AD-A114 364 PURDUE UNIV LAFAYETTE IN SCHOOL OF INDUSTRIAL ENGINEERING F/G 5/5
DEVELOPMENT OF A METHODOLOGY FOR ASSESSING AIRCREW WORKLOADS.(U)
NOV 81 J R BUCK, J W BARANY, M L LEHTO F33615-78-D-0617

UNCLASSIFIED AFAMRL-TR-81-50 NL

1 OF 1
AD A
114364

END
DATE
FILMED
6-82
DTIC

(12)

AFAMRL-TR-81-50

DEVELOPMENT OF A METHODOLOGY FOR ASSESSING AIRCREW WORKLOADS

JAMES R. BUCK
JAMES W. BARANY
MARK L. LEHTO
DAVID M. INGS
DAVID R. PAYNE
ROY D. NIXON
WILLIAM H. GROSSE

SCHOOL OF INDUSTRIAL ENGINEERING
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907

NOVEMBER 1981

AIR FORCE AEROSPACE MEDICAL RESEARCH LABORATORY
AEROSPACE MEDICAL DIVISION
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433

NOTICES

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Please do not request copies of this report from Air Force Aerospace Medical Research Laboratory. Additional copies may be purchased from:

National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161

Federal Government agencies and their contractors registered with Defense Documentation Center should direct requests for copies of this report to:

Defense Documentation Center
Cameron Station
Alexandria, Virginia 22314

TECHNICAL REVIEW AND APPROVAL

AFAMRL-TR-81-50

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

CHARLES BATES, JR.
Chief
Human Engineering Division
Air Force Aerospace Medical Research Laboratory

AIR FORCE/56780/25 February 1982 — 100

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFAMRL-TR-81-50	2. GOVT ACCESSION NO.: AD-A114 364
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): DEVELOPMENT OF A METHODOLOGY FOR ASSESSING AIRCREW WORKLOADS	5. TYPE OF REPORT & PERIOD COVERED: Final Report, October 1978 - December 1980
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): James R. Buck, David M. Ings, William H. Grosse, James W. Barany, David R. Payne, Mark L. Lehto, Roy D. Nixon	8. CONTRACT OR GRANT NUMBER(s): F33615-78-D-0617 (0005)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: School of Industrial Engineering, Purdue University, West Lafayette, Indiana 47907	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 62202F, 7184-14-12
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Aerospace Medical Research Laboratory, Aerospace Medical Division, AFSC, Wright-Patterson AFB OH 45433	12. REPORT DATE: November 1981
	13. NUMBER OF PAGES: 95
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Communication; divided attention; human engineering; information; motor tasks; network models; perceptual-mediational processes; queueing models; regression analysis; simulation; standard time systems; switching synthetic time systems; task activities; task interference; time study; tracking; workload; work sampling

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The adaptability of industrial methods for setting job and time standards to workload assessment requirements was investigated. Methods considered included direct and indirect time study, synthetic time systems, standard data systems, information content analysis, work sampling and job evaluation. Conventional methods were found to be deficient in accounting for task time variability, divided-attention effects, and cognitive demands which are regarded as

(Continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

20. ABSTRACT (Continued)

critical to effective air crew workload assessment. A combination of synthetic time and standard data system methodologies was proposed as an effective approach to the problem.

Three experiments were conducted to evaluate the feasibility of developing a Synthetic Data System (SDS) consistent with workload assessment needs. Switching, communication and perceptual-mediational tasks were paired with tracking to create divided-attention demands characteristic of pilot workloads. Regression analyses showed that significant amounts of variance in task time requirements and error rates could be accounted for in terms of task and man-machine interface design variables. Development of an SDS on the basis of the performance of aircrew members in aircraft simulators is recommended.

SUMMARY

This report describes a feasibility study of aircrew workload assessment methodologies using engineering approaches. The scope of this particular study was restricted to the single pilot case.

Three initial phases of this research consisted of (1) a literature search, (2) identification of representative pilot tasks and necessary conditions of workloads, and (3) an evaluation of alternative methodologies for workload assessment in light of these pilot tasks and conditions. Results of these three phases clearly indicated that what was needed was some combination of synthetic time and standard data systems which is called herein a Synthetic-Data System (SDS). Because of the tasks and USAF needs, existing industrial systems of this type would not serve. Phase IV of this study focused on identifying methodogical deficiencies of existing SDS and plans were made for testing the feasibility of developing such a system for the USAF. Three empirical experiments were then planned to address switch activation tasks, communication tasks, and mediation tasks which are reported below respectively as Phases V, VII, and VIII. Phase VI was conducted in parallel as a demonstration of SDS with computer simulation for workload evaluation. Later in the project a final Phase IX was added to determine the learning effects in the tracking task.

In the laboratory study of Phase V on activating switches, it was found that: (1) a number of factors of this task could be identified which accounted for 50-60% of the time for people to activate various types of switches in numerous locations, (2) component activities in activating switches demonstrated near statistical independence with each other, and (3) tracking errors varied with many of the same task variables which affected performance time in switch activations. These results indicate that a SDS is feasible for USAF applications for perceptual motor tasks with attention divided between these tasks and aircraft control.

Phase VI describes the effects of an SDS methodology with task priorities embedded. The results here also demonstrated some of the statistically related variables and helped us identify further assumptions implied in an SDS.

Experimental tests conducted in Phases VII and VIII are directed on vocal communication tasks and mediational tasks with vocal communications. The first test showed that few variables other than the length of the communications and individual differences in vocal communications have much effect on the performance time. This result clearly showed that individual differences need to be described in the SDS model but that otherwise the communication task is easily handled by such a system. The Phase VIII experiment showed large performance time differences due to various mediational task types and degrees of difficulty but relatively small individual differences. Over 50% of the performance time variability was described by the task and degree of difficulty compared to about 2% by individual differences. In general, these results support the feasibility of an SDS for the pilot tasks examined. Some of the tracking task learning effects, showing up in Phases VII and VIII, proved to show no consistent

trend in Phase IX but rather indicated a normal distribution.

In total, these results clearly indicated that a SDS for USAF needs was feasible and it would enhance other needs in computer simulation as well as other uses (e.g., Human Engineering Computer-Aided Design).

PREFACE

This research was performed for the Human Engineering Division, Air Force Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio 45433, in support of Project 7184, "Man-Machine Integration Technology". Mr. Billy M. Crawford was the Task Scientist and contract monitor. The research was performed by the School of Industrial Engineering, Purdue University, West Lafayette, Indiana 47907, under Contract F33615-78-D-0617, Subcontract No. SCEEE-ARB/79-5, between October 1978 and December 1980. Professor James R. Buck was the Principal Investigator.

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The investigation described by this report was focused on the development of a methodology for assessing the workload of members of an aircraft crew based on job descriptions, operational requirements, and physical factors of the jobs. In this initial study, the focus was upon pilots with no crew interaction. Various methodological approaches from industrial and other engineering disciplines were examined, toward the goal of identifying and testing the feasibility of a workload methodology which would encompass workload measurements of physical and cognitive pilot activities. This report describes the finding of this investigation.

Our investigation officially started at the end of 1978 with the initial phase of surveying and reviewing the literature in an effort to identify potential methodologies which could be employed to measure aircrew workloads. The second phase of effort consisted of investigating representative pilot tasks during in-flight operations as well as investigating measurement criteria and evaluation techniques. Visits were made to aircraft simulator facilities as part of the second phase efforts. This second phase was started in January 1979 and it continued in parallel with the first and later phase efforts. Phase three of this study, starting in February, 1979, consisted of evaluating the identified methodologies in reference to representative pilot tasks and various criteria pertaining to the applicability of the various methodologies for assessing workload. During this third phase a few of the most promising methodologies were selected. The fourth phase of effort consisted of identifying methodological deficiencies of these promising methodological developments. Also, this fourth phase, which started in March 1979, required the development of plans for testing the feasibility of the most promising workload measurement techniques and preparing for laboratory tests. These laboratory tests were commenced in the fifth phase which focused on motor activities of pilots. Since feasibility of a given system of workload includes application procedures, a parallel sixth phase effort investigated various procedures and provided a demonstration of a potential method of application which could be employed. While the experimentation and analysis of these motor activities was ongoing, plans for phases seven and eight were respectively developed for investigation of communications activities and perceptual-mediational activities of pilots. Because of results obtained in the analysis of phase five, a ninth phase was interposed before phase seven in order to test out the effects of the aircraft control task which pilots must perform simultaneously with other motor communications and perceptual-mediations tasks. Experimental work was completed during the late summer of 1980, and final analyses were made during the fall of that year. This Final Project Report describes the research conducted in each work phase of this project.

PHASE I
LITERATURE REVIEW

Most classical techniques of industrial engineering involve the measurement and/or the prediction of the time required to perform sequences of defined physical tasks. The basis of time management for workload assessment is that there is limited time available and a collection of tasks to be performed within the available time. If the required time is less than the available time, then this difference is the time slack. Some slack is required for various reasons (e.g. personal needs, fatigue recovery, or delays), the amount of which depends upon the job. Almost all of these methodologies were developed for use in industry where machine or assembly operations are performed. The most prominent of these methodologies are discussed individually below along with commentaries on their potential applicability to U.S. Air Force needs. Some of these commentaries address the phase of applicability from the Research and Development (R&D) phase to operational phases. Other commentaries are directed to various types of aircraft systems.

Direct time study is the oldest of the traditional industrial engineering techniques (Barnes, 1937). Essentially this technique consists of observing the activities being performed and timing the component activities as they occur. These recorded time values provide the time intervals actually expended by the observed operator. End points of these time intervals must be observable rather than inferred, causing this technique to be limited to measuring time requirements of task elements which can be reliably timed (about three seconds). Direct time study necessitates a time-study operator and the physical presence of the subjects being time-studied along with the equipment operated or some reasonable facsimile of it. Consequently, direct time study is least appropriate during the early development phases for single-pilot aircraft systems. It should be noted that direct time study is currently being used in the Air Force for assessing the workloads of crew members in multiple crew systems, during the final development phase or retrofitting phases, based on the total active time required during certain mission phases (Geiselhart, 1978). In the opinion of the author, this traditional technique appears best suited to its current use because of the limited space available in many aircraft systems, observation limitations with aircraft or their simulators, and the difficulties in assessing cognitive and communicational activities with this method.

Indirect time study is similar to time study except that some form of observational record is made which is studied at a later time. Typically, cinegraphic or television recordings are made. Since cinegraphic records can be taken at regular speeds (24 frames/second) or faster, time intervals which are too short to be captured by direct time study can be found with indirect time study because the recording can be slowed down, stopped, backed up, or rerun when being observed for time measurements. If the behaviors being observed are very rapid (e.g., eye movements), then indirect time study is most appropriate (Fitts, Jones and Milton, 1950). However, photographic and television recording techniques require moderate to high light levels which may not be available in the environment of the tasks being studied, particularly so with piloting tasks. Light levels, unavailable viewing locations for cameras, and other constraints can preclude the use of either the cinegraphic or television recordings, but other forms of instrumentation can be employed to capture and record the time of event occurrences on magnetic tapes, or by other means, for later analysis. Photoelectric cells, strain gages, pressure cells, piezo-electric crystals, ultrasonics, and other sensor devices which detect changes in voltage, magnetic intensity, capacitance, etc., can be used to instrument a task for the purpose of gathering the event time data (Hancock and Foulke, 1961). However, instrumental data gathering systems sometimes require a considerable amount of internal logic in their design to assure that events recorded are those desired. Nevertheless, instrumented forms of indirect time study provide a promising means for airborne data collection, with the smaller aircraft systems, or for simulations of such systems. Some illustrative forms of instrumentation are described below in phase five.

Synthetic or predetermined time systems are techniques which specify the time required to perform elemental tasks and denote how to combine these elemental time values into a composite time value for an entire task unit. There are a variety of such systems where the Motion-Time-Measurement (MTM) system, the Work Factor system, and the Basic Motion Time system are the most common (Niebel, 1967). Differences between these systems include: 1) Definitions of the elemental motions (e.g., reach, grasp, move, or apply pressure), 2) Factors which affect the time values of the elements, 3) Definitions of the standard operator's degree of effort and/or learning, and 4) The manner in which the elemental time values are combined (i.e., simultaneous or sequential activity times). In fact, there are a variety of versions for each type of system measurement and for the effort required by the time-study personnel. However, all synthetic time systems have tabled time values for an identified elemental motion of a given class with correction for a change due to an affecting factor. For example, the reach motion in the MTM system has five classes of this motion depending upon the nature of the location of the reach (e.g., to an object in a fixed location or to an object which varies slightly in location where minor visual guidance is needed) and the time values change with respect to the distance of the reach or whether the hand was previously in motion. The purpose of denoting classes of the reach or other elemental movements is to denote the effects of sensory and perceptual activities which are performed concurrently with the motor activities. In practice, a job is described by a sequence of these elemental activities for each hand (and sometimes the feet or main body stem) and the elemental times are found from the tabled data depending on the situation and values of the affecting variables for each body member. At this point, the time-study operator must check those elemental motions which appear to be simultaneous by using a simultaneous motion chart that specifies those motions which are incompatible (i.e., elements which are very difficult to perform simultaneously or can only be performed together with a great deal of practice). When elemental tasks are compatible, then the greater time interval of the two elemental tasks controls in these simultaneous tasks. Otherwise, the sum of the two elemental time values is usually taken as the controlling time for serial tasks. Sometimes, when compatibility is practice limited, the time is established as the longer of the two time intervals plus a fraction of the shorter time value where the fraction is based on the learning rate. However, this practice is rare and artistically performed so almost all synthetic time studies view tasks either as simultaneous or sequential in the time requirements. The total task time is the sum of all the controlling time values.

Predetermined time systems provide an advantage over direct or indirect time study because time requirements can be established without the need for the operator or the physical apparatus being present. This feature is particularly appealing for Air Force needs during the early development phases of an aircraft system as a preliminary technique in evaluating the workload effect of proposed changes in task allocations. However, the past industrial developments of synthetic time systems has been limited to low cognitive tasks and the time estimation to only expected time values. Some preliminary experimental work has investigated the extension of synthetic time systems to include decision tasks as part of a manual activity and the resulting time distributions, where the experimental results look encouraging (see Sadosky, 1968; and Thomas, Hancock, and Chaffin, 1974). These preliminary experimental results and the advantages of synthetic time systems clearly suggest that this approach to pilot workload assessment needs further investigation.

Standard data systems are one of the more contemporary approaches used in industry to evaluate time requirements of jobs. Such systems are really macroforms of synthetic time systems which are prepared for a family of jobs where differences in the jobs within the family are denoted by the values of a set of variables. In practice, data on the past jobs within the job family are investigated through regression techniques to establish a polynomial that best describes the mean time requirements (Steffy, 1970). Typically, the criterion in devising the values of the polynomial coefficients is the minimizing of the least squared error but other criteria (e.g., minimizing absolute time errors) have also been used. The resulting regression polynomial is then used as a predictor of the time requirements under the assumption of continuity between the new and the old tasks within this task family. Standard data systems provide a clear advantage in avoiding the need to time-study jobs which are similar to those for which extensive data already exist. However, the precision of estimating the time requirements with standard data systems is often not as great as with direct or indirect time-study and the establishment of the variables to employ in the standard data system poses a leading difficulty in these systems. Although factor analysis could be employed to aid in determining the set of variables which should be included in a standard data system, most applications in industry are formed through trial-and-error tests with regression analysis. The regression equation selected is that which provides the least unexplained variance and is easiest to use with the available data. Also, industrial practice typically limits the regression polynomial to, at most, four variables and the prediction of expected time. These restrictions need not be carried into Air Force applications. Further, the use of standard data systems for families of pilot tasks appears to have promise, particularly for the more cognitive tasks involving perceptual, mediational, and communication processes coupled with motor activities.

Information Content Analysis (INCAN) is a technique which originated in the late 1950s but was never fully developed. The basis of this technique as the name implies, is Shannon and Weaver's (1948) information theory. Numerous psychological studies (e.g., Miller, 1956; Fitts, Peterson and Wolpe, 1965; Garner, 1974) provided theoretical support and the work of Ross (1960), Rosenstein (1955), Raouf (1972), and Raouf and Mehra (1974) were developmental beginnings into industrial applications. These beginnings showed considerable promise in the measurement of some cognitive activities required in industry. However, the measurement of information within many industrial jobs proved to be difficult and so INCAN never came into extensive industrial use. Other problems with INCAN included: 1) Changing probabilities with certain stimuli, 2) Interrelated probabilities (nonindependence) and their apportioning, 3) The theoretical indistinguishability of different types of errors, 4) Learning changes, 5) Identification of chunking, and 6) The lack of a means for handling changes due to information feedback (Ross, 1960). Little progress was made on these drawbacks and so INCAN was never generally accepted by industry. Nevertheless, there are a variety of appealing features of INCAN as a metric to those situations where the problems noted above are negligible as shown by Thomas, Hancock, and Chaffin (1974).

Work sampling is a technique of taking observations of an operating system at random times in order to determine the percentage of time spent on component activities, within a level of statistical precision (Tippett, 1935). Thus, work sampling provides an alternative to direct or indirect time study where Table 1 shows the comparative advantages and disadvantages. It is of note that work sampling possesses some advantages for some Air Force needs in workload measurement but not for all cases. Moreover, there are a variety of sampling procedures that can be employed (Konz, 1979; Buck and Tanchoco, 1974) including: 1) Simple random sampling, 2) Stratified sampling, 3) Cluster sampling, 4) Sequential sampling, 5) Bayesian sampling, 6) Fixed interval sampling, and 7) Combinations of some of the above. Regardless of the form of sampling, work sampling requires that clear observations can be made at the times specified and that the occurring activity or activities can be unambiguously classified. These requirements of sampling pose greater problems as the number of different types of activities increases and as the type of activity becomes less overt (e.g., problem solving compared to a motor activity). It is of note that the classical study by Christensen (1950) on aircrew activities in early polar flight was performed using fixed-time-interval sampling which has been subsequently referred to as "activity analysis" whereas Tippett, who developed the random technique, called it ratio-delay analysis.

Job evaluation is a traditional area of industrial engineering which involves systematic procedures for ranking jobs for purposes of setting remuneration (Brennan, 1963). Regardless of the type of job evaluation procedure employed, these evaluation techniques start with an accurate job description. Job analysis follows where job requirements are determined with respect to education, experience, training, skill, responsibility, and working conditions. Nonquantitative methods of job evaluation consist of ranking comparative jobs or in classifying them into ranked categories. In either case, only job ranking occurs. Quantitative job evaluation methods involve point systems which are added up or which use weighted factors with points within each factor. In either case, these quantitative job evaluation systems provide interval measurements of the jobs' difficulties. The unfortunate thing about these job evaluation procedures is that the measurements of job difficulty are subjective and arbitrary even though they are systematic and logical. Efforts by Rosenstein (1955) to incorporate information content measurements were never developed. Accordingly, the techniques of job evaluation do not appear to be highly applicable to pilot workloading except perhaps in a guidance sense. Perhaps, as Sheridan (1978) points out, subjective techniques may be a good basis for evaluating a pilot's workload. However, we feel that the other approaches have greater potential promise and so the job analysis techniques were to be given a lower priority in our investigation.

TABLE 1. SOME ADVANTAGES AND DISADVANTAGES OF WORK SAMPLING IN COMPARISON TO TIME STUDY

Advantages

1. Many operations or activities which are impractical or costly to measure by time study can readily be measured by work sampling.

2. A simultaneous work sampling study of several operators or machines may be made by a single observer. Ordinarily an analyst is needed for each operator or machine when continuous time studies are made.

3. It usually requires fewer man-hours and costs less to make a work sampling study than it does to make a continuous time study. The cost may be as little as 5 to 50% of the cost of a continuous time study.

4. Observations may be taken over a period of days or weeks, thus decreasing the chance of day-to-day or week-to-week variations affecting the results.

5. There is less chance of obtaining misleading results, as the operators are not under close observation for long periods of time. When a worker is observed continuously for an entire day, it is unlikely that he will follow his usual routine exactly.

6. It is not necessary to use trained time study analysts as observers for work sampling studies unless performance sampling is required. However, if a time standard or a performance index is to be established, then an experienced time study analyst must be used.

7. A work sampling study may be interrupted at any time without affecting the results.

8. Work sampling measurements may be made with a preassigned degree of reliability. Thus, the results are more meaningful to those not conversant with the methods used in collecting the information.

9. With work sampling the analyst makes an instantaneous observation of the operator at random intervals during the working day, thus making prolonged time studies unnecessary.

10. Work sampling studies are less fatiguing and less tedious to make on the part of the observer.

11. Work sampling studies are preferred to continuous time studies by the operators being studied. Some people do not like to be observed continuously for long periods of time.

TABLE 1. (Continued)

12. It usually requires less time to calculate the results of a work sampling study. It fact, IBM mark-sensing cards may be used, and the results obtained from standard IBM equipment.

13. No stop watch or other timing device is needed for work sampling studies.

Disadvantages

1. Ordinarily work sampling is not economical for studying a single operator or machine, or for studying operators or machines located over wide areas. The observer spends too great a proportion of his time walking to and from the work place or walking from one work place to another. Also, time study, elemental data, or motion-time data are preferred for establishing time standards for short-cycle repetitive operations.

2. Time study permits a finer breakdown of activities and delays than is possible with work sampling. Work sampling cannot provide as much detailed information as one can get from time study.

3. The operator may change his work pattern upon sight of the observer. If this occurs, the results of such a work sampling study may be of little value.

4. A work sampling study made of a group obviously presents average results, and there is no information as to the magnitude of the individual differences.

5. Management and workers may not understand statistical work sampling as readily as they do time study.

6. In certain kinds of work sampling studies, no record is made of the method used by the operator. Therefore, an entirely new study must be made when a method change occurs in any element.

7. There is a tendency on the part of some observers to minimize the importance of following the fundamental principles of work sampling, such as the proper sample size for a given degree of accuracy, randomness in making the observations, instantaneous observation at the preassigned location, and careful definition of the elements or subdivisions of work or delay before the study is started.

from Barnes (1937).

Other engineering approaches have emerged more recently. Some of these approaches include: 1) Queueing models (Schmidt, 1978; Walden and Rouse, 1978), 2) Scheduling models (Sheridan, 1978), 3) Simulation concepts (Siegel and Wolf, 1969; Buck, Deisenroth and Alford, 1978; Buck and Maltas, 1979), and 4) Control models (Kleinman, Barron and Levison, 1971). The queueing model concept views the pilot as a service system where tasks arrive in some random manner and the service system has a probabilistic service rate. When the situation at hand is modelled in this fashion and parameters are estimated for the model, then statistical moments can be found for the system's operations either in closed-form or through a computer simulation. Scheduling models are simply the fitting of time blocks into an available time space, subject to time constraints of tasks' initiation and completion, and precedence requirements. In the variety of scheduling models, some allow a task to be interrupted when another higher-priority task enters, where interruptions create a time loss with such task changeovers, in a similar way that multiple tasking is performed on a computer. However, there are a variety of assumptions which must be made on how the operator manages the schedule of tasks. Simulation models simply consider the task arrivals and completion times as random variables in a computer simulation program and impose prescribed behavioral rules of logic. Here again assumptions need to be made on human behaviors. Repeated simulation runs with various random number streams provide the statistics of probable overload at various times during a mission phase. The alternative scheme of man-in-the-loop simulation is similar to the current method of using aircraft simulators where a computer controls the environment of the aircraft subject to the operators' control. While the man-in-the-loop simulation has greater face validity than the purely computer simulations, they are also much more costly to operate. Finally, the manual control models are purely mathematical descriptions of control responses with feedback, which specify the aircraft's and the operators' characteristics over various forcing functions (i.e., changes in the needed control directions and/or disturbances). A very wide variety of these manual control models exist but they have limited usefulness in determining pilot workloads. In contrast, the optimal control models are more extensive models where elements of pilot workloads can be investigated provided that parameters of the optimal control model can be reliably estimated. These engineering models are discussed here because they constitute alternative methods of approaching the workload problem once the behavioral data can be obtained. A wide variety of other approaches to workload assessment are summarized in Williges and Wierwille (1979).

PHASE II
REPRESENTATIVE PILOT ACTIVITIES

Representative pilot tasks and task families were identified through four principal means: 1) Literature reviews, 2) Examination of the F-15 pilot task analysis and other time-line analyses, 3) Visits to simulator facilities, and 4) Direct observations of in-flight activities. Some papers in the literature (e.g. Christensen and Mills, 1967 or Geiselhart, 1978) show the spectrum of various pilot tasks and those of other aircrew members in a macro-sense. More micro elements of pilot tasks may be inferred from the time-line analysis and by direct observation of simulated and actual flight situations. Observations were made at simulators at Orlando, Florida, Wright-Patterson Air Force Base, and at Purdue University.

A review of the F-15 pilot task analysis reveals that there are a wide variety of pilot tasks, but the most frequent of these is communication with the aircraft equipment through switches, levers, and a joystick. These overt motor processes may vary from simple to complex and discrete to continuous but they occupy a very large percentage of the pilot's activities. It should also be stated that these switching/control tasks are often performed in conjunction with associated perceptual and mediational activities which are not overt nor clearly discernable from the F-15 pilot task analysis. Although there are cases, such as with checklist activities, that the associated mediational activities are minimal. During flight, the switching and craft control tasks are performed concurrently. Perceptual activities also occur with: locating the switch, switch activation, monitoring of the switch setting, observing associated display readings, and numerous other associated perceptual tasks including the perceptual requirements of complex tracking needed for joystick directional control. The lower end of these perceptual-motor activities is highlighted by the task analysis while the upper end must be inferred for the most part. However, the clearly denoted perceptual-motor activities in the F-15 task analysis vary with the type of switch involved. These inferences were confirmed in direct observations which were made during local flights near Purdue University and by inquiries of Air Force pilots. Families of tasks frequently include associated perceptual and mediational tasks where there are affecting variables that change performance time for various members of these task families. Switching tasks are such a family.

The literature on aircraft activities is large. A good summary (as of 1967) is reported by Christensen and Mills; however, the scope of activities is broader than those considered in this initial phase of our study. Their paper contains a useful taxonomy of activities, behaviors, and associated psychomotor processes which they adapted from Berliner, Angell and Shearer (1964). This taxonomy is shown in Table 2 for purposes of reference. It should be noted that precise definitions of specific behaviors are not given and some of the specific difficulties in using this taxonomy are noted by Christensen and Mills (1967). Nevertheless, the taxonomy indicates some of the scope and variation to be expected with piloting activities. One of the evident processes shown here which is not evident from the F-15 task analysis is communication (i.e. other than that which occurs between a pilot and his visual displays). It is clear that communications activities are additionally superimposed on the perceptual-motor activities and that a variety of mediational processes are concurrent, particularly those behaviors typically associated with information processing. In "pilot only" systems, the communications tasks are principally oral messages sent to and received from other aircraft and ground stations. In these forms of communications, the message sender does not know the work status of the message receiver when the sender elects to transmit the message, unlike the communications between pilot and co-pilot, and the message transmissions tend to be more time-random and workload-imposing.

Other parts of the literature denote activities in more operational terms such as navigation and fire-control. In fact, the navigational tasks can be very workload-intensive but this workload is shown to vary considerably depending on the type of craft, the form of mission, and the phase of the flight (e.g., Sanders, Simmons, Hofmann and DeBonis, 1977; Christensen and Mills, 1967). Also, many of the navigation tasks are

TABLE 2. CLASSIFICATION OF BEHAVIORS

Processes	Activities	Specific Behaviors
1 Perceptual Processes	1 Searching for and Receiving Information	Detects Inspects Observes Reads Receives Scans Surveys
	2 Identifying Objects, Actions, Events	Discriminates Identifies Locates
2 Mediational Processes	1 Information Processing	Categorizes Calculates Codes Computes Interpolates Itemizes Tabulates Translates
	2 Problem Solving and Decision Making	Analyzes Calculates Chooses Compares Computes Estimates Plans
3 Communication Processes		Advises Answers Communicates Directs Indicates Informs Instructs Requests Transmits
	1 Simple/Discrete	Activates Closes Connects Disconnects Joins Moves

TABLE 2. (Continued).

4 Motor Processes		Presses Sets Adjusts Aligns
	2 Complex/Continuous	Regulates Synchronizes Tracks

from Christensen and Mills, 1967, an adaptation of the classification by Berliner, et al. 1964.

essentially described by perceptual-cognitive tasks. These aspects of the literature denote highly varied tasking for pilots who are alone to perform communications and navigation concurrently with switching and craft control tasks.

The workload definition and criterion problems are complex. One of the more frequently cited definitions of pilot workload is, "the sum of the task demands which can be clearly specified, plus the operator's response (and effort) to satisfy these demands" (Gerathewohl, 1976). A necessary component of measuring the workload, according to this definition, is the time requirements of the tasks which the pilot performs. Another is the time variability because there is a risk in not completing required tasks over a time frame and this risk is described by the time variability. However, the sum of all the individual-task expected time values is not necessarily equal to the expected time required to perform that collection of tasks because some tasks can be performed simultaneously whereas others cannot. Also, the sum of the individual task time variances may not equal the variance of the task collection because it is not known whether these tasks will have independent time variations. Insofar as the pilot's effort is concerned, it is extremely difficult to obtain effort measurements except through subjective elicitation or physiological measurements and then the validity and reliability may be questionable. Accordingly, time requirement statistics still appear to be the primary criterion of workload measurements even though there is question as to the sufficiency of this criterion.

A number of studies have shown that within-person task time variabilities are initially gamma distributed but with learning these time values tend toward a Gaussian (normal) distribution (Sadosky 1968, Thomas et al 1974, Mills et al 1975). The elements of activity in these studies were short in time but included motor, perceptual, and cognitive tools over the activity sequences. As a consequence, time variances can be used to measure the risk due to the symmetry of the Gaussian (normal) distribution but care will be required to time only well learned tasks. Gamma time distribution would have necessitated using semi-variances of time or some similar correction for the distributional assymetry. It has also been shown by extensive studies by Presgrave (1945) that between-person performance times are Gaussian distributed where the performance time of the slowest persons in the population are approximately 2.25 times those of the fastest in the population. Accordingly, risk due to population differences can also be obtained through time variance measurements. However, the Presgrave data encompass a far wider population than that expected in the U.S. Air Force in terms of age and physical-mental abilities, and so the population effect on performance times within the Air Force would be expected to be considerably less than that found by Presgrave.

Contemporary psychological theories have been addressed to the problem of performance changes with effort changes, particularly when multiple-tasking occurs. Although there are variations between the theories of Broadbent (1971), Welford (1968), Kahneman (1973), Posner and Klein (1973), and others, these theories all describe information processing limits to human activities. When two activities are concurrent, more effort applied to one task can interfere with performance on the other provided that both activities draw from the same pool of effort resources and that the decrement in effort on one task is sufficient to draw down performance on the other task. Differences in these theories deal with different resource pools, fixed capacity limitations, and a number of other subtle issues. In the Norman-Bobrow (1975) theory, tasks are: 1) resource limited where more resources result in better performance; 2) data limited where increased performance is independent of the processing resources; or 3) some mixture of the two cases. Transitional processes occur where performance improves with additional resources up to some limit where no further performance increase occurs with added resources. If the two concurrent tasks are resource limited, then task interference can result because these two tasks are drawing resources from a common resource pool (Wickens 1979). However, the interference between tasks is not necessarily symmetrical but rather priority dependent. High-priority tasks tend to capture the available resources to approach the data-limited case whereas the lower priority task tends to be resource-bounded. Wickens (1979) further provides evidence of separable sensory, information processing, and response resource pools whereby two concurrent tasks which use the same sensory resource (e.g., vision) will exhibit greater interference than if the sensory mode of resources were different (e.g. vision and audition). This theory shows that the operator's criterion of task importance as well as the nature of the resources demanded by a task will affect the performance on the two time-shared tasks. Accordingly, there are a variety of potential problems in measuring workload from a theoretical effort point of view, even if the measurements of effort were easy to obtain. Also, this theory provides descriptive richness but grave predictive limitations at this stage in its development.

PHASE III
EVALUATING ALTERNATIVE METHODOLOGIES

A recent study on pilot workload was reported by Williges and Wierwille [1979] and provided an extensive survey of this field including a few of the approaches shown above but many other techniques including: 1) Subjective opinions (e.g., the Cooper-Harper scale), 2) Spare mental capacity [Rolfe 1973], 3) Simulation, 4) Information theoretic approaches, 5) Nonadaptive and adaptive arithmetic/logic, 6) Nonadaptive and adaptive tracking, 7) Time estimation, 8) Occlusion, 9) Single and multiple primary task measures, and (10) Mathematical modeling. Their conclusion in reviewing these studies is that "no one single technique can be recommended as the definitive behavioral measure of operator workload". They go on to say, "Probably the strongest research support exists for using subjective opinions and task analytic methods involving task component/time summation". A variety of the task analytic methods of time summation are shown in the Phase I discussion with commentaries to their applicability to pilot tasks. Of all of these methods, some combination of the synthetic (predetermined) time system and a standard data system seems most appropriate. Since synthetic time systems or standard data systems do not require direct observation, an applicable system for the Air Force can be used from the beginning of the developmental phases of a new system. Moreover, some elements of Information Content Analysis (INCAN) appear to have merit for inclusion in spite of several identified limitations. It is clear from this investigation that direct or indirect time study, work sampling and classical job evaluation methods are inappropriate for many aircraft systems. Accordingly, further examination was then made on the set of synthetic standard data and INCAN systems for use in workload measurement.

PHASE IV
IDENTIFYING METHODOLOGICAL DEFICIENCIES AND PLANNING FEASIBILITY TESTS

The applicability of the synthetic time system approach hinges upon some significantly different features in the Air Force from those found in industry. One major difference is that most industry situations rarely entail continuous control tasks with combined discrete tasks. Another significant difference between pilot tasks and many industrial jobs is that pilot tasks tend to be considerably more cognitive in nature. A third important difference between industrial and Air Force applications of "some form" of synthetic or standard data system is that industrial users are focused heavily on the average time situation with some but minor concern about variability whereas the Air Force has a very heavy commitment to aircraft safety and therefore is highly concerned about the task time variability. While not all of the assumptions of synthetic or standard data systems are clearly met in industrial practices, such as consistent time values in sequential task variations [Heising 1954], these inconsistencies are inconsequential in most industrial circumstances. This may not be true for U.S. Air Force needs. Accordingly, a beginning effort in testing the applicability of synthetic time or standard data systems is to see if piloting tasks meet principal assumptions of these systems when there is some cognitive variation, both with respect to mean time values and time variances, for discrete tasks superimposed on a continuous control task. These apparent deficiencies require tests in order to ascertain whether the deficiencies exist and are of a significant consequence. Also an effective synthetic/standard data system for Air Force needs must have identifiable and measurable factors which account for reasonably constant time variations over a wide variety of people. Existing synthetic systems and INCAN provide a partial aid to identifying some of the affecting factors, but some factors will clearly be specific to flying activities.

The taxonomy of pilot tasks by Christensen and Mills [1967], Table 2, provided a planning basis for breaking up the variety of tasks and making feasibility tests as an initial means of developing a Synthetic Data System (SDS), which is the name given to the hybrid form of synthetic time system and standard data system which evolved. While this taxonomy involves four major processes, we felt that tasks which were principally mediational would likely require strong perceptual processes, whereas communication and motor processes appeared to be separable. Accordingly, we determined that three principal tests could be made for the purpose of creating a tentative SDS based on: 1) motor tasks, 2) communication tasks, and 3) perceptual-mediational tasks. This sequencing order of testing was selected because background information was greatest for the motor tasks based on existing synthetic time systems and least for the perceptual-mediational tasks. In this way we could take advantage of our findings along the testing sequence. However, each of these task groups were performed by pilots when they were simultaneously controlling the craft. Therefore, it was determined that these initial feasibility tests would be performed with a two-dimensional tracking control task which could be reliably repeated for subjects in each of the tests.

Motor Task Testing

Since most of the motor processes of pilots consisted of activating switches of various types, the family of discrete switching tasks was selected for a test on the motor processes. This family of switching tasks does involve minor perceptual processes in identifying and locating the appropriate switch and activating swiches of different types in an appropriate manner. Based on existing industrial synthetic systems, the time required to perform a switch task would vary with the reach distance and the rotational change of the switch. However, these industrial predictions do not account for a simultaneous nature of time variances. Therefore, this first experimental effort on Phase V was planned to address the switching time statistics under externally paced and self-paced switching. The concomitant effect on the root-mean-square (RMS) on the tracking task was considered under two levels of tracking complexity. Some of the assumptions to be addressed in this first effort included sequence, pacing, and time variability independence. Also, analyses were planned to determine the relation of the discrete task variables to expected time requirements, time variations, task errors, and the effects on the steering control task.

Switching tasks consist of elementary component tasks performed by the hands and these components consisted of: 1) the hand reach to the switch, 2) switch adjustment to a specified setting, and 3) the return of the hand. With pilots, the switch task is further complicated by the specific hand that is required to perform the switching since the joystick must be constantly controlled and because the throttle control in pilot-only craft is traditionally on the left. Time intervals for each component of this switching task were collected simultaneously with the RMS error of the steering control task. Because of the asymmetry of pilot-only craft, right hand reaches were started when the left hand was removed from the throttle. A variety of toggle and rotary switches were used in alternative locations, some out of the normal view of the operator. Thus, different switch locations with both equal and different reach distances were employed for all types of switches used. This arrangement was devised so that tests could be made on different switches in the same location, and/or different locations, of the same distance relative to each component of the switching task. Also, the information content of some of the individual switches was varied in order to assess the effect on the switching time components due to actual or implied information content. Figure 1 shows the laboratory schematic for this experimental effort.

Time data were collected automatically through instrumentation and a NOVA 12-20 minicomputer. Since similar data collection can be done in a regular aircraft or an aircraft simulator, this experiment would also provide a means of showing how these discrete data may be collected.

The initial laboratory experiment on motor processes in activating switches contained a wide number of experimental factors which were indicated from previous research to be important variables in performance time. Some of the factors included: 1) Within and between person differences, 2) Reach distances, 3) Information content of the switch (both potential and actual), 4) Switch types, 5) The switch location in the visual field of the operator, and 6) The switch location side of the operator.

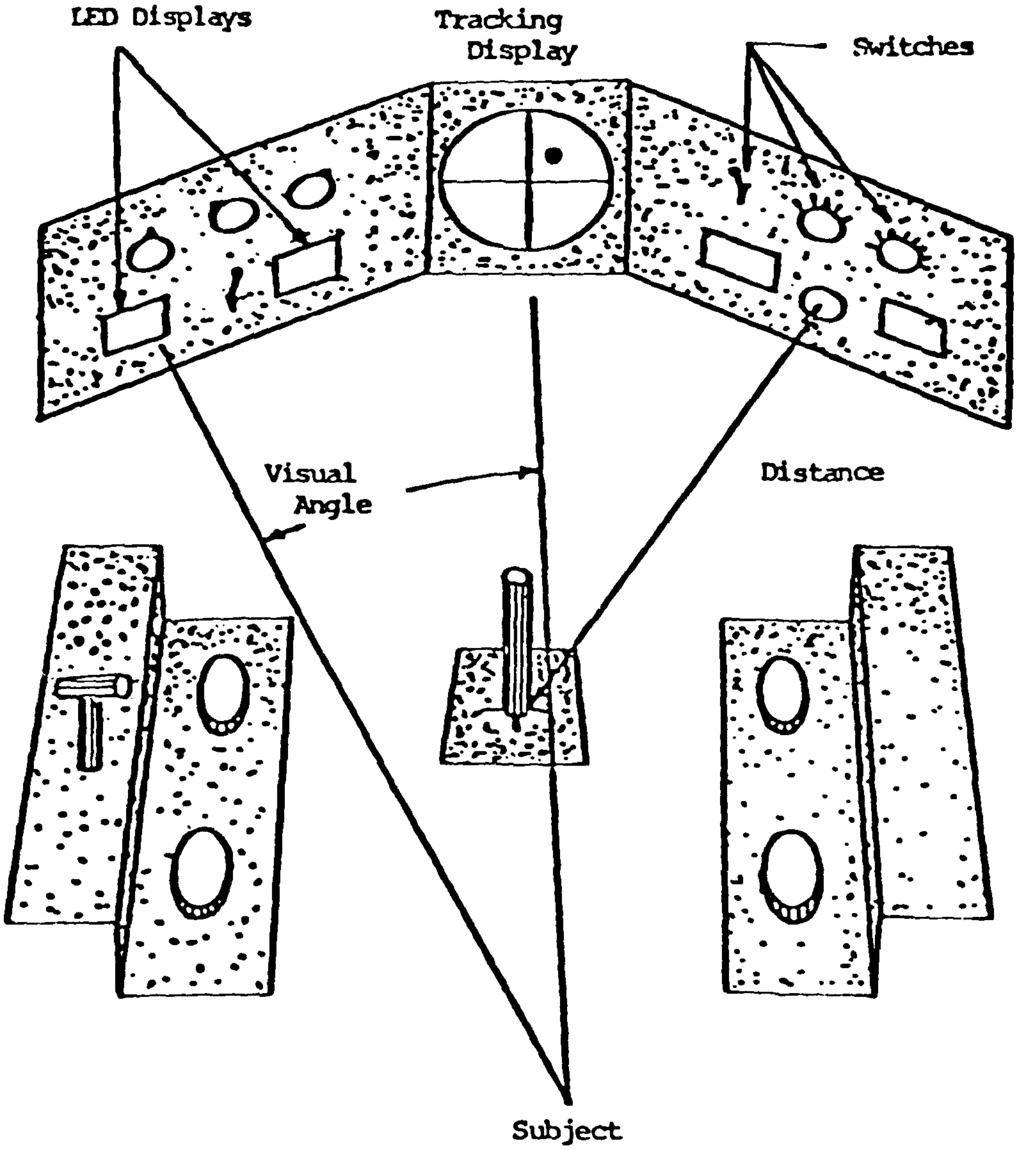

FIGURE 1. LAYOUT SCHEMATIC OF PHASES V, VII AND VIII

Plans were made to investigate these and other factors of the motor process by developing scenarios to vary these factors and then to examine the performance time effects in statistical analysis. Further details on this study are contained in the Phase V discussion.

Tracking Task

A standard tracking task was devised to operate simultaneously with the switching tasks as well as with evaluation of the communication and perceptual-mediational processes. This control task consisted of a two-dimensional compensatory tracking task with each dimension's forcing function consisting of a double sine wave as:

$$X = C_1 \text{ sine } W_x + C_2 \text{ sine } 2W_x$$

where:

C_1 and C_2 are arbitrary amplifying constants, and
W_x is the fundamental frequency of the x dimension

The Y-dimension was similarly configured so that the two fundamental frequencies were only slightly out of phase. Accordingly, the cycle time on the two dimensions was much longer than the maximum time of a subject's test session. For a lower level of complexity on the craft-control task, the fundamental frequency of both dimensions was set at approximately half of that at the higher-level of complexity. While this forcing function design is arbitrary in design, it is repeatable and of sufficient complexity to avoid anticipatory learning. Compensatory tracking was selected to further reduce subject learning of the forcing function. A joystick (two dimensional) control was provided to apply a voltage which was equal but opposite to the maximum of the forcing function. When the proper X and Y displacements were made on the joystick, the resulting voltage in each dimension equalized to give a zero error on the compensatory display. Accordingly, the resulting nonzero displacement on the display gave a positioning error which was recorded during the task; separately for each dimension.

Communication Processes

The second process to be examined was the oral communication made by the pilot to and from ground stations and other pilots. Geiselhart (1978) has shown that a significant portion of a pilot's time involves communications activities. Although some portions of the flight phases will encompass more communications than others (e.g. takeoffs, landings, and ground support missions), communication forms a sufficiently important role to be treated separately.

In the development of a communication workload measure several variables would appear to be of high interest. Attention must be focused on the interaction between communication and other time-shared tasks. Additionally, the intrinsic factors within the communication process must be isolated and evaluated.

Communication workload is easily estimated by calculating the proportion of time spent communicating. This approach has been applied when evaluating workload of air traffic controllers (Pasmoai et. al. 1976) but appears to be inadequate as a sole measure of workload. Two major deficiencies include the lack of consideration of information content of messages or the effect of time-sharing tasks or task components by the subject.

The mathematical theory of communication has been applied to psychological processes in attempts to quantitatively define behavior since the publication of Weiner's Cybernetics. This theory presents a mathematically tractable measure of the novelty and/or uncertainty of incoming stimuli. Message intelligibility has classically been shown to be dependent upon information content (Miller et al. 1951, Horves 1957) in a manner consistent with mathematical communication theory. Quite apparently, this variable is of interest when evaluating communication workload.

Other interesting variables that could influence the communication workload include time-sharing and the relative effect of response loadings. There is experimental and theoretical evidence that humans may have both common and individual capacity pools (Wickens 1979). Interactive effects may well be noted during time-sharing dependent upon whether responses draw from same or differing capacity sources. Similar effects may be noted dependent upon whether the subject is inputting or sending out information in the communication phase of experimentation.

The experiment consisted of manipulating information content, loading levels, types of communication, length of communication transactions, and levels of secondary loading by the tracking task. These factors were evaluated in terms of relative importance and their effectiveness in determining a quantitative measure of communication workload. Additional details on this study are shown in the discussion on Phase VII.

Perceptual-Mediational Processes

In pilot-only aircraft systems the navigation and many other perceptual-mediational forms of tasks must be performed by the pilot while flying. Such basic tools of categorizing, comparing, problem solving, etc. (see Table 2) are part of this workload category. However, the distinctions between many of these perceptual-mediational tasks are not free from ambiguities (e.g., categorizing, itemizing, and coding) nor is this Table 2 list assured of being complete. Even when several such tasks do form clearly distinguishable differences, it is not a priori clear that a simple pair of tasks cannot subsume several tasks (e.g., see Hitt 1961). While these uncertainties will remain even after the planned tests are complete, the primary purpose of the tests made on perceptual-mediational processes was directed toward investigating the feasibility of creating a SDS which includes such cognitive activities. However, it has been observed that perceptual processes almost always precede a mediational process which in turn precedes either a motor or a communicational task. Accordingly, a dimensional collapsing of tasks is eventually expected. For these reasons the tests on the perceptual-mediational process were not planned to be comprehensive in nature.

Four forms of perceptual-mediational tasks were selected for principal focus: 1) Identify-verify, 2) Itemize-categorize, 3) Code, and 4) Calculate-compute. These four task classes appeared to be reasonably distinct and variable in complexity along clear metrics. Also, these classes of tasks could be presented both orally and visually to the subjects of the experiment with clear audio-visual endpoints and with voice actuated detectors to register the time intervals for task presentations, the information gathering and mental processing time periods, and the duration of responses. Consequently, a similar experimental arrangement to that used in the communications experiment could be employed. Further details on this test are shown in the discussion of Phase VIII.

PHASE V
THE SWITCHING TASKS FEASIBILITY STUDY

Discrete switches used in this study consisted of a variety of types of rotary and toggle switches on each side of the subject; an equal number of types was placed on each side. These different forms of switches had different numbers of switching positions which accounted for the potential switch information as denoted by the maximum number of switch positions. Panels were prepared so that a switch, contained in a standard switch box, could be positioned or repositioned in any of a number of locations in front of or to either side of the subject. Specific locations of these switches constitutes a particular arrangement of the switches relative to the tracking control (joystick) and the speed control (see Figure 1). As mentioned before, these switches and the speed control were instrumented to mark the time when the speed control was released or touched or when any switch was touched or released, to identify the switch touched and the position of any switch at any point in time. Our NOVA 12-20 scanned the task every 36 milliseconds (i.e., every Time Measurement Unit TMU based on the MTM measurement system). In addition, the X and Y error in the compensatory tracking display was also recorded every 36 milliseconds. Details on this experiment are contained in Payne (1981).

Events occurring during a switching task cycle occurred when the subject's hand left or returned to the speed control and when he touched or released a discrete switch. Time intervals between the events of releasing the speed control until a switch was touched was classified as a "reach". Switch "activations" constituted the time interval between the events of touching and releasing a switch and the hand return to the speed control was termed a "return". Time data were collected on these components to the switching task.

The collection of these three successive motions constitute a switching "cycle" as illustrated in Figure 2 of reach, activate, and return motions. Also the squared tracking errors in each dimension were collected within these event intervals for each of the three classes of motion on each switch and for the time intervals between successive cycles; separately for the X and Y dimensions. These accumulated tracking positioning errors were then divided by the time intervals to give an RMS error per second over the reach, activate, and return portions of the switching cycles as well as between switching cycles. Figure 3 illustrates this data collection procedure.

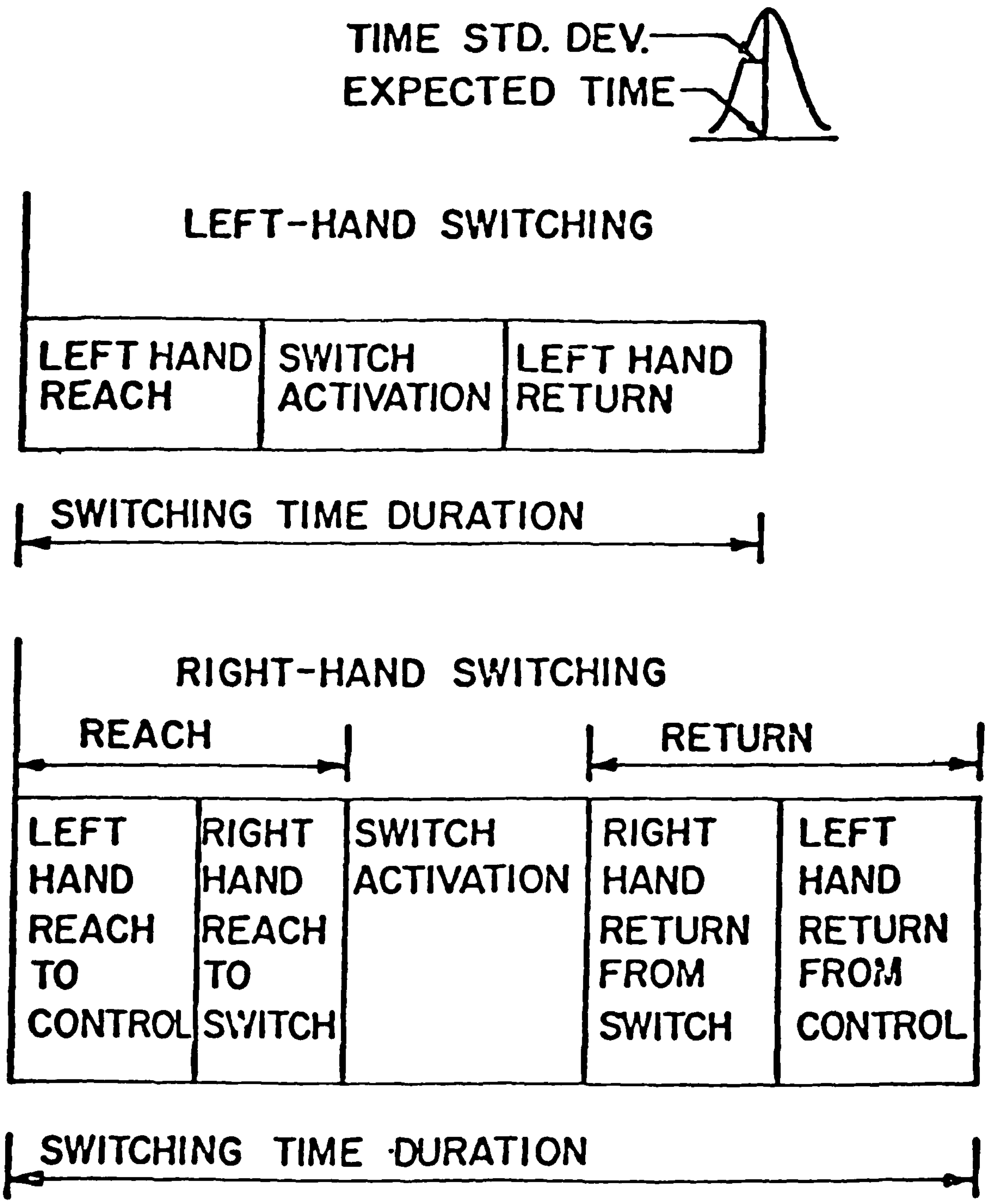

FIGURE 2. LEFT-AND-RIGHT HAND SWITCHING CYCLES

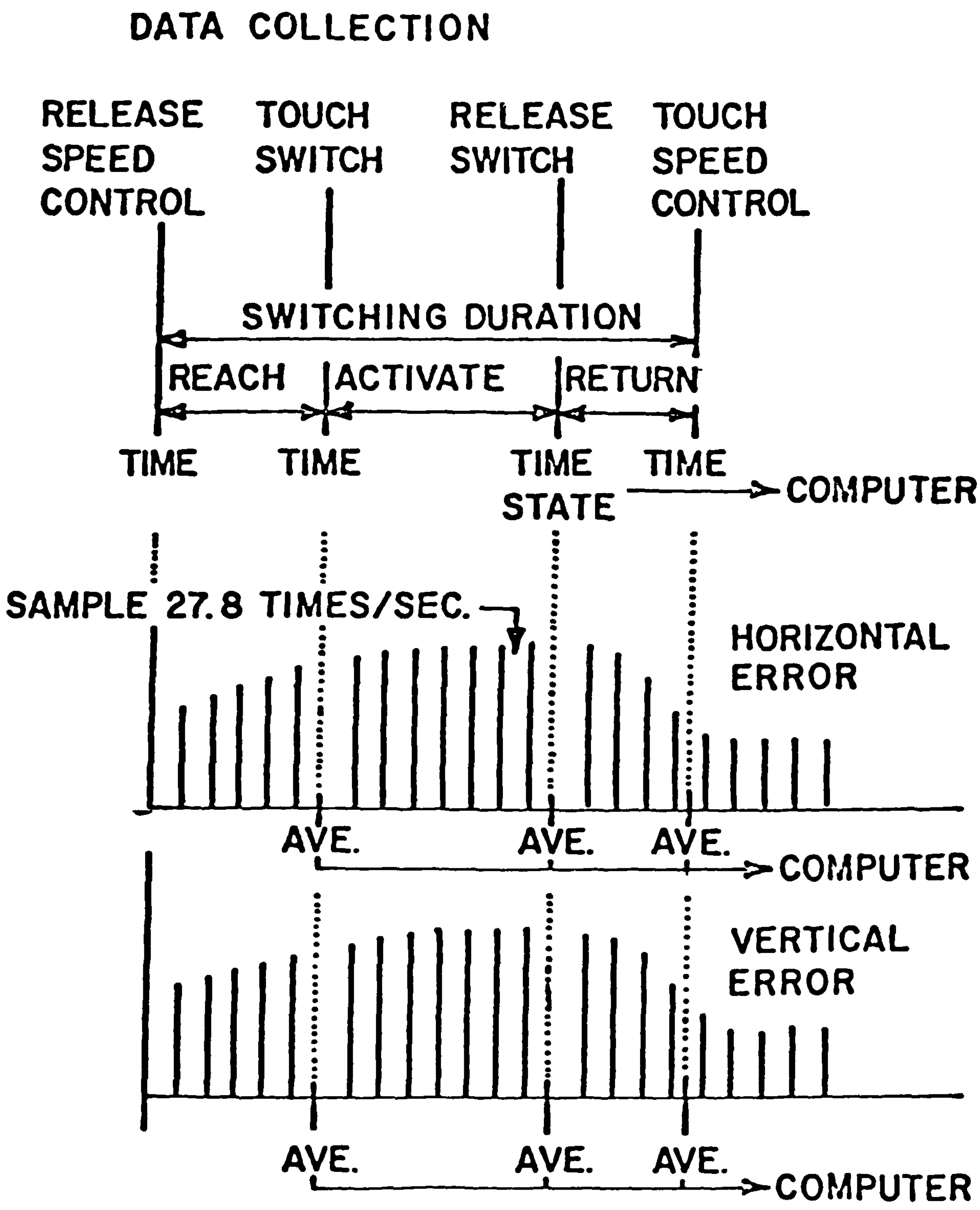

FIGURE 3. THE DATA COLLECTION PROCEDURE WITH TRACKING ERROR DATA

Subjects

Twelve male subjects acted as paid subjects in this study. All were naturally right-handed and all possessed normal uncorrected vision. Their ages were 20 to 26 years; corresponding to those of newer U.S.A.F. pilots. However, none had previous pilot experiences.

Experimental Procedures

The specific experimental procedure for each subject of the switching task study consisted of: 1) Two sessions of tracking only, 2) Two sessions of the discrete switching only, 3) Twelve sessions of the combined switching and tracking task, 4) Two sessions of only the discrete switching task, 5) Two sessions of tracking only, and 6) A repeat session of the first combined switching and tracking task. This total series of 21 sessions took place over approximately 2.5 hours. It was expected that the first four sessions would allow the subjects to learn the discrete switching and control tasks sufficiently well that they could complete the 12 experimental sessions which followed with only minor learning effects. Controlled experimental factors were manipulated during these experimental sessions. The final five sessions were added to track learning effects.

Four experimental variables were specifically controlled during the twelve sessions where both the discrete switching tasks and the tracking task were employed (i.e. sessions 5 through 16). These variables consisted of: 1) Three specific arrangements of switch types in the potential switch locations, 2) Two scenarios for the sequence of switching tasks which were presented orally to the subjects, 3) A forced-pace run through the senario, and 4) A regular frequency forcing function on the tracking task or at approximately half-frequency. It is of note that the different arrangement variable changes the reach distance and visual angle of the various types of switches during a session. Also the particular senario changes the rotation angle changes or the information value of switch changing. There was one frequency level of the forcing function used within a subject; thus, accounting for the twelve sessions per subject. Half of each subject's sessions were given at each level of tracking. All subjects performed under all levels of the other variables. Combinations of the experimental variables were presented in random order.

Analysis and Results

The first stage in the analysis consisted of obtaining statistics of time required for the component reaches, switch activations, and hand returns. These statistics are reported on Table 3 where the reaches and returns for the two hands are reported separately. While there were a large number of experimental conditions for estimating average time values, estimates of the standard deviations of time were restricted to identically repeated conditions; thereby reducing the number of cases for these time variability measurements. In both of these statistical estimations, there were both within and between subject differences as well as switch arrangement differences and sequential switching differences. Table 3 data show that the time variability within a given type of motor movement amounts to 23% to 39% of the average reach and return time and about 35% to 58% of the average switch activation time.

TABLE 3. TIME STATISTICS OF ALL SUBJECTS ON ALL SESSIONS WITH COMBINED DISCRETE SWITCHING AND TRACKING.

ACTIVITY	CASES AVE. S.D.	TIME STATISTICS AVERAGE	STD. DEV.	STD. DEV./AVE.
Left-Hand Reaches	1249-146	22.5	8.7	0.39
Left-Hand Returns	1271-140	19.7	5.3	0.27
Right-Hand Reaches	1215-89	52.0	14.1	0.27
Right-Hand Returns	1152-97	49.4	11.3	0.23
Rotary Switch Activations	2185-524	36.9	12.9	0.35
Toggle Switch Activations	959-114	15.8	9.1	0.58

Average time values for all data but standard deviation statistics include only completely repeated situations within the category.

Next, regression analyses were run on these component motor activity data. All of the experimental controlled variables plus those inferred from the changes in arrangements in switch locations and the scenarios were allowed to enter the build-up form of regression as first order terms, quadratic terms, or as pair-wise interaction terms. While the resulting regression equations become rather complex, they began to better describe the time differences within the groups of reaches and returns and the types of switches activated. For the data statistics shown in Table 3, the regression equation summary results are given in Table 4 where the resulting multiple correlation statistics from the regression are given along with the regression F statistic and the level of statistical significance. These results show that a moderate to a fair amount of the time differences can be explained by the variables in spite of within and between subject differences.

TABLE 4. REGRESSION FITTING STATISTICS OF SWITCHING ACTIVITIES

ACTIVITY	Multiple Correlation Coeff. AVE.	Multiple Correlation Coeff. STD. DEV.	F Statistics and (Significance) AVE.	F Statistics and (Significance) STD. DEV.
Left-Hand Reaches	0.259	0.225	4.9(0%)	1.9(12%)
Left-Hand Returns	0.456	0.225	19(0%)	1.9(12%)
Right-Hand Reaches	0.294	0.203	6.3(0%)	1.2(31%)
Right-Hand Returns	0.170	0.283	2.4(0%)	4.1(2%)
Rotary Switch Activations	0.521	0.437	45(0%)	6.3(0%)
Toggle Switch Activations	0.304	0.369	6.4(0%)	2.1(1/2%)

(See Table 3 for Time Statistics)

Regression models were run on the left and right-handed reaches and returns along with the rotary switch activations for the criteria of: 1) Expected switching time, 2) Standard deviation of switching time, 3) Horizontal tracking error (RMS/sec.), and 4) Vertical tracking error (RMS/sec.). These five component activities and four criteria gave twenty conditions for consideration. Experimental or inferred variables which had the greatest incidence of being more significant by virtue of the fact that they entered the buildup regression earlier, either as a first or second order term or as a member of a two-way interaction, were tabulated. Table 5 shows those variables that entered most frequently, in the 20 situations noted, as an important variable which denoted differences between the average time requirements, their standard deviations, or the horizontal and vertical tracking error. The data used in this analysis were the same as that reported in the two previous tables. It is of importance to note that the variables which showed the greatest differences in these component motor activities turned out to be primarily the inferred variables which were alluded to by the standard synthetic time studies, but not explicitly measured and those variables explicitly cited in INCAN, which was not fully developed.

TABLE 5. REGRESSION VARIABLES WITH THE GREATEST CRITERIAL EFFECTS IN ACCOUNTING FOR TIME OR ERROR DIFFERENCES IN REACH RETURN, OR SWITCH ACTIVATIONS OR VERTICAL OR HORIZONTAL TRACTING ERRORS DURING THESE ACTIVATIONS.

VARIABLES IN REGRESSION.	TIME CRITERIA	TRACKING ERROR CRITERIA	TOTAL
Change Information	10	7	17
Tracking Level	7	10	17
Visual Angle of Switch	8	7	15
Switch Type	8	7	15
Switch Information	5	8	13
Hand Movement Distance	6	3	9

The above reported analysis and that information embedded in the analysis led us to the following observations:

1) Horizontal tracking errors were slightly more highly correlated with the tracking level complexity than were vertical tracking errors, indicating either that there was more randomness in the vertical tracking errors or less ability to respond to the vertical variations due to more damping in the vertical tracking (i.e. greater movement mass).
2) There was greater correlation between the hand move distances and the tracking errors in the reach moves than the return moves and this correlation was greater for the left hand moves than for the right hand.
3) Correlations between the visual angles of the switches (see Figure 1) and the tracking errors were greater for return moves than for reach moves and greater for left-hand switching than right-hand switching. That is, the farther the pilot has to look away from the central tracking task, the greater the tracking errors but the shorter left-hand switching tasks took more of the visual resources per unit of time than did the longer right-hand switching tasks. This phenomenon was probably due to the fact that subjects tracked with their left hand only during right-hand switching events. Subsequently, right-hand switch activation required a switchover from right- to left-hand tracking which consumed visual resources.
4) There were generally greater tracking errors during the hand return activities than during the reach activities but there was little or no difference between left and right-hand switching; indicating that the switching task draws attention away from the tracking task but the amount of the attention withdrawal is about the same for switching to either side.
5) The correlation between the amount of switch change and the tracking error is positive and stronger during the hand returns

than during the hand reaches; indicating that the information processing activity due to the switch-changing uncertainty has a greater effect after the change than before.

These observations are certainly consistent or at least not discordant with synthetic time studies and in accord with the findings of Aume (1963).

The sequential time for a 20 inch left-hand reach to a rotary switch where a 45° rotation is made and the hand return was computed from the regression equations to be about 73 TMUs (i.e. 2.62 seconds). Using the standard MTM system, the estimated time was about 36 TMUs (i.e. 1.31 seconds) or about half of the time estimated from our regression data. While this is an improper use of MTM because of the information processing requirements in the switching task and because the operator's attention was clearly divided between the switching and the tracking tasks (i.e. eye movement times were not included in the MTM measurement), the approximate magnitude of difference is clearly most significant.

Many of the switching activities in the scenarios were multiple switch activations in the same cycle rather than full cycle single-switch activations. While such instances are realistic in flying, this is really a different situation than the full-cycle cases and this situation was identified as a possibly strong source of unaccounted for variance. Accordingly, the full-cycle single-switching data were extracted from the data bank and analyzed separately by regression. Since there was some suspicion of embedded learning effects in these data where the time distribution would likely be gamma distributed, regression analyses were made on the logarithmic transformed time data separately from nontransformed data. Also, regression analyses were made of the component reach, switch activation, and return activities as well as the full cycle. Table 6 shows the multiple correlation statistics obtained for these regression polynomials. These statistics, in comparison to those in Table 4, are generally stronger, indicating better predictions of the activity time. It should be noted though that within and between subject differences were <u>not</u> removed before the regressions were made. Table 6 data also show that: 1) Switch activation time variations were described by the regression equations better than the reaches or returns, 2) Left-hand time variations were better described than those of the right hand, and 3) Multiple correlation statistics for the cycle time (or log time) regression equations were generally greater than the time-weighted averages of the component multiple correlation statistics; indicating that the whole cycle is better described than it is inferred from the parts. There is another important difference between the regression analyses on only full cycle data and the regression analyses reported earlier and that difference is that switch activations by the right hand were separated from those made by the left hand. This difference accounts for, in part, the improved multiple correlation values shown in Table 6 over those in Table 4.

Table 6. Multiple R Statistics from Regression Analyses on Component Activities and Cycles for Only Full Cycle Single Switch Activations with Time Data and Logarithmic Transformed Time Data

LEFT HAND

CRITERION	REACH	SWITCHING	RETURN	CYCLE	TIME WEIGHTED MEAN
Time	0.281	0.680	0.320	0.562	0.472
Log Time	0.316	0.740	0.444	0.620	0.541

RIGHT HAND

CRITERION	REACH	SWITCHING	RETURN	CYCLE	TIME WEIGHTED MEAN
Time	0.319	0.606	0.223	0.477	0.347
Log Time	0.291	0.648	0.192	0.467	0.334

Time statistics were also computed from the regression analyses on the full cycle data. Table 7 shows the average time and standard deviations of time values for the component activities as well as for the switching cycles separately for the left and right hands. Also shown in this table are the correlation coefficients between the reach, switch activation, and return component activity times, within a cycle, where the coefficients are all modestly negative. This result indicates that the subjects were acting in a compensatory fashion where a longer time in one component activity was compensated by a shorter time on the others; a result which is consistent with the "par hypothesis" described years ago by Helson [1949]. Table 7 also shows that the sum of the time variances of the component activities within a cycle compared to the cycle time variance. In the case of both left and right switching, the cycle time variance is greater than the sum of the component time variances due to the mildly negative correlations between the component time values. These differences in standard deviations are 4% and 15% of the cycle standard deviations, respectively, for the left and right hand switching. This result indicates that the time variations in the component activities are not statistically independent but that errors resulting from an independence assumption will be rather small, particularly when within and between subject differences are accounted for and corrected.

Tracking error rates were also investigated during the switching cycles, within component activities of the switching cycles, and between successive cycles. Statistics of these data are reported in Table 8, separately for the right and left hand switching tasks. Since the discrete switching tasks had a randomly varied time interval between successive cycles in the forced pace situation, the variation in timing was assumed to average out the degree of tracking difficulty over the many switching cycles. In the self-paced condition the subject could defer the switching cycle more easily if the tracking task was in a difficult phase. However, the pace experimental variable generally entered the regression equations late in the buildup of the equation (typically pace was about the 8th

TABLE 7. A COMPARING OF TIME STATISTICS OF CYCLES AND COMPONENT CYCLE ACTIVITIES WITH COMPONENT ACTIVITY TIME CORRELATIONS

LEFT HAND SWITCHING

	REACHES	SWITCHING	RETURNS	SUM	CYCLES
Average	26.38	40.69	22.09	89.15	89.15
Std. Dev.	19.71	33.91	15.23	-----	43.93
Variances	388.66	1150.20	231.84	1770.72	1929.84

$\sqrt{\text{Variance Sum}}$ = 42.08 43.93

	SWITCHING	RETURNS
Reaches	-0.258	-0.282
Switching	------	-0.268

Correlation Coefficients Within Cycles

RIGHT HAND SWITCHING

	REACHES	SWITCHING	RETURNS	SUM	CYCLES
Average	52.02	30.53	51.78	134.34	135.34
Std. Dev.	17.92	21.61	20.51	------	41.13
Variances	321.14	466.92	420.73	1208.79	1691.68

$\sqrt{\text{Variance Sum}}$ = 34.77 41.13

	SWITCHING	RETURNS
Reaches	-0.349	-0.416
Switching	---------	-0.340

Correlation Coefficients Within Cycles

variable to enter). However, possible subject effects due to simultaneous tracking interferences, although expected to be small, were investigated further in Phase IX. The existing tracking error data shown in Table 8 do indicate a buildup of tracking error during the discrete switching cycle, generally peaking during the switch activation activity, with little or no decrease during the return activity; clearly indicating a divided attention effect. Tracking error variability followed a similar trend. A comparison of the tracking error statistics within a discrete switching cycle and between switching cycles show considerably less error and error variability between cycles when the subject was free to concentrate on the tracking task alone. Tracking errors in the right reaches and returns were lower than those in the left reaches and returns because the right reaches and returns include the time for changing hands on the tracking control during which visual attention can be maintained on the tracking task; hence lower mean errors and error variances during the reach and return components. However, the tracking errors and error variances during the switch activation component was nearly the same for both hands. These results clearly indicate the divided attention effects on the time requirements which need to be accounted for in developing a Synthetic Data System (SDS) for use in determining pilot workloads.

TABLE 8.
TRACKING ERROR RATES AND STANDARD DEVIATIONS BETWEEN AND WITHIN THE DISCRETE SWITCHING CYCLES AND FOR COMPONENT ACTIVITIES WITHIN CYCLES

Left Hand Discrete Switching

	Statistics	Reach	Switch	Return	Cycle	Between Cycles
VERTICAL	Average	.425	.580	.568	.530	.338
ERROR	Std. Dev.	.375	.505	.494	.383	.226
HORIZONTAL	Average	.426	.554	.562	.548	.300
ERROR	Std. Dev.	.366	.484	.529	.424	.220

Right Hand Discrete Switching

	Statistics	Reach	Switch	Return	Cycle	Between Cycles
VERTICAL	Average	.388	.560	.457	.441	.338
ERROR	Std. Dev.	.272	.502	.332	.278	.226
HORIZONTAL	Average	.348	.547	.419	.394	.300
ERROR	Std. Dev.	.255	.500	.334	.261	.220

Analyses of Variance (ANOVAs) were made on the specific cases of tracking without switch activation for the fast tracking condition separately from the slow tracking condition. Table 9 shows the results of these ANOVAs. Half of the subjects performed in each condition where the root-mean-square (RMS) error was the measured criterion. These data were taken either preceding or following those other experimental sessions where both tracking and switch activation tasks were performed. The purpose of this experimental variable (denoted as T in Table 9) is to examine the learning effect in tracking over the course of the experimental sessions. Two other variables examined in these ANOVAs were: 1) The tracking error direction (D) which was either horizontal tracking error or vertical tracking error, and 2) Subject differences (S). Tests in these ANOVAs were made at the 95% level of significance. The results of these tests showed that all main effects were statistically significant at the fast tracking condition; only the D x T, and the D x S x T interactions were not significant. Accordingly, more vertical error was made than horizontal error, but not with all subjects. Some subjects had a higher tracking error than others, and there was a subtle learning effect observed over the 21 experimental runs of this study. With the slower tracking condition, only the variable T failed to obtain statistical significance and so little or no learning effect was found on this simpler tracking task for most of the subjects.

TABLE 9. ANALYSES OF VARIANCE OF THE ROOT-MEAN-SQUARE DURING THE TWO TRACKING ONLY SESSIONS BEFORE AND AFTER THE COMBINED TRACKING AND SWITCH ACTIVATION TASKS

Faster Tracking Condition

Source of Variance	SS	df	MS	$F_{obs.}$	$F_{0.05}$
Direction of Error	537.54	1	537.54	41.00	4.26*
Subjects	2690.71	5	538.41	41.05	2.62*
Time	309.12	1	309.12	23.58	4.26*
D x S	251.80	5	50.36	3.84	2.62*
D x T	4.58	1	4.58	0.35	4.26
S x T	369.96	5	73.99	5.64	2.62*
D x S x T	72.36	5	14.47	1.10	2.62
error	314.68	24	13.11	---	---

Slower Tracking Condition

Source of Variance	SS	df	MS	$F_{obs.}$	$F_{0.05}$
D	70.59	1	70.59	30.25	4.35*
S	410.15	5	82.03	35.15	2.71*
T	3.65	1	3.65	1.56	4.35
D x S	274.31	5	54.86	23.51	2.71*
D x T	173.31	1	173.31	74.27	4.35*
S x T	327.97	5	65.59	28.11	2.71*
D x S x T	52.47	5	10.49	4.50	2.71*
error	46.67	20	2.33	---	---

* denote statistical significance 95%

A particular arrangement of the switches among the possible switch locations (see Figure 1) was selected because it had data from seven of the 12 subjects, maximum switch information content, and a minimum of missing data. Unbiased smoothing was made for those data missing and the degrees of freedom were correspondingly reduced for a further ANOVA test. In this case the criterion was the time required to reach to, position, or return the hand from a switch as measured by TMUs. Three main variables examined in this ANOVA were: 1) Subject differences, 2) Direction of hand movement toward a switch (reach) or away from a switch (return), and 3) Location of the switch within the panels. Table 10 shows the results of this analysis. Subject differences are shown to be very significant statistically and so are the effects of different switch locations. But the time to reach to a switch was not significantly different from the hand return time for each particular switch. That is, ignoring differences in switch type (e.g., toggle or rotary), the hand reach and return time depends on the switch location. Another important result shown here is the lack of any significant interaction effects of subject by direction or subject by location of switch. This result indicates that a relatively constant effect on reach and hand return time exists within a person due to the switch type and location but there are differences between people on how fast they perform the reach and return activities.

TABLE 10.
ANALYSIS OF VARIANCE RESULTS ON THE HAND REACH AND RETURN WITHIN A SWITCH ACTIVATION CYCLE FOR ONE ARRANGEMENT OF SWITCHES

Source Variance	SS	df	MS	$F_{obs.}$	$F_{0.05}$
Subjects	4487.3	6	747.9	11.70	2.27*
Direction	1.4	1	1.4	0.02	4.02
Location	5197.8	11	472.5	7.40	1.97*
S x D	668.8	6	111.5	1.80	2.27
S x L	4766.7	66	72.2	1.10	1.54
D x L	635.5	11	57.8	0.90	1.97
error	3505.0	55	63.7	---	---

* denotes statistical significance 95%

Further testing was made on the 16-position rotary switch in a fixed location. A series of t-tests were performed to contrast number of variables. Table 11 summarizes the results of these tests. Specifically, these test results show that the performance time on hand reaches or returns was not statistically different for: 1) The fast tracking task compared to slow tracking, 2) Forced pace switching compared to self-paced, 3) The reach time only differed from the hand return time during forced-pace switching under the slower tracking condition, and 4) Two different switch activation sequences in this study. The results further establish the stability of the reach and hand return components of the switch activation task cycle.

TABLE 11. STUDENT t-TEST RESULTS ON VARIOUS CONTRASTS OF CONDITIONS AFFECTING THE TIME REQUIRED WITH THE 16 POSITION ROTARY SWITCH

	$t_{obs.}$	df	$t_{0.025}$
FAST VERSUS SLOW TRACKING			
Reaches, Self-Paced	0.198	43	2.017
" Forced-Paced	0.945	44	2.015
Returns, Self-Paced	1.374	44	2.015
" Forced-Paced	1.782	43	2.017
Sequence 1, Self-Paced+	0.882	19	2.093
" 1, Forced-Paced+	0.394	24	2.064
Sequence 2, Self-Paced+	0.663	22	2.074
" 2, Forced-Paced+	0.925	18	2.101
SELF-PACED VERSUS FORCED-PACE			
Reaches, Slow Tracking	0.849	44	2.015
" Fast Tracking	0.218	41	2.020
Returns, Slow Tracking	1.075	45	2.015
" Fast Tracking	1.263	42	2.019
Sequence 1, Slow Tracking+	0.513	20	2.086
" 1, Fast Tracking+	0.050	23	2.069
Sequence 2, Slow Tracking+	0.360	23	2.069
" 2, Fast Tracking+	0.168	17	2.110
REACHES VERSUS RETURNS			
Self-Paced, Slow Tracking	1.934	45	2.015
" Fast Tracking	1.179	42	2.019
Forced-Paced, Slow Tracking	2.510	44	2.015*
" Fast Tracking	0.259	41	2.020
SEQUENCE 1 VERSUS SEQUENCE 2			
Self-Paced, Slow Tracking+	0.263	22	2.074
" Fast Tracking+	1.421	19	2.093
Forced-Paced, Slow Tracking+	0.102	21	2.080
" Fast Tracking+	1.252	21	2.080

* denotes statistical significance 95% level

+ switch activation only; others include reaching and hand returns

Conclusions from Phase V

A number of variables affect the required time and time variances for operators to activate different types of switches in different locations. Some of the most prominent variables are shown in Table 5. However, the industrially-observed variables used in traditional synthetic time systems are not sufficient for use in pilot tasks. Principal reasons for this insufficiency include the asymetry of the cockpit tasks in switch activations, information processing requirements in switch activation, and the divided attention effects due to the need for simultaneous craft control. Switch activation times in our study were nearly twice that expected from the traditional MTM analysis without eye-motion effects considered.

Regression analyses of the discrete switch activation cycles were able to account for about 25% of the time variability without excluding within and between operator differences but accounting for task factors. This result clearly shows the potential feasibility of developing a Synthetic Data System for these motor activities to meet U.S. Air Force needs, but such a system must be based on the USAF population of operators in real or simulated conditions of divided attention. To illustrate a simplified case of a Synthetic Data System, one of the better-fitting three-term regression equations for the switch activation time was:

$$Y = 8.311 + 0.249(\text{angular change}) + 0.142(\text{visual angle}) + 417\ (\text{switch information content}),$$

where the predicted mean time was the same as the observed mean time within the above level of precision.

PHASE VI
INVESTIGATION OF PROCEDURES AND APPLICATIONS OF A SYNTHETIC DATA SYSTEM

The Phase VI effort consisted of looking ahead to see how a Synthetic Data System (SDS) could be used and some potential benefits which could be derived from it. Also included in this effort was a demonstration of the potential use for SDS.

If all of the necessary conditions of an SDS can be tested and the variables identified which predict performance time differences, then a full SDS can be obtained from instrumented simulators with verifications from actual flight conditions. This feature would assure population fitting and acceptance with confidence. Since simulator data would contain repeated macro-activity modules, the micro-activities can be cross-checked against the component micro-activity buildups in the manner shown above. This development procedure would provide an easier use of the macro-activity modules where applicable and the use of micro-modules when larger modules are inappropriate due to composite errors. Also new macro-activity modules can be revised when there are procedural changes that improve earlier procedures.

Assuming the existence of a suitable computerized SDS, one use of the system would consist of testing procedural scenarios, such as time line analysis, for work overload possibilities. Task sequences may be inputted to the computer along with parameters of the switching arrangement, the control attention level required during the scenario, and any precedence time requirements along the task sequence. The SDS calculations of mean time requirements and time variances could then be determined in the computer for each of the precedence requirements in order to measure the probability that the average pilot would not complete the tasks within the precedence time frames. Figure 4 illustrates the statistical nature of such a situation. With population correction capability built into SDS, similar tests could be addressed to pilots who are at the lower 5% or 10% of this population in order to assess population problems. Of course, any scenarios which are found to have an unacceptable probability of overload would need revision or automated assistance to alter the design. Since these tests could be performed very early in the R & D phases of the aircraft system, potential problems with pilots or other aircrew personnel can be spotted and rectified in the design concept stage without physical simulation testing. Figure 5 illustrates a design application schematic which could be used. Since revised interface layouts merely change the parameters associated with the tasks to be performed, revised layouts can be examined for each and all scenarios tested simply by a computer program which converts the altered layout into parametric changes of the tasks and then reruns the SDS analysis. Tradeoffs in the interface design can be quickly evaluated in this manner. This potential application follows the Human Engineering Computer Aided Design (HECAD) concept of Topmiller and Aume (1978). In addition, various forms of optimization could be added to further enhance the system use.

Another application of SDS is in computer simulation. This analytical tool has been invaluable to systems design and the simulation technique appears most useful to the study of workload. To illustrate and to examine this conjecture, part of the efforts of this project were directed to a simulation demonstration. Since this demonstration was started and executed independently of other project efforts, time values and other assumptions were made with only minor empirical support. A synopsis of this demonstration is described below with a prelude discussion. It should be stated that another reason for performing this demonstration was to explicitly determine other features of SDS which would require testing.

A Queueing Simulation Demonstration

The simulation model described below described pilot tasks during in-flight refueling operations as a queueing system. There was a probability distribution for task arrivals to a pilot. In turn, the task completions were described by another probability distribution which could be described by a developed SDS. A service discipline was used to describe the priorities of a task relative to others awaiting. While the discipline is critical to a faithful simulation and the particular discipline mode used in the simulation is merely assumed here for demonstration purposes, other disciplines can be involved to provide reasonable simulations. A set of reasonable disciplines would need to be identified. However, the current demonstration provides a backdrop for viewing what might be done and it provides a basis for viewing deficiencies which need correction through further research.

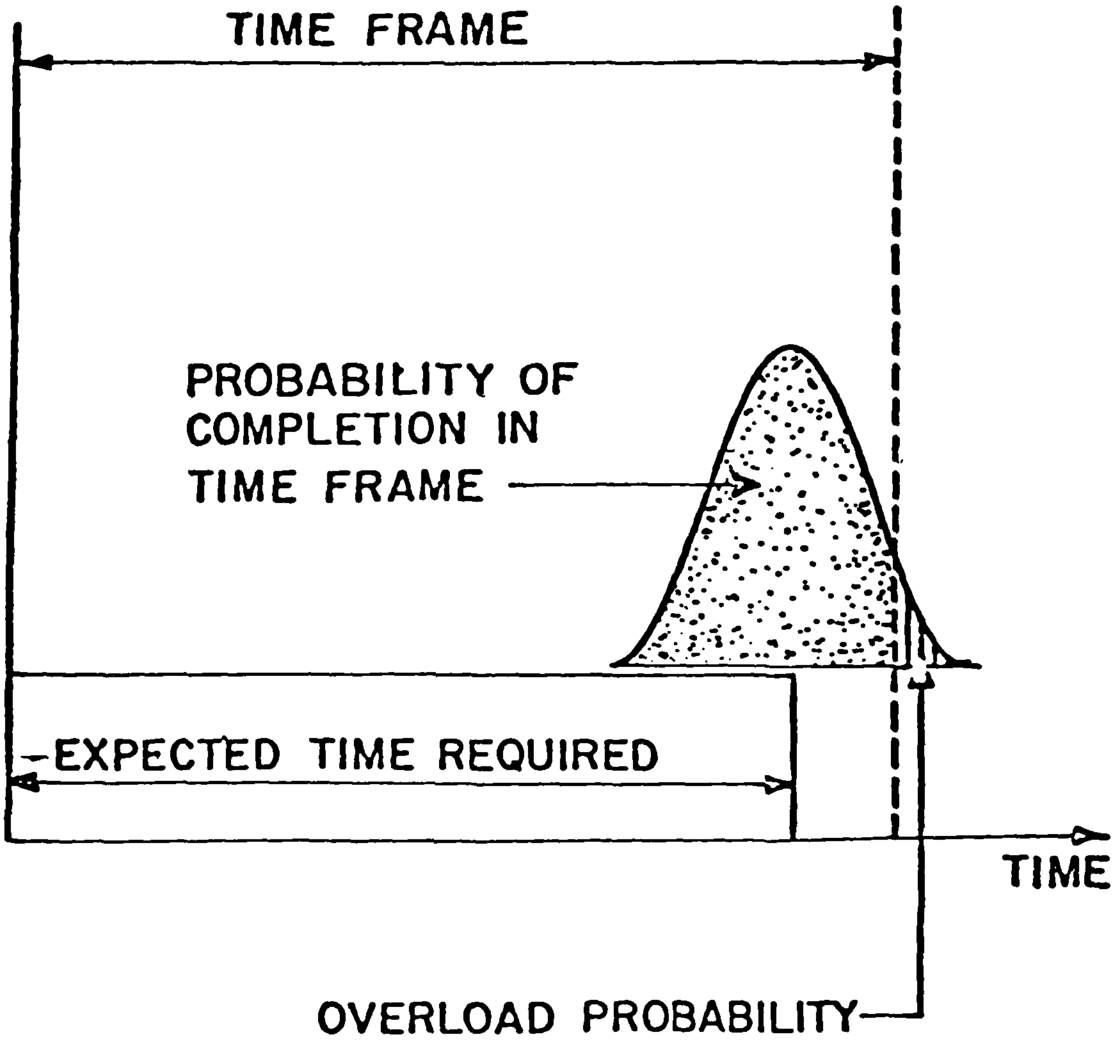

FIGURE 4. WORK LOAD ASSESSMENT IN TERMS OF AN OVERLOAD PROBABILITY

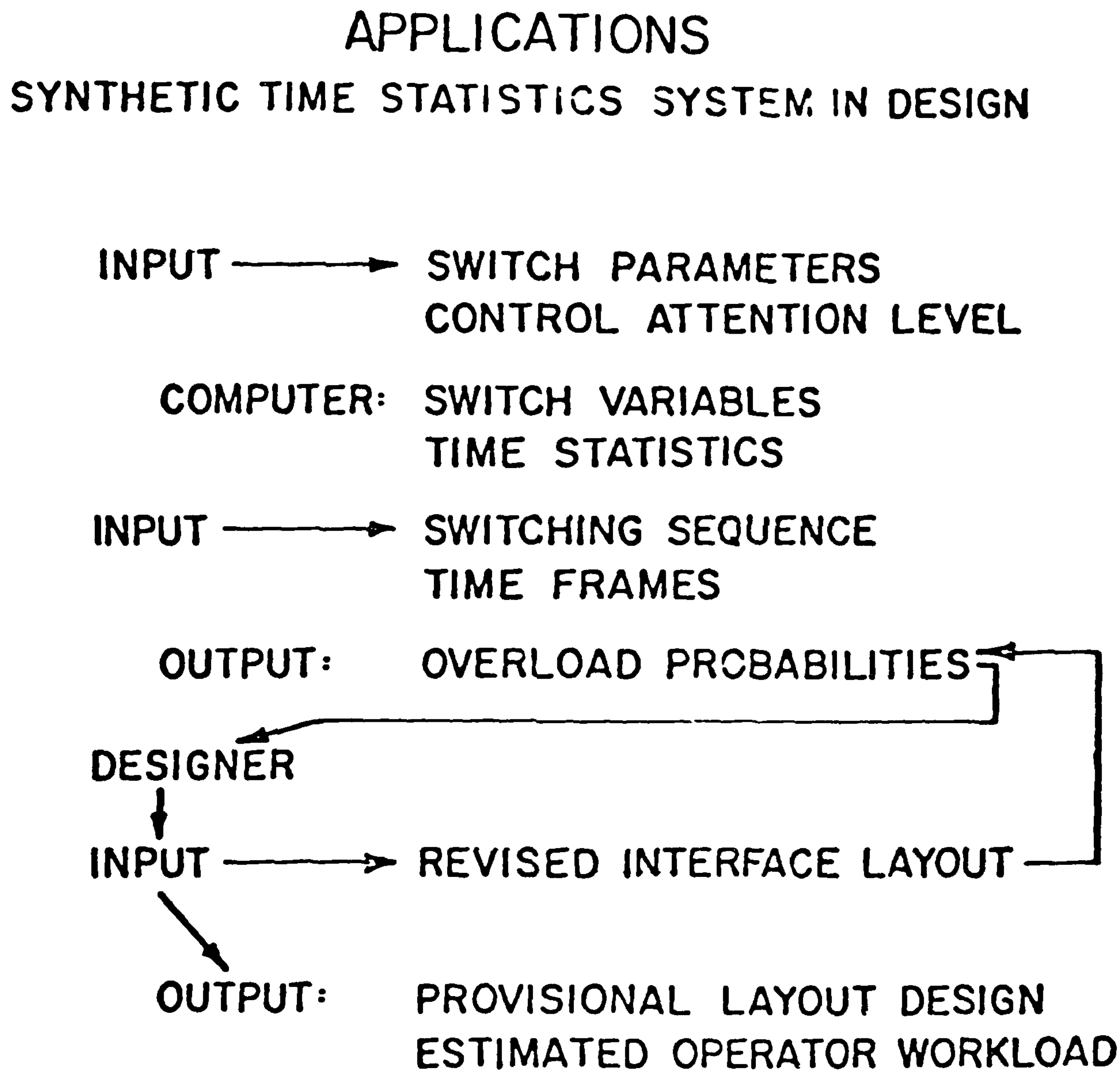

FIGURE 5. A SYNTHETIC DATA SYSTEM IN DESIGN APPLICATIONS

Carbonnel's [1966] concept was employed as the priority setting basis in the simulation demonstration to follow. Essentially, a task has a constant priority when entering the service queue. As time goes on, the priority of the task increases uniformly at a prescribed rate. For the immediate demonstration, the priority constants were randomly set by a triangular distribution and the frequency of task arrivals was randomly determined following an exponential distribution.

This computer simulation program was written using a recently developed simulation language called "SLAM" [Pritsker 1979]. Essentially this language is network based and the particular program developed for this demonstration consisted of two parallel networks, one for the continuous task of craft control and the other for a series of discrete tasks which a pilot performs. In the case of the continuous task (e.g., tracking), threshold levels were set so that a corrective action is made in the tracking task when the error level exceeds the threshold level, temporarily preempting effort on the discrete task. State variables were arbitrarily used in this simulation to describe the continuous tracking task. A variety of discrete tasks were similarly handled using the other network. For this particular case, the discrete tasks consisted of those shown in the upper part of Table 2 and the task was selected and introduced by random choice. Task completion times were assumed to be Gaussian distributed with known average times and standard deviations. While arbitrary time statistics were used in this demonstration, the actual time values could be obtained from SDS when available. Figure 6 shows a SLAM diagram of this simulation.

The entire philosophy of this simulation is that the pilot is a resource which is allocated among the tasks to be done. While the pilot is constantly at the tracking tasks, only when the tracking error exceeds an imposed threshold is the pilot resource claimed by the tracking task. At other times the pilot resource is free to accept the discrete tasks which arrive randomly and await in a queue until the pilot is free to work on these switching and other discrete tasks. When these tasks arrive faster than the resource is able to complete them, the queue builds up. The selection of the discrete tasks is made in this simulation, based strictly on the task priorities which are time updated to account for their period of waiting in the queue when a discrete task is selected. The service time to complete a task is determined by the statistics of the type of task and a generated random number. While the service times in this demonstration were arbitrarily set as shown in Table 12, these statistics could be generated by a SDS to be appropriate to the task and the level of tracking control difficulty. When the queue of discrete tasks is empty, the pilot resource is placed in a free mode.

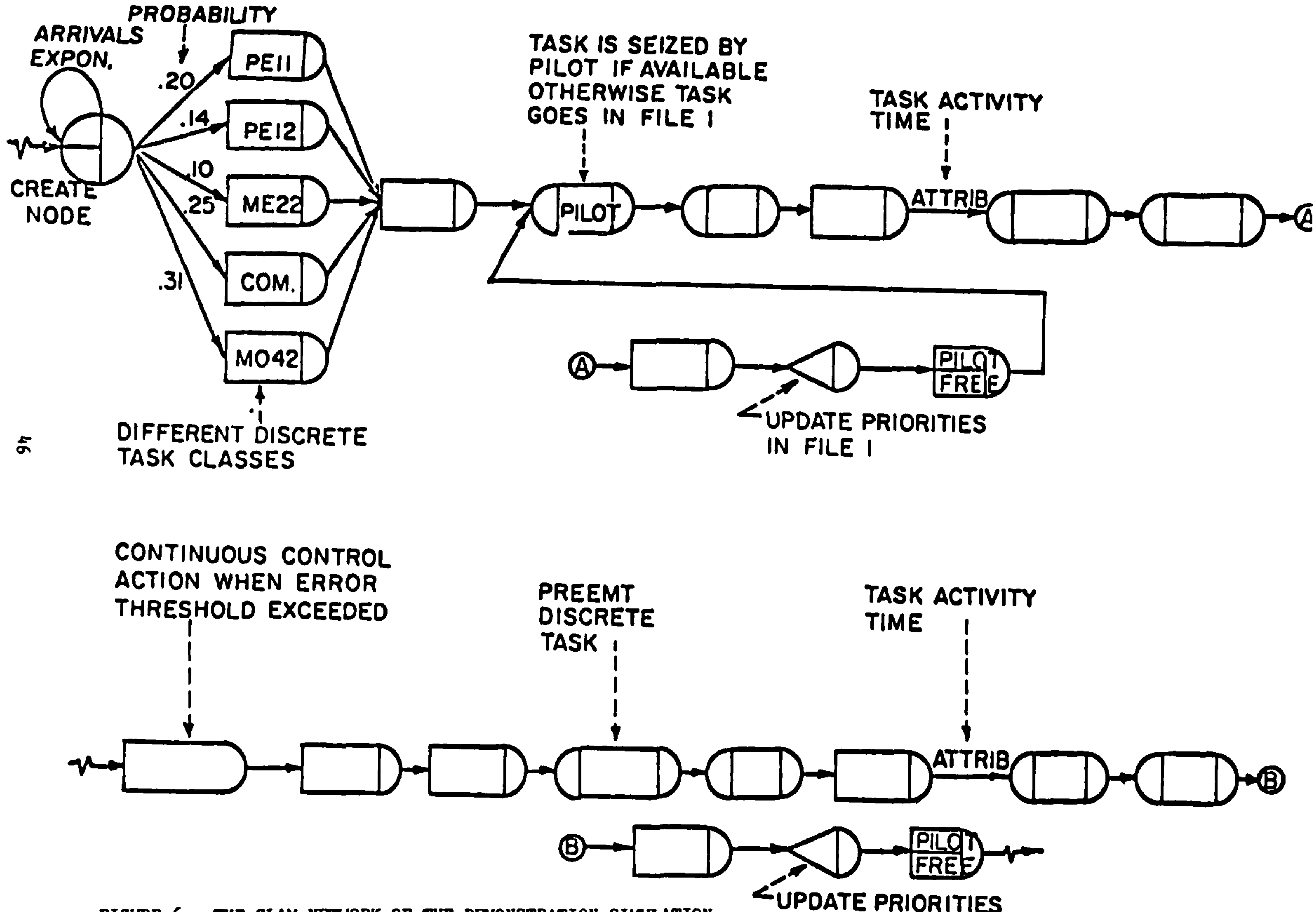

FIGURE 6. THE SLAM NETWORK OF THE DEMONSTRATION SIMULATION

TABLE 12.
ASSUMED PILOT ACTION SERVICE TIME STATISTICS FOR THE SIMULATION DEMONSTRATION [SIEGEL AND WOLF 1969]

PILOT ACTION	AVERAGE TIME (SEC.)	STD. DEV.
Set Toggle Switch	1.1	0.76
Set Rotary Switch	8.6	3.00
Push Button Switch	4.2	1.02
Set Thumbwheel Switch	3.0	0.48
Joy Stick Correction Setting	3.8	0.48
Read N Instrument Displays	0.6N+0.6	0.2N+0.2
Communicate N Words Orally	0.66N+0.6	0.34N+0.4

After the completion of every pilot task, an event is created which causes a subroutine to update the task priorities, another to collect the data generated, and a third to free the pilot to select another task. State variables are constantly updated, as in multitasking, and whenever the tracking error update exceeds the present threshold, then the tracking task preempts any discrete task by the tracking-correction service time (a similarly generated random variable to the discrete service times). Upon completion of the tracking correction, the interrupted discrete task may be selected and performed in the remaining time required or a new discrete task is selected, depending upon the task priorities.

During the execution of the simulation, there were a variety of collected statistics including: 1) Statistics of a discrete task waiting in the queue, 2) Statistics of the number of discrete tasks awaiting over time, 3) Time statistics of the discrete tasks leaving the queue for servicing by the pilot, and 4) The pilot's busy-time statistics. These statistics provide a basis for viewing the workload situation of pilot scenarios both in a static and a dynamic sense and for correcting deficiencies. Also these statistics can be collected and reported for sequential time frames over the simulation run in order to identify groups of tasks which tend to have high waiting times; indicating that the pilot needs an automatic assistance or an aid with some of those tasks. Situations where there are a large number of discrete tasks in the queue are candidates for pilots failing to perform them due to memory limitations.

This demonstration simulation was operated under the conditions shown in Figure 6 where the discrete task interarrival time was negatively exponential with a mean rate of 10 tasks per time unit and the continuous control task at a high level. Table 13 shows the statistical summary of data from this simulation. The statistics at the top show the discrete tasks' waiting time in the queue before being serviced by the pilot, the time being serviced by the pilot and the total time in the system. In this

TABLE 13. STATISTICAL RESULTS OF THE DEMONSTRATION SIMULATION

Statistical Results

Statistic for Variables Based on Observation

	Mean Value	Standard Deviation	Coeff. of Variation
Waiting Time	19.02	14.39	.76
Time in System	25.86	13.33	.52
Time in Service	5.88	3.24	.55

Statistics for Time-Persistent Variables

	Mean Value	Standard Deviation	Maximum Value
Number in the System	3.33	2.36	9

File Statistics

File Number	Average Length	Standard Deviation	Maximum Length	Current Length	Average Waiting Time
1	2.44	2.23	8	5	4.18

Resource Statistics

Resource Number	Resource Label	Current Capacity	Average Utilization	Standard Deviation
1	PILOT	1	.8929	.3092

case the discrete tasks took an average of 26 minutes to arrive and to be completed with a standard deviation of 13.3 minutes; indicating that 5% of the tasks were in the system as much as 48 minutes. Time-persistent statistics show the number of discrete tasks in the system (i.e., in the queue or being serviced). File statistics show queue characteristics of queue length statistics, where the current length is at the end of the simulation, and the average time interval between tasks leaving the queue for pilot servicing. These file statistics indicate a rather long buildup of tasks in the queue where memory failures may pose severe difficulties. Finally, the resource statistics denote the pilot's activity level. These data show that the pilot was busy 89% of the time on the average and with a rather high standard deviation, indicating a very heavy workload on the pilot.

To further demonstrate the potential of a simulation, a small experiment was performed using the simulation model as the experimental vehicle. One of the variables examined was the probability mixture of mediational and motor tasks where the probabilities leading to modes ME22 and MO42 in Figure 6 were as shown or reversed. Two other factors in this experiment which were varied were the interarrival rate of discrete tasks entering the queue, which was modelled as being from a negative exponential distribution with means of 5 or 10 tasks per minute and the continuous control task at the current rate of the complexity or half that rate. These eight (2^3) conditions were simulated using each of two random number streams and analysis of variance (ANOVA) was performed on two measures from the simulated output. One of these measures was the maximum number of discrete tasks in the queue. In that ANOVA only the interarrival rate of discrete tasks to the queue proved to be a statistically significant variable (95% level of significance). This same experimental variable was found to be statistically significant on the other ANOVA run on the pilot's percentage of utilization. In addition, the interaction effects between the discrete task arrival rate and the discrete task mixture also proved to be statistically significant. This ANOVA on the pilot's workload, as measured by the percent of busy-time, showed that the workload depended not only upon the arrival frequency of discrete tasks but <u>also on the mixture of task types</u>. Table 14 shows the results of these two ANOVAs.

Perhaps the greatest benefit achieved from this demonstration was to highlight deficiencies in this simulation model and the provision of information which would be needed to do more realistic simulations. A principal deficiency of the existing simulation model was the random ordering and timing of the arrival of pilot tasks. While some tasks will arrive in a nearly random fashion, others will be highly ordered (e.g., preflight checkout), and others will exhibit some randomness over order. Additionally, provision should be made for allowing related tasks to arrive simultaneously in a way that approximates certain aspects of flying. Another deficiency in the current simulation model was the built-in single channel processor between discrete and tracking tasks [Welford 1968, Norman and Bobrow 1975]. This deficiency can be modified using the state variables of the tracking task to modify the time functions for the discrete tasks. Data analyzed and described in Phase IX provided some information on how this time function modification can be properly done but these data were not available during this Phase VI effort. Another important deficiency of this current simulation model was the priority system and the assumption that a

TABLE 14. RESULTS OF THE TWO ANALYSES OF VARIANCE PERFORMED ON THE SIMULATION MODEL

ANALYSIS OF VARIANCE - MAXIMUM NUMBER OF DISCRETE TSKS IN THE QUEUE

Source of Variation	Degrees of Freedom	Mean Square	Test F Ratio	Comment
Control Task Rate = C	1	45.56	1.29	NS
Discrete Task Arrival Rate = D	1	1501.56	42.53	significant
Discrete Task Mixture = M	1	18.06	0.51	NS
C x D	1	52.57	1.49	NS
C x M	1	33.07	0.94	NS
D x M	1	7.57	0.21	NS
C x D x M	1	18.05	0.51	NS
Error	8	35.31		

ANALYSIS OF VARIANCE - PERCENT UTILIZATION OF THE PILOT

Source of Variation	Degrees of Freedom	Mean Square	Test F Ratio	Comment
Control Task Rate = C	1	0.003	2.25	NS
Discrete Task Arrival Rate = D	1	0.0536	39.98	significant
Discrete Task Mixture = M	1	0.0062	4.51	NS
C x D	1	0.0016	1.16	NS
C x M	1	0.0005	0.36	NS
D x M	1	0.0091	6.62	significant
C x D x M	1	0.0001	0.07	NS
Error	8	0.00138		

$F_{.95;1/8} = 5.32$

NS denotes not statistically significant partially-performed interrupted discrete task can be completed in the remaining time not previously expended on the discrete task. However, further research will clearly be required to obtain a realistic priority system and a basis for determining the remaining completion time of interrupted tasks. These and other limitations of the current simulation model can be corrected by the development of a more sophisticated model. This first effort toward a viable model pointed out a number of the corrections needed to achieve reasonable validity.

PHASE VII
THE COMMUNICATION WORKLOAD FEASIBILITY STUDY

Based on the study by Geiselhart (1978), it is clear that pilots spend a large percentage of their time in communications. Therefore, Phase VIII of this project was focused upon assessing the feasibility of a Synthetic Data System (SDS) for evaluating the communications workload. Only oral communications with minimal mediational contents are considered in this phase of the project as Phase V dealt with pilot-to-machine communications through switch activations and Phase VIII was directed to mediational tasks.

A Federal Aviation Administration study (Hunter, Blumenfield, and Hsu 1974) provided an analysis of messages between pilots and air-traffic controllers. This analysis indicated that these oral messages were generally short time duration (about 3 seconds), and were largely (70%) single message transmissions. The variational character of these messages was also shown. Since Miller, Heise, and Lichten (1950) found that intelligibility and attention demandingness of a message was associated with the information content of the messages as given by the Shannon (1948) measure, this characteristic of the communication was examined as one of the experimental variables of this study phase. Information content is measured by the theoretical formula of:

$$U = \sum_{i=1}^{n} - p_i \log_2 p_i \quad \text{(in units of bits; binary digits)}$$

where: 1) p is the relative frequency of the communicated message, 2) the index i refers to different messages, and 3) the units of measure are bits (binary digits). Messages were constructed to consist of an average information content of either 1.5 or 2.5 bits per message. Large numbers of past psychological studies and some of those discussed above on the topic of Information Content Analysis (INCAN) provide support for the selection of this variable as a potentially effective variable pertaining to communications workload and for the use of this metric of description. The variable loading level refers to the percentage of the total operating time which the operator spends in the communications process. In effect, the loading level describes the communications rate which varies considerably for pilots in different phases of flight. Three levels of loading were investigated (approximately 16%, 28% and 35%). Both speaking and listening types of communication were separately considered because the mental workload of listening was suspected to be more passive than speaking and this conjecture was tested. Also speaking involves motor responses which were expected to interfere with the tracking task while listening did not involve this source of interference; another conjecture to test. Another feature of communications which was suspected of creating an effect on workload was the number of messages communicated in a single transaction and this variable is referred to below as the length of communication transaction. Three levels of lengths were studied, with 2, 4, and 6 messages per transaction corresponding to these levels. The final experimental variable investigated in this study was the tracking level which was defined in Phase V.

The basis for selecting the loading level as an experimental variable were the studies of Pasmooij et al., (1976) and Senders (1964). Another reason for chosing this experimental variable was that the variable information content in bits per message "together with the variable loading level in terms of the percentage of time spent in communication" described the rate of information processing (Raouf and Mehra 1974). If the essential feature of the communications effect on the workload was a processing rate rather than the variational nature of the messages or the amount of time in communication, then the interaction effect of these two experimental variables would confirm this suspicion. These bases for variable selection and other details on this experiment are given by Lehto (1980).

Subjects

Twenty-four students were recruited from Purdue University and trained in operation of experimental apparatus. Initial training procedures included the provision of information relevant to the nature of the experiment such as: 1) An explanation of tracking task, 2) Instructions pertaining to locations and functions of the various displays, nature of the communication task, and 3) Other miscellaneous procedural instructions. Additional training consisted of both practice on tracking until their learning curve was approximately asymptotic and an arbitrary number of communication trials to give each subject experiences with each message type and their relative frequencies.

Equipment

The same two-axes compensatory tracking task, as described in Phase V, was used. However, the equipment for this task was modified so that alpha-numerical displays could be superimposed around the edge of the circular tracking display using a two-way mirror in front of a Hazeltine CRT and an oscilloscope projecting downward.

Eight displays were used, consisting of 4 LED displays and 4 displays presented via a Hazeltine CRT. All displays were driven by a computer so that the updated numerics followed different trends over time but were of the same hardware. The experimental apparatus was similar to Figure 1 except LED displays were used in the lower front panels instead of switches. Also the central display was changed to that shown in Figure 7.

Tracking error and message-time data were collected at 72 millisecond intervals by the Nova 3-12 computer. Tracking error was collected by summing the squares of the tracking errors during each between event interval. Message time data were obtained through the use of a voice-actuated device (vox) which denoted the initiation and termination of messages and responses to the computer. These times of message initiations and terminations to and from the subjects constituted the events of this study.

A communication transaction was initiated by inputting a brief signal to a Coxco tape recorder from the computer. Prerecorded messages (3 seconds in duration) were then presented to the subject via headphones.

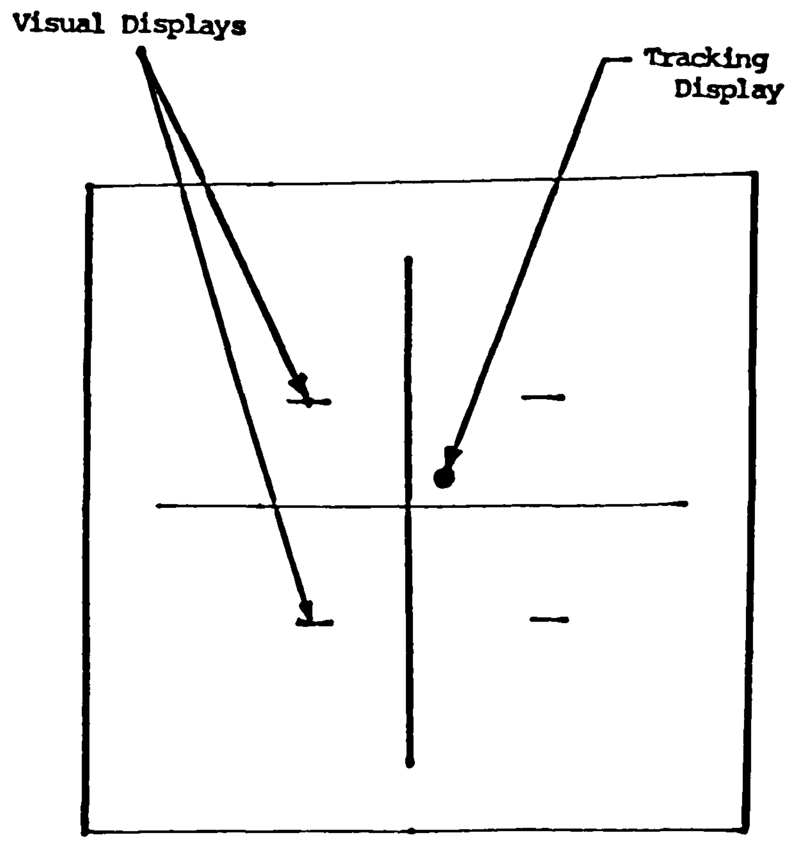

FIGURE 7. CENTRAL DISPLAY LAYOUT

After attaining desired competence in tracking ability and completing familiarization procedures, the collection of data on the subjects' performance under experimental conditions began. These experimental conditions consisted of four blocks in which the tracking task level *remained constant* while the other variables were manipulated factorially. Table 15 describes the experimental design of this study. Experimental variables were totally randomized with the exception of the control conditions (pure tracking) that preceded and followed each experimental block. The number and nature of the experimental variables in this study necessitated this rather complex experimental design.

As shown in Table 15, the subjects variable was nested within loading and information levels. In other words, each subject saw different levels of secondary loading, lengths of communication transactions, and types of communication and monitoring tasks while under the same communication loading level. Since the number of messages presented in each experimental session was held constant at 72, variations in the session duration was made to vary the different loading levels of communication. Table 16 describes the combinations of the length of transactions and the number of messages presented for the three levels of communication loading, with the resulting session duration. Each session was then divided into four experimental blocks.

TABLE 16. EXPERIMENTAL SESSION DURATIONS UNDER VARIOUS EXPERIMENTAL CONDITIONS

		Number of Messages Presented	Number of Replicates
Length of Transactions	2	24	24
Number of Messages Within Communication Transactions	4	24	12
	6	24	8

Where 16 % loading results in 45 minute sessions

25 1/3% loading results in 28.4 minute sessions

34 2/3% loading results in 20.8 minute sessions

As previously mentioned, the subjects variable was nested within communication loading levels, while the other factors were manipulated factorially. This design was accomplished by presenting messages, whose relative frequencies have been predetermined at random time intervals within each block. Table 17 illustrates a paradigm of this procedure. Each message *required the subject to indicate* the value of a specific display. These messages differed in that each message had a different key word (the first word in the message) and in that they refer to different displays. Messages corresponding to the different displays explicitly gave the

TABLE 15. THE EXPERIMENTAL DESIGN OF THE COMMUNICATIONS STUDY

				Information Content (bits per message) 1.5			2.5		
				Loading Level			Loading Level		
				16⅔ sub	28 1/3 sub	34 2/3 sub	16⅔ sub	28 1/3 sub	34 2/3 sub
Low level	L=Listening Type of Communication ↓ S=Speaking	Lengths of Transaction	2 4 6	1, ,4	5÷8				21÷24
		Lengths of Transaction	2 4 6						
tracking task	L	Lengths of Transaction	2 4 6						
High level	Type of Communication S	Lengths of Transaction	2 4 6						

TABLE 15. Continued

$$Y_{ijklmnopq} = M + I_i + C_j + IC_{ij} + O_k + IO_{ik} + CO_{jk} + ICO_{ijk} + S_{(ijk)l} + R_{(ijkl)m} + T_n + T_{int} + P_o + P_{int} + N_p + N_{int} + Z_{(ijklmn)q} + E_{(ijklmnopq)}$$

where:

M = mean	
I = information content	i = 1,2
C = communication loading proportion	j = 1,2,3
O = central display percentage	k = 1,2
S = subjects	l = 1,2,3,4
R = replicate (blocks)	m = 1,2
T = tracking level	n = 1,2
P = type of discrete task	o = 1,2,3
N = length of communication transaction	p = 1,2,3
Z = replicate (within cells)	q = 1,...,6
E = error term	

location of the display, in an attempt to reduce any learning effects. Individual messages were three seconds in duration and prerecorded to reduce statistical noise in the obtained data.

TABLE 17. A PARADIGM OF MESSAGE PRESENTATION

Block (1)

Initiation Time	Specific Messages Types	The Message Type	Relative Frequencies Of Each Type
.95	5, 8	1	.04
2.11	5, 3	2	.01
3.23	5, 3, 5	3	.43
5.08	5, 3, 5	4	.01
5.85	3	5	.42
6.45	5, 3	6	.01
7.72	5	7	.01
8.95	7	8	.06
9.22	3		
9.82	1		
10.63	3		

Table 17 further shows that the individual messages were presented in blocks of 1, 2, or 3. These blocks correspond to the respective lengths of communication transactions. If one message is presented to the subject, this message and the subjects' response corresponded to a two message communication transaction. In an obvious manner, this analogy carries to 4 and 6 message communication transactions. This experimental paradigm shows the relative frequency of individual messages which specifies one level of the information content; whereas the other level is achieved by modifying these relative frequencies of the message types.

Throughout the course of experimentation, a subject's performance was evaluated on the basis of the tracking error, accuracy of response, and response times. Tracking error measurements were made on the vertical and horizontal squared deviations of error on the compensatory tracking task and modified with respect to time and the input signal as discussed in the section on Phase V. Also, the tracking error standard deviation was collected. This dependent variable shows the changingness of the error within an event. Accuracy of the subject's response was monitored by the experimenter to see that the subject scanned the desired display. The likelihood of significant error seemed remote for this simple task, but this monitoring was performed to remove any experimental artifacts. Response times consisted of the time interval between the presentation of a message and the cessation and/or initiation of the appropriate response. These data were obtained through the use of voice actuated device that indicated when verbal communication was taking place.

Data Analysis

The information on raw data from this experiment was analyzed using two methods. An Analysis of Variance (ANOVA) provided conclusions as to the significance of the main effects and interactions of the variables that were considered here. Due to the structure of the experimental design, all of the experimental variables were tested for statistical significance. Following this ANOVA, these variables and interactions which obtain statistical significance were fitted to the data by regression methods in order to obtain predicting polynomial equations of the response time corresponding to the communication variables examined in this study.

Results of the Analysis

Mean performance measurements were obtained in this analysis which are reported in Tables 18 and 19. The symbols T, L, M, and S in the task type variable are tracking (only), listening to the recorded message, monitoring the data required on a display, and speaking out the value shown on that display. Central display percentages refer to the proportion of requested display information on the CRT display (which also shows the tracking information) compared to data on peripheral displays (see figure 1).

Results from the ANOVA on the reaction time criterion are reported in Table 20. The reaction time is the interval in TMUs between the end of the recorded message and the start of speech. These results show that none of the main effects tested obtained statistical significance in this study except the effects of different subjects and the communications transaction length (i.e. the number of sequential messages with minimum time durations between them). Also, the reaction time significantly decreased with the length of the communication transmission.

Table 21 shows the results of regression analysis on these communications data. As one can see from these data, about 27% of the reaction time variability was due to differences between subjects. Information content, communications loading percentages, and information rate variables also were significant factors in the regression analysis.

In the ANOVA on the reaction time standard deviation criterion, the only effects that obtained statistical significance were the two three-way interactions 1) ION x I x R, and 2) ION x T x R (see Table 20 for symbols). Also, in the regression equation only the information content was a predictive variable, not considering subjects, and then the multiple r value was only 9%.

The response time, or equivalently the average amount of time that a person was speaking, was examined by an ANOVA and the results are shown in Table 22. Only the subjects' main effects were statistically significant. Two two-way and one three-way interactions were statistically significant: 1) tracking level by replicate, 2) information content by eyes-on-display percentage, and 3) eyes-on-display percentage by tracking level by communication transaction length. When there was a larger percentage (75%) of central displays the response time decreased with greater information content but the reverse was true when the central and peripheral displays were equal; thus verifying the second two-way interaction. That interaction

TABLE 18. MEAN PERFORMANCE ON THE COMMUNICATION TASKS

Independent Variable	Level	Dependent Variable: RMS error	RMS error variance	reaction time	response time
Info Content	Hi	77.1	1115.2	45.8	49.2
	Lo	77.8	1156.5	42.2	55.9
Loading Percentage	16%	83.6	1312.4	47.4	51.0
	28%	74.7	1056.8	39.7	58.7
	35%	74.1	1038.4	44.8	48.0
Task Type	T	63.0	1190.1	-	-
	L	65.1	971.3	-	-
	M	87.7	1162.8	44.0	-
	S	79.7	1273.5	-	52.6
Tracking Level	Hi	91.6	1626.9	44.4	52.2
	LO	63.4	644.9	43.6	52.9
CT Length	2	75.4	1005.5	46.1	53.3
	4	79.2	1210.4	43.7	52.5
	6	77.8	1191.7	42.3	51.9
Central Display Percentage	50%	80.3	1294.1	44.7	51.8
	75%	74.6	977.6	43.3	53.4
Replicate	1	77.9	1107.8	45.5	51.7
	2	77.0	1163.9	42.6	53.5

TABLE 19. OTHER MEAN PERFORMANCE DATA FROM THE COMMUNICATIONS STUDY

Independent Variable Level		Dependent Variable: Reaction Time Std. Deviation	Dependent Variable: Response Time Std. Deviation
Info Content	Hi	16.6	9.4
	Lo	15.4	9.2
Loading Percentage	16%	17.1	11.1
	28%	15.7	9.1
	3%	15.2	7.7
Tracking Level	Hi	16.1	10.0
	Lo	15.9	8.6
CT Length	2	16.6	9.2
	4	16.1	9.6
	6	15.3	9.0
Eyes on Display Percentage	50%	15.8	8.7
	75%	16.2	9.9
Replicate	1	16.2	8.8
	2	15.8	9.8

TABLE 20. ANOVA FOR REACTION TIME

Source	MSE	F	df	Significance Level
Info Content (I)	5822	.33	1	-
Loading Percentage(L)	8847	.51	2	-
Central Display Percentage (ION)	924	.05	1	-
Tracking Level (T)	215	.36	1	-
CT Length (CT)	2127	5.0	2	.025
Subjects (S)	17399	-	12	-
Replicate (R)	1813	-	24	-
I x L	60224	3.46	2	.10
T x R	769	2.03	24	.01
L x ION x T	2047	3.51	2	.10

TABLE 21. REGRESSION ANALYSIS OF REACTION TIME

Independent Variables	β mean	β std. error	Significance	R^2 change	Simple R
RMS error variability	.0022	.0022	.000	.069	.26
Information content x communication loading percentage	-.124	.02	.000	.027	-.04
Message position in communication transaction	-2.23	.941	.000	.009	-.09
Experimental block	-2.64	.936	.000	.006	-.04
Visual angle	2.01	.984	.000	.005	.09
Communication loading percentage	2.11	.408	.001	.004	-.06

TABLE 21. Continued

RMS error	-.026	.01	.023	.002	.07
Message block number	1.65	.709	.018	.002	.06
Information content (average)	2.58	.602	.033	.002	.07
Information content of specific messages	27.95	8.97	.054	.001	.05
Message relative frequency	-.134	.052	.019	.002	-.06
Subjects	-	-	.000	.270	.52
Constant	.249	12.69	n.s.	-	-

Total R^2= .40

TABLE 22. ANOVA FOR RESPONSE TIME

Source	MSE	df	F	Significance Level
Info Content (I)	20240	1	1.31	-
Loading Percentage (L)	17684	2	1.2	-
Central Display Percentage (ION)	1098	1	.07	-
Tracking Level (T)	249	1	.19	-
CT Length (CT)	320	2	.45	-
Subjects (S)	15375	12	-	-
Replicate (R)	694	2	-	-
T x R	474	1	1.8	.01
I x ION	88465	1	5.75	.05
ION x T x CT	669	2	3.67	.05

is substantive while the first interaction effect is practically insignificant. However, all response time effects were insignificant, practically, except that of the differences between individual subjects.

In the regression analysis on the response time, the subjects factor accounted for 47% of the time variance. Several other variables showed contributions of about 2% of the remaining variance of response time. What seems particularly important here is the high constancy in response time by a given individual regardless of the factors and a wide variation in response time by different people. It even appears that the response time standard deviation is nearly constant; only the tracking level had even a very minor effect on this second-order statistic.

Since the variables of this communications task, other than individual subject differences, appeared to have only a very minor effect on the execution time, the data examination focused upon the tracking error analysis. Table 23 gives the ANOVA summary for the average tracking error (i.e. RMS). As expected, the tracking RMS was significantly greater with the more difficult (faster) tracking level and more so within different parts of the communication tasks (i.e. listening, monitoring, and speaking). Subject differences and the replicate effect (learning) were also statistically significant, as expected. The significant effect on the tracking error due to the task type or portions within the communications sequence was primarily due to the tracking error increase during the reaction time period when the subject was monitoring the display containing the requested information. During the monitoring, the subject moved his or her eyes off the tracking display and tracking error increased. With the more peripheral displays, the eyes were away from the tracking display longer and the tracking RMS was greater. Since the speaking portion of the communications task always occurred after the monitoring task, high tracking errors occurred during the speaking event when the subject was regaining tracking control. Figure 8 shows the typical change in RMS tracking error over a communication.

The standard deviation of the tracking error or error variability exhibited only very minor changes over the typical communication sequence as Figure 8 illustrates. An ANOVA on this criterion confirmed the observation that the different task types (listening, monitoring, or speaking) did not affect the tracking error variability. Subject differences and the tracking level strongly affected this error variability, as one would expect. However, it was also found through this ANOVA, that this error variability was significantly affected by the communication transaction length. With more than two connected transactions, this error variability increased. However, there was virtually no difference in tracking error variability with four or six sequential message transactions during the discrete task. A regression analysis on the tracking error variability revealed that only about 17% of the variability of this criterion could be accounted for by experimental factors and the two factors of tracking level and subject differences accounted for about two-thirds of the total.

It can be stated that there was a small effect of most of the experimental variables on the various criteria examined in this study. Table 24 shows the intercorrelation coefficients obtained between independent and dependent variables and between the mean performance time

TABLE 23. ANOVA FOR AVERAGE TRACKING ERROR

Source	MSE	df	F	Significance Level
Info Content (I)	729.8	1	.0047	-
Loading Percentage (L)	48416	2	.31	-
Central Display Percentage (ION)	42258	1	.27	-
Tracking Level (T)	1031325	1	75.6	.001
Task Type (TP)	227170	2	44.6	.001
CT Length (CT)	6590	2	3.0	.1
ION x TP	16323	2	3.2	.1
T x TP	10726	2	5.5	.025
Subjects	156210	12	-	-
Replicate	6067	24	-	-

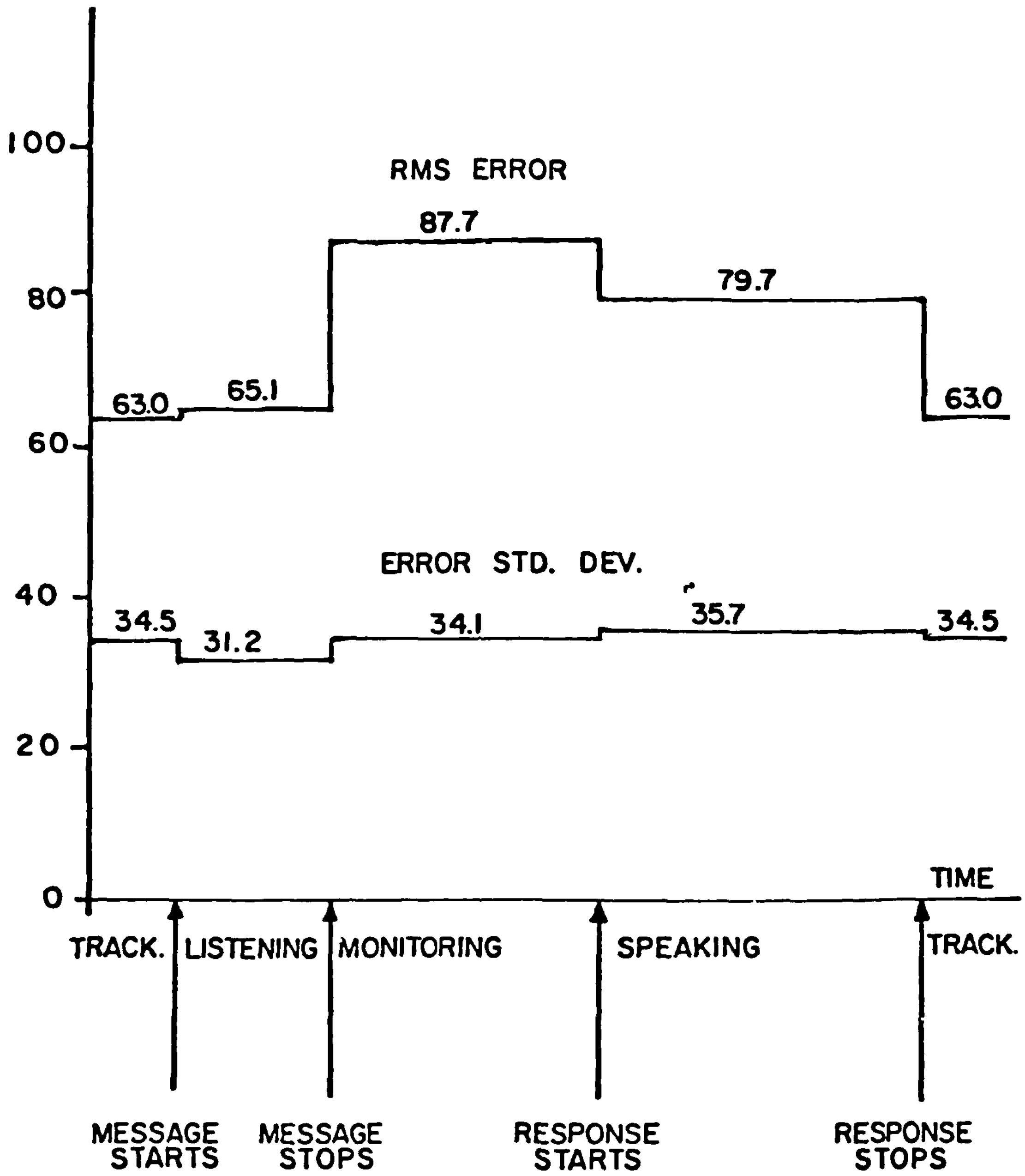

FIGURE 8. TRACKING ERROR STATISTICS DURING COMMUNICATION

TABLE 24. INTERCORRELATIONS OF INDEPENDENT AND DEPENDENT VARIABLES

Independent variable						
Information content	.09**	.09	-.15**	.02	-.01	-.01
Loading percentage	-.06*	-.03	-.03	-.06	-.08**	-.05**
Information rate	-.04	.04	-.13**	-.04	-.08**	-.06**
Tracking level	-.01	.01	-.02	.10	.28**	.22**
Number of messages within a CT	.06*	-.08	-.03	-.01	.02	.02
Visual angle	.09**	-	-.03	-	.17**	.18**
Task type	-	-	-	-	.12**	.06**
Replicate	-.06*	.02	.04	.07	-.00	.01
Event duration Monitoring	-	-	-	-	.07**	.26**
Speaking	-	-	-	-	.01	.13**
Message position in CT						
Monitoring	-.03	-	-	-	.03	.03
Speaking	-	-	-.03	-	.00	.02
Dependent variable	Reaction Time	Reaction Time Std. Dev.	Response Time	Response Time Std. Dev.	RMS Tracking Error	Tracking error Std. Dev.

* Significant for $p \leq .05$

** Significant for $p \leq .01$

and the tracking error average and standard deviation. Except for the most obvious effects, these coefficients are generally low.

Summary of Phase VII

A primary finding in this study on the oral communication task without a mediational task component is that the required time needed to perform it is nearly constant except for individual differences. The performance time variability is also nearly constant except for individual differences. Two general features tend to improve these statistics of performance time (i.e. reducing them) and those features are the number of quick communications in a row and the percentage of time spent in communications. Both of these features appear to affect the state of arousal in the human operator so that the communications are shorter and more consistent in time duration.

Tracking performance was found to be affected by some elements of the communication task during the monitoring (reaction phase) and speaking (response phase) of the oral communication. While the visual angle of the display to be monitored had only a modest effect on the time statistics, this variable had a more substantial interference effect on the tracking task. This finding corresponds to Wickens' (1979) concepts on dual tasking.

PHASE VIII
THE PERCEPTUAL - MEDIATIONAL WORKLOAD STUDY

Many of the features of the Phase VIII study followed similarly to those discussed in the Phase VII section. Therefore, the discussions below focus briefly on important differences between these two studies. Details on this study are given by Ings (1980).

Perceptual-Mediational Tasks

The principal distinction of the perceptual-mediational study from the communications study was the tasks examined. Four different classes of tasks were investigated in Phase VIII, as denoted in the discussion on Phase IV, consisting of: 1) Identifying-verifying tasks, 2) Itemizing-categorizing tasks, 3) Coding tasks, and 4) Calculating-computing tasks. All of these task families are present in pilot navigation and other operational pilot tasks. While there is no universally accepted metric which describes the degree of difficulty associated with each of these task families, paradigms were constructed which would be accepted as ordinal measures of increased complexity. In the identifying-verifying family of tasks the complexity rises with such necessary features as the number of elements and the amount of Boolean logic required. Accordingly, three levels of complexity were established in this perceptual-mediational task family. These three levels of identifying and verifying were devised by strings of alpha-numerical digits and Boolean operators where the length of the string and the number of operators specified the complexity level. Strings of 3 or 5 digits and 0 or 1 operators were used so that the least complex case consisted of 3 digits and no operators and the most complex case consisted of 5 digits and an operator. The calculating-computing task family was similarly defined using numerical digit strings and various numbers of additive and/or multiplicative operators. Task complexities were varied from the low level of three digits and two additive operators to an intermediate level of three digits and both additive and multiplicative operators or four digits with strictly additive operators up to four digits with mixed operators. Coding tasks required the subject to translate two numerical digits into a color coded response based on correspondence rules. These rules consisted of a red or green response respectively for 10-14 and 15-19 and a blue or yellow response respectively for 20 to 24 and 25 to 29. The three levels of complexity were varied by increasing the range of digits from 10 to 19, 10 to 24, and 10 to 29. Itemizing and categorizing tasks in this study consisted of displaying strings of digits where the subject was required to count the number of each different digit-type in the string. In this task family the complexity level was varied by the length of the digit string and the number of digit types. Strings of 5 to 7 digits were used with 2 or 3 digit types. Accordingly, all four task families were presented during this Phase VIII study with two levels of task complexity. Figure 9 illustrates these task families and the levels employed in this phase of the study.

Experimental Procedure

The experimental procedure of this study was simplified greatly from the one employed in Phase VII. Experimental sessions followed the initial familiarization session where the tracking alone task was presented. A representative set of the discrete perceptual-mediational tasks was presented alone and finally the combined tasks were presented. There were three experimental sessions consisting of no tracking, low level tracking, and high level tracking, presented in a balanced random order for twelve subjects (two subjects per sequence). Six different sequences of the discrete task were devised to cover the full spectrum of the discrete perceptual-mediational tasks, using different identifying, coding, computing, and categorizing problems, and complexity levels in a near random sequence. The time between successive discrete tasks was a constant plus a uniform random variable which always assured at least 5 seconds between discrete tasks. A random assignment of the discrete task sequence was made with the constraint that each subject received a different sequence on each session, including the two learning sessions. Each experimental session lasted about 8 to 10 minutes to cover some initial tracking without the discrete tasks and to provide a replication of the discrete tasks.

Each discrete task contained elemental component tasks where: 1) The problem was presented, 2) The solution finding element, and 3) The solutions response element. Beginning and ending times of the oral problem presentations were marked by the voice identifying device (VOX) connected to a tape recorder. Alpha-numerical messages were then displayed on the CRT beside the tracking task display to give the subject specific perceptual-mediational task information needed for identifying, coding, computing, or categorizing. Since the perceptual-mediational task could not be started until this alpha-numerical display was presented, the start of this display presentation was time-marked by the computer. There was also a time mark made for the beginning and end of the subjects' response and an experimenter's notation as to whether the response was correct or incorrect. At the experimenters' response, the computer program randomly selected a time interval before the presentation start of the next sequential problem. This procedure continued over each experimental session. Data collected during these sessions were logged onto the computer disc-pack and at the end of the sessions these data were sent to the computer center and stored in an

MEDIATIONAL TASKS

ARITHMETIC

EASIER 3 + 8 + 7 =

HARDER (6 × 8) + 3 =

IDENTIFICATION

HOW MANY LETTER TYPES ?

EASIER OCOOGGCCO

HARDER OCOQGCDOO

CODING

EASIER ODD NUMBER PRESENT ? (RED)

OTHERWISE ? (GREEN) 2638

HARDER IF G AFTER K, THEN GOLD

OTHERWISE BLACK CDFL

VERIFICATION

EASIER T/F DISPLAY CONTAINS 2 Es? EFRE

HARDER T/F DISPLAY CONTAINS 2 Os OR 2 Ds?

DODQ

FIGURE 9. ILLUSTRATIONS OF MEDIATIONAL TASKS

experimental data bank. These data included the modified tracking error statistics during each component task and those between successive discrete tasks.

Analysis

The time between the start of the discrete mediational task and the start of the subjects' response was called the latency (or reaction) time and an ANOVA was made on this criterion. Results from this ANOVA showed that the type of mediational task (coding, verifying, arithmetic, and identifying) and the degree of difficulty both were statistically significant at or above the 99% level. Since there were repetitions in each type of task and level over the experimental sessions but with different discrete tasks, the repetition of the task was also found to be significant above the 99% level. A regression analysis verified this ANOVA result. Table 25 gives a summary of the regression results where about 52% of the reaction (latency) time variance was explained by the task type and difficulty and only about 2% was explained by individual subject differences.

TABLE 25. REGRESSION RESULTS ON THE REACTION TIME DURING DISCRETE-MEDIATIONAL TASKS

Variable	Coefficient	R-Square	R-Square Change
Task Type	23.12	.341	.341
Task Difficulty	35.82	.523	.183
Subjects	2.32	.543	.020
Repetition	1.95	.547	.0001
Others	---	.548	.001
Constant	-53.45		

A similar analysis was performed on the verbal response time (i.e. the time interval from the beginning to the end of the response). Only the task type and difficulty obtained statistical significance at the 95% level in the ANOVA. However, time differences in the responses were relatively small; reflecting primarily individual differences and the type of response required by the nature of the task type.

The prompting time for the mediational tasks was computer paced and highly constant. Human reaction and response time varied and Figure 10 describes the mean time of these three sequential time periods for each of the four types of mediational tasks. This figure shows that the reaction time accounts for the lion's share of the mediational task time.

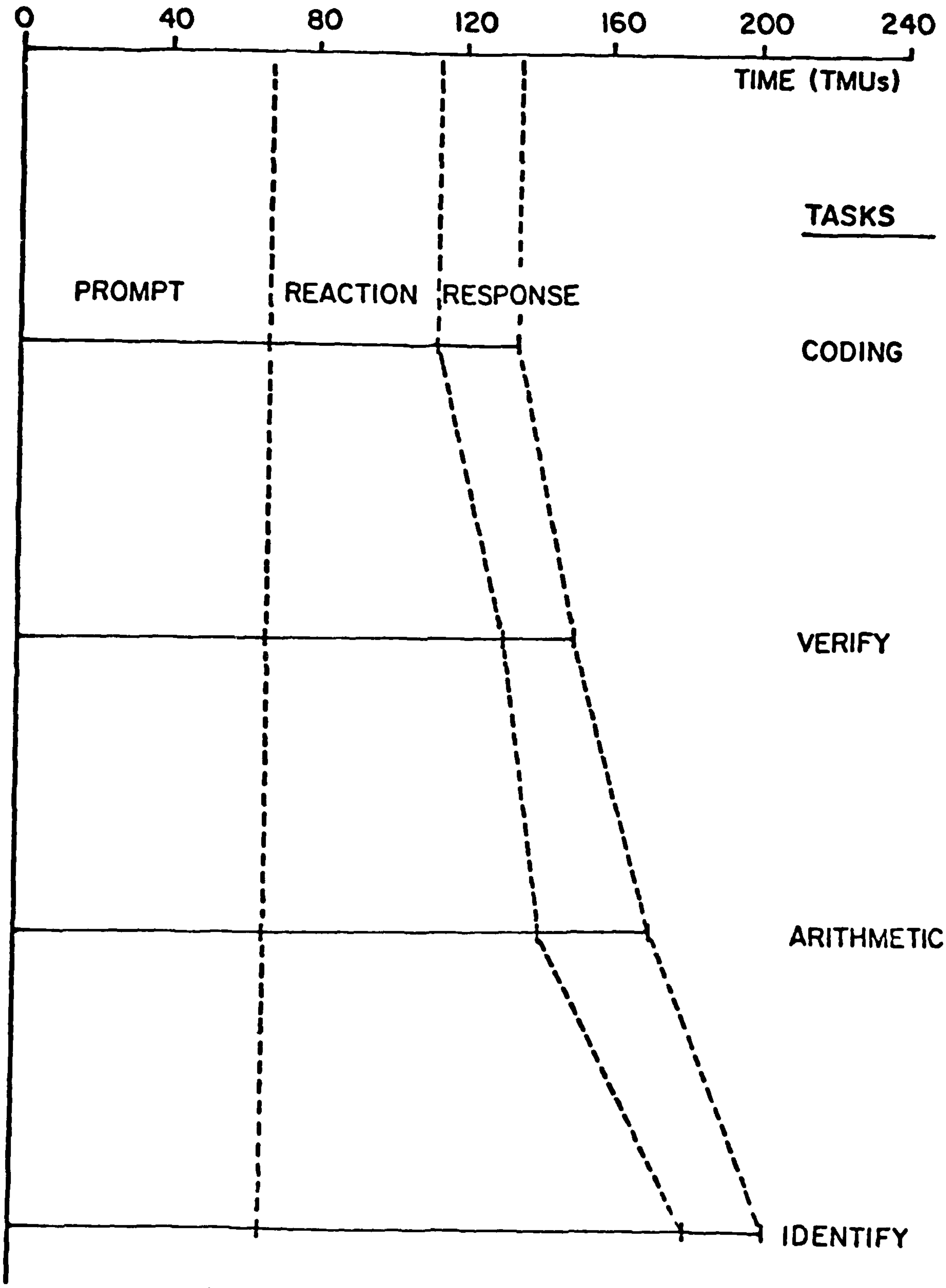

FIGURE 10. PHASE TIME EFFECTS OF MEDIATIONAL TASK TYPES

An ANOVA was performed on the RMS tracking errors and error standard deviation over the four phases of the tasking (i.e. tracking only, prompt, reaction, and response phases). Table 26 gives a summary of the results. These results show that the type of mediational task affected both the average tracking error and the variability of this error but that these criteria were affected differently by: 1) Different subjects, 2) Different tracking levels, and 3) Different phases of this dual tasking (i.e. tracking only, prompt phase, latency, and response phase).

TABLE 26. ANALYSIS OF VARIANCE SUMMARY OF MEDIATIONAL TASKS OVER ALL PHASES

Criterion	Sources of Variance			df	Significance Level
	Task	Type		3	.995
Average	"	"	x Phase	6	.975
RMS					
Tracking	"	"	x Subject	24	.975
Error	"	"	x Tracking Level	3	.950
Std. Dev.	Task	Type		3	.995
of	"	"	x Phase	6	.900
Tracking	"	"	X Subject	24	.995
Error	"	"	x Tracking Level	3	.950

Since the different phases of the mediational tasks affected the time and tracking criteria, analyses were performed for the data within the various phases. Table 27 shows the ANOVA summary for the reaction phase on the time and tracking error statistics. With all these criteria, the task type and difficulty were statistically significant. The tracking level was also significant for the tracking criteria but not the performance time criterion. A regression analysis on the reaction phase is summarized in Table 28 which shows the strong effect of the task type on all criteria. These regression data during the reaction phase also show that subject differences play a minor role in describing the performance time (2%) but a large effect on the average RMS in tracking (33%) and the error variability (12%).

An ANOVA was also performed on these criteria during the response phase and the results are summarized in Table 29. The task type has significant effects on the performance time and RMS error criteria but only an interaction effect on the tracking error variability was significant. It is of note, however, that the tracking level strongly affected the tracking error statistics but not the performance time.

Tracking error statistics tend to behave over this dual tasking situation as shown in Figure 11. A small increase in the average RMS occurs during the prompt phase while the error variability drops. During the reaction phase when the subject was performing the mediational task, the RMS increases markedly and the error variability increases slightly. When the subject gives the verbal answer to the problem, the RMS rises further but the error variance drops. Since the problem data for these tasks are all on the central display, there is little effect due to eye motions away from and back toward the tracking task during the reaction and response phases. Therefore, the imposed mediational task appears to strongly affect the tracking performance until the mediational task is almost fully executed.

Each of the mediational tasks was presented at two levels of difficulty which the above analysis described as statistically significant for all criteria. Figure 12 shows the magnitudes of the harder and easier level of these mediational tasks on these criteria. These types of mediational tasks were ordered in this figure for an increasing average reaction time. It is of note that the mean tracking error statistics do not follow the same trend. Our belief is that the reason for this difference is that different amounts of visual (sensory) resources are required in these mediational tasks separately from the cognitive resources. Where both sensory and cognitive processes require time, the use of more sensory resources interferes more with the tracking task.

<u>Summary of Phase VIII</u>

Several noteworthy conclusions resulted from this study:

1. Different classes of mediational tasks require different performance time which is moderated by the degree of task difficulty.
2. Different classes of mediational tasks and their degree of difficulty interfere with a dual tracking task. However, the varying of tracking task difficulty was not observed to substantially interfere with the time requirements for the mediational tasks.
3. The RMS tracking error tends to increase over the duration of the discrete mediational task while the tracking error variability does not.

These conclusions show that a mediational task can be described in a synthetic data system because of the distinct time requirement changes with the different mediational tasks. Also the lack of interference by variations in the dual tracking task on the mediational task time requirements shows that the SDS will not require the parameter estimation of modifiers for the mediational task.

Although there appeared to be an implicit priority shift from the tracking task to the discrete mediational tasks, this feature cannot be verified on the basis of data obtained here. Further investigation is needed to ascertain the priority assessment and its effects on dual tasking

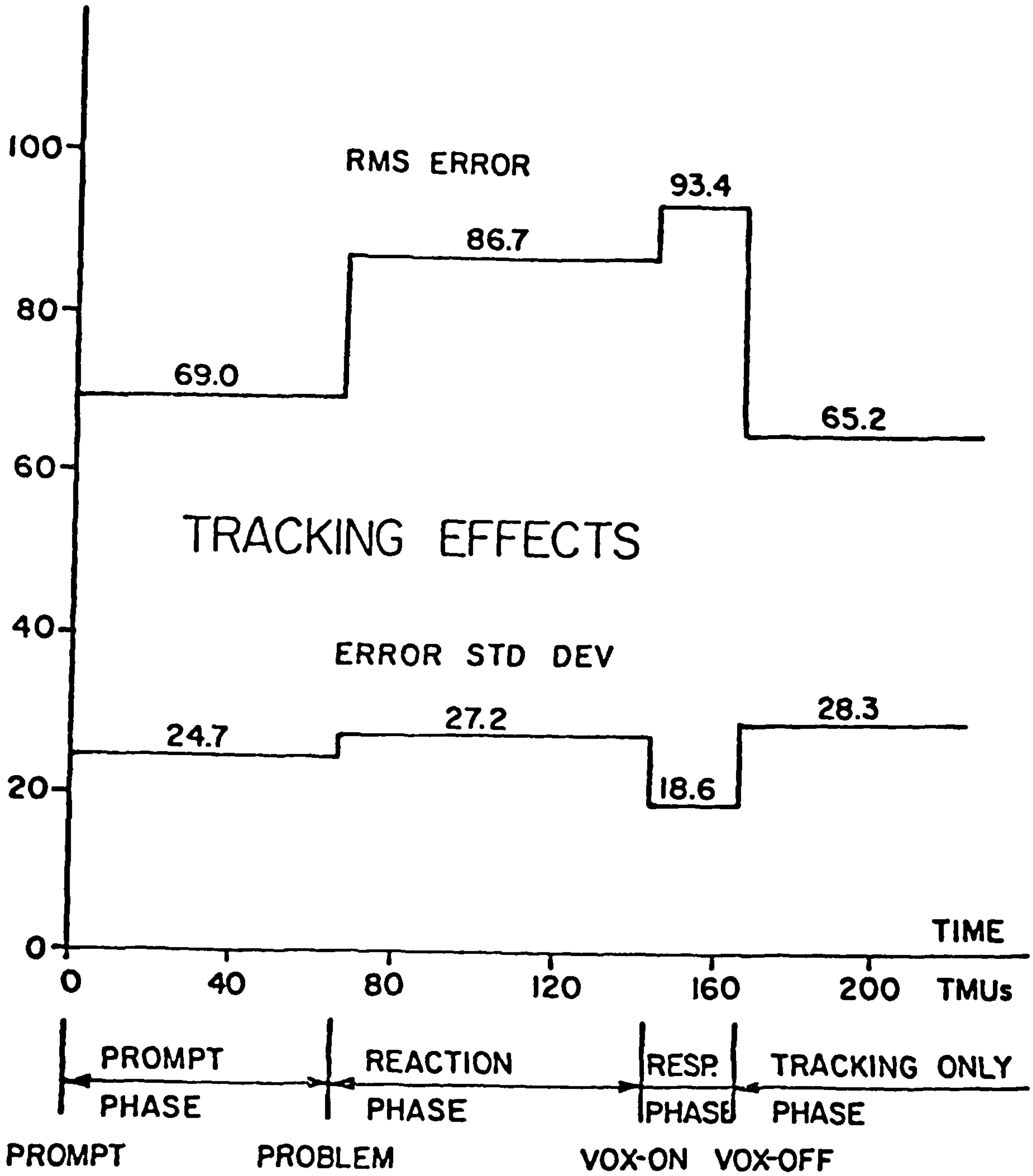

FIGURE 11. TRACKING ERROR STATISTICS DURING MEDIATIONAL TASK STUDY

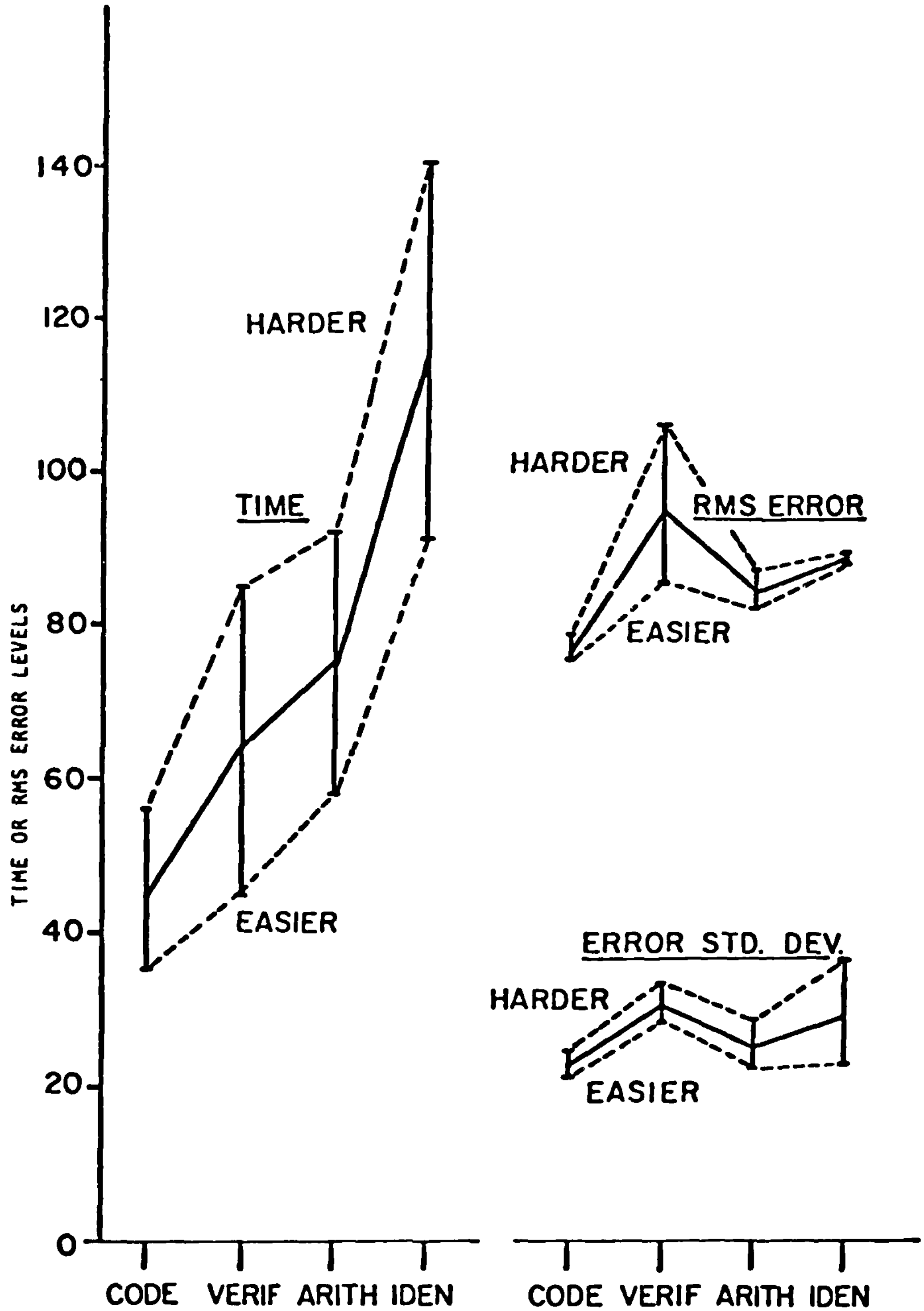

FIGURE 12. TASK AND DIFFICULTY EFFECTS ON REACTION TIME AND TRACKING ERROR

TABLE 27. ANOVA SUMMARY OF REACTION TIME DURING MEDIATIONAL TASKS

REACTION PHASE

SIGN. ANOVA VAR:	SIGN. LEVEL
TIME	
TASK TYPE	.995
TASK DIFFICULTY	.995
PROBLEM SEQUENCE	.995
TRACKING RMS ERROR	
TASK TYPE	.975
TRACKING LEVEL	.990
TASK DIFFICULTY	.950
TRACKING ERROR STD. DEV.	
TASK TYPE	.990
TRACKING LEVEL	.995
TASK DIFFICULTY	.995

TABLE 28. REGRESSION SUMMARY OF RESPONSE TIME DURING MEDIATIONAL TASKS

REACTION PHASE

REGRESSION VARIABLES	R-SQUARED	CHANGE
TIME		
TASK TYPE	.341	
TASK DIFFICULTY	.523	.183
SUBJECTS	.543	.020
TRACKING RMS ERROR		
SUBJECTS	.333	
TRACKING LEVEL	.387	.054
TASK TYPE	.399	.012
TRACKING ERROR STD. DEV.		
SUBJECTS	.115	
TASK DIFF. x TRKING LEVEL	.229	.114
TASK TYPE	.256	.026

TABLE 29. ANOVA SUMMARY OF RESPONSE TIME DURING MEDIATIONAL TASKS

RESPONSE PHASE

SIGN. ANOVA VAR:	SIGN LEVEL
RESPONSE TIME	
TASK TYPE	.95
TASK DIFFICULTY	.95
TRACKING RMS ERROR	
TRACKING LEVEL	.99
TASK TYPE	.995
TASK DIFFICULTY	.95
TASK TYPE x TRKING LEVEL	.975
TASK DIFF. x TRKING LEVEL	.99
TRACKING ERROR STD. DEV.	
TRACKING LEVEL	.995
TASK TYPE x TRKING LEVEL	.90

time statistics.

PHASE IX
EXAMINATION OF THE TRACKING TASK

This study phase was not originally anticipated as part of the workload project. However, the experimental results in the earlier phases indicated a need to further investigate the tracking distribution and trends within an individual when no additional discrete tasks were required. Since most of the statistical tests used in earlier phases of this project carry the assumption of a Gaussian (normal) distribution, one of the objectives of this study phase was to examine tracking data performance measurements to see if this assumption was violated. It should be stated that the analyses of variance (ANOVAs) used in the earlier study phases are not sensitive to minor variations in this assumption. The reason for the trend examination was to see if learning and/or fatigue effects were involved in the tracking performance data.

Two measures of tracking performance were used in these study phases. The traditional measure of tracking performance consisted of the root-mean-square (RMS) of tracking error. That is, the horizontal and vertical errors at a point in time were squared to remove negative versus positive differences, summed to give a squared radius of error measurement, and then the square-root was taken for a radius measurement. This radius measurement denotes a circular error around the cross-hairs of the scope center within which the target was kept but without regard to the quadrant of error. Actual units of measurement are arbitrary as this measurement was used for relative rather than absolute assessment. In this project, these RMS measurements were taken every 108 milliseconds. Between two events, such as the start and stop of a given activity, the RMS measurements were recorded 555.56 times each minute or 33,333 times per hour. Three times as many measurements were taken but the analogue-to-digital converter had a natural capacitance effect which could be overcome by taking 3 readings in a row at a sample time and using only the last measurement. Regardless of the actual time interval between the events, statistics were computed for the RMS measurements over the time interval. The arithmetic mean of these measurements is reported as the RMS as it denotes the average tracking error during the activity. Also the square root of the RMS variance was computed and reported as the error standard deviation. This second statistic describes the variability of the error over the activity time interval.

In the Phase IX study the time intervals were arbitrarly set at two 10-minutes and two 5-minutes time intervals after a familiarization session. Data within each time interval were examined for trends. Exponential learning curves were tested (Pegals 1969 and Buck, Tanchoco, and Sweet 1976). Also, the within-session data were examined for a frequency distribution. In this frequency distribution examination the first four central moments were found (i.e. mean, variance, skewness, and kurtosis). Although a Gaussian (normal) distribution has an independent mean and variance, the skewness and kurtosis measurements of this distribution are strictly determined as zero skewness and three degrees of kurtosis or mesokurtic. Accordingly, blocks of tracking performance measurements can be tested for deviation from normality.

Data Analysis

Data taken in this phase showed no consistent nor significant trend within the sequential time blocks. Learning curve tests indicated a sequential cyclic effect which was likely due to the cyclic nature of the forcing function used in the tracking task. Since this task was compensatory tracking rather than pursuit tracking and it had a long two-dimensional cycle (about 15 minutes in slow tracking), pattern recognition of the nature of the forcing function is highly unlikely. Hence, it can be concluded that the tracking task alone has little or very subtle learning, boredom, or fatigue effects within a session.

Table 30 shows the frequency distribution analysis data over the four sequential sessions within this phase IX study. If x_i is the RMS measurement at a sample observation, then the four statistics for n observations of these data are:

$$\bar{x} = m_1 = \frac{1}{n} \Sigma x_i$$

$$m_2 = \frac{1}{n} \Sigma x_i^2 - \bar{x}^2$$

$$m_3 = \frac{1}{n} \Sigma x_i^3 - \frac{3}{n} \bar{x} \Sigma x_i^2 + 2 \bar{x}^3$$

$$m_4 = \frac{1}{n} \Sigma x_i^4 - \frac{4}{n} \bar{x} \Sigma x_i^3 + \frac{6}{n} \Sigma x_i^2 - 3 \bar{x}^4$$

Only the first two moments of these data are shown along with the coefficient of variation or $\sqrt{m_2}/\bar{x}$. Higher statistical moments shown in Table 30 are given in dimensionless form. These skewness and kurtosis measurements are given respectively as:

$$GM_1 = m_3/\sqrt{m_2^3} \quad \text{and} \quad GM_2 = m_4/m_2^2$$

With a Gaussian (normal) distribution these dimensionless measurements should be respectively zero and three. A positive or negative GM_1 denotes the direction of the predominant tails of the distribution. Values of GM_2 greater than 3 denotes a leptokurtic or more-peaked distribution than the Gaussian for the observed mean and variance. Conversely, lower values denote flatter or more platykurtic distribution. These statistics of tracking performance are shown both for the RMS measurements directly or the error standard deviation; each over 28 observations so that 99 subdata blocks were used for each session. Also shown in Table 30 are the average values of these statistics over the four sessions. These average values provide a basis for trend evaluation over the sessions.

It can be seen from these data that no consistent between-session trend tends to exist. These data also show that:

1) Means of the RMS data do not differ essentially between the low and high level of tracking but the means tend to be more consistent at the higher tracking level.
2) The average RMS data have extremely little skewness but the

TABLE 30. FREQUENCY DISTRIBUTIONS OF TRACKING PERFORMANCE ONLY DURING PHASE IX STUDIES

		LOW TRACKING LEVEL					HIGH TRACKING LEVEL				
		$\bar{x}$	M_2	$M_2/\bar{x}$	GM1	GM2	$\bar{x}$	M_2	$M_2/\bar{x}$	GM1	GM2
10 min	RMS	160	472	0.14	0.04	2.71	163	156	0.08	0.09	2.64
	ESD	20.3	40	0.31	0.74	3.37	12.1	13	0.30	3.20	11.22
10 min	HMS	170	311	0.10	-0.01	2.60	170	176	0.08	0.07	2.65
	ESD	20.5	50	0.35	0.71	3.22	12.5	24	0.39	0.85	3.81
5 min	RMS	174	377	0.11	0.01	2.21	163	1373	0.23	-1.15	3.06
	ESD	19.6	39.4	0.32	0.43	2.64	11.9	18.5	0.36	0.47	2.52
5 min	RMS	191	326	0.09	-0.05	2.45	193	153	0.06	-0.19	2.27
	ESD	19.7	48.6	0.35	0.68	2.82	11.0	11.1	0.30	0.86	3.62
Ave.	RMS	174	372	0.11	-0.01	2.49	172	162*	.11	-0.05	2.66
	ESD	20.0	44.5	0.33	0.64	3.01	11.9	16.7	.34	1.35	3.32**

* Outlier value removed; without its removal the average is 465

** Outlier value removed; without its removal the average is 5.29

distribution is slightly platykurtic at both the low and high level of tracking.

3) Error standard deviation means and variances tend to be higher at the low tracking level but the coefficient of variation is about the same.

4) There appears to be positive skewness in the error standard deviation data and more so with the faster tracking level but the data are mesokurtic (i.e. of about the same peakedness as a normal distribution).

Since these data do not appear to differ significantly from a Gaussian (normal) distribution, the small violations from normality are not expected to have influenced the ANOVA results reported earlier. These results also suggest that the tracking task is readily learned when done alone. The trend effects shown in earlier phases on the tracking performance are expected to be a result of learning in the dual task situation rather than tracking alone.

It was our earlier intention to identify variables of the tracking task which affected the operators' tracking performance. In this way, a measure of tracking task difficulty could be obtained to adjust for real-time changes imposed by this task. However, that quest proved to be extremely difficult. With the project time overrun and numerous technical problems yet to be solved, that quest was temporarily abandoned.

PROJECT SUMMARY AND CONCLUDING REMARKS

The purpose of this project was to review possible engineering techniques with reference to workload assessment for USAF pilots, identify the most promising technique, and to assess the feasibility of the identified technique. Numerous concepts and techniques have been reviewed in early phases of this project and the synthetic (or predetermined) time system seems to best meet USAF operational needs. The basis of this recommended approach is that the personnel have sufficient time to perform the tasks to be done within the work time frames. A synthetic time system is merely one which predicts the time requirements for identified tasks with known parameter values. However, the question of feasibility with USAF pilot tasks required considerable investigative effort because there is no existing synthetic system which covers many of the tasks performed by pilots.

Three experimental efforts were undertaken to test the feasibility of a Synthetic Data System (SDS) for USAF pilots with regards to task families on: 1) Switch activations, 2) Oral communications, and 3) Mediational activities within communications. The first and last of these three studies identified variables of these task families which described a substantial amount of the time variation in the task performance time. In the second study it was found that communications of the same length had time variations which were principally described by individual differences. Since current synthetic time systems specify the predictive time requirements for a standard operator where adjustments have to be made for individual operators, the findings of these three studies are consistent with this situation. The three task families which comprise a large percentage of pilot tasks appear to have predictive variables of performance time statistics or constancies within operators. Therefore many necessary conditions of an SDS are met and feasibility is clearly indicated.

A fourth study was undertaken to explore potential applications of a developed SDS for an approximate assessment of projected benefits. Some of the applications cited were:

1. Testing time-line scenarios for work overloads,
2. Checking procedures for population effects,
3. Modifying equipment layouts in conjunction with the current HECAD system,
4. Improving the mathematical models of supervisory pilot roles such as with the Optimal Decision Model now under investigation, and
5. Simulation developments and improvements such as the current SAINT which is in USAF use.

Part of this phase consisted of a demonstration simulation of pilot dual tasking using the SLAM simulation language and artificial performance time formulae. SDS could be used in lieu of these artificial time formulae and other information from the development of SDS (e.g. distribution information beyond the statistics) would serve other modeling needs. Also this demonstration illustrates how statistical analysis can be used for sensitivity testing for a dynamic form of workload assessment.

Finally, a small empirical test was made on the tracking task to assess the adequacy of meeting assumptions of the analysis and learning. Results from this test indicated that normality assumptions of the ANOVAs and regression analyses were well within acceptable levels for these analyses. No distinctive learning effects could be quantatively assessed.

This initial project on engineering approaches to estimating pilot workloads has met the intended purpose of identifying an approach and assessing the feasibility of that approach. Accordingly, further development of an SDS for USAF needs is recommended. Further testing needs to be performed on task mixtures, priority strategies, task modularization, interference metrics of dual tasks, task interruption effects, and procedures for data collection.

REFERENCES

Aume, N.M., 1963, "The Effect of Reach Distance on Reach Time in a Button Pushing Task", AMRL Memo. P.-58, Systems Research Branch, Human Engineering Division, Behavioral Sciences Lab., Aerospace Medical Division, Wright-Patterson Air Force Base, Ohio.

Barnes, R.M., 1937, Motion and Time Study: Design and Measurement of Work, John Wiley & Sons, Inc., New York.

Berliner, C., Angell, D. and Shearer, J.W., 1967, in Simulated Environments", *Symposium and* Workshop on the Quantifications of Human Performance, Albuquerque, New Mexico.

Brennan, C.W., 1963, Wage Administration, R.D. Irwin, Inc., Homewood, Illinois.

Broadbent, D.E., 1971, Decision and Stress, Academic Press, New York.

Buck, J.R., Ings, D.M. and Lehto, M.R., 1980, "Predetermined Time Statistics of Discrete Tasks During Dual Tasking", Intl. Ergonomics Assoc./Nordic Ergonomics Soc. Conference Proc., Oslo, Norway, August.

Buck, J.R. and Maltas, K.L., 1979, "Simulation of Industrial Man/Machine Systems", Ergonomics, V. 22, N. 7, pp. 785-797.

Buck, J.R. and Payne, D.K., 1979, "Synthetic Switching Time Statistics", Proc. of the Human Factors Society, 23rd Annual Meeting, Boston, MA, October.

Buck, J.R., Deisenroth, M.P. and Alford, E.C., 1978, "Man-in-the-Loop Simulation", Simulation, pp. 137-144.

Buck, J.R. and Hancock, W.M., 1978, "Manual Optimization of Ill-Structured Problems", International J. of Man-Machine Studies, V. 10 pp. 95-111.

Buck, J.R., Tanchoco, J.M.A., and Sweet, A.L., 1976, "Parameter Estimation for Discrete Exponential Learning Curves", AIIE Trans., V. 8, No. 2, pp. 184-194.

Buck, J.R. and Tanchoco, J.M.A., 1974, "Sequential Bayesian Work Sampling", AIIE Trans., V. 6, N. 4, pp. 318-326.

Carbonell, J., 1966, "A Queueing Model of Many-Instrument Visual Sampling", IEEE Trans. on Human Factors in Electronics, V. HFE-7, pp. 157-164.

Christensen, J.M. and Mills, R.G., 1967, "What Operators Do in Complex Systems", Human Factors, pp. 328-340.

Christensen, J.M., 1950, "A Sampling Technique for Use in Activity Analysis", Personnel Psychology, V. 3, pp. 361-368.

Crawford, B.M., Pearson, W.H., and Hoffman, M.S., 1978, Multipurpose Digital Switching and Flight Control Workload, AMRL-TR-78-43, Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio.

Fitts, P.M., Peterson, J.R. and Wolpe, G., 1963, "Cognitive Aspects of Information Processing: II. Adjustments to Stimulus Redundancy", J. Expt. Psychology, V. 65, pp. 507 ff.

Fitts, P.M., Jones, R.E. and Milton, J.L., 1950, "Eye Movements of Aircraft Pilots During Instrument-Landing Approaches", Aeronautical Engineering Review, V. 9, N. 2, pp. 1-6.

Garner, W.R., 1974, The Processing of Information and Structure, Lawrence Erlbaum Press, Potomac, Maryland.

Geiselhart, R., 1978, "Crew Factors, Workload and Performance", Proceedings of the Symposium on Man-System Interface: Advances in Workload Study, Air Line Pilots Association, Washington, D.C., July 31 and August 1, pp. 102-114.

Goldman, J. and Hart, L.W., Jr., 1965, "Information Theory and Industrial Learning", The J. of Industrial Engineering, V. XVI, No. 5, pp. 306-313.

Hitt, W.D., 1961, "An Evaluation of Five Different Abstract Coding Methods - Experiment V", Human Factors, V. 3, pp. 120-130.

Hancock, W.M. and Foulke, J.A., 1966, "A Description of the Electronic Data Collector and the Methods of Its Application to Work", Research Information Paper No. 1, MTM Association.

Heising, K.W., 1954, "A Study of the Inter-Relationship of Therblig Times", Unpublished MS Thesis, Purdue University.

Helson, H., 1949, "Design of Equipment and Optimal Human Factors", American Journal of Psychology, V. LXII, No. 4, pp. 73-497.

Horves, D., 1957, "On the Relation Between the Intelligibility and Frequency of Occurrence of English Words", J. Acoustical Soc. Amer., V. 2, pp. 296-305.

Hunter, J.S., Blumenfeld, D.E. and Hsu, D.A., 1973, "Modeling Air Traffic Performance: Volume I. Message Element Analysis and Dictionaries", Federal Aviation Administration Report FAA-RD-73-147,I.

Ings, D.M., 1980, "Perceptual-Mediational Phases in Competition with a Compensatory Tracking Task", MS Project Report, School of Industrial Engineering, Purdue University, West Lafayette, IN.

Kahneman, D., 1973, Attention and Effort, Prentice-Hall, Inc., Englewood Cliffs, NJ.

Kleinman, D.L., Baron, S. and Levison, W.H., 1971, "A Control Theoretic Approach to Manned-Vehicle Systems Analysis", IEEE Trans. Automat.

Control, V. AC-16, No. 6, pp. 824-832.

Konz, S.A., 1979, Human Factors of Work, GRID Publications, Columbus, Ohio.

Konz, S.A., Dickey, G.L., McCutchan and Koe, B., 1967, "Manufacturing Assembly Instructions: Part III. Abstraction, Complexity, and Information Theory", The J. of IE, V. XVIII, No. 11, pp. 658-667.

Lehto, M.R., 1980, "An Investigation of Discrete Communication Tasks Within a Dual Task Environment", MS Thesis, School of Industrial Engineering, Purdue University, West Lafayette, IN.

Miller, G.A., Heise, G.A. and Lichten, W., 1950, "The Intelligibilty of Speech as a Function of the Context of the Text Materials", J. Exper. Psychology, V. 41, pp. 329-335.

Miller, G.A., 1956, "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information", The Psychological Review, V. 63, No. 2.

Mills, R.G., Bachert, R.F. and Hatfield, S.A., 1975, "Quantification and Prediction of Human Performance: Sequential Task Performance Reliability and Time", AMRL-TR-74-48, Aerospace Medical Research Laboratory, Wright-Patterson AFB, Ohio.

Niebel, B.W., 1967, Motion and Time Study, R.D. Irwin, Inc., Homewood, IL.

Norman, D.A. and Bobrow, D.G., 1975, "On Data-Limited and Resource-Limited Processing", Cognitive Psychology, V. 7, pp. 44-64.

Pasmooij, C.K., Opmeer, C.H.J.M. and Hyndman, B.W., 1976, "Workload in Air Traffic Control", in Monitoring Behavior and Supervisory Control, Plenum Press, New York, NY and London, England.

Payne, D.R., 1981, "On the Applicability of Synthetic Time Measurements of Switch Activation Tasks During Flying", MS Thesis, School of Industrial Engineering, Purdue University, West Lafayette, IN.

Pegels, C.C., 1969, "On Startup or Learning Curves: An Expanded View", AIIE Trans., V. 1, No. 4, pp. 300-305.

Presgrave, R., 1945, Dynamics of Time Study, McGraw-Hill Book Co., New York, 2nd Edition, Chapter 10.

Posner, M.I. and Klein, R.M., 1973, "On the Functions of Consciousness", in S. Kornblum (Ed.), Attention and Performance IV, Academic Press, New York.

Pritsker, A.A.B. and Pegdon, C.D., 1979, Introduction to Simulation and SLAM, Halstead Press & Systems Publishing Corp., West Lafayette, IN.

Raouf, A., 1972, "Information Processing Rate and Prediction of Decision Times: An Experimental Investigation", The J. of Methods-Time Measurement, V. XVIII, No. 1, pp. 10-14.

Raouf, A. and Mehra, M.L., 1974, "Experimental Investigations Related to Combined Manual and Decision Tasks", Intl. J. of Prod. Rsch. V. 12, N. 2, pp. 151-157.

Rolfe, J.M., 1973, "The Secondary Task as a Measure of Mental Load", Measurement of Man at Work (W.T. Singleton, J.G. Fox, and D. Whitfield, editors), Taylor and Francis, pp. 135-148.

Rosenstein, A.B., 1955, "The Industrial Engineering Application of Communication-Information Theory", The J. of IE, September-October, pp. 10-21.

Ross, D.K., 1960, "Information Content Analysis", ScD Thesis, Washington University, St. Louis, MO.

Sadosky, T.L., 1968, "Prediction of Cycle Time for Combined Manual and Decision Tasks", Ph.D. Dissertation, University of Michigan, Ann Arbor, MI.

Sanders, M.G., Simmons, R.R., Hofmann, M.A. and DeBonis, J.N., 1977, "Visual Workload of the Copilot/Navigator During Terrain Flight", Proc. of the Human Factors Society, 21st meeting, San Francisco, CA, pp. 262-269.

Schmidt, D.K., 1978, "A Queueing Analysis of the Air Traffic Controller's Work Load", IEEE Trans. on Systems, Man, and Cybernetics, V. SMC-8,No. 6, pp. 492-498.

Senders, J.W., 1964, "The Human Operator as a Monitor and Controller of Multi-Degree of Freedom Systems", IEEE Trans. on Human Factors in Electronics, HFE-5, n. 1, pp. 2-5.

Sheridan, T., 1978, "The Changing Role of the Pilot from Manual Controller to Computer Supervisor", Proc. of the Symposium on Man-System Interface: Advances in Workload Study, Air Line Pilots Assoc., Washington, D.C., July 31 and August 1, pp. 132-143.

Siegel, A.I. and Wolf, J.J., 1969, Man-Machine Simulation Models, John Wiley & Sons, Inc., NY.

Steffy, W., 1970, "The Use of Standard Data Systems for the Prediction of Operation Time Standards", Notes from the Advance Work Analysis Short Course, University of Michigan.

Thomas, M.U., Hancock, W.M. and Chaffin, D.B., 1974, "Performance of a Combined Manual and Decision Task with Discrete Uncertainty", Intl. J. Prod. Rsch., V. 12, No. 3, pp. 409-425.

Tippett, L.H.C., 1935, "Statistical Methods in Textile Research. Uses of the Binomial and Poisson Sampling", J. of the Textile Inst. Trans., V. 28, pp. 51-55.

Topmiller, D.A. and Aume, K.S.M., 1978, "Computer-Graphic Design for Human Performance", Proc. of the 1978 Annual Reliability and Maintainability Symposium of The Institute of Electrical and Electronics Engineers, pp.

385-388.

Topmiller, D.A. and Sharp, E.D., 1965, "Effects of Visual Fixation and Uncertainty on Control Panel Layouts", AMRL-TR-65-149, Aerospace Medical Research Laboratories, Wright-Patterson AFB, Ohio.

Walden, R.S. and Rouse, W.B., 1978, "A Queueing Model of Pilot Decisionmaking in a Multitask Flight Management Situation", IEEE Trans. of Systems, Man, and Cybernetics, V. SMC-8, N. 12, pp. 867-875.

Weiner, N. and Schade, J.P., 1965, Cybernetics of the Nervous System, Elsevier Publishing Company, Amsterdam, Netherlands.

Welford, A.T., 1968, Fundamentals of Skill, Methuen, Ltd., London, England.

Wickens, C.D., 1979, "The Structure of Attentional Resources", Attention and Performance VIII (R. Nickerson, editor), Erlbum Press, Hillsdale, NJ.

Wickens, C., 1978, "Brain Electrical Activity as an Index of Workload", Proc. of the Symposium on Man-System Interface: Advances in Workload Study, Airline Pilots Association, Washington, D.C., pp. 21-34.

Williges, R.C. and Wierwille, W.W., 1979, "Behavioral Measures of Aircrew Mental Workload", Human Factors, V. 21, No. 5, pp. 549-574.

U.S.Government Printing Office: 1982 -- 559-007/6022