PILOT WORKLOAD IN THE AIR TRANSPORT ENVIRONMENT: MEASUREMENT, THEORY, AND THE INFLUENCE OF AIR TRAFFIC CONTROL

Jeffrey G. Katz

May 1980
PILOT WORKLOAD IN THE AIR TRANSPORT ENVIRONMENT:
MEASUREMENT, THEORY, AND THE INFLUENCE OF
AIR TRAFFIC CONTROL

JEFFREY G. KATZ

Flight Transportation Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

May 1980
ABSTRACT

The operating environment of an air transport crew is characterized by multiple interrupting tasks, these tasks being composed of a mixture of purely physical control and purely mental planning processes. Measurement of crew workload is thus a difficult undertaking due to the necessity to resolve workload contributions imposed by several sources. These sources include physical efforts, mental efforts, random task interruptions, and emotional disturbances. A multiattribute, subjective opinion rating scale is presented for use as an effective measure for this air transport cockpit environment.

An analysis is performed which indicates that a major component of workload is induced by the federal air traffic control system. Mechanizations of this loading include speed and altitude restrictions imposed by regulation, confinement and restraint imposed by the structure of the National Airspace System, and loads induced by a stochastic interruption process associated with ATC voice communications. In fact, the analysis of a routine transport arrival into Boston's Logan airport indicates that the (primarily system induced) workload levels in the terminal area, may be higher than the (primarily aircraft induced) workload levels on final approach.

A fixed base, Boeing 707 simulator was employed to investigate the consistency, sensitivity, and acceptability of the subjective rating scale. Four airline pilots and four general aviation, IFE pilots flew a series of routine, IFE arrivals from high altitude cruise into Boston's Logan airport, each arrival terminating with a standard instrument approach. Consistent ratings were achieved across the airline subjects for all segments of the arrivals. In general, all subjects seemed receptive to the subjective assessment methodology.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td></td>
</tr>
<tr>
<td>Motivation</td>
<td>1</td>
</tr>
<tr>
<td>Objectives and Organization</td>
<td>2</td>
</tr>
<tr>
<td>II. CONCEPTS OF WORKLOAD THEORY</td>
<td>4</td>
</tr>
<tr>
<td>Workload and Mental Workload</td>
<td>4</td>
</tr>
<tr>
<td>Functional Definitions of Mental Workload</td>
<td>11</td>
</tr>
<tr>
<td>Workload Capacity and Related Theory</td>
<td>17</td>
</tr>
<tr>
<td>Summary</td>
<td>22</td>
</tr>
<tr>
<td>III. MEASURES OF MENTAL WORKLOAD</td>
<td>24</td>
</tr>
<tr>
<td>Physiological Measures</td>
<td>25</td>
</tr>
<tr>
<td>Brain Electrical Activity</td>
<td>26</td>
</tr>
<tr>
<td>Pupil Dilation</td>
<td>29</td>
</tr>
<tr>
<td>Heart Rate Variability</td>
<td>31</td>
</tr>
<tr>
<td>Behavioral Measures of Mental Workload</td>
<td>33</td>
</tr>
<tr>
<td>Primary Task Measures</td>
<td>33</td>
</tr>
<tr>
<td>Spare Mental Capacity</td>
<td>36</td>
</tr>
<tr>
<td>Subjective Opinions</td>
<td>40</td>
</tr>
<tr>
<td>Summary</td>
<td>48</td>
</tr>
<tr>
<td>IV. PILOT WORKLOAD AND THE INFLUENCE OF AIR TRAFFIC CONTROL</td>
<td>50</td>
</tr>
<tr>
<td>Loading Mechanisms</td>
<td>51</td>
</tr>
<tr>
<td>The Airspace Loading Mechanism</td>
<td>52</td>
</tr>
<tr>
<td>Regulatory Loading Mechanisms</td>
<td>53</td>
</tr>
<tr>
<td>Discussion of the Regulatory and Airspace Mechanism</td>
<td>73</td>
</tr>
<tr>
<td>ATC Voice Communication Mechanisms</td>
<td>75</td>
</tr>
<tr>
<td>Explicit Loading</td>
<td>75</td>
</tr>
<tr>
<td>Implicit Loading</td>
<td>77</td>
</tr>
<tr>
<td>V. WORKLOAD ANALYSIS OF A DCSTON ARRIVAL</td>
<td>91</td>
</tr>
<tr>
<td>The Scenario</td>
<td>92</td>
</tr>
<tr>
<td>Analysis Tools</td>
<td>94</td>
</tr>
<tr>
<td>Cockpit Activity Timelines</td>
<td>95</td>
</tr>
<tr>
<td>Task Precedence Maps</td>
<td>97</td>
</tr>
</tbody>
</table>
Chapter I
INTRODUCTION

Pilot workload has long been a concern for the operators and manufacturers of aircraft. This concern stemmed from theory which states that beyond some threshold of workload, a pilot's performance will be degraded to unsafe levels and his ability to cope with unexpected emergencies will be seriously impaired. Unacceptable levels of workload, therefore jeopardize safety in a system where the costs associated with error are intolerably high.

1.1 MOTIVATION

The late 1970s saw a series of major airline accidents that brought increased public attention to issues of air safety. Similarly, the early 1980s will witness the introduction of at least three new technology, transport category aircraft, the Douglas DC9-80, the Boeing 767, and the Boeing 757. These events are all significant contributors to recent, heightened research efforts in areas related to the human factors of cockpit design and operating safety. For, at this time of increased public concern, there remain few acceptable answers to the following questions:
1. Should an aircraft have a two or three man crew?

2. Against what standards should the safe use of new cockpit displays and instrumentation be certificated.

3. Against what standards should new flight procedures be evaluated.

The answers to the above questions require the ability to understand and measure the workload of an air transport crew. At present neither can be done to the satisfaction of the aviation community.

1.2 OBJECTIVES AND ORGANIZATION

This thesis research was conducted as part of a multidisciplinary effort at MIT over the two academic years, 1978-1980. It addresses several aspects of the workload problem. First, it addresses the state of recent workload theory as applied to the duties and environment of an air transport crew operating under present-day Instrument Flight Rules (IFR). Related to this the literature on workload measures is discussed with a particular emphasis on how these measures are adapted for use in the air transport environment.
Next, the linkage or interface between cockpit and air traffic control system is examined. The activity in each of these two systems strongly influences activity in the other. A framework for analyzing this interface is described that, in particular, is useful in examining the significant influence of the air traffic control system on pilot workload.

Finally, a complete analysis of a Boeing 707 arrival into Boston is described. This scenario serves as an analytic baseline for a series of flight simulator experiments which were devised to evaluate a pilot workload 'rating scale' developed at MIT. These experiments also were designed to explore some of the more theoretical constructs examined in the initial portions of this work. The flight simulator experiments and their results are discussed.

Pilot workload is, today, a 'topical' subject. It is, however, very complicated subject to analyze or to understand conceptually. The fundamental objective of this work is to place pilot workload in perspective by identifying and examining key aspects of the problem. In this thesis a combination of theory, analysis and experiment are employed in this task.

------------------

The aircraft systems includes crew, aircraft, and aircraft dynamics. The AIC system includes system architecture, the controller, and other aircraft.

- 3 -
Chapter II

CONCEPTS OF WORKLOAD THEORY

This chapter lays a framework which serves as the basis for the discussions of later chapters. The concepts of workload and workload measurement are elaborated by collating material from existing literature and from research performed in conjunction with the present thesis. In this chapter the theory and conceptual description of workload is explained in a general sense. The following chapter discusses workload measurement techniques for the IFE, air transport application.

2.1 WORKLOAD AND MENTAL WORKLOAD

Historically the major study of operator workload was concentrated on physical workload. That is, researchers and designers were mostly concerned with anatomical restrictions and limits of human capabilities in terms of strength, effort and physical exertion. Later it became apparent that purely mental, non-physical effort could also seriously impair an operator's performance or his potential to err. Consequently the literature has more recently recognized the importance of mental workload.
This thesis will deal almost exclusively with either total workload or mental workload. It is convenient to state that

$$\text{Total Workload} = \text{Physical Workload} + \text{Mental Workload}$$

Physical workload is quantifiable and research continues today into physical loads on the human operator. Physical workload can be measured in terms of hard physiological parameters such as heart rate, perspiration rate and carbon dioxide expiration rate.

However, in today's air transport environment the mental workload component is becoming increasingly the most important component of total workload and, for this reason, it alone is discussed in this thesis. Today's cockpits are increasing in simplicity and automation. The pilot's job has been transformed into that of a 'flight manager' (see Lavesson, 1978). As the pilot manages his complex system through the air, indeed physical levels of effort are quite low and continually seemed to be reduced. Although physical loads cannot be ignored, mental loads have become increasingly more complex and crucial to understanding pilot

*See Hay, et al, 1978, for an example of physical task reductions in the DC9 series aircraft.*

- 5 -
workload.

To this end, it must be clearly stated that mental workload is not performance nor is it directly proportional to the number of tasks being performed. Mental workload is an 'intervening' variable (see Simpson and Sheridan, 1978), it is not directly observable. In general it can be said that mental workload is some combination of mental effort, information processing, and emotion in response to task demand. Mental workload has to do with a sense of mental effort, how hard one feels one is working. As Simpson and Sheridan (1978) point out 'one person can claim to have a feeling of great mental effort, while another can claim to be exerting no mental effort at all -- and both be performing at the same level.

Sheridan and Stassen (1978) show in schematic form how the human is interfaced to his system and postulate the linkages through which total workload can influence task performance. Figure 2.1 is a reproduction of this concept which shows, in system form, the various interactions that take place when a human controller (operator) works on a task. The figure theorizes how mental workload is involved in affecting task performance and it is clear, from the figure, that it need not be directly related to task performance. The figure also shows external variables which may be measured (M1, M2 and M3) to be representative of workload.
It should be noted, though, that Sheridan's model of workload and performance shows the M variables to be only indirect measures of workload.

For a theoretical and conceptual understanding of workload it is necessary to have a framework for defining mental workload or a paradigm for analysis. Sheridan and Simpson (1978) have constructed a paradigm for the pilot workload case which is reproduced in figure 2.2. They note that, in the piloting task, there are three distinct functions which affect workload and which also affect performance.

1. There is a judgement function that is dependent on pilot experience and that will contain his long term planning efforts.

2. There is a supervisory function which allocates time and effort to be spent on tasks. This is a real-time, short term planning function.

3. There is the more fundamental sensory motor function that contains in-bred task skill and represents almost reflexive action of the operator.
Figure 2.1: Human Operator Control Paradigm
Figure 2.2: Qualitative Paradigm for Pilot Mental Workload
The paradigm of figure 2.2 fits many of the known characteristics of pilot workload and pilot performance. The paradigm also suggests that at some skill levels, sensory motor functions will require virtually no mental effort. Thus some pilots will feel flying an ILS approach is simple and requires little effort. Yet pilots on a low skill level will expend considerable effort controlling the aircraft and will not be able to adequately perform the long term planning or short term supervisory functions adequately. For this less skilled pilot, then, the workload may well be very high.

This formulation of mental workload also suggests that the long term planning function can load the pilot, as can the supervisory function, even though sensory motor tasks do not. This is a matter of great practical importance as automation relieves the pilots of many physical, sensory motor tasks (see Yoeger, 1978).

Finally, it should be noted that Sheridan and Simpson's framework theorizes that both the aircraft systems and the operating environment affect the mental workload of the IFR pilot. The human factors design of aircraft systems reflect the importance of this influence. However, the influence of the operating environment is difficult to predict, a priori, and has been generally ignored in workload studies. The influence of the operating environment, specifically the air traffic control system is the subject of Chapter 4.
2.2 **FUNCTIONAL DEFINITIONS OF MENTAL WORKLOAD**

There is no generally accepted definition of mental workload despite years of research in mental workload and operator workload. A recent FAA Task Force study (see Hay, et al, 1978) of the state of the art, reviewed some common definitions. This report cites two quite different approaches to a definition:

"Level of pilot workload ... is determined by ... the aggregate of the task demands placed on the pilot by the system ... coupled with the actions required of the pilot to satisfy those demands."

"Level of workload equals the sum of somatic energy expended and task difficulty multiplied by the duration of this activity."

In this thesis the functional definitions described in chapter 2.1 will be used. It is beyond the scope of the present research to go further than this in an attempt to coin a new definition for workload or mental workload. It is useful, however, to discuss in more detail the theoretical form of the relationships between task performance, mental effort and mental workload. The ensuing discussion draws heavily on the previously cited work of Simpson and Sheridan (1978)."
Formulation of a Workload Model

The IFF piloting environment is characterized as a system where the pilot is continuously working on many tasks. The tasks occur randomly (at time $t_i$ for task $i$) and often tasks must be worked on simultaneously. This is essentially the paradigm Tula (1978) studied in his thesis on supervisory control behavior. In Tula's paradigm each task required a different amount of time or effort to complete and for each task, $i$, there was a deadline, $t_d(i)$, by which time the task must have been completed. This scenario is quite representative of the piloting task and it is shown schematically by figure 2.3.

In general, it is assumed that task performance, $P_i$, on task $i$, is a function of the behavioral effort expended on that task. Note that performance, in this model, is postulated to be a function not of mental effort, but of behavioral effort. The individual pilot places some utility on a performance level achieved on task $i$ and a desired utility leads him to expend the required amount of behavioral effort. In algebraic terms

$$P_i(B) = U_i = U_i(P) \text{ Desired}$$

---

Figure 2.3: Task Arrivals, Deadlines and Mental Effort
Figure 2.4: Task Performance Function, \( P_i(B_i) \)
To meet his own performance criteria the pilot will expend the required behavioral effort. This is shown by figure 2.4.

It is postulated that there is a monotonic relationship between behavioral activity and mental work. As figure 2.5 shows, increasing skill levels allow the same amount of behavioral activity to be performed with lesser expenditures of mental effort.

Mental effort then is a time dependent but instantaneous workload parameter \( M_p(t) \) for an individual pilot. Mental work is the time integral of mental effort

\[
\text{Mental Work} = \int_0^t M_p(t) \, dt
\]

This work is dependent on both the task, \( i \), and the pilot, \( p \).

The pilot works in a multiple task environment. In some time interval over which workload is measured, the pilot may work on tasks simultaneously, or nearly so*. The total mental effort level, summed over all tasks, and using a moving average of interval \( T \) gives

\[
ML(t) = \frac{1}{T} \sum_{i=1}^{p} \left[ \int_{t-T}^{t} M_i(t) \, dt + \int_{t-T}^{t} M_i(t-\delta) \, dt \right]
\]

*It is a matter of contention whether pilots work on tasks simultaneously or whether they time-share. This point does not affect the present argument but may affect the choice of a model for a human operator.
Figure 2.5: Behavioral Activity as a Function of Mental Work

Figure 2.6: Mental Effort Level and Instantaneous Mental Effort
As shown in figure 2.6 this moving average tends to smooth the instantaneous work effort so that instantaneous work rate may exceed the average effort level at times.

2.3 WORKLOAD CAPACITY AND RELATED THEORY

With these description of the concepts of workload theory it is now possible to take one additional step. This step relates workload to a workload limit or capacity and it relates workload to performance.*

Simpson and Sheridan state that there is mental work associated with each task, and with each psychomotor function. Total mental work is then the summation across all tasks and all functions.

$$\text{Total Mental Work} = \sum_{i,j}^{\text{Pilot } p} M_{ij}^{p}$$

Pilot $p$ $i$ - task $j$ - function

Associated with each psychomotor element is some 'maximum work level or capacity, $MCAP_j^p$, for a given pilot, $p'.$ An index of the utilization or loading of each psychomotor element is represented by

$$\rho_j^p = \frac{ML_j^p}{MCAP_j^p} \quad \rho < \rho_j^p \leq 1$$

*This section also draws heavily on Simpson and Sheridan, op cit.
Saturation of this element occurs when $P = 1$.

It is then theorized that an individual pilot will maximize the utility derived from a certain performance level less the disutility of the mental workload. Mathematically this can be stated as

$$\max_{P} \left[ \sum_{i} U_i(Z) - U_i(N) \right]$$

subject to $ML_j < MCAP_j$ for each psychomotor element $j$ and $P > P$ for all $i$ and task $i$ (min. performance level)

**Implication of the Model**

The most important implication of the previously discussed workload formulation to this thesis relates to the presence of multiple, randomly occurring tasks. The model suggests (see figure 2.7) that at different times different psychomotor elements are constraining the pilot's workload. Different types of task load affect different elements in a unique manner. Consequently, there will exist tradeoffs in performance on varying tasks that occur when maximizing the objective function cited in the formulation above. This task tradeoff occurs explicitly due to the multiplicity and simultaneity of tasks in the piloting environment (see figure 2.8). In fact, queueing theory would suggest (see Senders and Posner, 1976) that performance will degrade as the probability of tasks of higher priority than presently being worked on, increases. This is shown by figure 2.9, that is also taken from Simpson and Sheridan.
Pilot Mental Work on Task 1

Motor component, \( j=21 \)
Sensory component, \( j=1 \)
Motor component, \( j=22 \)

Figure 2.7: Mental Work Capacities for Two Simultaneous Tasks

Pilot Performance on Task 1, \( P_1 \)

Pilot Performance on Task 2, \( P_2 \)

Figure 2.8: Performance Level Tradeoff for Two Simultaneous Tasks
Overall Task Performance

Figure 2.9: Overall Task Performance vs. Workload Ratio

regular task arrivals
random task arrivals

Overall Workload Ratio
Simpson and Sheridan have identified three categories of task in the IFR piloting environment that define a hierarchy of task priority and precedence. This hierarchy recognizes the existence of a task multiplicity and interruption dimension to pilot workload. The task categories are

1. Operating tasks - tasks concerned directly with the operation of the aircraft which must be handled immediately. Examples include aircraft maneuvering and control, and auditing radio communications.

2. Monitoring tasks - tasks of critical importance that may be delayed for a short period of time while operating tasks are being performed. Examples include systems monitoring, out-the-window visual scan, and instrument cross checking.

3. Planning tasks - tasks deferrable into idle periods of time. Examples include retrieving ATIS information, planning approaches, computing landing data.
If a monitoring task is cued while a planning task is being performed, the planning task will be deferred until the monitoring task has been completed. Similarly, the occurrence of an operating task will preempt both planning and monitoring tasks. The rate and randomness with which the three categories of task occur will influence the workload level perceived by a crewmember.

2.4 SUMMARY

In the air transport environment the mental workload component is increasingly becoming the most significant portion of total pilot workload. Despite this importance, mental workload has no generally accepted definition. It can be thought of as some weighted combination of mental effort, information processing, and emotion in response to an operator's task demand. Theory would suggest that mental load is not a simple function of task performance.

The task load of an air transport crew is characterized by a randomly occurring series of multiple tasks. These tasks tend to occur simultaneously and to interrupt one another. Pilots are thus trained to prioritize these tasks, ranking each task with a certain relative importance. Certain types of task are easily deferrable for sustained periods (planning tasks). Others are deferrable for only brief periods (monitoring tasks), while those of highest priority (operating tasks) must be performed promptly. It is postu-
lated that the perceived workload of the pilot will be influenced by the magnitude of this interruption/queueing process.

These simple concepts of workload theory are a useful basis on which to build experiments and analysis. The paradigm of the IFR pilot presented in this chapter is used in the further theoretical discussions of Chapter 4, in the analysis of Chapter 5, and in the experiments presented in Chapter 6.
Chapter III
MEASURES OF MENTAL WORKLOAD

The total workload of a pilot is the sum of his physical and mental workload. In the air transport environment, physical workloads are becoming the insignificant component of total workload as cockpits transition to a state of automation and integrated, simplistic designs (see Hay, et al 1978). Indeed the pilot's duties in the late 1970's make him more of a flight manager and supervisor than the 'stick jockey' of an earlier era or aviation (see Babcock, 1976).

Physical workload is measurable by a number of methods. These include the indices of respiration rate, heart and perspiration rate. Mental workload, however, is a relatively new concern to the engineer and operator. As physical loads have reduced in magnitude due to the introduction of automated systems, mental load may have been increased due to new complexities and uncertainties in decision making. Mental workload is today the most significant component of total workload in many operational circumstances and it is a concept that is, as yet, not well understood.

In this chapter attention is focused on measures of mental workload. Candidate measures for the air transport
environment are discussed and evaluated in terms of the results of previous research and with respect to the conceptual theory outlined in Chapter 2. An attempt is made to show how suitable these candidate techniques are for measuring the workload of an IFR, air transport pilot.

This chapter is not meant to be a compendium of all possible measures of workload (see instead Wierwille, 1979; Williges and Wierwille, 1979). Rather, this chapter surveys measures which are most likely to yield promising results in the air transport application or which have seen very common usage in measuring aircrew mental workload.

3.1 PHYSIOLOGICAL MEASURES

Physiological measures of mental workload are those methodologies which seek to infer workload levels by monitoring quantifiable, physiological signals. Examples of these signals are heart rate, skin temperature, and brain electrical activity. This class of measure is promising because it portends to remove much of the uncertainty and subjectivity from workload measurement by tapping physiological signals and matching these to known physiological functions of mental workload. These types of measures are still not well understood but there is a great deal of basic research underway that should yield improvements in the use of physiological workload measures.
3.1.1 Brain Electrical Activity

These techniques utilize electroencephalograph records of brain electrical activity. It is theorized that the neurons of the brain emit an electrical stimulus (called an Event Related Potential or ERP) that corresponds to the information processing associated with a discrete external event.

The signal from an event related stimulus (also referred to in the literature as an evoked response) is low level and imbedded in the noise of an EEG record. Signal averaging techniques must thus be used to extract the ERP pulse which typically has a magnitude of 10 to 250 microvolts and pulse widths of 500 milliseconds (see Donchin, 1978).

Figure 3.1 shows a typical ERP trace that was extracted using signal averaging techniques. The signal peaks are labelled by their positivity and by the time after the triggering event at which they occur. For example, a P300 peak is a positive peak occurring 300 milliseconds after the triggering epoch. As figure 3.2 indicates, various regions of the evoked response are theorized to correspond to various types of mental information processing.

Of particular interest is the P300 peak whose magnitude and latency have been found to vary as mental tasks become more complex. Research has shown that the P300 is affected by both memory tasks and by purely cognitive processing tasks. It thus may eventually serve as a direct index of certain categories of mental loading.
Figure 3.1: An Evoked Response Potential (ERP)

Figure 3.2: Components of the Visual Evoked Response
Unfortunately, the use of brain electrical activity for workload measurement still has some practical problems that must be overcome. The primary problem is that the technique as applied to operator workload is still in an infant state of research. Three problems that must yet be solved are listed below.

1. Continuous task and multiple tasks (e.g. piloting tasks) are not presently amenable to this technique.

2. Different mental processes interfere in different ways with one another and the effect of this on the P300 is not well known.

3. Signal averaging techniques require multiple tests of the same event on the same subject.

The great advantage in using the ERP is that the signals that directly reflect cognitive processing and loading are being tapped rather than an inference made based on some variable external to the system. This offers a future possibility of directly and repeatably assessing the mental loading of a subject.
3.1.2 Pupil Dilation

The pupil dilation technique seeks to infer mental loading by monitoring the expansion and contraction of the pupil, relative to some baseline, as a subject works on some task. The technique is based on theory which states that, as information processing loads are increased, a forward surge of 'activation' results in the brain stem* (see Beatty, 1978). By coincidence, the brain stem also contains the nervous system mechanism that controls pupil dilation. Thus it appears that pupil dilation could serve as an index of information processing load.

An instrument called a Whittaker Pupilliometer can be used to accurately measure the pupil diameter. Using this device it has been found that as mental loads induced by a task increase, the diameter of the pupil increases (see Beatty). Pupil response to task loading generally yields pupil diameter increases of fractions of a millimeter within 10 seconds of a loading increment.

The pupil dilation method for assessing mental load may have immediate application in analyzing, one at a time, discrete cockpit tasks. This technique would allow the workload required for each of the component subtasks in a larger piloting task to be measured. These measures could provide very useful quantitative workload numbers for use in cockpit

*The brain stem is the uppermost portion of the spinal cord.
design or flight procedure analysis.

There are two serious deficiencies of the pupil dilation method. This measurement technique requires that the eyes' dilation due to changes in mental load be resolved. However, the eye also responds to changes in lighting intensity and these changes make it difficult to detect changes due to mental loading. Air transport cockpits are subject to spatially varying intensities of light and so are not a good testbed for the pupil dilation method. Also, new electronic flight instrument displays (like those of the Boeing 757 and 767) may accentuate this problem.

The pupil dilation method is also untried in continuous task type environments. As a direct measure of mental workload in an air transport cockpit, then, this technique is still inadequate although it may provide useful information as was mentioned above. Compared to the brain electrical activity techniques mentioned in 3.1.1, the pupil dilation technique may possess an additional advantage. The techniques of 3.1.1 measure the response to a task or probe external or secondary to the task of interest. The pupil dilation method, on the other hand, measures the direct response of the nervous system to increases in processing load from the primary task of interest. As more is learned about the technique this advantage may prove the pupil dilation data more valuable than other, more indirect physiological measures of mental load.
3.1.3 Heart Rate Variability

Heart rate variability (often referred to as cardiac or sinus arrhythmia) is a technique that has shown some usefulness in basic research (see Kalsbeek, 1973). Cardiac arrhythmia is calculated by measuring the duration between every two heart beats and from this computing an instantaneous estimate of heart rate. The variance of these estimates then serves as an expression of cardiac arrhythmia. Kalsbeek found lower heart rate variability at higher levels of mental loading.

It should be mentioned that heart rate variability is not equivalent to an averaged heart rate. Heart rate was one of the early methods for assessing mental loads. However, heart rate has been found to be an insignificant index of mental loading. Most recently Smith (1979) did an extensive, full mission simulation where heart rate was recorded throughout. Smith attempted to correlate percentage changes (from some baseline) in heart rate with other indices of high workload. These 'other' indices of workload included number of errors in performing duties, decision times associated with tasks, and vigilance. Smith found no statistically significant correlation between heart rate and any of the three variables. In fact, heart rate was found to be significant only with respect to whether the subject was acting as 'Pilot at Controls' during the full mission simulation.
In a recent comparative study of mental workload assessment techniques Hicks and Wierwille (1979) evaluated the sensitivity of the heart rate variability technique. Their experiment utilized a moving base driving simulator that was subjected to a side wind gust whose magnitude and location was changed to increment workload. The heart rate variability technique was found to have no statistically significant relationship to their operationally defined level of workload. The technique was also found (issues of significance aside) to be less sensitive to workload than either subjective opinion ratings or primary task performance as an index of mental load. In their paper, Hicks and Wierwille state that heart rate variability may have shown significant workload effects if the number of test subjects had been increased, but doubt that this would change its relative sensitivity as a workload index when compared with the subjective opinion and primary task performance techniques.

Heart rate variability does not show much promise as an index of mental workload for the IFR, air transport pilot. The fundamental problem of the measurement method probably lies in the fact that heart rate and heart rate variability are influenced by many factors besides mental workload and the piloting environment. These factors include physical effort requirements and physical deficiencies, neither of which may be separable from task related, cardiac arrythmia data.
Mulder (1978) has suggested breaking the heart rate data into its frequency spectrum components. Mulder has found that certain segments of a heart rate spectrum are affected by the degree of mental loading. Still, this has not been found to solve the primary problem of heart rate techniques. That is, a great many factors affect heart rate variability other than mental loading.

3.2 BEHAVIORAL MEASURES OF MENTAL WORKLOAD

Behavioral measures seek to infer mental workload from the observation of some behavioral output of the aircraft/pilot system. Williges and Wierwille (1979) have grouped fourteen behavioral workload measures into one of three major categories:

1. Primary Task Measures
2. Spare Mental Capacity Measures
3. Subjective Opinions

Measures from each of these categories are discussed below.

3.2.1 Primary Task Measures

It is hypothesized that as the mental workload of a crew-member increases, his performance may degrade. Primary task measures then, attempt to correlate changes in observed crew performance with changes in mental workload. It should be
noted that the hypothesis underlying the use of primary task measures is a point of some contention. Simpson and Sheridan (1978), for instance, contend that performance is not equivalent, nor simply related to mental workload (see Chapter 2). Nevertheless, primary task measures have been used extensively in workload analysis and, in general, have been found helpful in understanding what is a very complex problem.

Examples of primary task measures are numerous. They include tracking performance, dual tracking performance, and the tabulation of number of task errors. Generally, one of these is related to an operationally defined metric of workload (e.g. turbulence intensity, or no. of turns/second, etc.). From these relationships the magnitude of workload as measured by primary task performance is then defined.

Hicks and Wierwille (1979) performed a comparative study of five workload assessment techniques. In a six degree of freedom driving simulator they used five workload measures. These were: lane tracking performance, subjective opinion scales utilizing the method of equal appearing intervals, visual occlusion, heart rate variability, and secondary task performance. Workload was incremented by changing the magnitude and location of wind gusts on the vehicle. Of the five techniques, primary task performance appeared to be the most sensitive to variations in the operationally defined increments of workload.
Primary task measures of mental workload have limitations in their usefulness for the air transport pilot application. First, in many phases of flight there is no simple performance measure which the pilot accurately controls. For instance, in the airport terminal area where aircraft are generally being vectored, a pilot will be commanded to hold a heading, but this pilot may not be concerned about the accuracy with which he holds this heading, as he might be when tracking the localizer course on final approach. A problem related to this is that pilots have different personal error and accuracy criteria. For example, while being vectored, some pilots may consider it a matter of personal pride to hold their vector heading +/- 2 degrees, while another pilot holds his heading +/- 5 degrees with the knowledge that the controller will give him a new vector if the aircraft drifts off course.

Next, there is the problem of cockpit automation. In most phases of an air transport mission at least one control axis of the aircraft is controlled by an autopilot. For the example of vectors in the terminal area, the pilot will most likely set the autopilot heading bug to the commanded vector. The autopilot will then accurately track the heading, leaving the pilot free to attend to other duties. In the role of 'flight manager' pilots are trained to delegate more of the cockpit tasks to the automatic flight control system as workloads become high. With the autopilot in control, it
may be difficult to specify a primary task measure of pilot workload.

Finally, the most fundamental problem of primary task measures of mental workload lies in the fact that pilots can and do exert more mental effort to maintain a specified level of performance. An illustrative example of this is a pilot flying an ILS approach. Despite the occurrence of minor difficulties in the cockpit, the pilot's major objective is to guide the aircraft precisely down the ILS. He will disregard other tasks, or place the aircraft in auto-land if necessary, but he will keep the aircraft on the ILS course with strict precision. A pilot in this phase of flight is very likely exerting increased levels of mental effort but his performance is unvarying.

3.2.2 *Spare Mental Capacity*

Spare mental capacity techniques assume that the human operator is basically a single channel instrument or single CPU computer system. That is, he has some fixed upper limit of processing capability and tasks on which the operator works absorb some percentage of this fixed capacity. Although this model of the human operator has not, in fact, been verified, it is a useful construct that has served as the basic foundation of much human factors research.
The most common mental workload measures using the theory of spare mental capacity are task component/time summation methods, and secondary or side-task measures. Task component/time summation methods analyze the microscopic structure of an operator's task load, assume a fraction of total time or total mental capacity that each task requires, and performs some weighted average over the total number of tasks or over the total number of seconds. This averaging process results in a workload figure measured in percent of total operator capacity. Hay, House, and Sulzer (1978) provide examples of this process. Task analytic methods have also been tied to sophisticated computer simulations of a human operator, to aid in the design of cockpit and flight systems (see for example Wherry, 1978; Hay et al, 1978).

Task analytic approaches such as these have severe limitations in the rather absurd arbitrariness which must be applied to the weights used to arrive at a percentage workload figure. Moreover, even the accuracy with which times can be associated to tasks, and the detail with which the task composition can be defined, is subject to question. Task component methods, then, offer at best an early design stage measure of mental workload.

Secondary task measures are the second common classification within the spare mental capacity category of workload measurement. These techniques require the subject to per-
form a secondary task (also called side task). The subject's performance on the side task is measured as he works on a major or primary task set of interest. The hypothesis underlying these techniques is that, if the subject's performance has degraded on the secondary task, he must therefore be working harder on the primary task. That is, the execution of the primary task is requiring the use of a certain percentage of the subject's spare mental capacity.

The specific type of secondary task employed generally varies with the application. Examples of side tasks include tracking side tasks, adaptive tracking side tasks (difficulty varies with side task performance) *, light cancelling tasks **, and mental arithmetic tasks ***.

One side task that has shown promising results as an indicator of mental load is subjective time estimation (see Hart, 1978). These techniques require the subject to estimate 10 to 15 second time intervals while performing a primary set of tasks. The intervals are estimated actively (i.e. have subject tap foot or repeatedly sing bars of a song), or retrospectively (i.e. by comparing the number and complexity of events within an interval with the remembered duration of intervals similarly filled). Time estimation

*See Ogden, 1977
**See Ephrath, 1975
tion techniques measure the degree of reliability and accuracy of the subject's time estimates while the primary tasks are performed and compare these with their counterpart baseline time estimates. Hart (1978) examines variations in shape, central tendency and variability of time estimates as task demand changes.

Secondary tasks, though commonly used in workload measurement have some fundamental problems in their use. Dominant among these is that secondary tasks are hard to differentiate from primary tasks. An experimental subject is very likely to devote much of his effort to working on the secondary, rather than the primary task. The use of adaptive secondary tasks, to some extent, reduces the magnitude of this problem. As with any measurement system, however, the intrusiveness of the measure itself will affect the measured results. The intrusiveness of secondary tasks seems to be a serious limitation of spare mental capacity methodologies.

For the air transport application, secondary tasks suffer yet another limitation. An IFR pilot works in a multiple task environment. He may work on none or on five tasks at any given time in a flight mission. The secondary task must thus be 'trained' into the priority structure with which a pilot performs other piloting duties. The introduction of a secondary task, then, may not indicate the 'true' percentage of spare mental capacity in such a time shared, queuing-type task environment.
3.2.3 Subjective Opinions

Subjective opinions are perhaps the most common measure of mental workload in use. Sheridan (1980) would suggest that mental workload be defined as 'a person's private subjective experience of his or her own cognitive effort.' Indeed most experimenters do calibrate their 'objective' measures against a subjective scale or interview. Like the other measures discussed in this chapter, subjective opinions are generally used in conjunction with other workload measures and have been historically useful in this regard.

Subjective opinion techniques utilize two basic formats:

1. Interview/Questionnaire

2. Rating Scales

Interviews and questionnaires are generally loosely structured, information gathering devices. Their usefulness as a workload metric is thought to be rather limited. Rating scales, on the other hand, may be rigorously structured and implemented according to the body of literature known as psychometric theory (see Nunnally, 1967).

Gartner and Murphy (1976) state that subjective opinion rating scales offer one of the most promising methodologies for assessing aircrew mental workload, but note also that none of the scales which they surveyed were constructed according to the proper techniques of attitude scale con-
struction (see Edwards, 1957). Such scales generally are ambiguously worded, unanchored, or are subject to rating intransitivities.

The Cooper-Harper scale, used by test pilots to rate aircraft handling qualities, is perhaps the best known rating scale in use for the test and development of aircraft (see figure 3.3). Figure 3.4 is a workload rating scale that appears to have been modelled after the Cooper-Harper scale. Unfortunately, the model is too similar and the scale reads more like a handling qualities/performance scale than a workload rating scale. Also, the scale seems to emphasize, primarily, the physical components of workload, ignoring mental contributions.

Figures 3.5a and 3.5b depict a multi attribute scale for measuring the workload of an IFR pilot that was developed at MIT. This scale is also modelled after the Cooper-Harper scale but is based on the theory developed by Simpson and Sheridan (1978) on mental workload in the air transport environment. The scale of figure 3.5 uses the concepts of perceived busyness, interruptions, stress, and simultaneous occurrences of operating, monitoring, and planning type tasks (see Chapter 2.). Figure 3.5b, in turn, seeks the reason why a specific rating (from 3.5a) is chosen. It weighs the magnitude of fraction of time busy, intensity of information processing, and the intensity of feeling or emo-
<table>
<thead>
<tr>
<th>Acceptable</th>
<th>Satisfactory</th>
<th>Unacceptable</th>
<th>Unsatisfactory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capable of being controlled or managed in context of mission, with available pilot attention.</td>
<td>Meets all requirements and expectations, good enough without improvement.</td>
<td>Deficiencies which require mandatory improvement. Inadequate performance for mission even with maximum feasible pilot compensation.</td>
<td>Reluctantly acceptable. Deficiencies which warrant improvement. Performance adequate for mission with feasible pilot compensation.</td>
</tr>
<tr>
<td>Acceptable may have deficiencies which warrant improvement, but adequate for mission. Pilot compensation, if required to achieve acceptable performance is feasible.</td>
<td>Clearly adequate for mission.</td>
<td>Unacceptable deficiencies which warrant improvement. Performance adequate for mission with feasible pilot compensation.</td>
<td>Some minor but annoying deficiencies. Improvement is requested. Effect on performance is easily compensated for by pilot.</td>
</tr>
<tr>
<td>Unacceptable</td>
<td>Unsatisfactory</td>
<td>Uncontrollable</td>
<td>Uncontrollable</td>
</tr>
<tr>
<td>Unacceptable</td>
<td>Unacceptable</td>
<td>Uncontrollable</td>
<td>Uncontrollable in mission.</td>
</tr>
</tbody>
</table>

**Figure 3.3: Cooper-Harper Scale**
Figure 3.4: A Workload Rating Scale
tion. Chapter 6 of this thesis describes a flight simulator evaluation of the rating scale depicted in figure 3.5.

Like other mental workload measures, subjective opinion methodologies have limitations. For instance, the workload rating a subject provides is likely to vary with the training, skill, and emotional state of that subject. Second, there is the problem that the subject may not be able to perceive his own workload. That is, he may become so immersed in his duties that he forgets how 'hard' the test interval was, after the fact. Finally, there are questions related to scale sensitivity, and to consistency within and across subjects that may impose limitations on the use of subjective opinion rating scales.

Proper scale construction techniques, training of subjects in the use of the rating scale and a structured experimental scenario should lessen the magnitude of the above problems. Hicks and Wierwille (1979), in fact, found that a subjective rating scale was the second most sensitive indicator (next to a primary task measure) among five they examined. The scale used by Hicks and Wierwille utilized the method of equal appearing intervals.

Aside from their sensitivity, subjective opinions have another advantage in that they are almost completely non-intrusive to the test. As such, these techniques are far simpler to implement than secondary task or physiological type
methodologies. Subjective methods, then, should interfere to a much lesser extent with the final or measured outcome of a workload experiment than other assessment techniques.
IFR TRANSPORT PILOT WORKLOAD SCALE

Second Version, May 1979, MIT FTL/MML

Figure 3.5a: MIT Scale, Part I
<table>
<thead>
<tr>
<th>1</th>
<th>FRACTION OF TIME BUSY</th>
</tr>
</thead>
<tbody>
<tr>
<td>rarely have anything to do</td>
<td>often</td>
</tr>
<tr>
<td>have free moments of time:</td>
<td>occasionally</td>
</tr>
<tr>
<td></td>
<td>very rarely</td>
</tr>
<tr>
<td>fully occupied every single instant</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2</th>
<th>INTENSITY OF THINKING/INFORMATION-PROCESSING</th>
</tr>
</thead>
<tbody>
<tr>
<td>activity is completely automatic; no conscious thinking or planning required</td>
<td></td>
</tr>
<tr>
<td>cerebral effort and planning required due to problem complexity, uncertainty, unpredictability, unfamiliarity, etc., is:</td>
<td>low level, occasional</td>
</tr>
<tr>
<td></td>
<td>moderate</td>
</tr>
<tr>
<td></td>
<td>high level</td>
</tr>
<tr>
<td>supreme mental effort and concentration are absolutely necessary</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3</th>
<th>INTENSITY OF FEELING</th>
</tr>
</thead>
<tbody>
<tr>
<td>experience is relaxing, nothing to be concerned about</td>
<td></td>
</tr>
<tr>
<td>emotional stress, anxiety, worry, frustration, confusion, etc., are:</td>
<td>mild, occasional</td>
</tr>
<tr>
<td></td>
<td>moderate</td>
</tr>
<tr>
<td></td>
<td>high level</td>
</tr>
<tr>
<td>severe and intense psychological stress</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.5b: MIT Scale, Part II
3.3 **SUMMARY**

It can be truthfully stated that there is, as yet, no measure of mental workload that is well adapted for use in the air transport environment. All of the measures currently under study are in an infant state of research and far from well-understood, production type use. Physiological measures such as brain electrical activity show great promise as 'hard' indices of mental load, tapping signals directly related to the brain's information processing. These techniques are, however, the furthest of any of the techniques surveyed from production type use.

Behavioral measures of mental load are numerous and it may be that, for a small set of task demands, or for some range of workload, a currently available behavioral methodology is useable. These techniques are limited by the fact that many of the behavioral methods are based on human operator theories that have been shown to be inaccurate. An example of this is the use of primary task measures of mental load despite the fact that pilots can and do, on occasion, exert increasing levels of effort to maintain a specified constant level of performance.

Two behavioral workload assessment techniques appear to show promise for use in the air transport application. These are subjective time estimation and subjective opinion rating scales. Their use also awaits the results of further
research, but results to date indicate that these two are viable techniques.

It cannot be stated, at present, if there will ever be one universal measure of mental load, even for the piloting application. As in the past, it is more likely that a variety of assessment techniques will be used for various applications and workload ranges. It can, however, be stated that the available methodologies will be improved over those presently available to the aviation community, as theory and experimental research in this area progress to a more mature stage.
Chapter IV

PILOT WORKLOAD AND THE INFLUENCE OF AIR TRAFFIC CONTROL

The literature deals at length with the analytic and theoretic description of piloting tasks and their affect on pilot workload. The cockpit, however, is not an isolated unit consisting of pilots and flight controls. The cockpit is a component of a larger system that encompasses the pilots, the aircraft controls and dynamics, and the airway/air traffic control system. Workloads are imposed not simply due to physically operating the aircraft but also due to the air traffic control and National Airspace System itself. For example, an instrument approach is a series of turns, descents, and speed control maneuvers, the execution of which require some physical and mental effort. An FAA approved approach procedure further constrains this series of maneuvers to be executed within certain lateral and vertical limits and with a certain accuracy. Given an aircraft's approach speed, specific tasks in the cockpit are triggered by the constraints of the approach procedure. The 'approved' approach procedure is an integral component of what is referred to in this thesis as the 'ATC system.'

The cockpit procedures and the air traffic control system are interrelated and the interface between these two compo-
nents directly affects the pilots workload. In the example above both the system structure and the cockpit procedure could be changed to effect a change in the timing and execution of piloting tasks, and thus workload.

The ATC/cockpit interface is an important issue in understanding workloads in the air transport environment. This chapter explores the issue in some detail. First the mechanisms through which the cockpit and the ATC system interact are discussed in relation to the workload concepts of Chapter 2. Next, several tools for quantifying cockpit procedures are described that are useful in workload analysis. The tools are particularly useful in that they specifically address the ATC/cockpit relationship. In the next chapter a routine arrival into Boston's Logan airport is analyzed using the concepts introduced in this chapter.

4.1 LOADING MECHANISMS

There are several mechanisms through which the air traffic control system can be seen to influence pilot workload. These mechanisms can, for convenience, be grouped into three categories:

1. Airspace mechanisms

2. Regulatory mechanisms

3. ATC voice communication mechanisms
The airspace structure, the presence of Federal Aviation Regulations, and the character of air traffic control communications are assumed to be integral parts of what is referred to in this thesis as the ATC system.

4.2 THE AIRSPACE LOADING MECHANISM

The airspace loading mechanism relates to pilot loading that is invoked by the structure of the airspace system. Examples of cockpit tasks that are cued by the airspace system are numerous.

1. Task: Change radio frequency. Required as aircraft passes an air traffic sector boundary or bend in an airway.

2. Task: Climb to an enroute altitude. Aircraft generally must maintain an altitude less than their planned cruising altitude until the aircraft is clear of restrictions imposed by the terminal airspace structure.

3. Task: Instrument approach procedure. There exist well defined lateral and vertical limits of protected airspace within which the approach maneuvers must be executed
(see Terminal Enroute Procedure Standards). These limits impose constraints which tend to restrict all landing/approach tasks to be executed within a small time window.

Instrument approach procedures also define navigation fixes over which approach related tasks are executed. Figure 4.1 provides an example of specific physical cockpit tasks that are associated with navigation fixes.

Summarizing, the airspace loading mechanism operates by constraining or invoking tasks. Because the pilot's tasks are affected by the system so must be the pilot's workload, by definition (see Chapter 2). Table 4.1 lists tasks evoked by common elements of the U.S. National Airspace system.

4.3 **REGULATORY LOADING MECHANISMS**

Certain aspects and operating restrictions associated with the U.S. National Airspace system have been regulated into law. There also exists a large body of nonregulatory, but recommended set of operating procedures that have been
defined to standardize aircraft movement within the system for air traffic control purposes. This body of rules tends to restrict and predefine piloting tasks, and thus loads the pilot above the conditions encountered in an uncontrolled environment.

Regulatory loading mechanisms are a subset of the airspace loading mechanism discussed in section 4.1. It is, however, convenient to separate out this subset for organizational purposes. Table 4.2 lists the documents which specify rules and operating procedures for aircraft. It should be noted that although some of the documents contain nonregulatory information, the Federal Aviation Administration stresses adherence to these procedures, nonetheless. Table 4.3 lists common tasks invoked by the regulatory constraints, as defined by these documents.

A simple example will serve to illustrate the regulatory loading mechanism and its affect on cockpit activity. Figure 4.2 is an annotated schematic of an instrument arrival into Presque Isle, Maine. Presque Isle, though not typical of most domestic transport operations, is presently a terminal into which Delta Airlines schedules three Boeing 727 arrivals.
CONTROL SECTOR A

VOR Changeover Point
- Switch Nav to VOR C

Compulsory Reporting Point
- If non-radar environment, report position and give estimate to next fix

CONTROL SECTOR B

VOR C

ARTCC Sector Boundary
- Change Communications Frequency
- Transmit Messages

Figure 4.1: Flight Along Federal Airways
<table>
<thead>
<tr>
<th>Airspace Component</th>
<th>Associated Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>VOR</td>
<td>o Turn</td>
</tr>
<tr>
<td></td>
<td>o Navigation Frequency Change</td>
</tr>
<tr>
<td></td>
<td>o VOR Course Selector Change</td>
</tr>
<tr>
<td>Airway Intersection (Bend in Airway)</td>
<td>o Turn</td>
</tr>
<tr>
<td></td>
<td>o Reset Course Selector</td>
</tr>
<tr>
<td>VOR Changeover Point</td>
<td>o Change Navigation Frequency</td>
</tr>
<tr>
<td>Control Sector Boundary</td>
<td>o Change Communication Frequency</td>
</tr>
<tr>
<td></td>
<td>o Ident Transponder</td>
</tr>
<tr>
<td></td>
<td>o Transmit Message</td>
</tr>
<tr>
<td>Terminal Area Boundary</td>
<td>o Climb/Descend</td>
</tr>
<tr>
<td></td>
<td>o Accelerate/Decelerate to Designated Speed for Air Traffic Control</td>
</tr>
</tbody>
</table>
Table 4.2: Documents Containing Flight Operating Rules

<table>
<thead>
<tr>
<th>Document</th>
<th>Contains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Federal Aviation Regulations</td>
<td>o Part 91 - contains general rules pertaining to all aircraft</td>
</tr>
<tr>
<td></td>
<td>o Part 135 - contains additional operating rules for scheduled air taxi operations</td>
</tr>
<tr>
<td></td>
<td>o Part 121 - contains additional operating rules for scheduled air transport category operations</td>
</tr>
<tr>
<td>Advisory Circulars</td>
<td>o Contains nonregulatory advisories and practices which pertain to specific parts of the Federal Aviation Regulations</td>
</tr>
<tr>
<td>Airman's Information Manual, Part I</td>
<td>o Contains nonregulatory information regarding recommended operating practices and other general information concerning operations under Instrument Flight Rules.</td>
</tr>
</tbody>
</table>
### Table 4.3: Common Tasks Invoked by the Regulatory Mechanism

<table>
<thead>
<tr>
<th>Trigger</th>
<th>Tasks Required by Law</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descend Below 10000ft</td>
<td>o Slow to less than 250 knots</td>
</tr>
<tr>
<td>Cross 18000ft</td>
<td>o Change altimeter setting</td>
</tr>
</tbody>
</table>
| Hold at a fix                    | o Slow below maximum authorized holding speed (as specified in Airman's Information Manual, Part I.)
|                                   | o Fly holding maneuver as per AIM, Part I.                      |
| Cleared for the approach         | o Fly approach procedure as prescribed in Federal Aviation Regulations, Part 97. |
Procedure Turn Must be Executed Within 10nm of VOR

Figure 4.2: The Presque Isle Approach Procedure
From figure 4.2 and Table 4.4 it is clear that the procedural requirements of the approach constrain certain cockpit tasks to be executed at predefined points. It is not entirely clear, however, to what extent these constraints affect total workload during the approach. To measure the full impact of regulatory and airspace constraints on crew workload it is necessary to enumerate, in detail, a chronological record of cockpit activity during the approach. This record of cockpit activity is referred to as a cockpit activity timeline.

Figure 4.3 is a cockpit activity timeline* for a Boeing 727's approach to runway 19. In constructing the timeline, a priori, some assumptions were necessary. The 727 aircraft was assumed to begin its approach over the Presque Isle VOR at 2500 feet, on a heading of 359 degrees. The subject aircraft was also assumed to be at the speed and in the aircraft configuration specified by the Delta Operations Manual for the initial approach segment. The procedures shown in the timeline are taken from the Delta Airlines, Boeing 727 Pilot's Operating Manual, Book I. Cockpit event times, on the figure, were estimated using speeds called for by this manual. Wind velocity was taken to be zero.

*For readers unfamiliar with timeline analysis, a discussion is included in the appendix.
The task timeline for the Presque Isle approach shows that, in addition to the tasks that are defined in Table 4.4 by the procedure, other tasks associated with the safe operation of the aircraft must be squeezed in between the procedural constraints. For example, on Panel 3 of figure 4.3, the timeline indicates that the checklist is being read and performed by the first and second officers. This checklist must generally be completed in the 100 second period before the VOR is reached. If a difficulty should occur (e.g. a landing gear problem), the crew would have little time for decision making in this high task load situation.

Panel 4 of the timeline indicates another problem area of the procedure. According to figure 4.2, upon reaching the VOR the pilot must step down, first to 1260 feet and then, 2.5 miles later, to 1140 feet. The timeline shows that the second step down begins only 20 seconds after the first step has been completed. Moreover, if the pilot is to maintain 140 knots throughout the phase, he will be intermittently applying speed control techniques in addition to the continuous tracking tasks and the quick step-down descents. Again, the nature of the procedure has invoked this relatively higher task demand situation. Theory would suggest that at this higher loading, if difficulties should arise in the cockpit, there is a greater probability that the pilot will make an error.
Figure 4.4 is useful for identifying the constraints which the Presque Isle approach imposes on cockpit activity of the arriving 727. The boxes of figure 4.4 represent discrete events and are positioned along event paths with which they can be associated. Time, on the figure, increases from left to right so that an event to the left of another (on the same path) is constrained to precede that task.

A solid line connecting two events indicates that some task is being performed between these two and the number of seconds between such events is indicated by the number in parentheses. A dotted line indicates a soft precedence constraint. That is, there is no task being performed between these two events, but the event on the left must precede the event on the right end of the dotted line. It is implied that the time between two events which are connected by a soft precedence constraint is greater than or equal to zero seconds. (See the appendix for a detailed explanation of task precedence maps in workload analysis)

The task precedence map of figure 4.4 possesses a unique capability in that it displays the precedence with which certain events trigger certain other cockpit events. For example (see Panel 1 of figure 4.4) event A1 is constrained to precede event D1 by the dotted soft-precedence constraint. Crossing the VOR outbound, event A1, cues the occurrence of event D1, the beginning of outbound tracking.
Event A1 also cues event D2, a slowing maneuver. Event D2, in turn, cues event C1, the extension of 5 degrees of flap.

As another example, Panel 2 of the figure indicates that event A4 cues tasks C3, D8, and D9. Event D8 then cues event B1. The implication is that one position event, A4, is constraining five other events. Further examination shows that almost all events are triggered (or constrained) by events along path A, the position event path. Referring now to figure 4.2 it can be seen that all events on Path A, except event A2, are predefined by the instrument approach procedure itself.
<table>
<thead>
<tr>
<th>Location</th>
<th>Major Control Tasks Triggered</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (at VOR)</td>
<td>o Begin outbound timing (1 minute)</td>
</tr>
<tr>
<td>B (1 min. from A)</td>
<td>o Turn to 314 degrees</td>
</tr>
<tr>
<td>C (45 sec. from B)</td>
<td>o Turn to 134 degrees</td>
</tr>
<tr>
<td>D (intercepting &quot;final&quot;)</td>
<td>o Descend to 2000ft</td>
</tr>
<tr>
<td></td>
<td>o Turn to 179</td>
</tr>
<tr>
<td>E (VOR)</td>
<td>o Descend to 1260ft</td>
</tr>
<tr>
<td>F (2.5 DME)</td>
<td>o Descend to 1140ft</td>
</tr>
<tr>
<td>F (4.8 DME)</td>
<td>o Make GO/NO GO Decision</td>
</tr>
</tbody>
</table>
Figure 4.3: Cockpit Activity Timeline, Presque Isle
<table>
<thead>
<tr>
<th>Heading</th>
<th>Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>2) Turn to 134</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Altitude</th>
<th>Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>H2500 ft</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Speed</th>
<th>Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>H150 kts</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Navig. and Planning</th>
<th>Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Set CDI to 179</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Communic.</th>
<th>Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1: Course Inbound</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Speed</th>
<th>Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>120</td>
<td>140</td>
</tr>
</tbody>
</table>

P2
3) Turn to 179

4) Track 179 Radial

1) Descend to 2000 ft

5) Gear Down

6) Flap 25

7) Slow to 140

8) P3: A/S and EPR Bugs

9) P2: Set

10) P3: No Smoke

11) P2: On

12) P1: Gear Down Flap 25

13) P2: Gear Down Flap 25

14) P3: Before Lodq Complete

P3
4) Track 179 Radial

4) Maintain 1140 ft (+/- 50 ft)

TL 2-6

TL 5-7

TL 3, 4

40 kts

17) P2: MAP
Figure 4.4: Task Precedence Map, Presque Isle
4.3.1 Discussion of the Regulatory and Airspace Mechanism

It has been argued in this chapter that the air traffic control system has a strong influence on the task demand rate of the cockpit crew. The discussion to this point has shown how the structure of the airspace system tends to constrain and invoke cockpit tasks. Related to this, there exists a body of rules, which instrument pilots must follow, that have been instituted to control the movement and behavior of aircraft in the system. There is no doubt that the ATC system has an affect on crew workload in the air transport environment. As cockpits become more advanced it may be that it is the system structure, not the operating requirements of the aircraft, which induces the predominant pilot loading. Some further discussion, though, is necessary on the nature of this ATC/cockpit interaction.

An important factor in the ATC/cockpit interface that has been omitted in the previous discussion is flight operations procedure. Flight procedures also influence the workload level of the crew, though they are designed to minimize such loading effects. An example of poor procedure design will clarify this point.

Suppose that an aircraft's operations manual specifies that full flaps should be extended upon intercepting the glideslope. Suppose further that the subject aircraft type tends to pitch nose-up as full flaps are extended (as is the
case on many aircraft). This procedure induces a relatively high workload situation at the time of glideslope interception. The high workload would result from the pilot having to compensate for the upward pitching moment while at the same time effecting the aircraft's smooth transition downwards onto the glidpath. It should be noted that most aircraft operations manuals call for flap extension prior to glideslope interception (see Delta Airlines Boeing 727 Operating Manual and American Airlines B707 Operating Manuals).

Flight procedure can also affect cockpit task demand through the same constraint/task triggering method as the airspace and regulatory mechanisms. In the example of figure 4.4, event D1 is seen to trigger event D2 and this then triggers event C1. It is the flight procedure as specified in the operating manual that imposes these constraints.

Most often the design of airspace or the design of air traffic control procedure is performed in coordination with the design of safe flight procedure. Nevertheless some federal procedures have been implemented which are questionable in their workload implications. Noteable in this regard are the steep, decelerating noise abatement arrival procedures which have been implemented at some terminal areas.

It is difficult to conceive of the National Airspace system changing or of regulated procedures changing markedly. For this reason it may seem unimportant to analyze the
influence of the air traffic control system on crew workload. Despite this, it is useful to recognize this influence. Understanding the complexities involved in the workload of an air transport crew can only help in designing systems that are workload tolerant. Also as systems such as RNAV (point to point navigation) and ATC systems based on DABS radar are implemented, the system will indeed change. At that time it will be important to have analyzed the airspace and regulatory loading mechanisms.

4.4 ATC VOICE COMMUNICATION MECHANISMS

4.4.1 Explicit Loading

There are essentially two different mechanisms through which air traffic control voice communications affect pilot workload.

First, there is an explicit mechanism. Every ATC message which is directed to an aircraft invokes, as a minimum the tasks:

1. Grasp microphone
2. Respond to message

In addition most messages directed to an aircraft invoke more complex tasks. These messages come in the form of ATC clearances and advisories. Table 4.5 lists some common piloting tasks that are triggered by ATC messages.
ATC messages that invoke piloting tasks are often triggered by the airspace mechanism discussed in section 4.2. For instance, all inbound aircraft must reduce their speed to 310 knots or less when crossing a particular sector boundary, in order to reduce delays in the terminal area. The pilot, however, is unaware of this and is also unaware of the location of the sector boundary. A message (see no. 2, in Table 4.5) causes the pilot to execute the speed reduction task, but it is the sector boundary (a component of the airspace structure) that defines where this task must be initiated.

The ideas of interruptions, constraints, and triggers are important concepts in workload analysis (see Chapter 3.4). Figure 4.5 is a task precedence map* which is useful for visualizing the constraints that one event imposes on another. It is also helpful in identifying when one event triggers another to occur. In this context the task precedence map is especially useful for identifying how air traffic control messages cue piloting tasks.

The dotted precedence constraint connecting box C2 to box E2 indicates that event E2 is cued by the ATC message of box C2. The Begin Slowing event of box D1 is also cued by a communications event, C1. However, event C1 is in turn cued by a sector boundary event, A1. Summarizing, as the subject

*See section 4.3 for a brief explanation of precedence maps, or the appendix for more detail.
aircraft crosses a sector boundary the controller initiates a speed control command which in turn evokes a task in the cockpit. Figure 4.5 thus highlights the difference between a task invoked by the airspace mechanism and task invoked by the communications mechanism.

4.4.2 Implicit Loading

There exists a second mechanism through which ATC messages tend to load the pilot. This is referred to as the implicit ATC loading mechanism.

Transmissions on the air traffic control frequency load the pilot whether they are directed to him or not. This occurs because trained instrument pilots divert some attention from other tasks when a transmission occurs. Experiments (see Morgenstern) have shown that pilots construct a mental image of the air traffic situation in the terminal area, based on air traffic control voice transmissions. These experiments found that transport pilots frequently could recite the identification and altitude of multiple aircraft, both preceding and trailing the subject aircraft.

An example will illustrate another form of the implicit mechanism. Suppose a flight is arriving at the terminal area expecting an ILS to runway 4 and is presently about fifteen minutes from landing. Suppose now that the controller contacts this aircraft with the message, 'Amend clear-
Table 4.5: Explicit Loading Due to ATC Messages

<table>
<thead>
<tr>
<th>Message</th>
<th>Type</th>
<th>Tasks Invoked</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) &quot;Descend to 18000&quot;</td>
<td>Clearance</td>
<td>o Descent Maneuver</td>
</tr>
<tr>
<td></td>
<td></td>
<td>o Descent Checklist</td>
</tr>
<tr>
<td>2) &quot;Reduce Speed to 310 knots.&quot;</td>
<td>Clearance</td>
<td>o Speed Control Maneuver</td>
</tr>
<tr>
<td>3) &quot;Turn left 080, intercept the ILS Runway 4R.&quot;</td>
<td>Clearance</td>
<td>o Turning Maneuver</td>
</tr>
<tr>
<td></td>
<td></td>
<td>o Navigation Maneuver</td>
</tr>
<tr>
<td></td>
<td></td>
<td>and approach checklists</td>
</tr>
<tr>
<td>4) &quot;Reports of moderate turbulence at FL240 by a 707.&quot;</td>
<td>Advisory</td>
<td>o Planning Request</td>
</tr>
<tr>
<td></td>
<td></td>
<td>for New Altitude</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or for Speed Reduction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>to Penetration Speed</td>
</tr>
<tr>
<td>5) &quot;Descend pilot's discretion to 8000ft&quot;</td>
<td>Clearance</td>
<td>o Plan Descent Rate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>and Top-of-Descent Point</td>
</tr>
<tr>
<td></td>
<td></td>
<td>to Allow Idle Thrust Arrival</td>
</tr>
<tr>
<td>6) &quot;Expect 10 minute delay at MILIS intersection.&quot;</td>
<td>Advisory</td>
<td>o Locate MILIS intersection.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>o Slow enroute or prepare to hold.</td>
</tr>
</tbody>
</table>
Figure 4.5: Communications vs. Airspace Mechanisms
ance. Expect VOR-DME approach to runway 27. This clearance causes a large planning task to be executed in a relatively short period of time. Such a clearance can also implicitly introduce confusion and arouse anxiety in the cockpit (see chapter 3 for the effects of anxiety on mental load). Thus, this type of ATC message could induce large mental loads upon the crew even though the associated physical tasks are rather minor.

All ATC transmissions mentally load the pilot. In fact, it is hypothesized that at some workload level a pilot cannot control the aircraft within strict tolerances and interpret ATC messages at the the same time. The effect of radio messages on pilot workload is summarized by the schematic representation of figure 4.6. The figure suggests that workload, on an ordinal, subjective scale is lowest in the long descent phase of an arrival and higher in the terminal area vectoring and final approach portions. These are workload ratings measured with the vectoring and altitude clearances as the only ATC communications.

Overlayed onto the workload rating of figure 4.6 is a plot of ATC channel utilization (Channel Util. = Busy Air Time/Total Air Time, averaged over 20 second intervals). The channel utilization curve was sampled during a peak hour on an IFR day at Boston, Massachusetts. The data represent samples from one center sector, one transition approach sec-
Figure 4.6: ATC Communications and Pilot Workload
Figure 4.7: High Rate Profile for a 'Busy' LAX Arrival
tor, and one local tower control position. The curve possesses a shape common to communications activity at many terminals in the U.S.

Figure 4.6 indicates that the channel utilization during terminal area vectoring and final approach segments of the arrival is double that for the long descent segment. Moreover, in this sample, channel utilization on final approach is slightly higher than that during the vectoring portion. What the figure suggests is that although, for experienced pilots, the nominal workload due to flying the aircraft on final approach may be low, the mental loading induced by high ATC channel utilization may add a significant increment to the total pilot workload level. High channel utilization during the long descent phase of an arrival could significantly affect the measured workload on what is generally a low workload segment. Figure 4.7 is an example of the communications profile during a 'busy' visual arrival into Los Angeles.

Air traffic control voice communications load the cockpit crew. This loading is generally applied in three ways:

1. Clearances or advisories require the pilot to execute some physical or mental piloting task.
a) ILS-DME APPROACHES, 0830-0930

b) ILS-DME APPROACHES, 1815-1915

Figure 4.8: Communications Transmission Histogram
2. Transmissions divert attention away from other tasks as the pilot monitors the message, whether the message is addressed to him, another aircraft, or the controller.

3. A message may invoke tensions, confusion or anxiety in the crewmembers.

Because the occurrence of voice communications affects crew workload it is worthwhile discussing some important features of ATC communications. First, the occurrence of voice messages is a stochastic process. Messages occur with some probability and associated with this probabilistic distribution is a mean message arrival rate and variance. Figure 4.8 is an example of a typical transmission histogram for message occurrences (i.e. probability of n messages occurring in a given time interval). It possesses a characteristic Poisson shape.

Because messages occur in a probabilistic manner, the mean message rate is important. It is one system parameter that is useful in quantifying the communications loading mechanism. It should be noted, though, that probability distributions such as that of figure 4.8 vary by airport, weather, approach in use, and hour of the day. Figure 4.9, for example, shows a set of transmission histograms for a
'busy' arrival into Los Angeles. The stochastic nature of air traffic control messages is quite complex and deserves special attention in pilot workload analysis. The appendix of this thesis contains a more detailed discussion on this subject.

An ATC 'message profile' can also be constructed as a useful characterization of the communications mechanism. Figure 4.10a and 4.10b are examples of this for the same Boston arrival discussed previously. Figure 4.10a displays the number of messages sampled in an 80 second interval versus the time into the flight mission. Figure 4.10b, on the other hand, indicates channel utilization versus time into mission for this same message data.

It is useful to use figure 4.10a and 4.10b in a complementary manner, to gain additional information about the nature of the 'communications profile.' For example, in figure 4.10a, the number of messages/second during the vectoring portion of the arrival is about 50% of the message rate during final approach. However, the channel utilizations (see figure 4.10b) for these two arrival segments are nearly the same (30-33%). What is happening during final approach is that tower control messages are more numerous, but of short mean duration. A typical Tower dialogue might appear as below:

Controller: Delta 11, clear for take off
Figure 4.9: Transmission Histograms for a 'Busy' Visual LAX Arrival
Figure 4.10: A Communications Profile
Delta 11: Delta 11 is rolling.
Controller: Delta 409, Position and hold 4R.

This is contrasted to the terminal vectoring segment where there are fewer messages/second but these, on average, tend to be longer in duration.

The communications time profile, like the transmission histogram, varies with day, weather, airport and communication frequency. Figure 4.11, for example, shows how different the profile is for a Boston departure control frequency for data collected during the same time period as that for figure 4.10. The difference between figure 4.10 and 4.11 is quite striking and reiterates the importance of communications in workload analysis.

Summarizing, air traffic control communications tend to load the pilot. They are characterized by a stochastic process which interrupts, commands and distracts the crew. In workload studies it is therefore important to identify the magnitude of this loading source. It is suggested that mean message occurrence rate, averaged message rate versus mission time, and averaged channel utilization versus mission time are system parameters useful in describing this mechanism.
Figure 4.11: Communications Profile for a Boston Departure Frequency
Chapter V

WORKLOAD ANALYSIS OF A BOSTON ARRIVAL

Up to this point, the critical elements of workload in the air transport environment have been discussed at some length. To comprehend the wide reaching implications of these elements it is useful to examine, in detail, an example of workload analysis. This chapter, then, examines the arrival of a Boeing 707 aircraft into Boston's Logan airport, breaking the arrival down into analytic detail.

Three forms of analysis are considered in this chapter. These are cockpit activity timeline analysis, task precedence analysis, and an analysis of the air traffic control communications. Each of these approaches provides slightly different information pertinent to crew workload during the arrival. Each method yields evidence as to the source of pilot loads and, a priori, as to whether the workload level is relatively higher or lower than some baseline level. In the analysis of this chapter, particular attention is given to the influence of the air traffic control system on crew workload.

The scenario examined in this chapter is the basis for a series of flight simulator experiments that were designed to
evaluate candidate measures of air transport pilot workload. It was, however, necessary to have a theoretical basis for comparison in order to understand the results of the experiments and in order to comment on the validity of their results. The analysis that follows, then, serves a dual role:

1. It will illustrate the concepts outlined in the first chapters.

2. It will serve as a reference for the simulator experiments described in later chapters.

5.1 THE SCENARIO

The scenario involves a Boeing 707 aircraft (identified as Clipper 54) that begins its arrival into Boston at 36000 feet over the Albany, New York VOE. Clipper 54 is given a series of clearances that determine its arrival profile into Boston. The arrival terminates at the Category II missed approach point after the airplane has executed an ILS approach to Logan airport's runway 4R. Figure 5.1 depicts the scenario graphically.

It is convenient to segment the Boston arrival into three portions. The limits of the three segments are as indicated below with separate references for the cockpit activity timeline and the task precedence map.
Albany, NY
VOR
118 deg.
magnetic course
Gardner
VOR
MANJO
Intersection
Boston
Logan
035
MILT
LOM

a) Plan View (Courses shown are magnetic)

36000 ft

22000

14000

100 ft

2000

6000

3000

b) Altitude Profile

Figure 5.1: A Boston Arrival
### Segment | Timeline | Precedence Event
---|---|---
Cruise Descent | 0-840 seconds | A1-A3, B1-B5, C1-C2, D1-D7, E1-E3

Vectoring | 840-1360 | A4-A6, B6-B10, C3-C9, D8-D23

Final | 1360-1600 | A7-B11-B13, C10, D24, E4

This segmentation is performed because, as will be shown, each of these segments has quite a distinct composition of required cockpit activity. It is hypothesized that each segment will have a significantly different level of average workload than the other two.

#### 5.2 ANALYSIS TOOLS

The following analysis makes use of three analytic tools each of which yields different information pertinent to crew workload during the Boston arrival. These tools are, respectively,

1. Cockpit Activity Timelines
2. Task Precedence Maps

- 94 -
3. ATC Communications Analysis

Because the figures on which these analyses are based are rather long and complex, they have been placed in Appendix B of this thesis so as not to break up the body of the text of the chapter. Each page of the figures in Appendix B has a panel number of the form P-XX. The panel numbers are used to refer to particular sections of the timeline and precedence map. Data used in constructing the figures in this chapter were gathered on numerous flight simulations of the scenario discussed above.

5.2.1 Cockpit Activity Timelines

Task timelines are a historically useful tool for quantifying cockpit activity. A timeline is essentially a graph that identifies cockpit tasks, gives estimates of task duration, and tabulates this information in chronological order.

Although task timelines have been an historically useful tool in workload analysis, they have some limitations which should be noted. First, not all cockpit tasks are identifiable or measurable. For example, most mental planning tasks cannot be identified and must therefore be omitted from the timeline. Second, some degree of arbitrariness exists in the detail with which tasks are broken down. For instance, a turn maneuver can be described in microscopic detail as a sequential combination of many hand, eye, and foot move-
ments, each of which require fractions of a second to execute. Alternatively, this maneuver can simply be described as a 'turn' of seven seconds total duration.

The precision of timelines is another limitation (see Appendix A.2). The time required to perform a task will vary between pilots, vary with the conditions of flight, and vary with the existing workload level. Timelines, then, are best suited to describing only a nominal flight. There will be wide variations in practice from this nominal timeline.

The usefulness of task timelines, despite the above limitations, should not be overlooked. Timelines can provide a description of cockpit activity to a reasonable degree of detail and accuracy. It is also possible to derive rough estimates of physical effort requirements from the timeline. Moreover, timelines can readily be expanded to reflect the following:

1. The uncertainty associated with each task's time of occurrence.
2. The simultaneity of multiple tasks in the cockpit.
3. The random interruption process associated with ATC and intercrew communications.
4. The presence of identifiable planning tasks.

The timeline format used in this thesis includes enhancements that are designed to highlight the workload parameters mentioned above. These features are explained, as they are used, in later sections of this chapter.

5.2.2 Task Precedence Maps

A task precedence map is another type of graph which can be constructed from cockpit activity timeline data. It represents a map or schedule of discrete events that occur in the cockpit where an event is

1. The beginning or end of some physical cockpit task (e.g. a descent maneuver.

2. The occurrence of a physical event (e.g. the arrival of an ATC message or the passage of a VOR)

The task precedence map displays events, time of event, uncertainty of these event times, and the precedence constraints that must be maintained between events.

Task precedence maps differ greatly from cockpit activity timelines both in their construction and in their use. The task precedence map is most useful in a very macroscopic or
aggregate form where only major events are displayed. They are useful in identifying the way in which certain cockpit events constrain other events. There exists an unwritten system of task priority among instrument pilots. The precedence map maps out these priorities and shows how the timing and execution of a series of tasks is constrained by the requirement to execute some other series of tasks. The task precedence map displays this hierarchy as a schedule of events, constraints, and event flexibility or slack.

Activity timelines, on the other hand, are most useful in the minute detail which they convey. The timeline can be referred to in analyzing the microscopic detail of cockpit activity, interruptions, busyness, and physical effort. It will be shown later that the timeline and the precedence actually complement one another in workload analysis.

Events, on a task precedence map, are denoted by boxes and are positioned along paths with which they are associated. There is generally some physical significance to an event and the path to which it has been assigned. Five aggregate event paths have been defined for the purposes of this thesis:

1. Position Event Path (the geographic position of the aircraft)

2. Altitude Event Path
3. Communications Event Path

4. Manual Control Event Path

5. Planning Event Path

Event paths may be independent, physically connected or constrained in temporal order only. An event on one path may be constrained to precede an event on another path.

A constraint is denoted by a solid or a dotted line. A solid line between any two events indicates that a physical task occurs between the two events and the duration, in seconds, of this task is indicated by the number in parentheses located on the constraint, approximately midway between the two events (see for example figure 5.2, events B1 and B2).

A 'soft' precedence constraint is indicated by a dotted line connecting two events (see events C1-D1 of figure 5.2). No physical task is being performed between two such events, but the event to the left is constrained to precede the event on the right. It is implied, then, that the time between two events that are connected by a soft precedence constraint, must be greater than or equal to zero seconds.

The implication of all the constraints that the precedence map displays graphically is that there is an 'earliest' and a 'latest' time* at which each event can

*Event times on the precedence map are taken to be deterministic. With some simplifying assumptions (i.e. normal-
occur. These limit times appear, in seconds, as the numbers in parentheses beneath each event box. The 'earliest' time is the leftmost number, the 'latest' time is the rightmost number.

The difference between the earliest and latest times for an event is referred to as the slack. The slack, strictly speaking, is a measure of the uncertainty in the event time. It is more pragmatically thought of as the flexibility a pilot or crewmember has in executing the tasks associated with an event. Slack is that flexibility allowed the crewmember by all the interacting constraints shown on the precedence map. The existence of slack can be depicted on a timeline by allowing tasks to slide from their nominal (or earliest) time to a Task Limit defined by the slack value of a precedence map.

Given some degree of task flexibility an operator working on multiple tasks is likely to perform these tasks in such a manner as to minimize some reward variable (see Tulga). Instrument pilots are trained to perform tasks on a priority basis that maximizes safety and that minimizes workload (or task demand rate). If no slack exists for a particular event, then its associated tasks must be worked on promptly, increasing task demand rate and thus instantaneous workload. A path along which the slack is always zero is called the

----------------------

(ity) stochastic event times can also be used.
Figure 5.2: A Task Precedence Map
'critical path' and is denoted on the precedence map by a double line (see Path A of figure 5.2).

To summarize, the task precedence map is useful in workload analysis because of its capability to highlight precedence constraints, task priorities, and task flexibility. The precedence map, though derived from timeline data, is much more macroscopic in its scope and in its use.

5.2.3 Analysis of ATC and Intercrew Communications

Section four of chapter four discusses the mechanisms through which ATC communications may tend to load the crew. In a similar manner intercrew communications may tend to interrupt and distract the crew and thus induce higher workload levels. Due to such effects the statistical properties of voice communications messages will be examined for the Boston arrival. Particular attention will be devoted to

1. Interruptions due to messages

2. Tasks triggered by messages

3. Effects of Channel Utilization (i.e. Busy Air Time/Total Mission Time) and message rate

4. Comparison of message properties between mission segments
5.3 AN ANALYSIS

In this section, six parameters are examined that (by the theory of Chapter 2) characterize the nature, and that indicate relative levels of pilot workload. These are:

1. Busyness - i.e. the amount of time spent working on cockpit tasks and not idle or performing only monitoring tasks.

2. Task Multiplicity and Interruptions - The degree to which multiple tasks are present and the magnitude of task interruptions which then results.

3. Task Simultaneity - The degree to which multiple tasks occur simultaneously.

4. Task priority and Slack - Examine the priority structure for executing different tasks and determine the time flexibility or deferrability which each task is then allowed.

5. Constraints and Triggers - The priority structure defines a set of
constraints between tasks. These constraints identify which elements of the ATC/cockpit system strongly influence crew activity and also identify which events trigger other events to occur.

6. ATC and intercrew communications -
The statistical properties of messages and the activity messages invoke.

Each of the above parameters are investigated for the three scenario segments (cruise descent, terminal vectoring, and final approach). The section that follows contains a concluding synopsis of the analysis.

5.3.1 Busyness

Chapter two categorized tasks into three broad classifications:

Operating tasks - Tasks which must be handled immediately, such as responding to an ATC message or controlling the aircraft.

Monitoring - Tasks which can be deferred for a small time period, e.g. systems monitoring.

- 104 -
Table 5.1a) Number of Tasks for the Boston Arrival
(Excluding all "Hold" category and communications tasks.)

<table>
<thead>
<tr>
<th>Segment</th>
<th>Heading</th>
<th>Altitude</th>
<th>Speed</th>
<th>Nav/Planning</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cruise</td>
<td>7</td>
<td>18</td>
<td>3</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>Descent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terminal</td>
<td>9</td>
<td>16</td>
<td>4</td>
<td>11</td>
<td>40</td>
</tr>
<tr>
<td>Vectoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>8</td>
<td>21</td>
</tr>
<tr>
<td>Approach</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Total is not the summation in this case, since many tasks occur simultaneously.

Table 5.1b) Busy Time (seconds)
(Excluding all "Hold" category and communications tasks.)

<table>
<thead>
<tr>
<th>Segment</th>
<th>Heading</th>
<th>Altitude</th>
<th>Speed</th>
<th>Nav/Planning</th>
<th>Total*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cruise</td>
<td>820</td>
<td>121</td>
<td>37</td>
<td>75</td>
<td>840/840</td>
</tr>
<tr>
<td>Descent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terminal</td>
<td>174</td>
<td>104</td>
<td>42</td>
<td>75</td>
<td>309/520</td>
</tr>
<tr>
<td>Vectoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final</td>
<td>246</td>
<td>173</td>
<td>32</td>
<td>38</td>
<td>233/240</td>
</tr>
<tr>
<td>Approach</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: Total is not the summation in this case, since many tasks occur simultaneously.
### Table 5.2: Mean Task Interarrival Time (seconds)

<table>
<thead>
<tr>
<th>Segment</th>
<th>Heading</th>
<th>Altitude</th>
<th>Speed</th>
<th>Nav/Planning</th>
<th>Mission*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cruise Descent</td>
<td>1#</td>
<td>43</td>
<td>390#</td>
<td>550#</td>
<td>31</td>
</tr>
<tr>
<td>Terminal Vectoring</td>
<td>86</td>
<td>32</td>
<td>97#</td>
<td>46</td>
<td>17</td>
</tr>
<tr>
<td>Final Approach</td>
<td>13#</td>
<td>16</td>
<td>33#</td>
<td>35</td>
<td>16</td>
</tr>
</tbody>
</table>

Notes: * Includes all tasks except communications tasks. 
# Indicates datum unreliable due to small sample size.

### Table 5.3: Mission Mean Task Interarrival Times (seconds)

<table>
<thead>
<tr>
<th>Segment</th>
<th>Excluding Communications Tasks</th>
<th>Including Communications Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All Tasks ¹</td>
<td>All Tasks for Pilot Flying ²</td>
</tr>
<tr>
<td>Cruise Descent</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>Terminal Vectoring</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>Final Approach</td>
<td>16</td>
<td>18</td>
</tr>
</tbody>
</table>

Notes: 1 - Excludes communications tasks.  
2 - As specified by flight manual, also excludes comm. tasks.  
3 - Includes intercrew communications and ATC messages to/from CL 54.  
4 - Same as above, but includes only intercrew messages which pilot must command or respond.
Planning - Tasks which can be deferred for a longer period, such as approach planning.

The nature of the piloting tasks in an IFB, air transport environment is such that there are phases of flight when the crew is nearly idle (performing mostly monitoring tasks) and there are phases of flight when the crew is extremely active (working on all three categories of task). Busyness, then, is one parameter which characterizes the workload level of a crew. It may be quantified to some extent, by measuring task load (in terms of number of tasks active), task type (an indication of relative difficulty), and busy time (i.e. the percentage of segment time during which a crewmember is working on at least one task).

The Cruise Descent Segment

At fourteen minutes in duration, the cruise descent segment is the longest in the scenario. Figure 5.1 indicates that this segment includes a long descent with altitude steps at 22000, 14000, and 8000 feet. There are two course changes during the descent. The segment length seems to be the most notable feature of this portion of the scenario. It is an important feature because, if workload is measured as an average over the 14 minutes, the instantaneous peaks in workload will be smoothed out by the averaging process.
Panels 1-9 of figure B.1 contain the task information for the cruise descent segment. A brief examination of these panels shows that the tracking task associated with flying to a VOR is the primary task and that it persists throughout the segment. From the timeline it is also clear that there are relatively few other tasks occurring and that these other tasks generally have a duration of less than 20 seconds. Table 5.1a was compiled from the data of figure B.1 and verifies this observation quantitatively. Excluding communications tasks (and counting the VOR tracking task as one task), there are 30 tasks occurring in 840 seconds.

The number of tasks is, by itself, misleading. Viewing Panel 4, for example, some tasks are continuous (e.g. the heading category tracking task), while other tasks are of short duration, but occur in rapid succession with other tasks (see tasks 6, 7, 8 in the altitude category). Thus, the small number of tasks does not necessarily imply that the crew is idle. From Table 5.2 it can be seen that the pilot is busy working on at least one task, 100% of the time during the cruise descent. Moreover, Table 5.2 does not include hold category tasks. These tasks require intermittent crew attention and load the pilot to some extent, but are not included in the data of Table 5.1 or 5.2.

The Terminal Vectoring Segment

- 108 -
The terminal vectoring segment is approximately eight minutes in duration or slightly more than half as long, in duration as the descent segment. Panels 9-14 of figure B.1 describe the cockpit activity for the segment. A cross comparison of these panels with Panels 1-9 shows that the vectoring segment is much more complex, much busier than the descent segment.

Figure 5.1 shows that the vectoring portion of the scenario is dominated by four turns, a turn to 160 degrees, a turn to 130, a turn to 080, and a turn to intercept the final approach course. Moreover, this segment includes three required altitude changes (see figure 5.1b). These are descents from 8000 to 6000 feet, from 6000 to 3000 feet, and from 3000 to 2000 feet. This aircraft maneuvering seems to dominate the activity on the timeline of Panels 9 through 14. Table 5.1a indicates that there are 25 heading and altitude category tasks for the vectoring segment as compared to 25 heading and altitude category tasks for the descent segment. However, these 25 aircraft maneuvering tasks are squeezed, now, into half the time frame of the cruise descent.

Navigation and planning tasks compose a major percentage (28%) of the 39 total tasks occurring during the vectoring segment. This is in comparison with only two navigation and planning tasks in the descent segment. Six of the 11 plan-
ning tasks occur in rapid sequence on Panel 10 of the timeline and are associated with the performance of the 'before landing' checklist.

Again, the number of tasks, by itself, is somewhat misleading. Table 5.2 indicates that (assuming one crewmember performs all tasks) the pilot is working on at least one task 60% (309/520) of the time during the descent segment. This is compared to 100% busy time for the cruise descent segment. Panels 9-14 of the timeline, however, indicate that the composition of tasks is much different than that for the descent segment. There is a greater variety in the number and type of tasks than those which present themselves in Panels 1-8.

Tasks which occur during the vectoring portion of the mission are more numerous, but of shorter mean duration than those during the cruise descent. From observation of the timeline it is not immediately obvious which type of situation tends to induce higher workload levels. The theory of Chapter 2 would suggest that the case of many, short duration tasks, rather than few, long duration tasks, induces a higher level. This is due to the fact that in the former situation, tasks are continually interrupted or preempted by tasks of higher priority. Thus, theory suggests that, during the vectoring segment, induced levels of workload may be higher than during the cruise descent segment, despite the
The busy time for this segment is lower (60% busy time for the vectoring segment) than that for the cruise descent (100% busy time).

**Final Approach Segment**

The final approach is characterized by a straight, constant descent-rate maneuver. The duration of this final mission segment is four minutes, half the duration of the vectoring segment. Examination of Panels 14-16 shows that the 'track localizer' and 'track glideslope' tasks dominate the activity of this segment.

As indicated in Table 5.1a, there are 21 total tasks which must be performed during the segment. This can be compared with 40 tasks for the vectoring segment which has a duration that is twice that of this final approach segment. The composition of these tasks appears similar to the composition of tasks for the vectoring segment. Again 28% of the tasks (6/21) are navigation and planning tasks.

The number of tasks should not be used alone in quantifying cockpit activity on final approach. This is because the 'Number of Tasks' metric weighs the two simultaneous tracking tasks ('track localizer' and 'track glideslope') equally with other, less difficult tasks (e.g. Seat Belt Sign - On). The busy times of Table 5.1b complement the 'Number of Task' data by indicating how much of the time the crew is busy,
regardless of the number of tasks. The table shows that, on final approach, the crew is working on at least one task 95% (233/240) of the segment time.

Further examination of Panels 14-16 of the timeline and Table 5.1 show that the composition of tasks during the final approach segment is a combination of that for the cruise descent and vectoring segments. There appears to be many tasks (21 in 240 seconds) some of which are long duration (the tracking tasks) and some of which are short duration (the Navigation and Planning tasks). Based on the busyness of this segment one would expect the measured workload level to be approximately the same or greater than that for the vectoring, but greater than that for the cruise descent segment.

5.3.2 Multiple Tasks and Interruptions

The Cruise Descent Segment

At any particular point in the activity timeline it is evident that the piloting task involves multiple, interrupting tasks. The timeline format of figure B.1 highlights the occurrence of this interruption process. Panel 7 provides an example of this feature.

In the time period 620 to 680 seconds, the pilot performs a course change, a level off followed by a further descent
initiation (see Alt. column) and he also must reselect his navigation course setting (see Nav/Planning column). By scanning the Panel horizontally along a line of constant time (660 seconds) the multiplicity of tasks becomes apparent. At 620 seconds a speed and altitude hold* task are being monitored while the heading category tracking task is being performed. At 670 seconds a left turn is being executed simultaneously with the execution of a short duration, three task altitude sequence.

Quantifying the magnitude of the interruption process, Table 5.2 gives mean task interarrival times. That is, the average time between the occurrence of two tasks (based on the timeline in figure B.1). Because the number of tasks is small (refer to Table 5.1a for the cruise descent segment) some of these computed interarrival times are unreliable. It does appear, though, that the altitude tasks contribute most to the interruption process with a mean interarrival rate of 1 task every 43 seconds. This is as expected based on the scenario structure portrayed in figures 5.1a and 5.1b.

-----------

*Hold category tasks, e.g. holding a descent rate, are indicated by the solid lines with arrowheads, that run through the center of task columns on the timeline.

*Mission mean interarrival time is computed by grouping all tasks in one category, maintaining the chronological relationships, and then computing mean interarrival times. Simultaneous tasks are counted as one task.
The Mission mean interarrival time, however, better emphasizes how multiple, interrupting tasks load the pilot. In Table 5.2, the Mission interarrival time of 31 seconds between tasks is 25% smaller than the 43 seconds between tasks, which is indicated for altitude tasks. In this long descent segment it is the altitude tasks which are most numerous (17 altitude tasks, see Table 5.1). However, the random occurrence of tasks in three other categories has reduced the Mission mean task interarrival time from approximately 43 to 31 seconds between tasks. This higher task arrival rate, induced by the multiple task environment, should tend to increase the workload level of the crew above that encountered in a single task mode.

Terminal Vectoring Segment

Panels 9 through 14 of the cockpit activity timeline indicate that the multiple task environment and its associated interruption process is much more evident in the vectoring segment than in the cruise descent segment. Scanning across the timeline at 940, 1240, and 1320 seconds makes this more apparent. At these mission times a combination of heading, altitude, speed, and Nav/Planning tasks are occurring either simultaneously or in a rapid, sequential manner.
Table 5.2 gives mean task interarrival times based on figure B.1. The altitude and the Nav/Planning task categories have the smallest task interarrival times (most frequent task occurrence) at 32 and 46 seconds between tasks, respectively. For the altitude category tasks this is smaller than the 43 second figure which obtains during the cruise descent segment. Heading tasks, with mean interarrival times of 86 seconds would seem to be the least demanding task category with respect to task interruptions, despite the fact that the vectoring turns are a predominant physical task during this segment.

The Mission mean interarrival times shown in Table 5.2 reflect the marked impact of the multiple task environment on crew task load during the vectoring segment. The Mission mean interarrival time of 17 seconds is 50% of the interarrival times for altitude tasks alone and 20% of the interarrival time for heading tasks alone. The interaction of the four task categories (communications tasks have not been counted) in quadruplex has the effect of dramatically reducing the task interarrival time perceived by a crewmember. The 17 second task interarrival time for the vectoring segment is roughly half that for the cruise descent. The crewmembers should thus perceive themselves as being much busier and find tasks interrupting one another more frequently during the vectoring segment than the descent segment. On the basis of perceived busyness due to interruptions, one would
expect workload to be relatively higher for the terminal vectoring segment of the Boston arrival.

**Final Approach Segment**

Comparison of the final approach and terminal vectoring segments on figure B.1 suggests that they are roughly equivalent in the degree to which multiple tasks and distractions caused by task interruptions occur. Scanning across figure B.1 at times 1380 and 1460 seconds, for instance, shows that tasks are occurring in most task categories simultaneously. Beyond time 1500 seconds there are two simultaneous tracking tasks and one 'hold' category speed control task being performed by the pilot as he flies the aircraft down the Instrument Landing System.

Table 5.3 indicates that altitude category tasks occur most often during final approach (ignoring the heading category due to the small sample size) at 16 seconds between tasks. Mission mean task interarrival time is also 16 seconds, very close to the value for the terminal vectoring segment. Because the Mission mean interarrival time is the same as the mean interarrival time of altitude tasks, the data suggest that task interruptions are not as pronounced as in the terminal vectoring segment.* However, another

---

*In the terminal vectoring segment the Mission mean task interarrival time was at least 50% of the value of the minimum task interarrival time for all task categories shown in table 5.2. This effective reduction in task
effect is taking place that is not reflected in the data of Table 5.3. This is the effect of task simultaneity which is discussed in a later section.

5.3.3 Task Simultaneity

The mean task interarrival times discussed above ignore the fact that some tasks occur simultaneously. For instance, on Panel 7 of figure B.1, it is apparent that a left turn to 111 degrees and a descent initiation sequence occur simultaneously at 667 seconds into the scenario. The presence of task simultaneity will thus affect crew loading but is not measured well by the task interarrival time metric. This section examines the degree to which task simultaneity is present in each of the three arrival segments.

The Cruise Descent Segment

The degree of task simultaneity during the cruise descent can be inferred from the Task Simultaneity histogram of figure 5.3. In this figure the number of occurrences of simultaneous tasks have been counted based on the nominal time of task occurrence shown in figure B.1. These data were then normalized to construct the histogram.
Simultaneous Tasks

a) Cruise Descent Segment—Excluding Communications Tasks

b) Cruise Descent Segment—Including Communications Tasks

Figure 5.3: Task Simultaneity Histogram for the Cruise Descent Segment
a) Vectoring Segment-Excluding Communications Tasks

b) Vectoring Segment-Including Communications Tasks

Figure 5.4: Task Simultaneity Histogram for the Terminal Vectoring Segment
Simultaneous Tasks

a) Final Approach Segment—Excluding Communications Tasks

b) Final Approach Segment—Including Communications Tasks

Figure 5.5: Task Simultaneity Histogram for the Final Approach Segment
Figure 5.3a indicates that there are frequent occurrences of two simultaneous tasks. Observation of the timeline indicates that the heading tracking task is continuous throughout the cruise descent. Thus, a task occurring in any other category must be performed simultaneously with this tracking task. Nevertheless, figure 5.3a shows few occurrences of three, four, or five simultaneous tasks and so task simultaneity should not significantly affect crew busyness during the cruise descent.

The Terminal Vectoring Segment

The terminal vectoring segment appears similar to the cruise descent segment in the degree to which simultaneous tasks occur. Figure 5.4a shows that there is a large probability that two tasks will occur simultaneously and a much smaller probability that three tasks will occur simultaneously. The major difference between these segments can be seen in comparing figures 5.3a and 5.4a. That is, during the vectoring segment there is a high probability that one task will occur by itself (not simultaneously with another task). This is perfectly consistent with the fact that tasks are more numerous and that mean task interarrival times are smaller during the vectoring segment. Tasks occur more often, but they simply do not occur simultaneously.

The Final Approach Segment
Task Simultaneity is a more important consideration during final approach than in the earlier mission segments. This is evident from figure 5.5a which indicates that the frequency of occurrence of two, three, and four simultaneous tasks is significant.

The discussion of section 5.3.2 showed that the final approach and the terminal vectoring segments have nearly identical task interarrival times. This suggests that the degree to which tasks interrupt one another is the same for these two segments. Figure 5.5a, however, shows that during 'final' there is a high incidence of multiple tasks occurring, which the task interarrival time metric ignores. Because of the incidence of two, three, and four tasks occurring simultaneously on final approach, the busyness perceived by the pilot may be greater on final approach than during the vectoring segment. Due to the joint influence of other parameters, however, it is uncertain whether measured total workload will be greater during 'final' than during the vectoring segment.

5.3.4 Task Priority and Slack

In section 5.1.1, it was explained that cockpit activity timelines are inaccurate because there is an uncertainty associated with the nominal task execution time shown on a timeline. That is (see Panel 7, figure B.1), the left turn to a heading of 111 degrees may be initiated at 627 seconds
or it may be initiated later if there is a more important task to be executed. The current workload level of the pilot can also affect the task initiation time. The pilot may elect to begin work on a task either earlier or later than the nominal task time shown on an activity timeline, this decision being based on what his experience, training, and workload level indicate to him.

The task precedence map discussed in section 5.1.2 uniquely determines the limits within which a task must occur. These task limits are indicated by the numbers enclosed in parentheses, located beneath event boxes of the task precedence map. The difference between the second number of the pair and the first number is referred to as the 'slack'. It is equivalent to the total time range (beyond the so-called 'earliest time') through which the task may be slid.

Slack, then is an indication of the priority of a task or of the time flexibility that a pilot has in executing a task. The trained IFR pilot would be expected to utilize event slack in a manner which allows tasks to be executed in a time sequence that minimizes his probability of error. In the timeline of figure B.1, all tasks are shown at their nominal time of occurrence (their 'earliest time'). Task limits for a task are denoted by dark bars and an accompanying TL 1-9 (Task Limit for task 1 through 9 of this column). The slack, on the timeline, is thus the difference between
the time at which the TL bars are shown and the time at which the associated task occurs.

**The Cruise Descent Segment**

For the long descent portion (see section 5.1 for the limits of each mission segment on the task precedence map) figure B.2 shows that the average, noncritical slack is very large. On Panel 1 of the figure, no slack (off the critical path) has a value of less than 100 seconds. Paths B, C, and E of the Panel have mean slack values greater than 300 seconds. The minimum noncritical slack value of 166 seconds (see event C2,D3) is again associated with altitude tasks which were shown in table 5.1 to be the most numerous task type for this segment. Panel 2 of figure 3.2 also indicates large slack values for events not on the critical path (see events B5, D4-D6, and E3).

What is most notable from Panels 1 and 2 of figure B.2 are the very large slack values associated with planning tasks. Events E1,E2, and E3 show slack values of 1600, 1220, and 940 seconds, respectively. As discussed in Chapter 2, the pilot will tend to start the execution of planning tasks early in the descent, but these tasks will be deferred as they are preempted by more important operating or monitoring category tasks. In the air transport environment, one pilot may execute operating and monitoring tasks while another executes planning tasks.
For the long descent segment the large slack values have important implications:

1. Though the planning tasks require substantial time to execute (figure B.1, Panel 2, indicates a nominal 70 seconds for one crewmember to review the approach procedure) the task flexibility is large enough so that its presence is not likely to induce high workloads.

2. Panels 1-9 of the timeline indicate a few occurrences of simultaneous tasks. Again, due to slack values that range from 166 to 1600 second, the occurrence of simultaneous tasks is also not likely to induce high workloads during the long descent segment.

The Terminal Vectoring Segment

A brief examination of Panels 2-5 of figure B.2 indicates that the terminal vectoring segment* has many more complexities than the cruise descent. Despite this, further

*On the task precedence map of figure B.2, events A4-A6, B6-B10, C3-C9, D8-D23, inclusive, contain the terminal vectoring segment.
examination of the figure shows that event slacks, off the critical path, are smaller than the slack values for cruise descent segment.

In all cases the event slacks for the vectoring segment are less than 415 seconds. The minimum nonzero slack in this segment is 46 seconds which occurs for events D17 and D18 (see Panel 5, figure B.2). There are numerous occurrences of slack values less than 120 seconds. The mean slack values for the segment (excluding events on the critical path) are listed below.

<table>
<thead>
<tr>
<th>Event Path</th>
<th>Mean Slack(seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>246</td>
</tr>
<tr>
<td>B</td>
<td>171</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>133</td>
</tr>
</tbody>
</table>

The mean event slacks for the terminal vectoring segment are much smaller than the values for the descent segment. Most notable are the 171 and 133 second values for events on path B and D, respectively. These can be compared to values of 587 and 356 seconds for the mean event slack on
paths 3 and D during the cruise descent segment. This suggests that the crew will have much less flexibility in the time in which they may execute required tasks. On average tasks may only be slid 30% of the time distance allowed during the cruise descent segment. This indicates that the intertask constraints will cause workload to be higher during the vectoring segment than during the cruise descent segment.

**Final Approach Segment**

Panels 6 and 7 of figure 8.2 describe the nature of the event slacks for the final approach segment. The most notable feature of this portion of the task precedence map is that most of the events lie on the critical path and thus have slack values of zero seconds.

Scanning across Panels 6 and 7 of the figure it can be seen that the only two events off the critical path (Events A9 and C10) have slack values of 123 and 213 seconds. These values are comparable to the average slack values for the vectoring segment and much smaller (on the order of 30% of the mean values), than the mean values for the cruise descent segment.
The cockpit activity timeline of figure B.1 provides a little more information on the effects of slack during the final approach portion of the scenario. Scanning Panels 14-16 of figure B.1 shows that there are numerous tasks occurring and these are occurring in multiple task categories (refer to sections 5.2 and 5.3). Since most of these tasks are associated with events on the critical path, however, they possess no time flexibility and are constrained to be executed promptly. Moreover, several tasks that nominally occur during the cruise descent and vectoring portion of the missions have slack values that allow their execution to be slid well into the final approach segment (see for example the Task Limits on Panels 14 and 16 of the figure). This suggests that if the terminal vectoring segment is unusually busy, some tasks may be deferred for execution during final approach. Because additional tasks must then be executed within the increasingly tight constraints imposed by the critical path, a higher task demand situation would result. This effect is caused by the low average slack value during final approach.

The final approach segment, then, is much more tightly constrained than the vectoring segment. Most of the final approach tasks are associated with events on the critical path and therefore must be executed promptly rather than afforded large degrees of time flexibility. Whether this
tightly constrained situation results in a higher workload level for the final approach segment will be a function of whether perturbations to cockpit activity (e.g., an emergency) during earlier flight segments cause tasks from these earlier segments to be deferred into the tightly constrained final approach segment.

5.3.5 Constraints and Triggers

In workload analysis it is useful to examine sources of crew loading. The task precedence map of figure B.2 is unique in this context as it indicates which elements of the ATC/cockpit system are constraining one another and which ATC/cockpit events trigger other to occur.

The Cruise Descent Segment

The discussion of section 5.2.2 explained that the critical path (the double lined path in figure B.1) indicates which events are imposing the primary constraints and task triggers. Panels 1 and 2 of figure B.2 show that for the duration of the cruise descent segment, path A is critical. That is, the aircraft's position along the airway is triggering most cockpit events. Following the arguments presented in chapter 4, it is the airspace loading mechanism of the air traffic control system that is triggering cockpit activity and inducing the predominant cockpit load.
The effect of the Regulatory Loading mechanism is also present on Panel 2 of the figure where at 10000 feet Event B5 cues the occurrence of a slowing maneuver to 250 knots (Event D6). However, path A suggests that it is the airway structure that is triggering most cockpit activity. If the aircraft altered its path of flight or changed its cruising airspeed, a change in the time distribution of cockpit activity would result.

Figure B.2 also indicates some event triggering due to flight procedure. In particular, flight procedure calls for two checklists to be executed one at 18000 feet and one at 10,000 feet. This effect is denoted by the soft-precedence constraint connecting Event E4 and E2 and that connecting B5 and E3. Observation of the cockpit activity timeline in the vicinity of these events (see P5, P8-9, figure B.1), however, indicates that these events do not greatly increase task loading at a busy time nor does this procedure induce the occurrence of many simultaneous tasks. Moreover, the task limits for these particular cockpit procedures are large enough (1220 and 940) so that high workloads are not likely to be induced.

Terminal Vectoring Segment
Observation of Panels 2-6, figure B.2, indicates that for the major portion of the terminal vectoring segment the critical path lies along Path C, the communications path. That is, during the vectoring segment most events are triggered by air traffic control communications events. Also the timing between these communications events constrains the execution and timing of tasks that are associated with events on non-critical paths.

The vectoring portion of the Boston arrival provides an example of what was referred to in chapter 4 as explicit loading of the pilot due to air traffic control messages. Event C4 (refer to Panel 2, figure B.2), for instance, triggers the occurrence of the Begin Descent event, D8. Associated with this Begin Descent event are timeline tasks 26, 27, 28 in the altitude category and task 5 in the speed category (see Panel 12, figure B.1). Panel 4 of figure B.2 shows how communication event C6 triggers a turning maneuver associated with events D13 and D14. In this segment, then, air traffic control messages exert a dominant influence on the nature and timing of the crew's task load.

There are other portions of the vectoring segment where tasks are triggered by events off the critical path. For example, the aircraft's arrival at Manjo intersection (see event A5, F2, figure B.2) triggers a turning event (Event
D10) Associated with event D10 are heading tasks 8,9 on Panel 10, of figure B.1. These tasks are triggered only indirectly by the critical path.

**Final Approach Segment**

The critical path for the final approach segment wanders from event path A to path D and then to path B (see P6-7, figure B.2). During the early stages of the final approach, cockpit tasks are triggered by the aircraft's position along the localizer. An example of this is event E4 on Panel 6 of the precedence map. This 'Begin Checklist' event is essentially triggered by the localizer interception event (note that event E4 and A6 have the same earliest time). Panel 14 of the cockpit activity timeline indicates that Nav/Planning tasks 14,15 and communications task 43 are associated with this event.

As the aircraft arrives at the point of glideslope interception, the altitude path (Path B) becomes critical. The aircraft's position on the glidepath then triggers cockpit events. Figure 8.1 provides the detail on how this occurs. On Panel 15 the call for and selection of landing gear extension and for the extension of 40 degrees of flap are both triggered by the aircraft's position in relation to the glidepath. Similarly, communications events 56-63
are also triggered by the aircraft's altitude, once the glidepath has been intercepted.

It should be noted that during final approach it is not entirely clear from the task precedence map whether the airspace structure (i.e., the instrument approach procedure, glidepath angle, etc.) is the primary cause of pilot loading. The structure of the instrument approach procedure definitely does affect crew workload and this is clearly indicated on the precedence map by the location of the critical path along the position and altitude event paths. However, again the flight procedure, as specified in the aircraft operations manual, plays a significant role in inducing task loads. If, in particular, the specified approach speeds were changed, the length of the critical path would be affected. In some circumstances the length of the critical path could affect the nominal workload level (see appendix A.2).

5.3.6 ATC and Intercrew Communications

The discussion of chapter 4 suggests that the structure of air traffic control communications can induce explicit and implicit loads on the crew. Similarly, required intercrew communications can cause interruptions and distraction to occur. The nature of these communications is thus an important element of this workload analysis.
Figure 5.6 was constructed using the ATC dialogue depicted in the communications column on the timeline of figure B.1. It represents a 'low' message rate relative to some of the data presented in chapter 4. Figure 5.7 represents a 'high' message rate profile that was also used in the flight simulator experiments described in the next chapter. The dialogue for this profile is not shown on the timeline of figure B.1, but is included (for reference) in Appendix C. The high rate profile is included here strictly for discussion and comparison with the 'low' rate profile.

The Cruise Descent Segment

Figure 5.6 represents a low ATC message rate relative to some of the data presented in Chapter 4. Figure 5.6a shows that during the long descent this profile averages only one message in an 80 second interval. This corresponds to a channel utilization (see figure 5.6b) of 8% during the segment. Examination of Panels 1-9 of the timeline indicates that messages are likely to cause few interruptions.

From the timeline it can be seen that there are also few intercrew communications occurring during the cruise descent. Table 5.3 (see columns 3 and 4) verifies that ATC communications are likely to cause few interruptions. The table indicates that the mean interarrival time between tasks is unchanged if communications events are included as
Figure 5.6: A Low Message Rate Communications Profile
Figure 5.7: A High Message Rate Communications Profile
tasks that the crew must monitor. Similarly, figure 5.3 shows that the presence of ATC and intercrew communications does not alter the degree of task simultaneity during this segment.

The Terminal Vectoring Segment

It is evident from figure 5.6 that the intensity of ATC messages is much greater during the vectoring segment than during the cruise descent. The average message rate is 7 messages in an 80 second period as compared to 1 per 80 seconds for the descent segment. Channel utilization (see figure 5.6b) has increased from 8% for the descent to approximately 30% for the vectoring segment. It should be noted, however, that the message rate is still quite 'low'. The communications profile of figure 5.7a shows a message rate of nearly 14 messages per 80 second interval, twice that for the 'low' rate profile. Recall that these figures include only ATC messages (messages to/from ATC by/for any aircraft) and no intercrew communications. Nevertheless, the loading effects of voice communications are expected to be much more significant during the vectoring segment than during the cruise descent.

Panels 9-14 of the timeline indicate that required intercrew communications associated with checklist execution are numerous but of small average duration. These intercrew
communications also tend to trigger an associated cockpit tasks such as 'Seat Belt Sign - On,' or 'Shoulder Harness - Fasten.' As intercrew communications become more numerous they, too, will tend to interrupt other tasks, disrupt mental planning tasks and thus implicitly load the crew through the same mechanism as ATC messages (see Chapter 4.4).

Table 5.3 shows the effects of voice communications (both ATC and intercrew) on the Mission mean task interarrival time. In this table each message is counted as a discrete task in addition to the other four task categories that are tabulated in Table 5.2. Voice communication messages are taken to be a separate category of task because the crew's attention is diverted from other tasks, in order to monitor a message. Also, messages to which a crewmember must respond require physical effort to grasp the microphone and to speak. Table 5.3 shows that the inclusion of messages as tasks reduces the Mission task interarrival time from an average of 17 seconds between tasks to 13 seconds between tasks. If only tasks and messages in which the pilot must participate (per standard flight procedure) are included, the interarrival time is 15 seconds between tasks. In either case, the presence of voice communications has a strong effect on the magnitude of the interruption process during the vectoring segment.
Another effect caused by the presence of ATC and inter-crew communications is seen by comparing figures 5.4a and 5.4b. The figure indicates that, if communications tasks are counted along with the other four task categories of the timeline, then the degree of task simultaneity is strongly affected. From figure 5.4b it is apparent that the inclusion of messages significantly increases the frequency of occurrence of three and four simultaneous tasks. The pilot, thus, may perceive himself to be incrementally exerting more effort due to the influence of voice messages during this segment.

Examination of figure B.2 (Panels 2-5) shows that the few ATC messages that are directed to Clipper 54 tend to explicitly load the crew. This is because the critical path lies along the communication event path for most of the vectoring segment. Most of these messages invoke turning or descent maneuvers to be executed. Explicit loading due to ATC messages is very evident, then, during the vectoring segment.

Following the above discussion, it is possible to infer how the high message rate profile shown in figure 5.7 will affect crew loading during the vectoring segment. The high rate will cause the interruption process induced by the arrival of ATC messages and the necessity to monitor them,
to be greatly intensified. The implicit mental loading of the crew due to the ATC Voice Communications mechanism will be greater than with the low rate profile that is depicted in figure 5.6 (and in the timeline). The magnitude of explicit loading due to ATC messages would not be affected because messages directed to the subject aircraft are unchanged in the communications scenarios described by figure 5.6 and 5.7. The dialogue which figure 5.7 represents is produced by introducing additional messages directed to traffic in the terminal area, other than Clipper 54.

The Final Approach Segment

Figure 5.6 indicates that the average message rate at 10 messages per 80 second interval is slightly higher during the final approach segment than during the vectoring segment. The channel utilization for these two segments is nearly identical at 30%. This suggests that messages encountered on the Tower control frequency are slightly more frequent, but of shorter mean duration.

According to Panels 14-16 of the timeline, few of the ATC messages occurring are directed to Clipper 54, the subject aircraft. This is substantiated by Panels 6 and 7 of figure B.2 which shows that, unlike the vectoring segment, the critical event path does not coincide with the communications event path. Explicit loading due to ATC messages is, therefore, likely to be minimal.
Implicit loading effects due to ATC and intercrew messages are expected to be significant during this segment of the arrival. Again Table 5.3 indicates that when communications are counted as tasks, the mission task interarrival time is reduced from 16 to 12 seconds between tasks. If the high rate profile of figure 5.7 were introduced into the segment, implicit loading due to messages would be greatly increased. This increase would tend to increase the nominal workload level for the segment as the magnitude of the interruptions due to messages increased. Note that these interruptions occur when there are two simultaneous tracking tasks being performed and when the timeline indicates that task slack values are very small (and thus tasks are tightly constrained to be executed promptly).

Again, the simultaneity histogram of figure 5.5 provides additional information with regards to voice communications. Comparison of figures 5.5a and 5.5b shows that the inclusion of voice messages as discrete tasks markedly changes the shape of the simultaneity histogram. The presence of the communications task category causes the frequency of three and five simultaneous tasks to be significantly increased. The frequency of occurrence of two and four simultaneous tasks is reduced. As in the vectoring segment, the presence of voice communications on final approach should tend to incrementally increase perceived workload. The magnitude of this increment is investigated experimentally by introducing
the higher message rate communications profile of figure 5.7 (see Chapter 6).

5.4 CONCLUDING DISCUSSION

The previous sections have described in some detail the nature, a priori, of cockpit task loads during a simulated Boston arrival. This task load was then analyzed in the context of the workload elements discussed in Chapter two. It was found that the task load and workload characteristics of the three arrival segments (cruise descent, terminal area vectoring, final approach) are quite different, necessitating the separate consideration of these segments in workload analysis. This section summarizes the implication of the preceding analysis and serves to tie the chapter together.

It should be noted that the phrase 'workload level' has been used in the relative sense in this chapter. The scenario portrayed in this analysis is modelled after a normal, IFR, transport category arrival. In this nominal case the workload is probably never excessive. However, compared to some baseline, the nominal workload level can be inferred to be relatively higher or lower over different time intervals of an arrival.

The Cruise Descent
The cruise descent was shown to be a rather long segment where task loading was minimal as were interruptions to this task load. Several major planning tasks do occur during this segment but these are afforded sufficient slack as to allow ample time in which to perform these tasks satisfactorily. Workload levels were inferred to be relatively low, with the exception that high ATC messages rates may significantly increase perceived pilot workload through the implicit loading mechanism discussed in Chapter 4.

The Terminal Vectoring Segment

The terminal vectoring segment was shown to be half the duration but with roughly twice the task load in terms of number of tasks performed per second. Interruptions caused by the occurrence of multiple and simultaneous tasks were shown to decrease the mean task interarrival time from 31 seconds, for the descent, to 17 seconds. Navigation and planning tasks were significant, comprising approximately 30% of the total number of tasks during the segment.

Voice communications significantly contribute to crew loading in the vectoring segment. The task precedence map of figure B.2 indicates that ATC communications explicitly trigger numerous tasks to be executed. Intercrew communications associated with checklist execution also introduces an
additional implicit loading effect by reducing the mission mean task interarrival time to 13 seconds from 17 seconds (see Table 5.3).

The nominal workload level would appear to be relatively higher than that for the cruise descent segment based on the analysis of the chapter and the theory discussed in chapter two. The slack values for the segment are still generous, though markedly less than the values during cruise descent. It is suggested that the presence of smaller average slack values may result in high workloads should a cockpit emergency or difficulty occur.

The Final Approach Segment

The final approach was shown to be very similar to the vectoring segment in terms of busyness and interruptions. The notable difference between the two segments is that most tasks which occur on 'final' are associated with events that lie along the critical path of figure B.2. These tasks thus have little time flexibility and must be executed promptly.

This segment contains the simultaneous tracking tasks associated with flying the Instrument Landing System. The analysis presented provides no information as to the difficulty of this task but does suggest that the presence of
high ATC message rates, intercrew communications, and other required operating tasks can induce relatively high workload levels. Because tasks are afforded little slack during final approach, extraneous tasks deferred until the final approach, or introduced by external sources could cause instantaneous task demand rates to rise above satisfactory levels. It appears that the workload level during 'final' is higher than that during the cruise descent. It is not clear from the data presented whether workload during 'final' is relatively higher or lower than the level for the vectoring segment.
Chapter VI

FLIGHT SIMULATOR EVALUATION OF A SUBJECTIVE RATING SCALE

This chapter describes a series of flight simulator experiments that were designed as a preliminary test of the MIT, subjective pilot workload rating scale.

6.1 EXPERIMENTAL OBJECTIVES

The experiments described in this chapter were preliminary in nature. Because the MIT workload scale is new and has undergone only elementary design iterations, the structure of the scale is expected to evolve, to some extent, as it is used by engineers and pilots. The properties of rating scales are strongly influenced by errors of construction and implementation. Experimental use is thus an important step in developing the rating scale into a useful and accepted form. Such 'development' is the primary objective of the experiments described herein.

Next, this thesis has described some relatively new concepts in the theory of pilot workload. The simulator test described in this chapter were designed to explore some of these concepts. In particular, the following issues were investigated experimentally:
1. Sensitivity of the scale versus the sensitivity of performance measures.

2. Consistency of ratings among and across subjects for the same nominal task load.

3. The relationship between busyness and perceived workload.

4. Sensitivity of the scale to the implicit loading of ATC communications.

6.2 FLIGHT SIMULATION FACILITY

Flight missions were simulated using a fixed base Boeing 707 simulator. Figures 6.1 and 6.2 show the exterior and interior of the aircraft cab. Although the cab mockup is a substantially accurate replica of the 707, there are some noticeable features missing. In particular, the co-pilot's station is configured without flight controls and there is no flight engineer's station. Again, it should be noted from figures 6.1 and 6.2 that the cockpit is fixed base and there is thus no physical cockpit motion. Moreover, there is no provision for out-the-window visual simulation so all simulated flights represented instrument meteorological conditions (IMC).
Figure 6.1: Cab Mock-Up

Figure 6.2: Cab Mock-Up Interior
**Flight Instruments**

The flight instruments were generated in real time by a computer and displayed on cathode ray tubes located in the center of the captain's and first officer's panel. The computer generated flight instruments moved in a realistic manner and were displayed essentially flicker free. As can be seen from figure 6.2, airspeed, altimeter, vertical speed, FMI, attitude director indicator, and horizontal situation indicator were generated. The ADI/HSI was modelled after the Collins FD-109 flight director. However, during the experiments only raw data (Flight Director Mode - Off) was displayed and so no command bars were generated.

**The Simulation**

The facility's simulation capability was based on an Adage AGT-30 minicomputer which was located in a room adjacent to the cockpit area. The Adage computer served three functions in the simulation process:

1. It simulated the equations of motion of the Boeing 707.

2. It sampled control inputs from the cockpit and stored flight performance data from the experiments.

3. It generated the flight instrument displays as the simulation ran.
Cockpit controls were sampled 15 times per second and a new aircraft state was computed 30 times per second. Flight instrument displays were updated 5 times per second.

Together the cockpit and the Adage computer form the closed loop flight simulation capability used in these experiments. High computation speeds allowed for good response to pilot control inputs.

Crew Stations

For this test series the captain was the test subject. Therefore, only the captain's station was outfitted to detail. From figure 6.2 it can be seen that the overhead panel switches, operating placards, and functional accessories were quite complete. It was felt that attention to such detail was important in simulating the cockpit environment for workload studies. Seemingly unimportant hand movements all comprise distractions that are important elements of the 'loading' process in the cockpit. Attention to detail was attempted in preparing this experimental study.

6.3 EXPERIMENTAL SCENARIOS

Two basic flight profiles were used in the rating scale evaluation tests. Each flight profile (see figures 6.3 and 6.4) involved an instrument arrival from a high altitude cruise configuration, through an instrument approach to Boston's Logan airport. Scenario parameters were adjusted to form five unique tests.
a) Plan View (Courses shown are magnetic)

b) Altitude Profile

Figure 6.3: ILS 4R Arrival Profile
Figure 6.4: VOR-DME 15R Arrival Profile
The basic profile shown in figure 6.3 consisted of an arrival from Albany, New York at 36000 feet. The flight was routed along federal airways until the simulated aircraft reached the Boston terminal area at which point it was vectored to the final approach course of the active runway at Logan airport. This first profile included an ILS approach to Logan's runway 4R. The mission was terminated when the subject aircraft reached the published Category II missed approach point.

The second flight profile (that of figure 6.4) was nearly identical to the first. It consisted of the same arrival from Albany, N.Y., but concluded, instead, with vectors to a VOR-DME approach to runway 15R. Again, the flight was terminated upon the arrival of the simulated aircraft at the published missed approach point.

**Workload Parameters**

Each of the above flight profiles was altered in one of three ways to increment the nominal workload level:

1. Increment the ATC message rate (i.e., the rate at which radio messages occur on the frequency). Figures 6.5 and 6.6 show the high and low rate message profiles which were used.
2. Execute the flight profile of figure 6.4 rather than that of 6.3, or vice-versa.

3. Introduce an unexpected change of the instrument approach, just prior to initiating the approach procedure.

Altering the workload level in this manner had the advantage that no artificial tasks (such as a secondary, light cancelling task) were introduced, while some control was still maintained over the workload increment. This technique eliminated the intrusion effects of secondary loading methodologies. The disadvantage of this technique, however, was that it was difficult to create workload levels that were incrementally 'higher'. This may have caused difficulties in exercising the sensitivity of the measure. For the present objective, though, the value of using realistic loading mechanisms was thought to be important in evaluating a workload measure for the air transport environment.

Five unique test scenarios were thus defined. The scenarios are summarized in Table 6.1 and a test matrix that is helpful for identifying these scenarios is shown in figure 6.7. These five experiments allowed the hypotheses previously outlined to be examined more closely.
Figure 6.5: Low Message Rate Communications Profile
Figure 6.6: High Message Rate Communications Profile
6.4 **EXPERIMENTAL PROCEDURE**

Test subjects completed a four step process as part of their participation in the rating scale evaluation.

1. Preflight Briefing

2. Warmup Flights

3. Data Flights

4. Post Flight Critique

These experimental flights were aimed at evaluating the MIT scale as a measure of the pilot's workload only. The scale's use for other crew stations was not tested here. All simulated flights situated the subject as pilot-in-command and pilot-at-controls in manual flight. No auto pilot or flight director information was provided. It was, however, necessary to simulate the captain's duties, his cockpit station, and the captain's interaction with his crew.

As has been mentioned, the captain's station was replicated to detail, the copilot's was less detailed and there was no flight engineer's station. The flight engineer duties with which a captain normally interacts were performed by the copilot. Primary examples of these duties were preparation of landing/approach data and the reading of checklists. Also, there are, in actual line operations, intercrew communications that take place between captain and
<table>
<thead>
<tr>
<th>Scenario No.</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Albany arrival to an ILS runway 4R. Low ATC message profile.</td>
</tr>
<tr>
<td>2</td>
<td>Albany arrival to an ILS runway 4R. High ATC message profile.</td>
</tr>
<tr>
<td>3</td>
<td>Albany arrival to VOR-DME 15R. Low ATC message profile.</td>
</tr>
<tr>
<td>4</td>
<td>Albany arrival with a planned approach to an ILS runway 4R. Approach switched to VOR-DME 15R, just prior to MANJO intersection. Low ATC message profile.</td>
</tr>
<tr>
<td>5</td>
<td>Albany arrival with a planned approach to a VOR-DME 15R. Approach switched to ILS runway 4R, just prior to MANJO intersection. High ATC message profile.</td>
</tr>
</tbody>
</table>
### Figure 6.7: Scenario Test Matrix

<table>
<thead>
<tr>
<th>ATC Message Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High</strong></td>
</tr>
<tr>
<td><strong>Yes</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Change in Approach?</strong></td>
</tr>
<tr>
<td><strong>No</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ATC Message Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High</strong></td>
</tr>
<tr>
<td><strong>Albany Arrival</strong></td>
</tr>
<tr>
<td><strong>ILS 4R</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Albany Arrival</strong></td>
</tr>
<tr>
<td><strong>VOR-DME 15R</strong></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
flight engineer. These communications were not simulated. The experiments thus replicated a two-pilot crew such as currently exists on the Boeing 737 and DC-9 aircraft.

Again, the pilot subjects acted as both pilot-in-command and as pilot-at-controls for the entire duration of the test scenarios. The copilot was told to perform duties called for by the flight operations manual and those commanded by the captain-subject. The copilot could not, however, activate the control wheel or rudder.

6.4.1 Preflight Briefing

Prior to flying the simulator, subject pilots were asked to fill out an experience questionnaire (see Appendix C). Basically this questionnaire identified the subject's currency and general experience in IFP and air transport operations.

Pilots were next briefed on the purpose of the experiment and given a detailed briefing on the use of the workload assessment scale. Upon completion of the rating scale briefing, subject pilots were then briefed and familiarized with the MIT simulation facility. Each pilot was supplied with a set of instrument enroute and approach charts for the New England area. Pilots unfamiliar with the area were given the opportunity to review these charts.
Test subjects were briefed on a standard set of crew procedures. These procedures were drawn from the operations manual of a U.S. airline that operates the Boeing 707-320 aircraft. The procedures were thus representative of standard operating practices. Each pilot was also briefed on his responsibilities and duties as pilot-in-command and as pilot-at-controls. Subjects were briefed, in addition, on the duties and responsibilities of the copilot, whose role was played by an MIT researcher experienced as an IFR pilot.

It should be noted that the majority of the participating subject pilots in this evaluation were Boston based. The subjects were, for the most part, very familiar with the Boston area and with the standard instrument approaches for Logan airport. None of the subject pilots were 707 pilots and so the characteristics of the simulated aircraft were unfamiliar, initially, to all.

The preflight briefing required an average of one hour to complete (excluding the questionnaire).

6.4.2 Warmup Flights

Test subjects flew a series of 'warmup' flights to accustom the pilots to the flight procedures and to the operating characteristics of the simulator. Table 6.2 lists the warmup scenarios in the order in which they were generally flown. In some cases the warmup flying period was reduced
as some of the pilots were particularly adept at flying the simulator. In other cases the warmup phase was extended until the subject pilot seemed at ease and performing to nominal IFR standards of acceptability.

TABLE 6.2
Warmup Flight Scenarios

i) Arrival from Gardner VOR at 14000 feet with vectors to an ILS approach to runway 4R.

ii) Two practice approaches on the ILS 4R. Approaches began at 2000 feet, 2 miles west of the final approach course, on an intercept heading of 70 degrees.


Note: A low ATC message rate was used during the warmup.

6.4.3 Data Flights

Subject pilots flew three arrivals during which data was collected. In some instances fewer than three arrivals were flown due to equipment malfunction. In all cases, however, data was sampled in a manner that allowed a variety of data comparisons to be made. Workload data that was sampled could be compared across measure types, across pilot groups, between flight segments, and across test scenarios.

Flight Data: Performance

- 162 -
The pilot's mean absolute deviation from glide path (on precision approaches) and frcm localizer (on all approaches) was sampled by the Adage computer during the final approach segment of each arrival. It has been shown that, in some cases, performance data is sensitive to variations in total workload. Performance data could thus be used in a comparative manner with the rating scale results for final approach.

The displacement of controls (physical control activity) was sampled and recorded on an analog chart recorder during data flights. Controls sampled were aileron, elevator, and pitch trim. It was hypothesized that the flight control activity (in manual flight) might serve as an index of workload. This data, then serves as another measure for comparative use.

**Flight Data: Subjective Ratings**

Each data flight consisted, logically, of three mission segments. These were the cruise descent, terminal area vectoring, and final approach segments. For the ILS 4R arrival, these segments had durations of 14, 8, and 4 minutes, respectively. Similarly, for the VCR-DME 15R arrival, these segments had durations of 14, 4 and 4 minutes. After each segment of an arrival, the simulation was frozen and the subject was asked to use the rating scale to assess his average segment workload. Freeze points were selected to minimize discontinuity effects.
In addition to the pilot's self-assessment, the copilot/observer also rated the subject pilot's workload. This was an attempt to investigate the ability of trained observers to rate workload. It also provided a cross-check on the ability of a crewmember to evaluate his own workload.

**Flight Data: Video**

A low light-level camera was used during all flights to record audio and video activity in the cockpit. The camera was focused over the captain's right shoulder so that the pilot's hands and flight instrument group were within the camera's field-of-view. Control movements outside the field-of-view of the camera (such as altitude alert, speed brake, and throttle) could also be noted.

Several subjects were asked to view the video replay of one of their test flights, after completion of this flight. Again the video was frozen at the end of a segment of interest, at which time the subject was asked to reassess the segment workload. This data was expected to yield information on the ability of pilots to self-assess their workload and to experiment with post-test, pilot calibration*.

*Calibration would involve telling a subject that this flight represented a two not a four.*
6.5 **DISCUSSION OF RESULTS**

In total 8 pilots participated in the rating scale evaluation tests. Of these four were airline pilots with over 4000 hours of air transport experience and four were general aviation pilots, two with less than 1000 hours of general aviation, IFR experience. This sample size proved to be adequate for identifying some properties of the scale and some problems associated with its implementation. As these tests were preliminary in nature, the sample was not meant to provide a data base with which to statistically test hypotheses. The discussion that follows highlights the major rating scale properties and problems that were noted in testing.

**The Learning Curve**

The primary difficulty encountered in these simulator experiments was a marked learning curve effect demonstrated by the subject pilots. That is, pilot workload ratings for a particular flight phase consistently dropped as the subjects became intimately comfortable with the simulator and the flight procedures. This occurred despite concerted efforts to circumvent learning curve effects by establishing a rather long briefing/warmup flight period as described in section 6.3. Pilots spent approximately six hours at the MIT facility during their participation. Of this approximately three hours was spent flying the simulator and one hour of this three was spent in warmup flights. The result
of the pronounced learning curve effects on the test data was that consistency, within a subject, was not evident in much of the rating data. Consistency was evident only in later runs, after the learning curve had 'flattened out'. It is evident from figure 6.8a that the same scenario consistently produced lower ratings in later runs, for a given subject pilot. Moreover, figure 6.8b indicates that, in several cases, these lower workload ratings were accompanied by better performance. Together, these data confirm the presence of strong learning curve effects.

These learning curve effects limit the conclusions which can be drawn from the rating scale data and thus are important to note. It is believed that such effects will strongly influence other experiments utilizing the rating scale. The following conclusions can be drawn with regard to learning curve effects and the use of rating scales:

1. If consistency across or within subjects is desired, the experiment will have to be repeated until such time as the ratings stabilize. Performance stabilization and verbal questionnaires are inadequate indicators of learning curve plateaus. Subjects may very well show excellent performance but assess lower and lower workload ratings (see figure 6.8).
Figure 6.8: Learning Curve Effects

- 167 -
2. If unfamiliarity or surprise is an issue in the measurement experiment, but consistency is desirable in the rating data, then subject pilots must be chosen with very similar experience and skill levels. It also appears that pilots in this kind of emergency workload measurement environment, should have extensive training in the use of the scale. Such training would necessarily calibrate the subjects to the rating scale.

Egocentricity

Second in importance to learning curve problems in the rating scale data is the apparent problem of subject egocentricity. Flight officers and pilots, as a sociological sect, have very protected egos. As a group they seem particularly proud of their skills and their flight positions and thus seem to be cautious in the extent to which they will truthfully assess their workload. In particular, these experiments encountered one subject who would not rate any test situation higher than the lowest conceivable rating (a one or a two).

Figure 6.9 shows that subject number 3 is indeed an 'outlier' for the final approach scenario that involved a VOR-DME approach to runway 15, with a low message rate. For this run the subject assessed his workload as being very low.
Figure 6.9: Effects of Egocentricity
and, yet, his performance was well outside the tolerances of instrument flight rules. Also, for this particular segment, the subject's multi attribute rankings state that the subject perceived only occasional moments of free time, moderate levels of information processing intensity, and mild levels of emotional stress and confusion. These perceptions seem inconsistent with a run where the subject was one half a mile off course and his rated workload as low as a 2.

In these experiments, egocentricity appeared to be a problem with only one subject and its affect on the data is therefore not as significant as the learning curve effect. Despite this, it is thought that egocentricity may pose a difficulty in future use of the pilot workload rating scale. These tests seemed to indicate that further ego problems could be eliminated by including the following steps as part of the preflight briefing:

1. Ensure that pilots understand that they are assessing the workload required by a set of flight procedures in a given cockpit and flight environment. It is the procedures that are being evaluated, not the subject pilot.

2. The subject pilots should not be made aware of how 'most' pilots rate a particular segment, unless the intent is to calibrate the
subjects to the scale. The subjects should also not be unintentionally made to feel as if they are being compared to a pilot of higher or lower skill/experience/prestige.

These steps seemed to have minimized further data integrity flaws due to subject egocentricity.

Workload Perception: Self-Assessment vs. Observer Assessment

One of the objectives of this test series and, indeed, question marks associated with the use of subjective rating scales, concerns the ability of a subject pilot to assess his own workload. Figure 6.10 indicates that there was, in general, excellent agreement between subject and observer*. The figure displays surprisingly little scatter, with most of this scatter indicating that the observer's assessment was lower than the subject's assessment. Although figure 6.10 certainly does not prove that subjects can assess their own workload, it does lend some supportive evidence to this hypothesis.

One subject pilot was familiar enough with the rating scale so that he was, at cued intervals, able to call out a nominal assessment as he flew. Figure 6.11 shows the results of this run. It should be noted that, again, there was good agreement between subject and observer**. Also, _____________________________

*Observer and subject assessments were made independently.

**Note that the observer's assessment was made upon viewing
b) General Aviation Subjects

Figure 6.10: Subject vs. Observer Ratings

---

the video replay.
of special interest, the jaggedness in the early portions of the rating profile (see figure 6.11) correspond well to the nominal task load depicted in the timeline of figure B.1 of the appendix.

Several test subjects were asked to view a video replay of their flights and then reassess their workload using the scale. In no case did these reassessments yield markedly different ratings and in most cases the original rating remained unchanged. Finally, each subject was asked, during debriefing, whether they felt it was possible to judge or perceive their own workload. The response to this inquiry was a unanimous affirmative, which was generally expressed with a strong desire to try the self assessment procedure, if there was some doubt as to its value. This was perhaps the strongest evidence in favor of the subjective assessment technique. It indicates that pilots do seem (over the range of workload levels tested) to feel that they are able to rate their own workload.

**Workload, Performance, and Skill Level**

Chapter 2 summarized Simpson and Sheridan's (1978) functional relationships between workload, performance, and skill level. Simpson and Sheridan would contend that there is no simple relationship between these parameters and that workload is not equivalent to performance. Figure 6.8b
Figure 6.11: Continuous Workload Assessment for an ILS 4R Arrival
shows several examples of perceived workload versus performance, all of which are consistent with the theory outlined in chapter 2.

The general aviation pilot (indicated by the rectangular points) assessed identical ratings for two runs with the same nominal task load, and yet his performance improved by a factor of 2. The two examples using airline pilot data (the circular data points) indicate both a workload reduction and a performance improvement in the later runs. Thus as the pilots' skill increased their performance improved and their workload was rated as lower. For the case of the general aviation pilot, as his skill level in performing the final approach increased, he was able to expend a comparable amount of effort and yet perform better. Notice in figure 6.8b that the general aviation pilot performed (on his 2nd run) nearly as well as the airline pilot (on his earlier run), but the general aviation subject rated his workload to be much higher than the airline subject. This, too, is consistent with theory which states that, if the skill level of one pilot is higher than that of another, the more skilled pilot will exert less effort to perform the same tasks.*

**Workload and Busyness**

*In this case the general aviation subject's experience, measured in hours, is a fraction of that of the airline subject's. Hours experience was taken here to be a proxy for skill, although flight hours is not always a good metric of relative skill.
It is interesting to note that 'percentage of busy time' appears to bear no simple functional relationship to perceived workload. That is, the data indicates numerous examples of low workload levels where both the task timeline of figure B.1 and the multi attribute rankings of the subject indicate high levels of busyness. The subjects in these cases were performing a fairly difficult instrument approach maneuver using almost reflexive, low level effort.

Conversely, there are also examples (see table 6.3) of relatively high workload during periods where the cockpit activity timeline and the multi attribute rankings indicate low levels of busyness. These data provide further supportive evidence to the theory outlined in chapter 2 which suggests that busyness, alone, is not a good workload measure, even in an aircraft under manual control.

**Rating Consistency within a Subject**

Due to the learning curve and egocentricity effects discussed previously, the simulator tests did not yield any reliable data on the repeatability of ratings for equivalent nominal task loads, within a subject.

**Rating Consistency across Subjects**

Ignoring the test subjects' earlier runs due to learning curve effects, there is some encouraging evidence that consistent ratings, across pilots, are attainable. Figure 6.12 shows rating data for the high ATC message rate, ILS
Table 6.3: Workload Ratings vs. Multi attribute Busyness Rankings

<table>
<thead>
<tr>
<th>Workload Rating</th>
<th>Busyness Attribute</th>
<th>Always Idle</th>
<th>Have Free Moments</th>
<th>Always Fully Occupied</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>often</td>
<td>Occas.</td>
<td>Rarely</td>
</tr>
<tr>
<td>i</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ii</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>iii</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>iv</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>GA Subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Airline Subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>ii</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>iii</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>iv</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>
approach to runway 4R. Observing only the airline subject data points (the circular points), it is apparent that these ratings are fairly consistent. The figure indicates a general consensus that the cruise descent phase is a 1, the terminal vectoring stage is a 2 or 3, and that the final approach is a 2 or 3.

Recall that these rankings are quite consistent with the results of the analysis of chapter 5. This analysis suggested that the cruise descent phase of the arrival should induce the lowest workload levels of the three arrival segments. The analysis also concluded that it was unclear whether the nominal workload on final approach should be higher or lower than that for the terminal vectoring segment. Note also that all segments seem to have a fairly low nominal workload. This, too, is consistent with the theory of chapter 2 and the analysis presented in chapter 5. It reconfirms what is already common knowledge. That is, nominal workloads on routine, air transport arrivals are not excessive, despite the complexity and busyness depicted in figure B.1 of the appendix.

Observation of the general aviation subject ratings indicates that consistency is less obvious for this class of subject. In part, this is due to the small sample size and in part, to a greater degree of data scatter. This increase in scatter is very likely due to the wider variation in
Figure 6.12: Rating Consistency for an ILS 4R arrival
experience and skill that is present among general aviation pilots, than is present among Part 121, airline pilots.

One trend that is evident in figure 6.12 is that the general aviation subjects seemed to have, on average, rated the workload for this scenario to be higher than have the transport pilots. This lends some further support for the skill/performance/workload relationships discussed previously.

To summarize, the data does provide evidence showing that consistent ratings are attainable. The small sample size, however, limits the conclusiveness of this evidence, particularly with regard to the repeatability of a subject’s rating.

Sensitivity of the Scale to ATC Message Rate

These tests provided no reliable indication that the rating scale was sensitive to an increase in workload due to the implicit loading effects of ATC messages. Once again, learning curve effects occurred across scenarios that were designed to investigate implicit loading, causing this data to be unusable for this purpose.

Pilots did, however, react on several occasions to the message scenarios. In one instance, a subject pilot insisted on knowing the distance of his aircraft from other simulated aircraft and several times prompted the controller
for this information. As discussed in chapter 4, this subject had a mental image of the local traffic situation and knew that there were aircraft not far behind him in the terminal area.

As another example of implicit loads, several pilots thought one of their ATC clearances to be ambiguous. From the clearance

*Clipper 54, depart MANJO at 6000 feet on a heading of 170 degrees*

it was not obvious to these pilots whether they were cleared down to 6000 feet, or not. Following this, these captains generally discussed the clearance itself and, in some instances, examined navigation charts for minimum altitudes or asked ATC to confirm their interpretation of the clearance. None of these extra tasks is noted on the task analytic description of a timeline and all contribute to the instantaneous workload of the crewmember. Moreover, the confusion aroused by this message was apparent to the observer, and to the subjects, upon replay of the video recording.

Despite these noted occurrences of implicit loading effects due to ATC messages, the rating data do not provide any reliable evidence as to the scale's sensitivity to such effects. Moreover, the control of the flight simulator experiments was such that the multi-attribute ratings also provide no evidence of these effects.
6.6 CONCLUDING DISCUSSION

To conclude, it is important to restate that these flight simulator evaluation tests scale were preliminary in nature. They were not intended to provide final evidence that demonstrated the sensitivity and universality of the scale. Indeed, the acceptance of a workload rating scale for the air transport environment will necessarily require experimentation in high fidelity simulators and in real aircraft. Such tests will require input from many pilots, manufacturers, and operator. Such acceptance was not the objective of these experiments.

The experiments described in this chapter did yield a great deal of valuable information on the proper use and on the properties of the MIT scale. In particular, it was noted that learning curve effects caused ratings to drop steadily as pilots became more familiar with the simulator. Performance, at the same time, showed little or no improvement. Pilots thus required longer test sessions, despite verbal statements and performance levels that suggested their mastery of the flight simulator/crew procedures.

Despite limitations on the data imposed by learning curve effects, some theoretical concepts were successfully examined. Of special interest were results which consistently showed that perceived workload is not equivalent to performance, nor is the level of busyness equivalent to perceived
workload. The data showed instances of low workload ratings at times of 100% busyness and it also showed high workload ratings at times of low busyness.

Surprising rating consistency, for an equivalent task load/scenario, was achieved across pilots. There was, however, a notable tendency for the general aviation subjects to assess higher workload ratings, for a given scenario, than their air transport counterparts. Due to the presence of pronounced learning curve effects, the test data did not yield any reliable information on the sensitivity of the MIT scale to implicit loading by ATC messages. It should be mentioned, though, that a number of implicit loading 'incidents' were observed to occur during the high message rate, test scenarios. It was simply not clear from the available data whether the scale was sensitive to the presence of this loading mechanism.

It can be concluded that there are indeed problems associated with the use of any subjective assessment technique for pilot workload. Yet, these experiments indicate that the MIT scale can provide valuable and consistent information. It is suggested that further research be undertaken using high fidelity flight simulation technology with experienced, line crews as subjects. These experiments must utilize structured scenarios that include the presence of air traffic control. ATC imposes perhaps the most influential
constraints and is the predominant source of stochastic disturbances to the nominal task load.

The results of these tests indicate that only the low end of the rating scale is exercised by routine, air transport arrivals. To exercise the full range of the workload scale, there is no doubt that a structured scenario set which introduces abnormalities and emergencies will be required. With this 'apparatus' as an experimental base, it would be possible to obtain the more conclusive evidence on scale consistency, sensitivity, and acceptability that the experiments described in this chapter suggest exist.

Finally, the overwhelming response of the subject pilots participating in these evaluation tests, was that they felt they could perceive their own workload. Moreover, subjects expressed an eager desire to try the subjective technique. This finding may be the most important result of the rating scale evaluation series.
BIBLIOGRAPHY


**Federal Aviation Regulations** Part 25.1523 and Part 25 Appendix D.


- 186 -


Yoerger, Dana F., Sheridan, Thomas B., Tulga M. Kamil, Daryanian, Bahman, Experimentation in Supervisory Control and Flight Management Progress Report, Man-Machine Laboratory, Department of Mechanical Engineering, Massachusetts Institute of Technology, September 1978.
Appendix A

SOME TOOLS FOR WORKLOAD ANALYSIS

This section describes several analytic tools that are available for quantifying cockpit activity and for characterizing cockpit workload. These techniques are useful for analyzing flight procedures in the design phase, and for a priori investigations of crew workload. They are particularly useful in that they serve to quantify what has historically been a very qualitative subject. Moreover, these analytic methodologies focus on the mechanisms through which the ATC system interacts with cockpit activity.

A.1 Cockpit Activity Timelines

Cockpit activity timelines represent the standard format for the quantification of cockpit activity. Timelines evolved from the late 1930's idea of scientific management and represent, simply stated, a time and motion study of crew activity. Figure A.1 provides an example timeline. The figure portrays 100 seconds of cockpit activity during the localizer intercept phase of a Boeing 707 arrival.
Figure A.1: 100 Seconds of a Cockpit Activity Timeline
The timeline format of figure A.1 groups tasks into five categories. These are

1. Heading Tasks
2. Altitude Tasks
3. Speed Tasks
4. Navigation and Planning Tasks
5. Communications Tasks (Includes both intercrew and ATC messages.)

Time into the flight mission is depicted, in seconds, along the bottom edge of the figure.

Discrete and continuous tasks are denoted on the timeline by the shaded boxes. The length of these boxes corresponds to the nominal time required to perform this task. Tasks, though placed at some nominal time of occurrence, may slide relative to one another along the dotted 'task tracks.' That is, pilots have some degree of flexibility in executing tasks and may choose to defer the execution initiation time for a task, as workload levels increase. Tasks may not slide across one another (i.e. task precedence is fixed) and they may only slide toward increasing mission times. Moreover, a task may only slide as far as its associated task limit (the darkened TL bars).
For example, in figure A.1, under the heading task category, Tasks 10-13 may slide as far as the Task Limit bars at 1340 seconds. Tasks 10 and 11(21) may not, however, cross over, and thus occur later than, Tasks 12 and 13. Similarly, the figure shows that Heading category Task 14 and 15 may not slide at all. (22) Tasks 14 and 15 are said to have zero 'slack'.

Uses of Activity Timelines

Because of the cockpit activity timeline's detail, it may be used in a variety of ways. First, the format used in figure A.1 is particularly useful in a visual sense for locating time intervals where busy periods are extended or where task interruption and multiplicity is pronounced. Notice, for example, at 1380 seconds on figure A.1 that five categories of task are occurring simultaneously. It is also readily apparent from the figure how the intercrew and ATC communications tend to interrupt other task activity during this portion of the arrival.

(21)*Tasks 10 and 11 are not on the figure. Their nominal time of occurrence is earlier than 1300 seconds.

(22)*Tasks may only slide toward increasing time since the nominal time of occurrence used here is equivalent to a task's earliest time of occurrence.

(23)*Timeline data is either collected from actual or simu-
The nominal data (23) depicted on the timeline can also be used to estimate certain workload indices. A list of these would include

1. Percentage of time busy working on tasks
2. No. of simultaneous tasks
3. Mean task duration
4. Mean task interarrival time

Cockpit activity timelines are also useful in the design of flight procedures to highlight areas of obvious extended effort or task multiplicity. Timelines can be constructed to include all cockpit tasks (as in figure A.1) or they may be constructed to include only the delegated tasks of specific crewmembers. Timelines can, thus, be used in the design of workload tolerant, coordinated crew procedures.

Finally, it is also possible to use the cockpit activity timeline to estimate physical crew loads or mental crew loads. Physical crew loads may be estimated by summing the known physical efforts required to perform the subtasks shown on the timeline. Mental efforts can be estimated by inputting the activity data into human operator simulation models (refer to Hay, et al, 1978 for descriptions of these lated flight tests, or estimated using a set of operating assumptions.)
Limitations of Timelines

Cockpit activity timelines, and task analytic methods related to timelines, have limitations which should be noted. First, timelines, despite their detail, cannot be made accurate to any microscopic standards. That is, there is a great deal of uncertainty in the construction of timelines which must be recognized. Task execution time, for instance, varies with pilot skill, pilot workload, and pilot personality. Similarly, there will be wide variations in practice from the procedure noted in a timeline. Some tasks may never be performed and still others, not shown on the timeline, may be included.

The activity timeline is subject to arbitrariness. There is arbitrariness in the detail to which activity is described. A turn, for example, may be described in complicated detail as a series of hand, eye, and foot movements, or it may be more simply described as 'a turn,' of several seconds duration. Related to this, it is not possible to identify all tasks which must be performed by the crew. Many mental planning tasks occur that are not readily identifiable and must therefore be omitted from the timeline.
Finally, the timeline format of figure A.1 does not indicate how much effort each task requires. (24) It is a fairly common practice to use some weighting scheme to determine a percentage workload from timeline data based on either time occupied or on effort for each task type (see Hay, et al., 1978). These techniques seem counter to current theory (see Moray, 1980) and are not used in this report.

Summary

Cockpit workload is a very complex subject. The issues involved in any quantitative work in this area are still at the forefront of theoretical and experimental research. The detail and the nominal information that timelines provide is a bare necessity, without which an understanding of the crew workload is impossible.

A.2 Task Precedence Maps

------------------------

(24) It does make the distinction between a tracking task, i.e., a task which must be worked on continuously with strict precision, and a Holding task, i.e., an 'easier' task which requires only intermittent attention and that has large error tolerances. Hold tasks are indicated in figure A.1 by solid lines with arrowheads, e.g., H200kts = Hold 200 knots.
Instrument pilots are trained, in general, to prioritize the tasks of their multiple task environment. Because of this, there exists an identifiable hierarchy of task priority and precedence among piloting tasks. Related to this, certain cockpit events tend to cue or trigger the execution of other cockpit events. For example, on a descent into the terminal area, as an aircraft approaches 10,000 feet, the pilot will slow the aircraft to 250 knots or less. The event, 'arrive 10,000 feet,' triggers the execution of the slowing maneuver.

Task precedence maps provide a graphical representation of the constraints, precedence, and triggers which exist in the piloting environment. An analog of the PERT(25) methodology, task precedence maps represent a schedule of discrete events that take place in the cockpit. Events, as used here, denote the occurrence of a physical event such as the passage of an ILS marker beacon or the arrival of an ATC message. The precedence map depicts events, time of the events, the uncertainty associated with the event times, and the precedence which must be maintained between events.

Elements of the Precedence Map

A precedence map (see figure A.2) consists of physical cockpit events (rectangular boxes), event paths, and event constraints. Constraints are indicated by solid, or dotted, lines connecting two or more events. A solid line connecting two events is referred to as a 'hard' precedence constraint. It indicates that some physical task is being performed between the two events, the task duration being denoted by the number enclosed in parentheses. Moreover, an event to the left must precede an event to its right, along connected paths.

A 'soft' precedence constraint (see figure A.3) indicates that Event 1 must precede Event 2, but there is no definite time after Event 2 at which Event 1 must occur, since no physical task separates the two. The time between Event 2 and Event 1, in figure A.3b, must be greater than or equal to zero seconds.

Associated with each event is an earliest time at which the event can occur and a latest time at which the event can occur. These values are indicated by the left and rightmost numbers, enclosed in parentheses beneath the event boxes. For example, for Event 1, figure A.3a, the earliest time at which the turn maneuver can begin is 120 seconds into the mission and the latest time is 323 seconds. The difference between these two numbers is referred to as the 'event slack.' Event slack represents the uncertainty in the time
Figure A.2: A Task Precedence Map
a) Hard Precedence Constraint

b) Soft Precedence Constraint

Figure A.3: Precedence Constraints
of occurrence of each event. It should be noted that the path along which the slack is always zero is referred to as the critical path.

**Constructing a Task Precedence Map**

A task precedence map may be constructed by following a simple four step procedure.

1. Break down the task components and cockpit events into a desired level of detail of events.

2. Set the nominal time of task or event occurrence, on the activity timeline, to the earliest time for an event.

3. Sweep forward through the timeline until all events have been placed on the map, constraints indicated, and earliest times assigned.

4. Beginning with the final scenario event, sweep backwards through the timeline, setting the 'latest time' to the smallest 'earliest time' imposed by the constraints.
An example using figure A.2 will illustrate this process. On the activity timeline, the intercept (associated with Event D22) nominally begins at 1387 seconds into the mission. At approximately this same time, but definitely after the intercept turn, the aircraft arrives at the localizer, 5 miles outside of the outer marker. The activity timeline indicates, also, that the intercept turn requires 15 seconds and so event D23 occurs at 1402 seconds (1387 + 15), at the earliest. ATC procedure dictates that when the aircraft arrives at, but not before, the 5 mile point, the aircraft will be told to contact the tower (Event C10). Flight procedure further dictates that the Mechanical checklist be initiated upon intercepting the localizer. Thus Event E4, is constrained to occur at or after Event A6, which in this case occurs simultaneously with Event A7.

Event C10 is assigned an 'earliest time' of 1387 since it is constrained to occur no earlier than event A7. Similarly, Event E4 may then be assigned an 'earliest time' of 1387. Should two or more precedence constraints affect a single event, C10 perhaps, then the 'earliest time' assigned should be the largest value imposed by the appropriate constraints.

'Latest times' are assigned by sweeping backwards from the final event. For instance, Event A8 (not shown) has a 'latest time' of 1457. Event A7 must therefore be assigned the 'latest time' of
Minimum (1457-70 or 1600 seconds as imposed by Event C10)

This process continues backward until the first event has been assigned a 'latest time'. Note, on figure A.2, that along the double line the slack is always zero. Thus, the double-lined path indicates the location of the critical path.

Uses of Precedence Maps

Precedence maps have several obvious uses in workload analysis or flight procedure design. The most obvious may be in the use of the slack values. These values indicate the time slideability of tasks defined on the timeline. Thus, for each cockpit task, a Task Limit may be identified and placed on the cockpit activity timeline. Figure A.1 shows the presence of Task Limits by darkened TL bars, where these bars are located at the 'latest time' associated with specific events.

The value of the slack, itself, may indicate the potential for high workload situations, or the workload tolerance of a set of flight procedures. That is, if slack values are large, tasks associated with these events may be deferred for long periods should tasks of higher priority be introduced. Similarly, if mean slack values are small, then tasks are not readily deferrable, and the introduction of
high priority tasks by exogenous sources may lead to potentially high workload situations.

The precedence map also highlights the existence of task triggers by indicating which events constrain other events. Thus, in figure A.2, the aircraft's arrival at the 5 mile point triggers the radio frequency command to occur. In other instances one event may trigger a long sequence of other events and their associated task to occur. It may, in some cases, be appropriate to change procedure if the task sequence becomes inordinately complex with very small average slack values.

Finally, the precedence map indicates the presence of a critical event path. The critical path is that path which constrains all other paths and events. Thus, if the critical path is changed the 'earliest' and 'latest' times on all other paths will be affected. Moreover, the critical path generally indicates which path or set of events is imposing the dominant constraints. If the critical path is thus lengthened, or events along the critical modified by restructuring flight procedures or ATC procedure, the nominal workload will be modified.

Summarizing, task precedence maps are a useful analytic tool for evaluating cockpit procedure in the context of pilot or crew workload. They are derived from cockpit activity timeline data but are different in that
1. They identify a critical path of events which constrains all other cockpit events.

2. They highlight formal precedence constraints.

3. They indicate which system elements are triggering other cockpit events to occur.

The precedence map is generally most useful when constructed with a more macroscopic level of detail as compared to the microscopic detail with which activity timelines are generally constructed. It is suggested that these two tools should be used in a complementary manner in workload analysis.

A.3 Stochastic Analysis of ATC Communications

It has been shown that ATC communications are a source of loading to an air transport crew. This loading acts through two types of mechanisms.

1. Explicit mechanisms - ATC messages in the form of advisories or clearances that cause the pilot to execute some set of physical or mental planning tasks.
2. Implicit mechanisms - ATC messages implicitly interrupt and distract the crew. Some messages, e.g. severe weather warnings can also arouse anxieties and confusion in the cockpit.

The extent to which explicit mechanisms affect the crew can be determined from cockpit activity timelines and task precedence maps. Implicit loading, however, is the result of the random occurrences of ATC messages. It is useful to examine some of the features of this stochastic process.

Each message distracts or diverts the attention of the crew. The probability that a message will occur, then, is one parameter which characterizes the magnitude of this interruption process. Given a probability distribution for 'message occurrence,' in a given time interval, the mean no. of messages or mean message rate can be computed. Dunlay (1975) has examined this in some detail.

The most comprehensive treatment of the statistical properties of ATC communications was performed by Hunter, et al (1974) in the development of an ATC communications simulation. Hunter defined three groups of data that characterize all ATC message activity. These are
1. Communication transaction (CT) - A dialogue between a controller and one aircraft.

2. Transmission (TR) - When a microphone is keyed.

3. Message element (ME) - One phrase or message of which there may be none, or many, in a transmission.

As figure A.4 and A.5 show, in one ATC sector an aircraft typically participates in several CTs. For every CT there may be one or more TRs, and there may be several message elements in each transmission.

CT/aircraft, TR/CT, and ME/TR are all random variables. In addition, the length of time of each transmission is a random variable. Hunter created 'dictionaries' which tabulated all message elements and listed the proportion of transmissions in which these elements occurred. (26) Moreover, he identified probability distributions that described CT/aircraft, TR/CT, and TIME/TR, given the number of aircraft in a control sector. For the New York area Hunter found the following distributions to representative of the data.

(26)*Hunter used data from the New York and Houston control areas.
Figure A.4: Air Traffic Communications Performance
Figure A.5: Decomposition of a Communications Transaction
CT/aircraft - Doubly shifted, negative binomial distribution
TE/CT - Truncated, negative binomial distribution
Time/TR - Gamma Distribution

It is suggested that in workload simulation approximations to Hunter's techniques should be used in constructing a realistic ATC communications scenario. Moreover, it is suggested that in the workload analysis of a mission phase, the statistical parameters that describe the ATC communications provide an indication of the magnitude of the implicit loads due to messages.
Appendix B

FIGURES FOR THE ANALYSIS OF A BOSTON ARRIVAL
COCKPIT ACTIVITY TIMELINES - LEGEND

Task #10
Duration=10 sec.

Task #11
Duration=11 sec.

Task Slide Limit for Task #11

Task Sliding Track.
Tasks may slide from their nominal time (shown) to right only.

Tasks retain precedence depicted, i.e. Task 10 may slide as far as TL 10, but it may not cross over Task 11.

Altitude Tasks

H2000fpm

Indicates a Hold type task, i.e. Hold 2000 feet per minute descent rate.

Hold type tasks require only intermittent attention from crew.

Time, in seconds, into mission.

20 40 60 80 100
Figure B.1: Cockpit Activity Timeline, ILS 4R Arrival
<table>
<thead>
<tr>
<th>HEADING</th>
<th>TASKS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALTITUDE TASKS</td>
<td>H500fpm (FL230)</td>
</tr>
<tr>
<td>SPEED TASKS</td>
<td>H320kts</td>
</tr>
<tr>
<td>COMMUNIC. TASKS</td>
<td>9) CL 54 clear 8000 cross GDN 14000</td>
</tr>
<tr>
<td></td>
<td>10) Response, CL 54</td>
</tr>
</tbody>
</table>

- **3)** Set Bug
- **4)** Turn to 118
- **5)** Track 118 Radial
- **6)** Pitch Change
- **7)** Trim
- **8)** H3000fpm
- **9)** H320kts
- **10)** H320kts
5) Track 118 Radial

9) Set Altimeters to field

TL 1-8

H2000 fpm (FL190)

TL1

H320 kts

NAVIG. AND PLANNING TASKS

COMMUNIC. TASKS
| TASKS | HEADING | | Altitude | | Navig. and Planning | | Communication |
|-------|----------|-------------------|------------|-------------------|------------------------|--------------------------|
| TL 14-18 | 25) Trim | | | | |
| H250kts or less | 4) Throttles to 1.6EPR | | | | |
| | | | 27) Initial Call-up, AA388 | 29) Resp. AA388 | 31) Resp. CL 54 |
| | | | | | |
| | | | | | |
| 1020 | 1040 | 1060 | 1080 | 1100 |
26) Set Alert
27) Pitch Down
28) Trim to 2000fpm

TL 19-21

H160deg.

H2000fpm

TL 2, 3

)Throttles Idle

H220kts

H250 or less

10) Tune 110.3
11) Select 035 CDI

NAVIG. AND PLANNING TASKS

12) CL 54,
Clear 3000

32) CL 54,
Clear 3000

33) Resp.,

1290

P12

ALTITUDE TASKS

SPEED TASKS

COMMUNIC. TASKS

13) 1120
14) 1140
15) 1160
16) 1180
17) 1200
2) Seat Belt

NAVIG. AND PLANNING

36) CL 54, clear 2000 ft

COMMUNIC.

35) Cklist #12

12) Seat Belt Sign On

'Soft' precedence constraint. B3 may occur at any time after event A2.
Figure B.2: Task P: precedence Map, ILS 4R Arrival
Leave 6000 ft

Begin Descent to 3000 ft

Clear to 3000 ft
PILOT EXPERIENCE QUESTIONNAIRE

Subject Name ____________________________

Part 121 or 135 Experience? Yes ___ No ___

Total No. of Hours ______

Total No. of Hours under Simulated or Actual IFR ______

RECENT EXPERIENCE

Last 6 Months Last 30 Days

Hours Total

Hours IFR
(Simulated or Actual)

No. of Instrument Approaches

Estimated No. of Non-Precision Approaches

Do you, on a regular basis, fly into the following terminals (check as many as necessary):

New York _____ Atlanta _____
Washington, D.C. _____ Los Angeles _____
Chicago _____ San Francisco _____

Type of aircraft you fly most often (list one, or at most, two)