JUL-21-98 14:22 FROM:USC CONTRACTS & G

Final
IN-31
358315

THE EFFECT OF INTERRUPTIONS ON PART 121 AIR CARRIER OPERATIONS

DIANE L. DAMOS

FINAL

NASA Cooperative Agreement NCC 2-910
University of Southern California
Los Angeles, CA

June 30, 1998

To: CASI

TABLE OF CONTENTS

INTRODUCTION

Background

Effective task management requires that cockpit tasks be prioritized correctly, executed in a timely manner, and allocated between the crew so that no one is overloaded. Effective task management has been a challenge for crews of both traditional and automated aircraft. It became a focus of concern when poor task management was implicated in several accidents and incidents in Part 121 air carrier operations (Chou, Madhavan, and Funk, 1996). In response to these concerns, several government laboratories initiated research efforts to understand how air carrier pilots perform task management (Funk, 1991; Latorellat, 1966; Rogers, 1996; Schutte and Trujillo, 1996).

These efforts have identified task prioritization as a critical component of effective task management. However, determining task prioritization in a two-person, automated aircraft is problematic. The environment is dynamic, changing frequently. Task assignments between pilots may change depending on the circumstances. Time constraints may dictate that some tasks be rescheduled, interleaved with other tasks, or omitted entirely, making it difficult to identify the relative priority of tasks at a given point in time. Additionally, collecting data during operational flying is difficult because of safety and operational considerations.

Consequently, most attempts to identify task priorities have been conducted using techniques that range from structured interviews (Rogers, 1996) to medium-fidelity simulations (Latorellat, 1996; Schutte and Truijillo, 1996). Some of these techniques ask the pilots about their task priorities in specific situations; others infer priorities from the order of task execution. The study reported in this paper takes a different approach to identifying task priorities. This study used interruptions to infer relative task priorities by assuming that if an ongoing task was interrupted by the arrival of a new task, then the new task had the higher priority. This research also used video tapes of Line Oriented Flight Training (LOFT) scenarios as the data source, ensuring high realism.

Describing interrupted activities and stimuli that have the potential to interrupt ongoing activities can lead to considerable confusion. Consequently, throughout the remainder of this document, "events" will refer to stimuli with the potential to interrupt ongoing activities. An activity is said to be interrupted only when it has stopped. Similarly, an interruption occurs when an event causes an activity to stop.

Purpose

The primary purpose of this study was to determine the relative priorities of various events and activities by examining the probability that a given activity was interrupted by a given event. The analysis will begin by providing frequency of interruption data by crew position (captain versus first officer) and event type. Any differences in the pattern of interruptions between the first officers and the captains will be explored and interpreted in terms of standard operating procedures.

Subsequent data analyses will focus on comparing the frequency of interruptions for different types of activities and for the same activities under normal versus emergency conditions. Briefings and checklists will receive particular attention. The frequency with which specific activities are interrupted under multiple- versus single-task conditions also will be examined; because the majority of multiple-task data were obtained under laboratory conditions, LOFT-type tapes offer a unique opportunity to examine concurrent task performance under "real-world" conditions.

A second purpose of this study is to examine the effects of the interruptions on performance. More specifically, when possible, the time to resume specific activities will be compared to determine if pilots are slower to resume certain types of activities. Errors in resumption or failures to resume specific activities will be noted and any patterns in these errors will be identified. Again, particular attention will be given to the effects of interruptions on the completion of checklists and briefings. Other types of errors and missed events (i.e., the crew should have responded to the event but did not) will be examined.

Any methodology using interruptions to examine task prioritization must be able to identify when an interruption has occurred and describe the ongoing activities that were interrupted. Both of these methodological problems are discussed in detail in the following section.

APPROACH

The major methodological obstacle to studying interruptions concerns determining when an interruption has occurred. Identifying an interruption in some situations is straightforward, i.e. a pilot stops talking in midsentence. Most situations are not as obvious, however. Several different methods for identifying interruptions were tried, including time-based techniques that examined changes in activities for fixed periods after an event. The most promising of these methods, which was subsequently adopted, uses the resumption of an activity as the primary criterion for an interruption, i.e., if the pilot resumed the activity, it was scored as having been interrupted. If the pilot did not resume the activity, the investigator had to distinguish among four alternatives: the activity was completed before the event was addressed, the activity was not completed but was no longer relevant, the activity was unimportant and did not need to be resumed (i.e., casual conversation), and the activity should have been resumed but was not. Clearly, distinguishing between these four alternatives is the most problematic aspect of this approach.

Expectancy was a secondary criterion for determining if an activity had been interrupted. An expected event was assumed not to cause as an interruption. For example, if a pilot contacted air traffic control (ATC) for information and was told to stand by, the subsequent ATC call was assumed not to interrupt any of that pilot's ongoing activities.

The time to resume an interrupted task (resumption time) was used, when possible, as a secondary measure of priority. That is, resumption time was used to confirm estimates of priorities. For example, if an activity was rarely interrupted and resumed quickly after the interruption, then the resumption time measure supported an inference of high priority for the activity.

The second major methodological issue in studying interruptions concerns the systematic classification and description of the pilots' tasks. Because flying has been conceptualized as a hierarchy of goals and tasks since at least 1947 (Williams, 1971), a task analysis is an appropriate tool for describing the pilots' activities in a systematic manner. A generic task analysis developed by the FAA for automated aircraft was used in this study (Longridge, 1995). This task analysis had six levels of activities, which was sufficient to provided a comprehensive analysis of the pilot's tasks. Because the task analysis was generic, it could be used for different automated aircraft and, with very minor modifications, for different air carriers.

The effects of interruptions on two levels of the pilot's activities were examined. The first level will be referred to as the *element* level. Elements generally consist of fine-grained activities that are easily defined and observed, e.g., read, manipulate, talk. The only exception to this are the two elements that represent periods of unobservable activity—listen and monitor.

The second level of activity (second-level activity) usually was the lowest level activity represented in the task analysis and was coded using its numerical designation. Examples of second-level activities are "Perform after takeoff checklist," "Select approach mode on mode control panel (MCP)," and "Select legs page." Generally, identifying and coding second-level activities was straightforward. The only two exceptions occurred when the pilots were monitoring the instruments or programming the systems. When the pilots appeared to be monitoring the instruments, it was frequently impossible to identify the second-level activity. In these cases, one of the higher levels of the task analysis, such as "Perform enroute cruise," was used as the second-level activity.

Programming the flight management computer (FMC) presented special coding problems. The pilots could be observed typing information into the FMC. However, because the video cameras were placed behind the pilots, none of the videos were clear enough to allow the displays to be read. Thus, the investigator could not determine precisely what step of the programming tasks the pilot was performing. In these cases, the investigator again used one of the higher levels of the task analysis to describe the pilot's second-level activity.

Some questions may be raised about the usefulness of the element data. This level of activity was recorded and analyzed for four reasons. First, no data had been collected at this level of detail when the study began. Thus, such data could be a valuable resource, particularly for investigators developing models of crew performance. Second, as noted earlier, task prioritization is a difficult topic to investigate. Fine-grained behaviors appeared to be more tractable to understanding and analysis than high-level cognitive tasks. Third, recommendations from a study focused on fine-grained behaviors may be more easily implemented in air carrier training curricula than recommendations concerning high-level tasks. Fourth, subsequent research will be concerned with the interruption of high-level tasks.

METHODS

Scenarios

The video tapes were obtained from three different sources. Each source used a different scenario. One scenario included a fuel management problem at cruise that was not covered by any procedure. Additionally, the crew was required to execute a missed approach, which was followed by a side-step maneuver during the second approach. The second scenario involved a critical passenger illness that required a diversion to an airport where the weather was at minimums. The third scenario involved a smell of undetermined origin that required a return to the departure point.

ATC, support personnel (dispatch, ramp, etc.), and flight attendants were simulated in different ways in the three scenarios. In the first scenario, retired air traffic controllers performed ATC functions while other confederates role played support personnel and flight attendants. In the second scenario, the LOFT instructor role played all of the personnel. In the third scenario, the LOFT instructor role played all personnel except ATC, which was simulated by an intern. None of the simulators were equipped with data link. Consequently, all communications between the pilots and ATC or support personnel were conducted using standard radio procedures.

The scenarios were performed in full-motion, Level C or D simulators. Each scenario involved a different model aircraft. All of the aircraft were produced by the same manufacturer and are considered "glass cockpit" aircraft. None of the scenarios was modified in any form for this study.

Participants

The data were obtained from 11 flight crews from three different air carriers. All crews consisted of line-qualified, current captains and first officers. All crews used the operating procedures and manuals of their own airline. The crews flew simulators of aircraft in which they were currently qualified. None of the crew members were instructors or management pilots. In ten of the crews, the captain was the pilot flying. In the 11th crew, the first officer was the pilot flying. Data from this crew was included in this study only after they were inspected to ensure that thye did not differ from the others.

Task Analysis

Few modifications to the FAA task analysis were necessary for this study. The few omissions that were noted usually involved procedural differences between the air carriers. For example, during an emergency some air carriers require the crew to execute a checklist as well as perform certain procedures from memory, whereas the FAA task analysis only included the procedures executed from memory. Thus, the task analysis omitted tasks for some air carriers. In such cases, an additional item was added to the task analysis at the appropriate point. Occasionally, a step was out of place for a given air carrier and simply was added at the appropriate point in the task analysis.

Scoring

Before the data could be scored, several arbitrary decisions were necessary. The first concerned the type of elements that could be interrupted. Much of commercial airline flying involves cognitive, rather than physical, tasks. These tasks—such as monitoring instruments, planning approaches, and listening to the other pilot—are unobservable. Their execution, consequently, must be inferred from other behaviors and information sources. The investigator decided that no "unobservable" elements would be scored as interrupted although they were recorded, i.e. a pilot's element might be recorded as "monitoring" at the time of an event, but no interruption would be scored.

The second arbitrary decision concerned identifying the time at which an event occurred. For visual and auditory warning signals, such identification is straightforward; these signals have a clear onset. Other types of events, however, do not have such a clear onset. For ATC calls, scoring began at the end of the call sign because some of the scenarios included ATC communications with other aircraft. For events involving the entry of the flight attendant into the cockpit, scoring began when the flight attendant began talking because the individual role playing the flight attendant never simulated using a key to enter the cockpit and frequently did not knock to enter the cockpit.

Seventeen elements were used to reflect the behaviors that were occurring at the time of the event (see Table 1). Second-level activities were represented using the numeric codes from the task analysis.

TABLE 1. LIST OF ELEMENTS USED TO CODE THE DATA

Element	
arrange	pause
don	point
fly	reach
input	read
laugh	scan
listen	talk
manipulate	tune
monitor	write
move	

Each video tape was examined for six types of events from the V_r (rotate speed) call to touchdown. These events were ATC calls, cabin chimes, appearance of the flight attendant, voice communications from the flight attendant, communications from support personnel, and auditory and visual warning signals. When an event was identified, 17 variables associated with it were encoded. Among the variables encoded were the time at which the event occurred, the time at which a pilot responded to the event, who responded, the type of event, the elements (see Table 2), and the second-level activities that were in progress for each pilot at the time of the event. The investigator determined if any of the elements or second-level activities were actually interrupted and, if so, the time at which they were resumed. If a second-level activity was resumed, the investigator also determined if it was resumed at the correct point by consulting the appropriate checklists and procedures.

The scoring allowed for concurrent performance of elements. For example, a pilot could be reaching for a chart while talking. In such cases, the effect of the event on each element was noted and scored separately. The scoring also allowed for concurrent performance of second-level activities although such instances were rare. Again, the effect of the event on each activity was noted and scored separately.

On occasion the investigator could not identify the second-level activity in progress at the time of the event. In these situations, the investigator consulted a subject matter expert.

RESULTS AND DISCUSSION

Because two of the techniques used to analyze the frequency data—logit (multi-way frequency) analysis and logistic regression—are relatively new, some discussion of their characteristics and limitations is appropriate. Logit analysis is an extension of the traditional two-way χ^2 test of independence to multiple categorical variables in which one is considered to be a dependent variable and the others are considered to be predictors. In logistic regression analysis, the dependent variable is categorical and the predictors may be either categorical or continuous.

Logit analysis has the advantage of permitting tests of the interactions among predictors as well of their individual effects on the dependent variable. However, logit analysis looses its sensitivity to reliable effects when the expected cell frequencies are insufficient. Logistic regression uses maximum likelihood methods that will fail to converge on a solution when the cell frequencies are too low or too discrepant (Tabachnick and Fidell, 1996). However, a form of logistic regression, conditional exact inference, sometimes provides a solution when the cell frequencies are low or discrepant. Thus, whenever expected cell frequencies were sufficient to permit sensitive prediction of interruptions, a logit analysis was used. If the cell frequencies were not sufficient, the conditional exact form of logistic regression analysis was used.

Regardless of the technique used, the captains' and the first officers' data were always analyzed separately. Separate analyses avoided dependencies in the data caused by having the same event represented in both data sets. For the same reason, analyses of elements were always conducted separately from analyses of second-level activities.

All of the analyses reported in this section are concerned with frequencies of interruptions. To determine if some types of elements have a significantly greater probability of being interrupted than others, the event was the unit of observation and was assumed to occur randomly relative to the elements being performed by the crew. Although this assumption may appear to be questionable, for five of the crews, the person generating the events could not see the pilots. For the other crews, the simulator instructors agreed to generate the events based on the position of the aircraft (i.e., ATC calls), not on the activities of the pilots. The assumption of independence, therefore, seems tenable. The same assumption was made for second-level activities.

Summary Data

A total of 324 events were scored. The overall frequency of each type of event is given in Table 2. One event was incorrectly recorded and is not shown in Table 2. In some cases the number of elements may exceed the number of events. In these cases, the pilot was performing two elements when an event arrived. The number of second-level activities also may exceed the number of events for the same reason. This table also gives the average time to respond to the event, which was calculated from the onset of the event to the beginning of the reply for the first five event types. Response time for the auto throttle disconnect warning was calculated from the onset of the signal to the time the pilot moved the auto throttle arm switch.

The data presented in Table 2 should be interpreted with caution for several reasons. Some of the events, particularly those generated by support personnel, are very infrequent. Any inferences about the probability of interruption should be made with extreme caution. The data pertaining to the appearance of the flight attendant and direct voice communication with the pilots may be particularly unrepresentative; the door between the pilots and the cabin crew normally is closed and locked during flight in all three of the simulated aircraft. Thus, the flight attendant could not appear in the cockpit without using a key or without knocking and having the pilots release the door. Additionally, a direct call from the cabin attendants to the pilots is not normally possible in any of the three aircraft included in this study; the flight attendant must ring the cabin chime to signal the crew to switch to the interphone mike.

Table 2 indicates that, for most events, the probability of interruption of a second-level activity is equal to or greater than the corresponding probability for an element. These results may reflect, however, an artifact of using a task analysis to structure the pilots' activities. Within this structure, a pilot must interrupt some level of activity to respond to an event. Because only two levels of activities were analyzed, the second-level activity was scored as interrupted if the element was not clearly interrupted and the pilot resumed either the element or the second-level activity. Thus, the higher frequency of interruptions for second-level activities may reflect a scoring artifact.

TABLE 2. PROBABILITY OF INTERRUPTION BY EVENT

Event Type	Freq.	Average Resp. Time [1]	Capt. Element (prob.)	Capt. Second Level (prob.)	F.O. [2] Element (prob.)	F.O. [2] Second Level (prob.)
ATC	275	6	.163 (47/288) [3]	.305 (84/275)	.177 (57/281)	.402 (111/276)
Flight Attendant						
Cabin Chime	20	14	.095 (2/21)	.286 (6/21)	.190 (4/21)	.550 (11/20)
Appears	11	10	.091 (1/11)	.818 (9/11)	.083 (1/12)	0 (0/11)
Voice [4]	2	13	0 (0/2)	0 (0/2)	0 (0/2)	0 (0/2)
Support						
Dispatch	11	8	.333 (4/12)	.727 (8/11)	.250 (3/12)	.636 (7/11)
Ground Support	1	4	0 (0/1)	0 (0/1)	0 (0/1)	1.00 (1/1)
Systems Main.	1	27	1.00 (1/1)	1.00 (1/1)	0 (0/1)	0 (0/1)
ARINC	1	3	0 (0/1)	1.00 (1/1)	0 (0/1)	0 (0/1)
Auto Throttle Disconn.	1	5	1.00 (1/1)	1.00 (1/1)	1.00 (1/1)	1.00 (1/1)

[1] Response times are rounded to the nearest second

[2] First Officer

[3] Denominators greater than the frequency of the event reflect concurrent elements or second-level activities. An event could interrupt one, both, or none of the concurrent elements or concurrent activities.

[4] Talking directly to the pilots without use of the interphone

Despite these caveats, several trends are evident in the data. First, events interrupt ongoing elements relatively infrequently for both the captain and the first officer. This result will be discussed in more detail in the following section. Second, events generated by the flight attendants appear to have a low probability of interrupting either the ongoing elements or the second-level activities of either pilot. Additionally, pilots are slow to respond to flight attendant events, confirming their low priority. Third, pilots respond promptly to ATC, which is in keeping with operational practice. Indeed, the true response time is actually shorter than given in Table 2 because, as noted earlier, the response time shown was measured from the end of the call sign, not from the end of the message. Fourth, ATC calls appear to be more likely to interrupt the second-level activities of first officers than captains. This result was anticipated because 10 of the 11 first officers were the pilot not flying in the scenario and, consequently, were responsible for radio calls.

Interruption of Elements

A major goal of this study was to determine if the frequency of interruption differed between elements for the same event. Because the frequency of interruption for a given element may differ under normal versus emergency conditions, the frequency should be compared between these two conditions. Fortunately, because all three scenarios involved emergencies, a sufficient number of events occurred while the captain and the first officer were performing emergency procedures to allow such a comparison.

The task analysis used to code the second-level activities had a group of numeric codes for emergencies. Activities that were performed during normal operations, such as copying ATIS or lowering the gear during an approach, were not included in the emergency numeric codes. An event, therefore, could occur when one pilot was performing under emergency conditions while the other appeared to be performing under normal conditions. Clearly, however, both pilots actually were working under emergency conditions. To avoid misleading results, all analyses counted elements and second-level activities as performed under emergency conditions if one or both pilots were operating under emergency conditions as indicated by the numeric codes of the task analysis.

As shown in Table 2, only ATC calls were sufficiently frequent for analysis. Table 3 shows the frequency of interruption for the most common elements under both normal and emergency conditions. Logistic regression techniques were used to analyze the probability of interruptions as a function of element type and condition (emergency versus normal). However, to avoid statistical problems from low cell frequencies, some of the elements had to be grouped. Consequently, the movement elements (move, reach, point) were combined into a movement group, as were the manipulation elements when the manipulated object was part of the automated flight system (electronic manipulation group). Thus, element type had six levels: movement group, electronic manipulation group, talk, input, write, and read.

Despite grouping some of the elements, the maximum likelihood methods would not converge because of the small cell frequencies. Therefore, conditional exact inference on the parameters of the logistic regression model were conducted using LogXact-Turbo software (Cytel, 1993), which offers an exact conditional scores test, distributed as χ^2.

TABLE 3. FREQUENCY OF ATC INTERRUPTIONS FOR SELECTED ELEMENTS

Element	Captain		First Officer	
	Emergency[1]	Normal	Emergency[1]	Normal
Fly	0/0	0/48[2]	0/0	0/1
Talk	2/6	28/54	3/4	10/26
Move	0/1	0/3	0/2	1/5
Reach	0/0	1/5	0/0	0/3
Point	0/0	0/0	0/0	0/1
Read	0/6	1/1	2/5	11/13
Write	0/0	1/1	0/0	0/3
Input to CDU	0/1	2/7	0/1	14/22
Manipulate				
MCP altitude	0/0	0/0	0/0	0/2
MCP autopilot	0/0	0/0	0/0	0/1
MCP heading select	0/1	0/0	0/0	1/3
MCP speed	0/0	0/0	0/0	0/1
MCP heading	0/0	0/1	0/0	1/1
MCP (general)[3]	0/0	1/7	0/0	0/2
Center CDU	0/0	0/0	0/1	0/1

[1]Emergency conditions were defined using the task analysis. The entries in this column were obtained when one or both pilots were operating under emergency conditions.

[2]The denominator represents the frequency of this element in the database. The numerator represents the number of times an ATC call interrupted the element.

[3]This classification was used when the investigator could not identify the counter or knob being manipulated by the pilot.

A model including both element type and condition significantly predicted interruptions among captains, $\chi^2(5, \underline{N} = 94) = 15.44, p < .01$. However, within the model, only element type significantly affected interruptions, $\chi^2(5, \underline{N} = 94) = 10.91, p = .03$; there was no significant effect of emergency versus normal conditions, $p = .14$. The same pattern of results was observed for first officers, with the two component model significantly predicting interruptions,

χ^2 (5, N = 97) = 26.54, $p < .01$. Type of element was a significant predictor, χ^2 (5, N = 97) = 26.27, $p < .01$; but condition was not, $p = .73$, $p < .01$.

These results indicate that the probability that an ATC call will interrupt any of the six element types does not differ between normal and emergency conditions. This is somewhat surprising given the urgency of many emergencies. That is, elements performed under emergency conditions might be assumed to have a higher priority and, therefore, a lower probability of interruption than elements performed under normal conditions. These data do not support such an assumption.

Post hoc comparisons were made among elements using the conditional scores test with data collapsed over condition. The large number of potential post hoc comparisons precluded an exhaustive determination of the source of the difference, and casual inspection of the data gave little hint of the source of the difference for either pilot. Consequently, Wickens' Multiple Resource Theory (1992) was used to guide the selection of comparisons to be tested. According to this theory, the maximum interference between two tasks occurs when they assess the same types of processing resources. Thus, an ATC call, which accesses verbal resources, should interfere most with a verbal task and, presumably, have the highest probability of causing an interruption. The probability of interruption should be lower for tasks that access other types of resources, such as spatial resources. Because these data were based on interruptions from ATC calls, contrasting a verbal activity (talking) with more manual activities (movement, input, electronic manipulation) seemed appropriate.

The Type I error rate was set to .0125 to adjust the family-wise error rate for the four comparisons, in which "talk" was contrasted with all other elements except "read" (another verbal activity). None of the comparisons for captains nor for the first officers were statistically reliable using the Bonferroni-type correction. Thus, at this time it is impossible to determine which element types differ significantly in terms of their likelihood of interruption.

One important element, fly, was not included in the logistic regression analysis because of its unique status, i.e. hand flying typically occurs under "sterile cockpit" conditions or under emergency conditions. Table 3 shows that 48 events occurred while the captain was flying. None interrupted flying. Because the captain was the pilot flying in 10 of the 11 crews, only one event occurred while a first officer was flying and, again, this event did not interrupt flying.

The results of this study correspond exactly to the anticipated results and reflect standard operating procedures. In Part 121 Air Carrier operations, one pilot is clearly designated as the pilot flying. When this pilot is hand flying the aircraft, the other pilot assumes essentially all other duties. Thus, no interruptions of flying should occur.

Concurrent Element Performance

Concurrent performance of elements is a relatively rare occurrence. Of the 324 events recorded in this study, 14 occurred when the captain was performing two elements under normal conditions and one under emergency conditions. The corresponding numbers for the first officer are seven and two. A logit (multi-way frequency) analysis was performed to determine if elements performed concurrently were more likely to be interrupted than elements performed alone. Because the probability of interruption may vary between normal and emergency conditions, two factors were included in the logit analysis: number of elements performed (one versus two) and condition (emergency versus normal).

Neither number of elements nor condition individually predicted interruptions for first officers, $p = .71$ and .96, respectively. However, number of elements and condition interacted in their effect on interruptions, $\chi^2(1, \underline{N} = 326) = 3.89$, $p < .05$. When the first officer was performing two elements concurrently, 50% of the elements were interrupted under emergency conditions and about 7% under normal conditions. However, when the first officer was performing a single element, 16% of the elements were interrupted under emergency conditions and 21% were interrupted under normal conditions. An analogous analysis for captains showed no reliable prediction of interruptions by number of tasks ($p = 70$), condition ($p = .65$), nor their interaction ($p = .36$).

The results of these analyses indicate that concurrent elements are no more likely to be interrupted than single elements (see Table 2) except under emergency conditions for first officers. The lack of a condition effect is difficult to explain; concurrent element performance under emergency conditions should reflect high workload or high stress. Under such conditions, events should be less likely to interrupt concurrent elements. The interaction demonstrated by the first officers' data appears particularly anomalous.

The effects of an event on concurrent element performance, however, may not be limited to the elements themselves. When a pilot is performing two elements concurrently, the second-level activity may be more likely to be interrupted than when the pilot is performing one element. Interestingly, visual inspection of the data revealed that under normal conditions, pilots who were performing two elements concurrently at the time of an event were never performing two second-level activities concurrently. In contrast, under emergency conditions, two of the three instances of concurrent element performance were associated with concurrent second-level activities. With no concurrent second-level activities under normal conditions, analyzing the data from both conditions in one analysis would have been problematic because of the low cell frequencies. Consequently, only data obtained under normal conditions were included in the analysis.

A χ^2 test of independence was performed for second-level activities. The test had two factors—number of elements (one versus two) and effect of the event (interruption versus no interruption) on the second-level activity. Neither the test performed on the captains' data nor the test performed on the first officers' data showed an effect of number of elements performed at the time of the event on the probability of an interruption of the second-level activity.

Interruption of Second-Level Activities

Only a few of the large number of questions that can be asked about second-level activities will be addressed in this report. This section will be concerned with the likelihood that second-level activities other than briefings and checklists will be interrupted. Briefings and checklists will be addressed in the following section.

Some of the most serious human factors issues in aviation today concern the types of errors that can occur in automated as compared to traditional cockpits (Wiener and Curry, 1980). Interruption of ongoing elements and second-level activities is one way in which errors may be introduced into the system. Although no data were obtained from traditional cockpits, activities that are common to both traditional and automated cockpits may be compared with activities that are unique to automated cockpits.

Consequently, selected second-level activities for normal operating conditions were combined into groups for the purpose of analysis. The first group consisted of procedural and "housekeeping" activities that are common to both traditional and automated aircraft. These activities included those found in climbing to cruise altitude (e.g., turning the landing lights off, setting the altimeter to 29.92" passing 18,000 ft, observing airspeed restrictions, etc.) and those used to configure the aircraft systems enroute (e.g., adjusting the cabin temperature, setting the anti-ice system, monitoring the warning lights, gauges, and messages, etc.). The second group consisted of crew communication and situational awareness activities that again are common to both traditional and automated cockpits. Examples are communicating with the cabin crew about turbulence, discussing weather changes, maintaining terrain awareness, and assessing de-icing requirements. The third group was unique to automated aircraft and consisted of operating and programming the FMS (see Table 4).

The χ^2 test of independence examined interruptions as a function of second-level activity group (procedural/housekeeping, crew communication and situational awareness, and operate/program the FMS) for normal operations only. Group did not predict interruptions for captains (p = .17), but was a significant predictor for first officers, $\chi^2(2, \underline{N} = 84) = 8.06$, $p < .05$. Ryan's post hoc procedure (Ryan, 1960) examined pairwise differences among the three groups. Only the difference between the operate/program the FMS group and the procedural/housekeeping group significantly predicted interruptions for first officers, $\chi^2(1, \underline{N} = 58) = 7.82$, $p < .05$.

TABLE 4. PROBABILITY OF INTERRUPTION OF THREE SECOND-LEVEL ACTIVITIES UNDER NORMAL CONDITIONS

	Captain	First Officer
Procedural/ Housekeeping	9/29	12/26
Crew Commun. And Situational Awareness	17/14	17/14
Operate/Program FMS	7/17	26/32

Visual inspection of the data indicates that the first officer is roughly twice as likely to be interrupted during opearte/program activities as during procedural and housekeeping activities. Two factors may account for this difference. First, programming tasks frequently are relatively long, whereas procedural and housekeeping tasks are brief. Programming tasks, therefore, have a larger window of opportunity for interruption than housekeeping and procedural tasks. Second, the computer processors of most aircraft are extremely slow and require significant amounts of time to execute many functions. While waiting for a command to execute, a pilot may "leave" the operate/program activity to perform other activities.

The high interruption rate may indicate that operate/program activities have large "windows of opportunity" for errors. The opportunity for error may be increased further by the fact that the majority of interruptions are caused by ATC calls. Much of the information contained in these calls is numeric with a format similar to that being programmed. Thus, the possibility of entering the wrong information after resuming the task appears to be relatively high.

The time at which a specific activity was interrupted was not recorded. However, the time at which the pilot responded to the event can be used as an approximation to the interruption time. The time at which the interrupted task was resumed was always recorded. The difference between these two is referred to as the "resumption time" and approximates the "true" resumption time (the difference between the time at which the pilot stopped an activity and the time at which he resumed it). The calculated resumption time should be less than or equal to the true resumption time.

Occasionally, an event interrupted a pilot's ongoing element or second-level activity although the pilot did not respond to the event (i.e., he did not answer the ATC communication). Such situations were included in the calculation of the resumption time for the pilot and were calculated from the response time of the other pilot. Again, this procedure probably underestimates the resumption time; the pilot probably interrupted his element or activity before the other pilot made a response. Resumption times, therefore, should be considered only as indications of the time to resume a task.

The median time to resume the housekeeping/procedural second-level activities was 10 s. The median resumption time for the operate/program activities was 14 s. However, the range of scores differed considerable. For housekeeping activities, resumption scores ranged from 3 s to 39 s; the range for operate/program activities was 2 s to 1084 s.

Interruption of Second-Level Activities—Briefings and Checklists

From an operational perspective, the interruption of briefings and checklists, particularly under emergency conditions, poses hazards to the safety of flight. A sufficient number of interruptions of both of these activities under normal and emergency conditions occurred in the video tapes to allow an analysis of both briefings and checklists (see Table 5). In the majority of instances, both the captain and the first officer were performing the same activity (briefing or checklist). Occasionally, other activities were interleaved if an omission were noted during the briefing or checklist. For example, during the instrument approach briefing, one pilot might realize that the missed approach procedure had not been entered into the computer and "leave" the briefing to program the missed approach. In such instances, an event could only interrupt the briefing of the pilot who was not programming.

Three types of briefings occurred under both normal and emergency conditions: initial descent briefings, instrument approach procedures briefings, and approach/landing briefings. For the purposes of the analysis, these were combined into one group. If these briefings occurred when one or both of the crew were operating under emergency conditions, they were scored as emergency briefings and combined with other briefings that occurred only under emergency conditions. Table 5 shows the frequency of interruptions for briefings under both normal and emergency conditions.

TABLE 5. PROBABILITY OF INTERRUPTION OF BRIEFINGS AND CHECKLISTS UNDER NORMAL VERSUS EMERGENCY CONDITIONS

	Captain		First Officer	
	Normal	Emergency[1]	Normal	Emergency[1]
Briefing	.800 (16/20)	.750 (3/4)	.765 (13/17)	.000 (0/3)
Checklist	.500 (6/12)	.100 (1/10)	.765 (13/17)	.500 (1/2)

[1] Emergency conditions were defined using the task analysis. The entries in this column were obtained when one or both pilots were operating under emergency conditions.

Interruptions during briefings were analyzed using the exact conditional scores test. Emergency versus normal conditions did not predict interruptions for captains ($p = .12$), but first officers were about 9.5 times more likely to be interrupted during normal than emergency conditions, $\chi^2(1, \underline{N} = 20) = 6.23$, $p = .03$, $\beta = 2.254$.

The average time for the captain to resume the interrupted briefing was 33 s; the corresponding time for the first officer was 26 s. Because briefings are somewhat unstructured, it is difficult to determine if information subsequently was omitted. Nevertheless, the investigator found no evidence that information was omitted from an interrupted briefing.

Three types of checklists were performed in the scenarios: the after takeoff checklist, the approach/descent checklist, and the before landing checklist. For the purposes of the analyses, data from the three checklists were combined. Data on checklist interruptions are given in Table 5. Interruptions during checklists also were analyzed using the exact conditional scores test. Emergency versus normal conditions did not predict interruptions during checklists for captains ($p = .12$) or first officers ($p > .99$).

The average time to resume an interrupted checklist was 26 s for captains. The first officers required an average of 24 s to resume a checklist. The interruptions did not cause the captains to miss any steps although two captains repeated the immediately preceding step when they resumed the checklist. The first officers did not miss any steps when they resumed the checklist nor did they repeat any steps. However, one before landing checklist was never resumed because, by the time the first officer completed the interrupting event, he was too far behind the aircraft to complete the checklist before the aircraft landed.

On the whole, these analyses support most assumptions about pilot performance under emergency conditions. Pilots concentrate on the most important tasks, which are executing the emergency checklists and briefings.

Failures To Resume The Second-Level Activity

Other than the examples given above, the data provide little evidence that pilots fail to resume second-level activities that are interrupted. Only three instances were found in which the captain failed to resume a second-level activity, all of which occurred after an ATC call. In all cases the second-level activity involved talking. In three instances the first officer failed to resume a second-level activity after an ATC call. One of these activities involved talking about the fuel status of the aircraft. In no case did the failure to resume the second-level activity result in any observable errors or problems.

Concurrent Second-Level Activities

Concurrent second-level activities were rare. Only one instance was found for a captain and one for a first officer. Both of these occurred under emergency conditions, and both occurred when the pilot was performing two elements concurrently. Neither the captain nor the first officer interrupted either second-level activity to respond to the event.

Missed Events

Missed events, particularly ATC calls, are a concern for flight crews. Inspection of the data revealed few instances of missed events. Under emergency conditions, only two events were missed. One of these involved not responding to a flight attendant who entered the cockpit and the other, to an ATC call. Similarly, only seven ATC calls and one dispatch call were missed under normal conditions. In two of these instances, the pilots appeared to hear the call and act on the information but failed to reply. These instances may reflect a more casual approach to radio communication than would be found in operational flying. The only similarity in the missed events was the frequency of talking; in four of the eight instances of a missed event, the captain was talking. Casual inspection of the tapes indicated that several of the captains had a tendency to talk "through" radio communications, making it impossible for the first officer to hear and respond to the calls.

Other Errors

The data give few indications of any performance errors. The lack of errors may be attributed, at least in part, to the crews' familiarity with the aircraft; no data were obtained from pilots transitioning to the aircraft. Nevertheless, many errors are relatively subtle and may be difficult to detect. This section describes some of the investigator's observations that are not reflected in previous analyses.

One of the most striking features of the tapes is the number of times that pilots question each other about the heading or altitude. Interpreting these questions may be problematic because the question may not indicate that the pilot has forgotten the information; these pilots may actually be using cockpit resource management (CRM) procedures to obtain confirmation of a setting. Only a few instances were found in which a pilot clearly could not remember a heading or altitude after being given a frequency change.

The data, however, do provide some indication of an informal approach to radio communications in the simulator. For example, six ATC calls asked the crew if they had received and understood the instructions. In another case, the crew did not give their call sign during a transmission. Another crew failed to respond to ATC although they clearly heard the transmission.

A few examples of operational errors were noted. One crew forgot to call the tower at the outer marker. Another forgot to reset the altimeter after climbing through 18,000 ft. Several crews forgot minor procedural items, like turning off the logo lights after climb out. One first officer told a captain that they had received clearance to land although ATC had only issued a clearance for the approach. Because no comparable data are available from revenue flights, it is not possible to determine how the frequency of the observed errors compares to their frequency in operational flying. The fact that such errors occur may testify to the realism of the scenario or again, it may indicate a casual approach to training.

SUMMARY

The major purpose of this study was to use patterns of interruptions to determine the priorities of various events and activities. The data demonstrate that most elements and second-level activities are interrupted relatively infrequently, implying that they have a higher priority than the events. Of the six types of events that were examined in this study (ATC communications, appearance of the flight attendant, cabin chimes, voice communications from the cabin attendant, warning signals, and communications from support personnel), communications from ATC and dispatch appeared to have the highest priority. Those unfamiliar with air carrier operations may be surprised at dispatch's relatively high priority. However, in revenue operations dispatch usually conveys flight critical information. In the three scenarios included in this study, the pilots contacted dispatch after an emergency had begun.

Events other than ATC calls occurred too infrequently in the three scenarios to allow statistical analysis. Consequently, all subsequent conclusions pertain only to ATC communications. One of the more surprising results of this study was that the probability of interruption for both elements and second-level activities did not differ under emergency as compared to normal conditions except in two conditions, which showed opposite effects. This lack of differences appears counterintuitive and may reflect the types of scenarios included in this study; only one (smell of unknown origin) had the urgency usually associated with emergencies. Thus, the lack of differences should be viewed with some skepticism until more data can be collecting using other scenarios.

An equally puzzling result concerns the lack of significant differences in the probability of interruption under dual- and single-task conditions for both elements and second-level activities. Again, these results seem counterintuitive since multiple-task performance in air carrier operations is often associated with high workload and a sense of urgency. Such conditions would appear to make the pilots less responsive to events. The most parsimonious explanation of these results is low statistical power; very few events occurred while the pilots were performing two elements or two second-level activities. Again, this result should be viewed with some skepticism until more data can be collected.

Perhaps the most interesting finding in this study concerned the high relative probability of interruption for activities associated with operating and programming the FMC as compared to more traditional housekeeping and procedural activities. These results may provide some insight into how errors are introduced into automated systems; investigators have observed the results of programming errors but generally have not identified the mechanism by which the errors were introduced into the system. Interrupting FMC programming, particularly to respond to ATC, may open a "window of opportunity" for error.

On the whole, the data showed little evidence of errors and reflect the expertise and professionalism expected in air carrier operation. Training organizations, however, may wish to review how they simulate ATC communications and emphasize communication procedures in their recurrent training.

REFERENCES

Chou, C., Madhavan, D., and Funk, K. (1996). Studies of cockpit management errors. *The International Journal of Aviation Psychology, 6(4)*, 307-320.

Cytel Software Corp. (1993). LogXact-turbo: Logistic regression software featuring exact methods. Cambridge, MA.

Funk, K. (1991). Cockpit task management: preliminary definitions, normative theory, error taxonomy, and design recommendations. *The International Journal of Aviation Psychology, 1(4)*, 271-286.

Latorella, K.A. (1996). Investigating interruptions: an example from the flightdeck. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting*, 249-253.

Longridge, T. (1995, October). Personal communication.

Rogers, W.H. (1996). Flight deck management: a cognitive engineering analysis. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting*, 239-243.

Ryan, T.A. (1960). Significance tests for multiple comparison of proportions, variances, and other statistics. *Psychological Bulletin, 57*, 318-328.

Schutte, P.C., and Trujillo, A.C. (1996). Flight crew task management in non-normal situation. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting*, 244-248.

Tabachnick. B.T., and Fidell, L.S. (1996). *Using Multivariate Statisics* (third edition). New York: Adidson-Wesley-Longman.

Wickens, C.D. (1992) *Engineering Psychology and Human Performance*. New York: HarperCollins.

Wiener, E.L., and Curry, R.E. (1980). Flight-deck automation: promises and problems. *Ergonomics, 23*, 995-1011.

Williams, A.C. (1971). Preliminary analysis of information required by pilots for instrument flight. In. S.N. Roscoe (Ed.), *Aviation Research Monographs 1(1)* (pp. 1-17). Urbana-Champaign, IL: Aviation Research Laboratory, Univ. of Illinois at Urbana-Champaign. (Reprinted from Interim Report 71-16-1, 1947).

ACKNOWLEDGMENTS

This research, Individual Differences in the Strategic Management of Workload for Air Carrier Pilots, was supported under Cooperative Agreement #NCC 2-910 from NASA-Ames Research Center to the University of Southern California. Dr. Everett Palmer was the technical monitor. I would like to thank the participating airlines and ALPA for their support of this research. The author thanks Dr. Barbara Tabachnick for her assistance with the data analyses.

OTHER PUBLICATIONS PRODUCED UNDER THIS GRANT

Damos, D.L. (1997). The effects of operational interruptions on flight crew performance. *Proceedings of the 1997 Flight Crew Training Conference*. Newcastle upon Tyne, UK: EAC Aviation.

Damos, D.L. (1997). *Using interruptions to identify task prioritization in Part 121 air carrier operations.* Paper presented at the Ninth International Symposium on Aviation Psychology.