to examine these sequences of events in the cockpit for future analysis.

All CVR transcripts contain a warning outlining the dangers of taking the contents of the CVR data out of context, which is a natural caveat given the reconstructive nature of the CVR transcription process. An obvious limitation is the nature of the recording medium itself. An auditory account of an unfolding aviation emergency is only a partial record of the complex interactions of the crew and the aircraft itself. The pilot's adjustment of a heading or the first officer's reducing airspeed may be difficult to capture unless a verbal acknowledgement accompanies the action or if it is correlated with the flight data recorder (FDR). The intensity of an aviation emergency further complicates any investigative process. As emotions run high and stress levels grow, meaningful verbal interactions may become unintelligible to the recorder. Despite these limitations, accurate interpretations and assessments of CVR transcript data are rendered difficult, but not impossible.

The initial difficulties in utilizing actual extreme aviation accident data lie in the data's general availability. According to NTSB analysis, the total general aviation flight hours for 1999 was approximately 29.7 million hours (NTSB, 2003) with over 11.3 million departures for U.S. air carriers (NTSBc). The number of Aviation Safety Reporting System (ASRS, 2001) self report incidents for 2002 was 34,831 (ASRS, 2001). These reports, which NASA maintains for aviation research, capture the next level of analysis, specifically those instances in which a performance error does not result in an accident. However, there is no timed recording of the event sequence in the cockpit in these instances. The number of extreme aviation situations is further reduced in the light of actual accident scenarios. In 1999, the number of recorded and investigated aviation accidents was 1,906 (NTSB, 2003), only 0.16% of the aforementioned departure figure. Further reducing the number of potential CVR transcripts is irreparable damage to the unit, improper maintenance of the recording following the incident, or failure of the unit to yield valuable information during the accident investigation. If these pitfalls are successfully avoided, the NTSB accident investigators can reconstruct the CVR data for transcription, and we now turn to those successes for our response-time analysis. The process used to search for these appropriate instances are outlined below.

Method

Procedure

A search of the NTSB's aviation accident synopses database over the last 16 years (1986-2002) yielded 39,856 total recorded accidents (NTSB*b). A subsequent search* was conducted for incidents that would generate CVR transcripts with auditory emergency signals such as alarms. Given the relatively scant recording of response times in real-world emergency situations, these transcripts are a distinctive source of information that provides the timing data necessary for the investigation of non-laboratory reaction times. The search terms used for the database included key words such as "alarm," "warning," "CVR," "reaction time," and "noise."

The synopses of the results of this search were then reviewed for their applicability given that further search criteria were met. The search criteria were those synopses in which an emergency situation existed and the likelihood of the timing of the emergency situations onset and the crew's verbal response to that emergency would be able to be captured within the transcribed CVR. We identified incident synopses for potential inclusion in the final analysis. The accident synopses were not limited to a specific type of fixed-wing aircraft or flight condition. Subsequent requests were made to the NTSB for the CVR transcripts of the identified flights. The actual audio recordings themselves were not available due to policies that provide for the recycling of the recording device after the NTSB has established the timing and other information contained within the CVR.

Once the 27 available ones were received (several were unavailable due to ongoing investigations or due to no transcription having been made because of technical problems with the CVR), we reviewed the transcripts to identify those instances in which the onset of an emergency situation was captured by the CVR. These emergency onset signals included numerous cockpit alarms as well as the recorded events of audible damage occurring to the aircraft. Another dimension was whether or not the CVR captured the timing of the crewmember's response to the emergency onset signal. Often the nature of the emergency onset signal was primarily visual and therefore not able to be recorded onto the CVR, or the timing of the crewmember's response was highly ambiguous.

Out of the 27 transcripts, 14 were found to contain discernible emergency signal onsets as well as identifiable crew responses to the emergency situation. The aircraft involved in these incidents range from large commercial airliners to smaller personal airplanes. No transcripts of accidents involving helicopters were included in this investigation. Emergency signals vary throughout the analysis. In some instances there is a specific cockpit alarm, while in other transcripts there are sounds such as an impact with no particular cockpit warning accompaniment. Crewmember responses were based on the callout for a specific action or the sound generated through the initiation of the proper response, such as an increase in engine noise resulting from a throttle increase.

Results

From the data obtained in the 14 usable CVR transcripts, an analysis of the time from the onset of the emergency signal to the recorded response of the pilot/crewmember to the signal yielded a mean response time of 9.57 s (SD = 10.56 s), with a distribution as seen Figure 1 and Figure 2. One particular transcript revealed an emergency auditory signal occurring with no response until 41 s later, which is contributing heavily to the positive skewing of the data. With the removal of this outlier, the data takes on distribution seen in Figure 3, with a recalculated mean

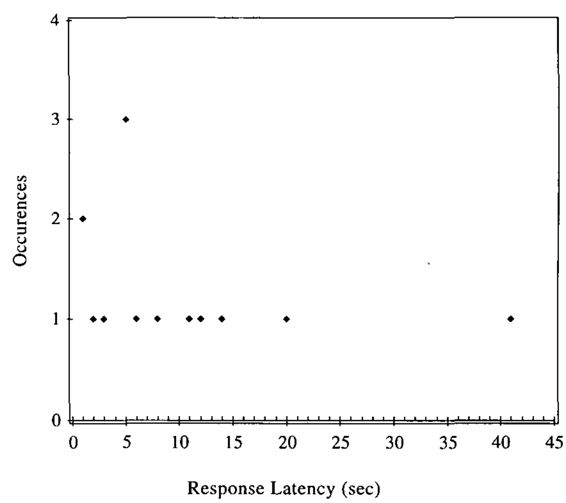

Figure 1. *Each point represents the number of occurrences of each response latency expressed in seconds (N = 14). The mean response time was 9.54 sec (SD = 10.56).*

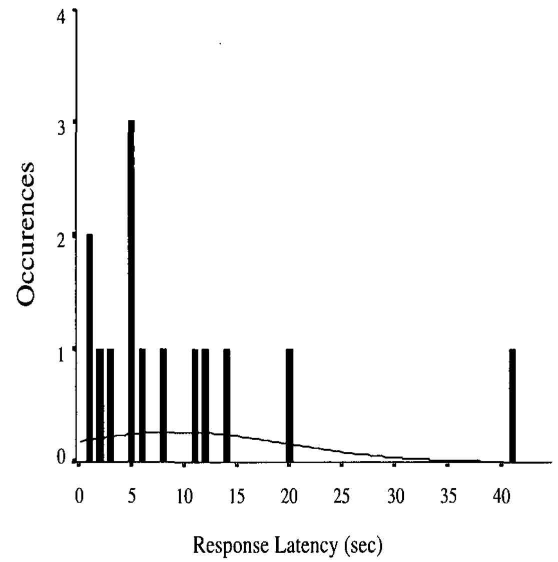

Figure 2. *Graph of number of CVR transcripts for each response latency with overlying distribution curve (N = 14).*

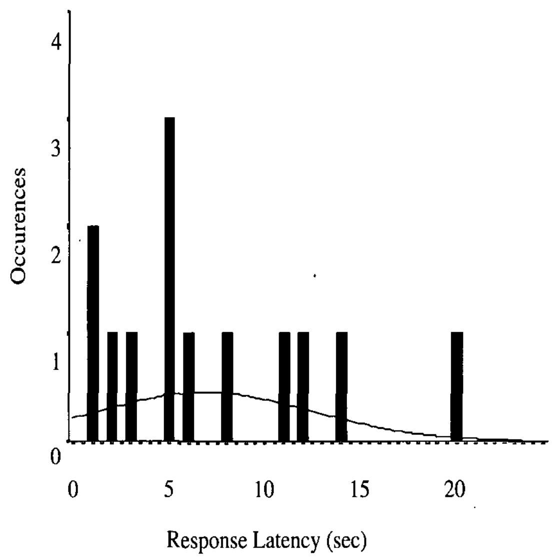

Figure 3. *Graph of number of CVR transcripts for each response latency with overlying distribution curve after outlier removal (N = 13).*

of 7.15 s (SD = 5.67). The distribution of the data with or without the outlier indicates an obvious push toward responding quickly to auditory emergency signals.

Discussion

Interpretation of the data is possibly limited because of the small sample. CVR transcripts afford a limited glimpse into the details of an aviation accident. However, the latencies found in this analysis are a clear departure from the surprise reaction time data (see Niepel, 2001, Meyer et. al., 1991, Niepel et. al., 1994). Further investigation must seek to unravel the implications of an "appropriate" aviation emergency response time for both cockpit and procedure design.

Currently, CVR's may not capture all stimulus and response parameters that can occur during an emergency. The movements of the flight crew must be discerned from secondary sources, such as verbal correlation with executed commands or changes recorded in aircraft characteristics during flight. Measures to generate comparisons of situations that resulted in an accident with those that did not result in an accident would aid in this process. If CVR data preservation procedures were amended to provide CVR data in conjunction with ASRS reports of non-accidents, there is the potential for more ecologically sound comparisons of similar accident and non-accident response times along critical factors.

Generally, the use of CVR transcripts suffers from their relative unavailability. There are several reasons for this scant amount of data. The number of emergency accidents in which CVR transcript data would prove useful is thankfully low. CVR data may also fall victim to (1) crewmember error in failing to trip the appropriate circuit breaker, (2) damage to the device, or (3) failure to recover the unit. Given the sum of the factors that may affect the usability of this wealth of data, the CVR should record as many parameters as possible. If there were multiple visual and auditory channels coupled with revised procedures for preserving meaningful data for both accident and "almost" accident situations, a uniquely detailed level of analysis would emerge for extreme aviation environments.

Many human performance models predict increased response latencies under circumstances in which the operator is under the general blanket of increased task demands. It is reasonable to assume that an extreme, potentially life-threatening emergency in an aviation environment qualifies as a situation with increased task demands on the flight crew. Multiple resource capacity models (Wickens, 1984) information processing (IP) models, and surprise reaction research all predict increased response times due to the complex demands (Sanders & McCormick, 1993) of an actual aviation emergency. These models fail to predict the amount of increase in response latency during the dire circumstances present on the CVR. Bliss (1997) found increased response time patterns to cockpit alarms, including false alarms, and is one of the few studies showing latencies in the ranges present in CVR transcript data. Further examination and expansion of this and related work may prove useful in future response time assessment of CVR transcript data.

In summary, response time data from CVR transcripts offer a unique window into the nature of emergency events as they unfold in the cockpit. These perspectives do not provide an empirical basis for the validation or invalidation of human performance models. However, our findings do indicate real-world response times that are well beyond simple RT's, both in actual and simulated environments. Caution is warranted as there are no recordings of emergency situations which did not result in an accident. Future researchers would benefit through the identification and use of real-world recordings of comparison emergency situations in order to generate new avenues of investigation and determine the applicability of current theoretical models. The implementation of cockpit recorders with expanded capabilities and appropriate safeguards for controlled use and application of the data will aid in the improvement and development of these accident research tools for extreme aviation environments.

References

Aviation Safety Reporting System. (2001). *The ASRS program briefing*. Retrieved September 16, 2003 from http://asrs.arc.nasagov/briefing/program_briefing_nf.htm.

Bliss, J. P. (1997). Alarm reaction patterns by pilots as a function of reaction modality. *International Journal of Aviation Psychology*, 7(4), 1-14.

Gilson, R.D. (2002). *Some observations on real-world aviation responses*. Unpublished report, September, University of Central Florida.

Graham, R. (1999). Use of auditory icons as emergency warnings: evaluation within a vehicle collision avoidance application. *Ergonomics*, 42(9), 1233-1248.

Green, R.C., Muir, H., & James, M. et al. (1996). *Human factors for pilots* (2nd ed.).Brookefield, VT: Ashgate.

Hancock, P.A., & Scallen, S. F. (1998). Allocating functions in human-machine systems. In R. R. Hoffman & M. F. Sherrick (Eds.),*Viewing psychology as a whole: The integrative science of William N. Dember* (pp. 309-330). Washington, DC: American Psychological Association.

Johansson, G., & Rumar, K. (1971). Drivers' brake reaction times. *Human Factors*, 13(1),23-27.

Meyer, W., Niepel, M., Rudolph, U., & Schutzwohl, A. (1991). An experimental analysis of surprise. *Cognition and Emotion*, 5(4), 295-311.

Muto, W. H., & Wierwille, W. W. (1982). The effect of repeated emergency response trials on performance during extended-duration simulated driving. *Human Factors*, 24(6), 693-698.

National Transportation and Safety Board. (2003). *U.S. General Aviation, Calendar Year 1999. Annual Review of Accident Data* (NTSB/ARG-03/02 PB 2003-105847). Washington, D.C.

National Transportation and Safety Board. (n.d.) Aviation Accident Statistics. Retrieved September 15, 2003, from http://www.ntsb.gov/Aviation/stats.htm.

National Transportation and Safety Board. (n.d.) Aviation Accident Database and Synopses. Retrieved September 30, 2002, from http://www.ntsb.gov/ntsb/query.asp.

National Transportation and Safety Board. (n.d.) *Cockpit Voice Recorders (CVR) and Flight Data Recorders (FDR)*. Retrieved September 15, 2003, from http://www.ntsb.gov/aviation/CVR_FDR.htm.

Niepel, M. Rudolph, U., Schutzwohl, A., Meyer, W. (1994). Temporal characteristics of the surprise reaction induced by schema-discrepant visual and auditory events. *Cognition & Emotion*, 8(5), 433-452.

Niepel, M. (2001). Independent manipulation of stimulus change and unexpectedness dissociates indices of the orienting response. *Psychophysiology*, 38(1), 84-91.

Sanders, M., & McCormick, E. (1993). *Human factors in engineering and design* (7th ed.). New York: McGraw-Hill, Inc.

Wickens, C. (1984) *Engineering psychology and human performance*. Columbus, OH: Merril.