INVESTIGATING INTERRUPTIONS: AN EXAMPLE FROM THE FLIGHTDECK

K. A. Latorella
Industrial Engineering Department
State University of New York at Buffalo
Buffalo, New York 14260

This study investigates an aspect of multiple-task management, interruption management, in an operational context. Fourteen commercial airline pilots each performed 16 approaches in a simulated commercial flightdeck. Air traffic control (ATC) clearances interrupted subjects as they performed three procedures during these approaches. Common ATC interruptions were found to be significantly disruptive to ongoing procedure performance on the flightdeck by producing significantly more procedure performance errors and increased flightpath management activity. These results corroborate, for the flightdeck, that which is true in laboratory experiments, and which is evidenced in aviation accident/incident reports.

INTRODUCTION

Human operators increasingly act as managers of multiple tasks in complex and dynamic environments. One aspect of multiple-task management (MTM) is the handling of interruptions, or interruption management. Research in attention management (*e.g.*, Broadbent, 1958; Schneider & Detweiler, 1988) and human error (*e.g.*, Reason, 1992) indicate that humans do not handle interruptions easily, or, often, very well. Previous research investigating interruption management takes four approaches: (1) development of a theoretical framework for MTM, including interruption management (2) laboratory studies aimed at understanding mechanisms of interruptions (3) human/machine interface evaluations using interruption-recovery as an evaluation metric and (4) identification of interruptions as a causal factor in accident/incident analyses and field investigations.

Adams, Tenney, & Pew (1995) describe problems associated with developing situation awareness in MTM environments. One problem associated with situation awareness is to accurately develop and retain a task queue in memory. Consistent with theories of limited working memory capacity, these authors suggest that interruptions may cause other tasks' representations to be deleted from the queue, and the tasks not performed. An interruption is not merely an additional task competing for a limited resource, it also redefines that which is resident in active memory. Knowledge structures associated with the interrupting task impose on those already resident at the time of interruption. Based on these two facets of the MTM problem, Adams, Tenney, & Pew (1995) develop a framework for the management of multiple tasks based on Neisser's (1976) expanded model of the perceptual cycle. Neisser's model includes an *explicit focus* memory bin and an *implicit focus* memory bin. Explicit focus describes a limited-capacity storage, and corresponds to working memory. Implicit focus relates to the knowledge structures that are related to those tasks represented in explicit focus. This framework provides the foundation for speculation on when interruptions might be handled most easily. Specifically, that interruptions related to the items in explicit focus should be easily integrated with ongoing activities, and, interruptions tangent to the immediate explicit focus should be relatively easily assimilated due to common representations in implicit focus.

While Adams, Tenney, & Pew (1995) postulate mechanisms affecting interruption management, laboratory experiments demonstrate deleterious effects of interruptions and empirically identify significant causes of these effects for simple tasks. Most research indicates that interruptions increase post-interruption performance times (Detweiler, Hess, & Phelps, 1994; Gillie & Broadbent, 1989; Field, 1987; and Kreifeldt & McCarthey, 1981) and error rates due to interruptions (Detweiler, Hess, & Phelps, 1994; Cellier & Eyrolle, 1992; Gillie & Broadbent, 1989; Field, 1987; and Kreifeldt & McCarthey, 1981). However, Cellier & Eyrolle (1992) found increased performance speed following an interruption. They attribute this to activation of previously untapped resources after interruption. Significant factors influencing the effects of interruptions include; task complexity (Cellier & Eyrolle, 1992; Gillie & Broadbent, 1989), similarity of interrupted and ongoing tasks (Detweiler, Hess, & Phelps, 1994; Cellier & Eyrolle, 1992; Gillie & Broadbent, 1989), memory load at interruption (Detweiler, Hess, & Phelps, 1994), ability to rehearse departure conditions (Detweiler, Hess, & Phelps, 1994; Gillie & Broadbent, 1989), and time constraints (Cellier & Eyrolle, 1992). Laboratory investigations identify factors that mediate interruption management performance in artificial settings.

In addition to research that explicitly studies mechanisms of interruption performance, studies using interruption recovery to evaluate human/machine interface investigations provide useful insights. In a comparison of reverse-polish notation and algebraic notation calculators (Kreifeldt & McCarthey, 1981), and in a comparison of different search strategies in a menu system (Field, 1987), users' performance was generally worse following the introduction of an interruption. Field studies and accident and incident reports indicate the significance and magnitude of deleterious effects of interruptions in more complex operational contexts.

Interruptions are cited as a contributing cause of power plant incidents (*e.g.*, Bainbridge, 1984). In addition, the frequency, types (Monan, 1979), and deleterious effects (*e.g.*, Madhaven & Funk, 1993) of interruptions on the flightdeck are well documented. A recent search of the Aviation Safety and Reporting System (ASRS) revealed at least 315 reported incidents due to interruptions on the flightdeck since 1986. Worse than this, interruptions are implicated in disastrous accidents which result in loss of life. An air traffic control (ATC) interruption during checklist procedures appears to have caused the aircrew of an airliner departing from the Detroit Metropolitan Airport forgot to lower the flaps before takeoff (National Transportation Safety Board, 1988). System failures can also be considered interruptions to ongoing flightdeck tasks. On an Eastern L-1011 aircraft, the crew became so distracted by an alert and performance of an irregular procedure that flightpath management duties went unattended (National Transportation Safety Board, 1973). Recognizing the effects of interruptions on the flightdeck, *task interruption* was included in a taxonomy of cockpit task management (CTM) errors (Funk, 1991; Chou & Funk, 1990). The is relatively little experimental research investigating interruptions on the flightdeck. Existing studies use frequency-of-interruption as a dependent measure to evaluate datalink and checklist usage (Williams, 1995; Linde & Goguen, 1987, respectively). Williams (1995) reports significantly greater resumption times for datalink than voice ATC interruptions, however does not quantitatively report the degree of disruption imposed by these interruptions.

The study of interruptions is motivated by the significant effects of interruptions, both in terms of their ubiquity and their consequences in operational environments, and the implications for extending basic cognitive theories to an understanding of MTM. Several laboratory studies cite the deleterious effects of interruptions, and field and incident/accident investigations evidence the consequences of interruptions in operational environments, the effects of interruptions on the flightdeck have not been experimentally quantified. This research investigates the effects of interruptions on the commercial flightdeck. The fundamental question answered by this research is, to what extent do interruptions disrupt pilot activity on the flightdeck? To address this issue, an experiment was conducted in which commercial airline pilots performed approach and descent procedures in a commercial flight simulation environment. ATC clearances were systematically inserted into the scenario to interrupt procedure performance.

METHODS

Subjects

Fourteen male commercial airline pilots from various carriers served as subjects. Subjects currently flying advanced Boeing aircraft with minimally one year glass-cockpit, flightpath management system / control display unit (FMS/CDU) experience, and 5,000 flying hours.

Apparatus

The experiment was conducted in NASA Langley's Transport Systems Research Vehicle (TSRV), a fixed-base simulator similar to a B-737. A remote confederate interacted with subjects for real-time ATC and airline operations personnel (company) contacts. Thirty-second continuous loop tapes provided Automatic Terminal Information Service (ATIS) information. A menu system on a touchscreen contained approach and final descent checklists. Another touchscreen display contained a simple datalink system for introducing visual interruptions. Datalink messages were introduced by the mechanized voice utterance, "incoming message". Subjects were not provided with an out-the-window view.

Scenario

The scenarios included three components: flightpath management (FPM), procedure performance, and interruption management. Flightpath profiles required subjects to manually fly a complex, step-down approach with multiple turns, and hard crossing restrictions at each waypoint in single crew member operations. Subjects flew in the attitude control wheel steering (ACWS) mode of the autopilot with no other autopilot functions or autothrottles. ACWS is a rate-controlled flight mode which retains an established attitude or lateral deviation. Flightpath profiles provided three procedural intervals, which required minimal flightpath management (FPM) activity. Natural procedural interval deadlines were imposed by creating difficult FPM intervals between the procedural intervals and extremely difficult FPM regions surrounding waypoints.

Three procedures, the top of descent (TOD) procedure, 18,000' (18K) procedure, and final approach fix (FAF) procedure were designed to be performed in the procedural intervals (Table 1). The TOD procedure was performed at 19,000', 290 knots calibrated, indicated airspeed (KIAS), level flight. The 18K procedure was performed in a stable descent from 18,000' to 12,000' at 240 KIAS. The FAF procedure was performed in a stable descent from 8,000' to 4000' at 150 KIAS. Field elevation was 3,500'. Tasks identified for inclusion in these procedures were obtained through extensive interviews with two retired United Airlines pilots and reference to several airlines' approach and descent checklists. Some tasks included in these procedures were somewhat artificially placed in order to define experimental conditions or satisfy experimental control, *e.g.*, that the 18K and FAF procedures be isomorphic.

Interruptions were interjected into this scenario in each of the three procedures. In any one run, subjects could experience as few as zero and as many as three interruptions, one *per* procedure. Interruptions were ATC-initiated clearances associated with one of five FMS/CDU interrupting tasks (IT); enter an initial runway, side-step to the parallel runway, program a standard holding pattern, alter a crossing speed restriction, or alter a crossing altitude restriction. While requests to enter initial runways could be either through the auditory or visual modality (via the datalink system), and only occurred in the first procedure, all other interruptions were voice-conveyed and occurred in both the 18K and FAF procedures. These interruptions were interjected into the procedures at specific intervention positions (IP), triggered by subject performance of pre-defined events constituent to procedural tasks.

Table 1. Procedural Tasks

TOD Procedure
Pre-tune Company Frequency
Pre-tune ATIS Frequency
Listen to ATIS
Pre-tune Tower Frequency
Obtain Status Information from FMS/CDU
18,000' Procedure
Set Altimeters in FMS/CDU
Contact Company
Obtain ETA-Zulu Time from FMS/CDU
Calculate ETA-local Time.
Turn on Seatbelt sign
Announce to Cabin (Seatbelt sign, gate, ETA)
Turn on Landing Lights
Turn on Anti-skid
Select appropriate Autobrakes
Perform Approach Checklist
FAF Procedure
Select EPR for Go-Around from FMS/CDU
Contact Tower
Obtain $\mathrm{Vref}_{30.}$ from FMS/CDU
Calculate Adjusted Target Speed.
Turn on No-Smoking sign
Announce to Cabin (No-Smoking, landing)
Lower Gear
Arm Speedbrake
Select Flaps 25
Perform Final Descent Checklist

Pairing ITs and IPs defined different experimental interruption conditions. These conditions operationalized several task factors hypothesized to affect flightdeck interruption management, *e.g.*, modality and semantic similarity of an interruption and ongoing task, embeddedness of an interruption in a procedure, and the relationship between procedural tasks severed by an interruption. Null IP and IT specifications provided uninterrupted procedure control conditions. All subjects received all experimental conditions. Since this paper addresses only the first stage of analysis for this investigation, *i.e.*, quantifiying the general effects of interruptions on the flightdeck, the above factors will not be elaborated on further.

Procedure

Each subject participated in the experiment for two days. On the first day, subjects received explicit descriptions of scenario performance goals and FPM, procedure performance, and interruption simulation training. To enable the introduction of interruptions at specific points in these procedures, and to provide a standard performance goal by which to evaluate performance errors, subjects were asked to perform procedural tasks in exactly the order and method trained. Subjects were told that the scenario would also include *incidental tasks* to increase scenario realism, and that these tasks were termed *incidental* only because they would occur non-deterministically throughout the scenario. Subjects were instructed that they must perform incidental tasks and must confirm any incoming ATC request prior to actually accomplishing the incidental task. The second day included three refresher runs and sixteen data runs. Each run was approximately seventeen minutes in length and was preceded by a three-minute reset period. Subjects received a break after the refresher runs and after each set of four data runs.

Dependent Measures

This experiment measured the deleterious effects of interruptions on the ensemble task; that is, the integration of procedure and interrupting task requirements. Three measures defined these deleterious effects: (1) procedure performance errors, (2) procedure performance time, and (3) ensemble FPM activity. Procedure performance errors were identified as procedural task omissions, misorderings, or redundant performance of procedural tasks. To ascertain the effect of interruptions on procedure performance time, ensemble performance times, for which interruptions occurred within a procedure, were compared to constructed *composite times*. Composite times were constructed by adding the average of uninterrupted procedure times and interruption performance times for those occurring before procedures. Composite times were defined for all possible (subject, procedure, interrupting task) triplets, to eliminate effects of these variables. The ensemble FPM activity measure counted the number of active attitude and lateral control inputs made during the ensemble interval, *i.e.*, from the first activity of the procedure or interruption to the last activity of either the procedure or interruption, and divided by the elapsed seconds in that interval. Since the scenario afforded hands-free FPM after deviations were nullified, and procedure intervals were time-constrained, reversion to active FPM during procedures was considered non-optimizing performance.

RESULTS

Procedure performance errors were analyzed using analysis of variance rather than a non-parametric statistic because error data was too sparse to calculate expected cell frequencies. Significant effects using this parametric assessment are conservative estimates of effects expected from an non-parametric analysis. Overall error rates were very low, less than one major error per procedure. However, interrupted procedures contained significantly more, on average 53% more, procedure performance errors than uninterrupted procedures, $F(1,13) = 25.809, p = 0.0002$ (Table 2). Generally, one would expect one error in every three uninterrupted procedures, but in one of every two interrupted procedures. Some task omissions were more consistently evident than others and seemed associated with the previous occurrence of an interruption (Table 3).

Table 2. Procedure performance errors.

condition	n	mean	std.dev.
interrupted	504	0.518	0.860
uninterrupted	168	0.339	0.716

Table 3. Frequently omitted tasks.

		% of omissions		
task omitted	% omitted	no IT	after IT	before IT
tune tower	9.56	4.17	66.67	29.17
obtain vref	3.10	0	71.43	28.57
descent check	4.93	9.10	90.90	0

The mean ensemble time for each subject's performance on each procedure and interrupting-task type was compared to its corresponding composite time using a paired t-test, matching on subject, interrupting task-type, and procedure. Marginally significant results indicated that composite times were slightly longer, 1.63 seconds on average, than ensemble times, $t(242) = -1.672, p = 0.0958$ (Table 4). One explanation for this result might be that interrupted conditions' average performance times were less than uninterrupted average performance times due to more omissions. To eliminate the possibility of this bias, a paired t-test was performed on the data after all conditions having either a procedural or interruption performance error were removed. On error-free data, the same trend exists, $t(132) = -1.665, p = 0.0984$ (Table 5).

Table 4. Procedure Time Comparison.

measure	n	mean	std.dev.
ensemble time	250	111.47	19.14
composite time	243	112.64	16.97

Table 5. Error-free Procedure Time Comparison.

measure	n	mean	std.dev.
ensemble time	136	114.09	18.27
composite time	133	115.87	13.21

On average, subjects made 17.25 active FPM control inputs during interrupted conditions and only 13.10 inputs during uninterrupted conditions. While the absolute difference between interrupted and uninterrupted conditions' average ensemble FPM *per* second is small, it represents a significant increase, $X^2(1) = 14, p < .005$, on average 10%, in the proportion of the ensemble interval devoted to FPM.

Table 6. Ensemble FPM Activity *per* second.

condition	n	mean	std.dev.
interrupted	467	0.160	0.142
uninterrupted	161	0.146	0.122

DISCUSSION

This research empirically quantifies the disruptive influence of interruptions on the commercial flightdeck. The effects of interruptions found in laboratory experiments and exemplified in field studies and accident/incident investigations were supported by this experiment's results. As is reported for more generic tasks, ATC interruptions significantly increased performance errors in flightdeck procedures. Several examples illustrate the attribution of task omissions to previous occurrence of an interruption. These omissions have operational consequences, and may have serious operational consequences if combined with other irregular occurrences. For example, a mis-tuned tower frequency minimally causes confusion and increased radio traffic, and maximally, if left uncorrected, may result in pilots' inability to receive life-saving instructions.

Interruptions did not degrade subjects' speed in performing ongoing procedural tasks, even when considering only error-free data. In fact, a marginally significant result indicated that the presence of an interruption actually slightly sped procedure performance time. Although contrary to findings of most previous research, this result was consistent with that reported by Cellier & Eyrolle (1992), and may have indicated that subjects' adopted a compensatory strategy after interruption. Consistent with results indicating a significant effect of time constraints (Cellier & Eyrolle, 1992), and strategic workload theory (*e.g.*, Raby & Wickens, 1991) subjects may have recognized that additional demands of an interruption might interfere with performance on impending high-FPM workload regions, and actively compensated by performing remaining procedural tasks faster. Given this, increased error rates in interrupted conditions could be, derivative of a speed/accuracy trade-off effect rather than directly due to an interrupt's imposition on working memory.

The ensemble FPM activity measure is not directly related to any dependent measure previously associated with studying the effects of interruptions. However it does implicate the disruptive effects of interruptions on the flightdeck. Given the scenario conditions, active FPM during procedures was unnecessary and therefore considered sub-optimizing behavior.

Having quantified the effects of ATC interruptions and datalink presentations degrade performance on the flightdeck, and in light of accidents and incidents attributable to flightdeck interruptions, it is evident that further research is required to identify specific task, environment, and operator performance-shaping characteristics that modulate interruption management behavior. Subsequent analyses of results from this investigation more sensitively study task factors hypothesized to affect interruption management on the flightdeck.

Understanding both the significance of and the factors modulating interruption management improves our understanding of how humans manage multiple tasks, and provides a means for more sensitively introducing and integrating interrupting tasks in MTM contexts. For the aviation domain, this information contributes to an understanding of human CTM (*e.g.*, Abbott, 1993), models of CTM (*e.g.*, Funk, 1996), and informs design of CTM aids (*e.g.*, Funk and Lind, 1992), however the the ubiquity of MTM contexts and interruptions suggests a wide range of applications.

ACKNOWLEDGMENTS

This work was conducted at the NASA Langley Research Center's Crew/Vehicle Integration Branch and funded by a NASA Graduate Student Researcher Fellowship, grant number NGT-50992. This work benefited significantly from the auspices of Mr. Paul Schutte and Dr. Kathy Abbott, the expertise of Captain Dave Simmon (ret.), the programming wizardry of Mr. John Barry, useful comments from Dr. William Rogers, and the assistance of TSRV hardware and software support personnel.

REFERENCES

Adams, M. J., Tenney, Y. J., and Pew, R. W. (1995). Situation awareness and the cognitive management of complex systems. *Human Factors, 37*(1), 85-104.

Abbott, T. (1993). *Functional categories for future flightdeck designs.* NASA-TM-109005. Washington, DC: The National Aviation and Space Administration.

Bainbridge, L. (1984). Analysis of verbal protocols from a process control task. In E. Edwards and F.P. Lees (Eds.), *The Human Operator in Process Control* (pp. 146-158). London: Taylor and Francis.

Broadbent, D. E. (1958). *Perception and communications.* New York: Pergamon Press.

Cellier, J. and Eyrolle, H. (1992). Interference between switched tasks. *Ergonomics, 35*(1), 25-36.

Chou, C. and Funk, K. (1990). Management of multiple tasks: Cockpit task management errors. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics.* Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Detweiler, M. C., Hess, S. M., and Phelps, M. P. (1994). *Interruptions and working memory.* Unpublished technical report. University Park, PA: The Pennsylvania State University.

Field, G. (1987). Experimentus Interruptus. *SIGCHI Bulletin, 19*(2), 42-46.

Funk K. H. (1996). A functional model of flightdeck agenda management. In *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting.* Santa Monica, CA: Human Factors and Ergonomics Society.

Funk, K. H. (1991) Cockpit task management: Preliminary definitions, normative theory, error taxonomy, and design recommendations. *The International Journal of Aviation Psychology, 1*(4), 271-285.

Funk, K. H. and Lind, J. H. (1992). Agent-based pilot-vehicle interfaces: Concept and prototype. *IEEE Transactions on Systems, Man, and Cybernetics, 22*(6), 1309-1322.

Funk, K. (1989). Development of a task-oriented pilot-vehicle interface. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics.* Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Gillie, T. and Broadbent, D. (1989). What makes interruptions disruptive? A study of length, similarity, and complexity. *Psychological Research, 50,* 243-250.

Linde, C. and Goguen, J. (1987). *Checklist Interruption and Resumption: A Linguistic Study.* NASA-CR-177460. Washington, DC: The National Aviation and Space Administration.

Kreifeldt, J. G. and McCarthey, M. E. (1981). Interruption as a test of the user-computer interface. In *Proceedings of the 17th Annual Conference on Manual Control.* JPL Publication 81-95. Jet Propulsion Laboratory, California Institute of Technology.

Madhaven, D. and Funk, K. (1993). Cockpit task management errors in critical in-flight incidents. *Seventh International Symposium on Aviation Psychology.* Columbus, Ohio.

Monan, W.P. (1979). *Distraction - A Human Factor in Air Carrier Hazard Events.* NASA-TM-78608. Washington, DC: The National Aviation and Space Administration.

Neisser (1976). *Cognition and reality: Principles and implications of cognitive psychology.* San Francisco: Freeman Press.

National Transportation Safety Board. (1988). *Aircraft accident report - Northwest Airlines, McDonnell Douglas DC-9-82, N312RC, Detroit Metropolitan Wayne Co., Airport, Romulus MI, August 16, 1987.* Washington, DC: National Transportation Safety Board.

National Transportation Safety Board. (1973). *Aircraft Accident report: Eastern Airlines Inc. Lockheed L-1011, N310EA, Miami International Airport, Miami, Florida, December 29, 1972.* Washington, DC: National Transportation Safety Board

Raby, M. and Wickens, C. D. (1991). Strategic behavior in flight workload management. *Sixth International Symposium on Aviation Psychology.* Columbus, Ohio.

Reason J. (1992). *Human error.* New York: Cambridge University Press.

Schneider, W. and Detweiler, M. (1988). The role of practice in dual-task performance: Toward workload modeling in a connectionist/control architecture. *Human Factors, 30,* 539-566.

Williams, C. (1995). Potential conflicts of time sharing the flight management system control display unit with data link communications. *Eigth International Symposium on Aviation Psychology.* Columbus, Ohio.