Spring 1991

Operator Strategies Under Varying Conditions of Workload

Ruth Justine Arnegard
Old Dominion University

OPERATOR STRATEGIES UNDER VARYING CONDITIONS OF WORKLOAD

By

Ruth Justine Arnegard
B.S. June 1986, Virginia Polytechnic Institute and State University
M.S. May, 1988, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

INDUSTRIAL/ORGANIZATIONAL PSYCHOLOGY

OLD DOMINION UNIVERSITY
May, 1991

Approved by

Glynn D. Coates, Ph.D. (Director)

J. Raymond Comstock, Jr., Ph.D.

Raymond H. Kirby, Ph.D.

Peter J. Mikulka, Ph.D.

ABSTRACT

OPERATOR STRATEGIES UNDER VARYING CONDITIONS OF WORKLOAD

Ruth J. Arnegard
Old Dominion University, 1991
Director: Dr. Glynn D. Coates

The present study attempted to operationally define and measure strategic behavior in a complex multiple task environment. The Multi-Attribute Task battery was developed to simulate various aspects of flight and consisted of an auditory communication task, monitoring tasks, a tracking task, a resource management task which allowed a wide range of responding patterns, and a scheduling window which allowed operators to predict changes in workload. This battery was validated for its sensitivity to strategic behavior and baseline measures for each individual task were collected.

Twenty-four undergraduate and graduate students then performed the battery for four 64-minute sessions which took place over a period of two days. Each subject performed the task battery under four levels of workload, which were presented for equal lengths of time during all four sessions. Results indicated that in general, performance improved as a function of experience with the battery but that performance decreased as workload level increased.

The data also showed that subjects developed strategies for responding to the resource management task which allowed them to manage the high workload levels more efficiently.

This particular strategy developed over time but was also associated with errors of complacency. These results are presented along with implications for the aviation field and areas of future research.

ACKNOWLEDGEMENTS

Many individuals contributed to the work presented here through professional and personal support. I was fortunate to work with and benefit from such highly talented and energetic people.

To my advisor, Dr. Glynn Coates, I owe a deep sense of gratitude. He has influenced me as a teacher and mentor throughout my graduate career and his teachings and influence will continue to be with me as I make my way in my professional career. I have learned so much from you.

Dr. Raymond Kirby and Dr. Peter Mikulka have also made important contributions to this dissertation and my professional development. Thank you for your support and guidance.

The contributions of Dr. Ray Comstock are directly responsible for the completion of this dissertation in many ways. First, he created all of the software for the Multi-Attribute Task Battery. Additionally, he provided me with an internship opportunity through which much of this work was completed and provided support along every step of the way. Thank you.

I must also thank the entire Human Engineering Methods group from NASA-Langley Research Center in Hampton, Virginia. You have all provided support and insight throughout the course of this project. Thanks especially to Mary McManus and Scott Crews for helping me with the graphs presented here.

Melinda Montgomery, Gwen Pearson and Coleen Thornton have encouraged me through both friendship and professional example. I don't think I could have made it without you. I am honored to be your colleague.

I am also grateful to my friends and family, who, I am fortunate to add, are too numerous to list here. You were always in my heart and I felt your encouragement and pride at all times. You mean more to me than anything I could put on paper.

Special thanks go to Scott for providing love and understanding throughout this entire process. As the one who was closest to me and my struggles, I know it wasn't easy. You deserve an honorary degree for patience and perserverance. I love you.

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont.)

TABLE OF CONTENTS (Cont.)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Cont.)

INTRODUCTION

The term "strategy" has been used in many different contexts throughout the literature in areas such as cognition, military psychology, judgment and decision-making. Human factors researchers have examined strategies involved in process control (Fuld, 1987; Morris & Rouse, 1985), route selection (Hayes-Roth & Hayes-Roth, 1979), risk assessment (Svenson, 1985), as well as complex dual and multiple task situations (Damos, Smist & Bittner, 1983; Wickens, Mountford & Schreiner, 1981). Furthermore, the term "strategy" is often used in relationship to other terminology such as "cognitive model", "internal representation", "planning behavior", and "tactics". Almost always, the term "strategy" appears in the literature without an accompanying definition. Many approaches to the topic of strategies (some of which overlap) have been taken. The major approaches to this topic are summarized below.

Strategies as predictors of workload and performance.

Welford (1978) has stated that workload and performance depend on task demands, the operator's capacity, the strategies used to relate those task demands to the operator's capacity, and the operator's skill in choosing the most efficient strategy. According to Welford, individuals learn generic strategies (e.g., time awareness, spatial or positional awareness, memory encoding) through education, training or experience. These generic strategies are then adopted and synthesized for use with specific tasks on given occasions.

Though strategies are situation specific, the generic strategies are utilized in many different situations and then tailored according to all of the perceived demands of any given situation. These task-specific strategies can be modified/selected without any overt action occurring, since an individual can assess the outcomes of possible actions or inactions through his or her mental representation of the situation at hand.

According to Welford (1978), skilled performance is a direct function of the efficiency of the strategy chosen in reaching the operator's goals. The most efficient strategy is defined as minimizing the discrepancy between aim and achievement at the least cost to the operator and system in terms of time and effort.

<u>Strategies as one step or plane in the planning process</u>.

Hayes-Roth and Hayes-Roth (1979) theorize that the planning process is a multi-directional rather than top-down process. In other words, they assume that individuals can plan low-level subtasks without having previously planned that task at higher levels. Additionally, sub-plans can be partially completed and "set aside" until a later time, while other planning activity is being carried out. The overall process may appear chaotic, rather than organized in a top-down fashion.

According to this approach, independent cognitive specialists (mental resources) generate decisions during the planning processes and these decisions are recorded in a common data structure which consists of five conceptual planes. These planes are: (1) the planning plane, representing the individual's intended actions; (2) the plan-abstractions plane, indicating the desired attributes of the intended actions, without identifying the actions themselves; (3) the knowledge-base, which contains all observations and information regarding relationships which might be relevant; (4) the executive plane, which determines the allocation of cognitive resources during the entire process; and (5) the meta-plan plane, which contains decisions about how to approach the planning process.

Within this model, the plan-abstractions plane is comprised of intentions, schemes, strategies and tactics, and as such is the only plane relevant to the present review. According to Hayes-Roth and Hayes-Roth (1979), <u>intentions</u> refer to the overall goal or outcome of the situation. <u>Schemes</u> are the individual's mental representation of all possible relationships between the tasks and sub-tasks necessary to reach the intended goal.

Strategies are defined as procedural decisions and tactics are the operational decisions about how to execute the strategies behaviorally.

Strategies and time sharing (allocation of resources).

Although in some cases, one task could be completed using a number of different strategies, complex environments present multiple tasks to an individual. When this occurs, a time-sharing resource must be called upon in order to complete all of the tasks successfully. Some researchers have addressed the question as to whether or not a time-sharing ability exists (Wickens, 1984), though there is not much evidence for this skill. However, individual differences in attention-switching (Keele, Neill & DeLamos, 1977) and the extent to which the individual tasks are under automatic processing (Jennings & Chiles, 1977) have been shown to account for differences in performance both between individuals and during various task combinations.

Damos, Smist and Bittner (1983) have inferred an attention-switching ability from the pattern of responding that their subjects employed in a dual-task scenario. They theorized that the pattern of responding also reflected the subject's choice of strategy. For example, each subject's method of responding fell into one of three categories: simultaneous responding, alternating responses or massed responding.

Simultaneous responding was defined as a response to both stimuli in the dual-task situation within an arbitrarily small time interval. Alternating response strategies were said to have occurred if the subject alternatively made one response to each task within a total time interval greater than that for the simultaneous strategy. Finally, massed responding involved more than two responses made consistently to one task before switching to another. These strategies are presented in order of their effectiveness, with simultaneous responding producing the best performance.

Damos and her colleagues (1983) found that each subject consistently utilized one of these strategies, and if a new strategy was attempted, large disruptions in performance occurred. For example, performance was degraded without recovery when those subjects with a natural massed strategy tried to adopt an alternating or simultaneous response pattern. These researchers concluded that the use of these different strategies by subjects reflected fundamental individual differences in information processing.

Simultaneous responding on two tasks probably reflects that the processing of the two sets of information is automatic to some extent. This "parallel-processing" has been found to produce better performance than rapidly alternating between two tasks (Wickens, 1984; Damos et al., 1983). Even though Damos and her colleagues found individual differences in responding that appear to be "fixed" within an individual, these differences may be task-specific and either depend on the degree of automatic processing involved or result from multiple-task practice.

Wickens (1984) has stated that if cognitive resources are divided between tasks, performance must drop off on one or both unless both tasks are data-limited. In other words, performance degrades unless, after a certain point, additional resources or information are irrelevant and will not affect performance. Previous research has indicated that tasks drawing upon common resources are time-shared less efficiently with regard to performance than unrelated tasks (Wickens, Mountford & Schreiner, 1981). Wickens (1984) also notes that optimum time-sharing of mental resources does not necessarily have to be an equal (50/50) allocation. He has found that if the tasks are different, and require different resources to some extent, a 60/40 (approximately) allocation of time produces optimum performance on both tasks.

In conclusion, time-sharing seems to be composed of two aspects: attention-switching ability and automatic processing. Experience with a task leads to automatic processing and a greater ability to perform other tasks along with the automated task. An individual's ability to alternate attention between tasks determines a response strategy which may be difficult to change for a certain multi-task situation. Finally, the similarity of the tasks, or degree to which they share mental resources, is another factor in an individual's ability to simultaneously perform multiple tasks.

Strategies resulting from instructional set/payoff matrix.

One standard principle regarding reaction time theory is the speed-accuracy trade-off. It is well accepted that if an individual is instructed or chooses to emphasize speed in responding, then accuracy will degrade. The converse is true when accuracy is the priority. The decision to emphasize speed, accuracy or both can be considered a strategy that an operator chooses in attaining his or her goal (Wickens, 1984). Of course, this decision is made based upon the operator's knowledge and perceptions about the task situation, including the costs and/or payoffs of adopting either speed or accuracy as a strategy. For example, if safety is critical, then accuracy will probably have priority over speed.

One outcome of emphasizing speed over accuracy is summarized in Yellott's (1971) fast guess model. This model asserts that under high speed responding, subjects do not wait to identify the stimulus on some trials. They instead execute a random guess immediately upon signal detection, which is likely to be the response with the highest probability of occurrence. Even though responding is fast, accuracy is reduced by the number of trials on which fast guessing is employed. The decision to emphasize speed, accuracy or moderate levels of both can differ between tasks or within the same task at different time periods.

Strategies and signal detection theory.

Research in signal detection theory provides another avenue for the examination of strategies. It is well accepted that human operators can change their performance by adjusting their criteria in response to changes in payoffs and probabilities. The decision to adopt either a strict or lax response criterion is a strategy that operators use and can be either a function of instructional set or situational requirements.

Though research has shown that humans do adjust criteria in response to the above strategy, laboratory studies suggest that the criterion is not adjusted as much as it should be to attain an ideal level of performance (Wickens, 1984). There is evidence that real world situations produce the same result, though it is difficult to draw conclusions when the variables cannot be precisely controlled. Specifically, individuals are less risky than they should be if the ideal criterion placement is low and they are less conservative than necessary if the ideal criterion is high.

Green & Swets (1966) have referred to this as the "sluggish beta" and have found it to be more pronounced when the criterion is manipulated by probability of signal occurrence than by payoffs. This sluggishness may occur because operators have a tendency to overestimate the probability of very rare events and underestimate the probability of very common events (Sheridan & Ferrell, 1974).

Another determinant of response criterion placement is experience with the task. For example, Bisseret (1981) has examined the application of signal detection theory to an air traffic controller's task. In this task, a decision must be made regarding the necessity of corrective action based on the projected courses of two airplanes. He has found that controllers are more willing to detect a conflict (lower criterion) as the difficulty and uncertainty of the problem increases. Furthermore, experts have been found to be more willing to detect a conflict and to call for corrective action than novices. This could be attributed to the fact that novices are more uncertain as to how to implement corrections and are therefore more reluctant to detect conflicts.

Overall, operators have response biases in various situations which determine the strategy selected. They may be biased towards saying "yes" and detect most of the signals but incur many false alarms as well. At the other extreme, operators may be conservative and make few false alarms but miss many of the signals. Criterion placement is dependent on signal probability, experience, cost of false alarms or misses, and individual differences.

Differences in strategies as a function of experience.

It has been found that there is a difference in the kinds of strategies employed by experts and novices of a task. One result of experience is the change in response criterion placement, as discussed in the previous section. However, experience is a determinant of strategy selection in complex systems as well. As an individual becomes more familiar and skilled with a task, he or she is better able to incorporate situational changes into the planning process and anticipate the need for interventions in a system (Wiener, 1989). These adaptive mechanisms are aids to performance and are attributable to experience.

Experienced operators have a greater ability to detect or to avoid their errors of action, known as slips (Fuld, 1987). Therefore, these types of errors are more commonly found among novices of a task. Some examples of slips are: forgetting an action which has already occurred, being distracted by future or past events, incorrectly perceiving the environment in unusual or ambiguous situations, or responding based on habit rather than features of the stimulus (Morris & Rouse, 1985). However, it has been found that skilled operators can still make errors of intention (mistakes), due to biases in judgment or planning (Reason, 1983).

Experienced operators can be prone to become overconfident about the correctness of their state of knowledge. Additionally, past successes may be an inappropriately large influence on an operator's choice of strategy for action. While these processes directly contribute to continued successful decision-making, there are times when biases or overconfidence can allow the operator to overlook alternative solutions or to oversimplify the problem, leading to mistakes. Morris and Rouse (1985) have stated that the use of past success as the sole consideration in problem solving (imperfect rationality), oversimplification of the problem (bounded rationality) and "jumping to conclusions" or reluctant rationality are three causes of mistakes that experts can make.

Slovic, Fischoff, and Lichtenstein (1982) list several ways in which experts may overlook important situational attributes due to overconfidence. These are: (1) failure to consider the ways in which human errors can affect technological systems; (2) overconfidence in current scientific knowledge; (3) failure to appreciate how technological systems function as a whole; (4) slowness in detecting chronic, cumulative effects; (5) failure to anticipate human response to safety measures; and (6) failure to anticipate "common-mode failures" which simultaneously affect systems that are designed to be independent.

Overconfidence leads to a failure to perceive risk in a system which has contributed to many major catastrophes. Examples of this phenomenon are the Three Mile Island incident, the early failure of the DC-10 due to decompression of the cargo area, which destroyed vital control systems, and the Brown's Ferry, Alabama, fire at the nuclear reactor in which all emergency core cooling systems were damaged by a single fire (Slovic, Fischoff & Lichtenstein, 1982).

In conclusion, novice operators are more likely than experienced operators to make slips, errors of action or other errors based on lack of knowledge or experience. However, experienced operators can still make mistakes based on biases or overconfidence. Another result of experience or learning is that responding can become automatic rather than controlled (Schiffrin & Schneider, 1977). When information processing is automatic, it is virtually unaffected by workload and is parallel in nature. As mentioned previously, when tasks in a multiple task situation are automatic in nature, overall performance is increased (Wickens, 1984; Jennings & Chiles, 1977), though complacency or overconfidence can still be a problem.

The role of mental representation.

An operator's cognitive representation of the task situation is yet another influence on strategy selection. Specifically, the individual's frame of reference affects the cognitive processes utilized in problem solving. Thorndyke (1980) has contended that at least two representations of navigational information exist: route and survey knowledge. The first of these, route knowledge, shares certain properties with an inside-out display in that the frame of reference involved directly corresponds with the operator's view of the environment.

On the other hand, survey knowledge, similar to an outside-in map, is completely independent of the particular view the operator has of the environment. One example of the difference between these two internal maps is that an individual with route knowledge conceptualizes navigation as a series of left and right turns while one with survey knowledge represents the journey with compass directions and is thus aware of euclidean distances as well.

Subjects who follow step-by-step instructions (route knowledge) have been found to use more short term memory space than those who have survey knowledge (Berg & Sheridan, 1984). This occurs because each step in route knowledge consists of new parameters that the individual must have in working memory and must be constantly updated. Survey knowledge consists of an overall picture of the situation which is only changed if novelty is introduced. Thus, when an emergency or abnormal situation arises, those operators with survey knowledge of the task can devote more short term memory space to problem solving.

Survey knowledge has an additional advantage over route knowledge in navigational tasks where the operator finds himself or herself off course. These operators can take alternative routes to guide them back to the correct course much more readily than those operators with route knowledge (Wickens, 1984). Berg and Sheridan (1984) have also found that when following a step-by-step procedure, operators rated their workload as significantly higher than during trials on which they were given an overall representation of the system.

In their research on pilot training, Braune and Trollip (1982) have found that experienced operators develop an internal representation of their task environment that includes a wide range of sensory stimuli as well as a thorough knowledge of the system's dynamics. This allows the pilot to have certain expectancies about future events, leading to a reduction in workload.

Individual differences in spatial abilities influence the adoption of one of these mental representations. Additionally, novice operators are more likely to utilize route knowledge until they gain experience with their environment and task. However, individuals who naturally adopt route representations of their environment can be trained to develop a survey representation. Thorndyke (1980) has reported that individuals trained this way can eventually perform as well or better than those who solely utilize survey knowledge.

Some common themes.

In summarizing the approaches to the topic of strategies, many common elements appear:

(1) Strategies differ between tasks because they are dependent on the goal or desired outcome.

(2) Skilled operators use different strategies than novice operators in the same situation and both are prone to unique errors.

(3) Strategies are based upon all perceived information from both the present situation and relevant past experiences.

(4) In multi-task situations, strategy selection can involve a prioritization of goals such as considering the probability of various future situations.

(5) In all situations, outcomes of any possible action can be weighed before final strategy selection is made.

(6) Strategy selection differs from standardized procedures in that the appropriate action depends on a number of factors and options.

(7) Strategies are subject to change based on new

information or a status change.

(8) Strategies have a decision-making component (if, for one task, there is more than one way to complete it) and a time-allocation or time-sharing component (if there is more than one task that must be completed).

While these different approaches do share common themes, there is still a need for a unifying theory in this area. Until the scientific community can agree on an overall definition of the term "strategy", it is necessary to formulate a description of strategic behavior that meaningfully describes the planning, decision-making and time-sharing processes of a particular task environment. The purpose of this review is to present the need to understand strategic behavior in one complex environment, namely the flight environment. Based on the available research, the following definition of strategic behavior is proposed:

> The action (or inaction) that an operator takes in order to change the task structure, sequence of responding, or allocation of mental resources with the purpose of achieving a more manageable workload, dealing with unexpected change in the environment or achieving one's goal safely and efficiently.

The task structure can be changed by delegating one task or part of a task to another individual or to automation. The response sequence can be altered when changes in the environment require a reprioritization of the tasks. The allocation of mental resources between tasks, or time-sharing, is yet another strategic behavior which is dynamic in a complex environment such as aviation.

The Flight Environment

Of all multi-task situations, the flight environment is one of the most complex. Pilots are faced with an abundance of information from a variety of sources: cockpit displays, the outside environment, other crew members, passengers and air traffic control. As such, the operator is required to practice strategic decision-making about the sequence of behaviors and the allocation of mental resources necessary for a safe journey. Added to this is the changing workload that different stages of flight present to the pilot. Problem-solving is also required under abnormal or emergency conditions, often in addition to existing high workload demands. Furthermore, the increased role of automation affects not only the pilot's workload but also influences his or her choice of strategic behavior.

Mental Workload in Flight.

The introduction of automation to the cockpit has probably altered the pilot's workload more drastically than any other factor. Automation offers the potential for both decreasing and increasing the operator's workload. For example, workload is decreased since automated devices perform tasks previously carried out by the human operator. However, the pilot is still responsible for operating the automatic device and this increases his or her monitoring requirements. The net effect is a reduction of workload, which can lead to extreme work underload or boredom during flight.

The psychological state of underload which results from increased automation has been termed "automatic complacency" (Ternhem, 1978) and has been subjectively rated one of the highest stressors encountered by various operators. In fact, boredom is often rated as more stressful than task overload (Mackie, 1987). While an operator is in an underload state, the probability of error in response to an unexpected or emergency change in flight status is greatly increased.

It is too simplistic, however, to state that a certain level of automation is either beneficial or detrimental to aviation safety because different phases of flight present different workload environments to the crew. Therefore, during takeoff and landing of the aircraft, which involve high levels of mental workload and attention, automation takes some of the burden of high information loads from the crew. However, during the rest of the flight, expert systems virtually fly the plane, leaving the crew with a task underload situation.

Of course, many other factors contribute to aviation safety and these are too numerous to discuss in the present review. Furthermore, the concept of mental workload is very elusive and has been defined in numerous ways. However, for the purposes of this review, a definition of pilot workload formulated by Hart (1987) is referred to. She asserts that pilot workload is defined as "the cost incurred by the human operators of complex airborne systems in accomplishing the operational requirements imposed on them." She further states that this cost reflects a combination of demands such as mission requirements, the amount and clarity of information and equipment provided, the flight environment, pilots' skills and experience, the strategies they adopt, the effort they exert and their emotional responses to the situation.

Strategies and mental workload

Overall, the research suggests an interaction between strategic behavior and mental workload. Hart (1987) suggests that strategy selection can change workload, with efficient strategies serving to lower mental workload. The link between strategies and mental workload also exists in that individuals may choose to perform a task in a qualitatively different way as workload (task difficulty) increases or decreases. Operators can decide to change performance along a variety of dimensions which have been summarized above, such as hits and false alarms, speed and accuracy, allocation of resources, or attention-switching strategies.

If mental workload is a catalyst of error in decision making, particularly at the two extremes (low and high mental workload), poor strategy selection should be more frequent during these two levels of workload. This should be the case particularly when environmental conditions are conducive to error. Examples of this include: the existence of an abnormal situation, salient cues which suggest an inappropriate solution, or the need for simultaneous consideration of more than three variables. Additionally, individual differences in attention switching, response strategy, and experience would moderate the effects of workload on strategy selection.

Strategic Behavior in Flight

Although an abundance of literature exists concerning mental workload in flight, decision-making and judgment in aviation systems have been relatively neglected in current research. This is surprising considering that these abilities are critical to air safety. Federal Aviation Administration reports suggest that errors in pilot judgment account for over 50% of pilot fatalities (Jensen & Benel, 1977). In spite of the growing literature on decision-making and information-processing in general, few applications have been made specifically to the aviation environment. Furthermore, few researchers have examined strategic behavior in flight situations. The research that does exist concerning strategies focuses on narrow aspects of flight rather than the overall process from beginning to end.

For example, Johannsen and Rouse (1983) have studied pilots' responses to normal, abnormal or emergency scenarios during a landing simulation. Abnormal scenarios involved procedural changes, such as the temporary closing of the runway, requiring a holding pattern. The emergency situation was characterized by the failure of one engine. Johannsen and Rouse examined depth of planning (level of detail) in each of the three scenarios and found that planning, as subjectively reported, did not differ significantly between abnormal and emergency scenarios.

However, it was found that pilots who flew the plane manually during the abnormal conditions and relied on automation during the emergency situation reported greater depth of planning than either of the other combinations (e.g., abnormal/autopilot or emergency/manual). The assumption of this research is that since the pilot is often a supervisor of automation in advanced aircraft, a deeper level of planning on the pilot's part leads to better system performance. The strategic decision to engage or to disengage automation is one way to structure the situation so as to create a better atmosphere for good planning.

Along a similar line of research, Giffin and Rockwell (1984) examined operators' problem solving skills in a diagnostic scenario of a flight simulation. Graphical aids, or pilot information plots, were generated for each subject's response strategy on four different diagnostic scenarios. These plots were essentially flow process charts which provided the following information: (1) the number of logic tracks (coherent lines of questioning) used by the operators; (2) the order of inquiries within and between tracks; (3) the amount of time between inquiries; and (4) the number of track returns or information resampling.

From this analysis, several conclusions were drawn. First of all, pilots follow a wide variety of search patterns during problem diagnosis. Secondly, pilots tend to use similar search strategies across all diagnostic scenarios. Finally, the behavior that distinguished the incorrect from the correct problem-solvers was often the omission or abandonment of a critical line of questioning. For example, in a broken magneto scenario, the correct group checked the ignition system, whereas the incorrect group did not (Giffin & Rockwell, 1984).

Many other studies have focused on response patterns to individual components of flight. Much work has been done with flight deck display configurations (Stokes and Wickens, 1988), crew response to automation (Wiener, 1989; Williges, Williges and Fainter, 1988), and crew and air traffic control interaction (Foushee and Helmreich, 1988), to name a few of the major lines of research. However, there has been a lack of research attention on strategic behavior as a function of workload in the flight environment.

The complexity of flight for the pilot is well known. The many subtasks of flight impose specific demands and require different types of effort from a pilot. These various subtasks are combined in many different ways throughout the course of the flight. As such, they may compete for common processing or response resources or, with experience, some of the subtasks may become automatic for the pilot. Another variable is the level of automation in the cockpit which requires the pilot to assume a monitoring role, even though he or she may control the degree to which systems are automated.

Therefore, the pilot's task is to take off, fly and land the plane successfully while monitoring many subsystems and communicating with the crew and air traffic control. Furthermore, if an abnormal or emergency situation should present itself, the pilot must then reprioritize the various tasks and make decisions about the next course of action.

Adding further to the complexity of flight is the range of workload demands that a pilot of modern aircraft faces. While the high workload demands (e.g., landings or emergency conditions) are undeniably stressful, many pilots cope with these conditions in an effective manner by employing strategies to handle the demands. On the other hand, these same pilots are faced with work underload during fully automated or transoceanic flight, and must utilize strategies that permit them to maintain performance effectiveness under these conditions.

Therefore, operators of modern flight systems are likely to be faced with the two extremes of the workload continuum, and in order to avoid performance decrements associated with these extremes, must develop strategies that serve to maximize performance. It is proposed in the present paper that identification of this strategic behavior and the modification of the strategies as a function of varying workload will serve several purposes. First, it will contribute to our overall understanding of the "mental workload" continuum and its effects on performance. Secondly, it will enable us to delineate more precisely the differences between successful and unsuccessful operators. This understanding will allow for better selection of pilots and/or enable us to specify interventions (e.g., training) to improve the abilities of operators to cope with the varying demands of the system.

Development of a sensitive multiple-task scenario

Given the need for research in strategic behavior under varying workload conditions, the first step involves the development of a task which simulates total flight requirements, represents multiple levels of workload and is sensitive to a wide variety of strategic behavior that an operator might choose. Several software packages have been developed and utilized for related research, particularly mental workload studies. Two examples of this software are Workload/PerformANcE Simulation (Window/PANES) and Strategic Control of Response Efficiency (SCORE).

The first of these, Window/PANES, was developed at NASA Ames Research Center (King, Keifer, & Hart, 1989). This software allows for experiments in which information is presented in four separate "windows" on the monitor. A three-axis (heading, altitude, and speed) tracking task is located in one of these windows. Digital readouts of these three types of information are displayed below the tracking task. Furthermore, a topographical map in another window shows the positions of the target path, ownship, and the intended goal. A gauge window can present up to four gauges in either analog or digital form. The meaning attached to the gauges and their values can be either related to the flight task or completely independent. Finally, the fourth window presents alphanumeric messages with the purpose of displaying any type of discrete task to the subject.

Window/PANES was designed to present a simple simulation of a flying task with the purpose of conducting research on the effects of complex task structure and subtask demands on workload, training and performance (King et al., 1989). This software has been utilized in research pertaining to operators' abilities to time-share between tasks and the relationship between performance and subjective workload ratings of individual task components (King, Hamerman-Matsumoto & Hart, 1989).

Strategic Control of Operator Response Efficiency (SCORE), also developed at NASA-Ames Research Center, allows the experimenter to utilize five windows: tracking, monitoring, scheduling, planning and a mode-of-operations indicator. The subject must integrate tracking, monitoring and scheduling (problem-solving). A planning window allows the subject to observe future taskload and plan accordingly. This feature permits the subject to structure his or her task load so as to be more manageable. With SCORE, the subject can also choose to "automate" certain task components within a given time interval.

SCORE has several features which allow for the observation of strategic behavior (e.g., the planning window and the automation option). For a multi-task scenario to be generalizable to strategic behavior in an actual flight environment, many features must be available. First, task load must vary over the course of the scenario, which is available with SCORE or Window/PANES. Second, sensory input must come from not only the visual field, but from the auditory field as well, simulating communication with air traffic control, which is not readily available with existing software. Third, including a complex process control task would allow for a variety of strategies for problem solving during an abnormal or emergency condition. The problem solving "windows" of Window/PANES and SCORE are not very complex, as these tasks require an answer which is either correct or incorrect.

Therefore, the first step in the proposed line of research into strategic behavior as a function of workload in a flight simulation is the development of a battery of tasks that would meet the above-mentioned criteria. The Multi-Attribute Task (MAT) Battery consists of monitoring, tracking, communication, and process control tasks, each of which has its own "window" area on the monitor. Furthermore, a scheduling window presents the operator's location in time as well as those time intervals during certain tasks are automated.

The tracking task is of a compensatory nature to which the subject must respond with a mouse- or joystick-controlled cursor. This task can be automated during certain sections of the flight simulation. The <u>monitoring</u> task consists of two parts: warning lights and probability monitoring. In the upper portion of this window, two lights appear. The first of these is a green light to which the subject must respond when that light is extinguished. The second of these occasionally turns red and the subject must respond when this occurs. The second part of the monitoring task (probability monitoring) involves four vertical, moving-pointer scales with arrows that normally fluctuate around the mid-point. The subject's task is to respond when these fluctuations deviate significantly from the center. All three of these monitoring tasks are required throughout the simulation.

Subjects are also required to respond to a <u>communication</u> task. Pre-recorded auditory messages are presented to the operator at selected intervals during the simulation. However, not all of the messages are relevant to the operator. The subject's task is to determine which messages are relevant and to respond by data input which is presented in the communication window as feedback to the subject. This task cannot be automated by the subject. A scheduling window indicates elapsed time, remaining time and those time intervals during which communication tasks and the manual tracking task are presented.

The final task is a <u>resource</u> (fuel) <u>management</u> task. The monitor presents two "active tanks" to the operator, each of which have two supporting tanks. These tanks are filled halfway at the beginning of the battery. The subject must choose to activate and deactivate the supplying tanks in order to maintain the active tanks at 2500 gallons each.

For each active tank, one of the two supporting tanks has a low capacity for fuel and a high rate of transfer and the other supporting tank has a high capacity for fuel and a slow transfer rate. Therefore the operator has a choice of strategic behavior to employ and this can be recorded for data analysis. Furthermore, the experimenter has a wide variety of potential manipulations with this subtask. For example, a fault (i.e., a leak) could be programmed to occur in one of the tanks, causing the operator to shift priorities and plan a way to solve the problem. The task scenario is described in detail in the Method section.

After development of this task scenario, baseline performance measures were taken for each task segment of the battery. The methodology and results of the baseline measures are described in the following two chapters as Experiment 1. The hypothesized results are as follows:

1) Baseline measurement will show performance on the individual tasks to improve as a function of blocks of time.

2) No differences are expected in performance as a function of gender.

In Experiment 2, the validation of the battery with respect to workload was researched in order to determine if the manipulations of taskload are indeed perceived as high, medium and low. This was accomplished by having subjects perform one 64-minute session of full battery performance and rate their workload for four different task combinations. The hypotheses for Experiment 2 are as follows:

1) Subjects will subjectively rate workload to be significantly different for each of the four task combinations, leading to the establishment of four different workload conditions.

2) Secondly, the validation study will ascertain that the simulation is complex enough to distinguish many different types of strategic behavior, through analyses of performance data.

The experimental design which follows the validation study, Experiment 3, more thoroughly examined the effects of amount of time on task (four 64-minute segments) and mental workload (four different task combinations) on performance and strategic behavior in a 4 X 4 within-subjects factorial design. This design as well as the results of the baseline and validation studies are described in full detail in the Method section.

Based on the previous review, the following hypotheses are proposed:

1) Certain sequences of responses (strategic behavior)

will lead to better performance than other patterns of responding.

2) These successful strategies are hypothesized to be a function of learning such that experienced operators will utilize more successful strategies than operators during the first phase of the battery.

3) During high workload conditions, all subjects will be less able to time-share many tasks efficiently and this will be reflected in lower performance. However, experienced operators (last two 64-minute segments) will be better able to do so than inexperienced operators (first two 64-minute segments).

5) During high workload conditions, all subjects will be less likely to detect relevant signals on the communication and monitoring tasks. Again, experience will moderate this effect.

6) The moderate levels of workload will result in better performance than the low or high levels of workload.

METHOD

Task Battery

As described previously, the task battery is composed of four separate segments, or windows: monitoring, tracking, communication, and resource management. A scheduling window is also presented to the subject. These five windows are described individually below and are graphically depicted as they actually appear to the subject in Figure 1.

Monitoring. The upper left window presents the monitoring task. There are two parts to this task: warning-light monitoring and probability monitoring. The two boxes in the upper portion of this window are the warning lights. The light on the left is normally "on", as indicated by a green light. The subject is required to detect the absence of this light by pressing the "F5" key when the light goes out. The light on the right is normally "off"; however, a red light does come on occasionally. The subject's task is to respond by pressing the "F6" key when he or she detects the presence of that red light. If the subject does not detect an abnormality, the situation reverts back to normal status after a preprogrammed timeout period (15 seconds).

The probability monitoring task consists of four vertical scales with moving pointers. In the normal condition, the pointers fluctuate around the center of the

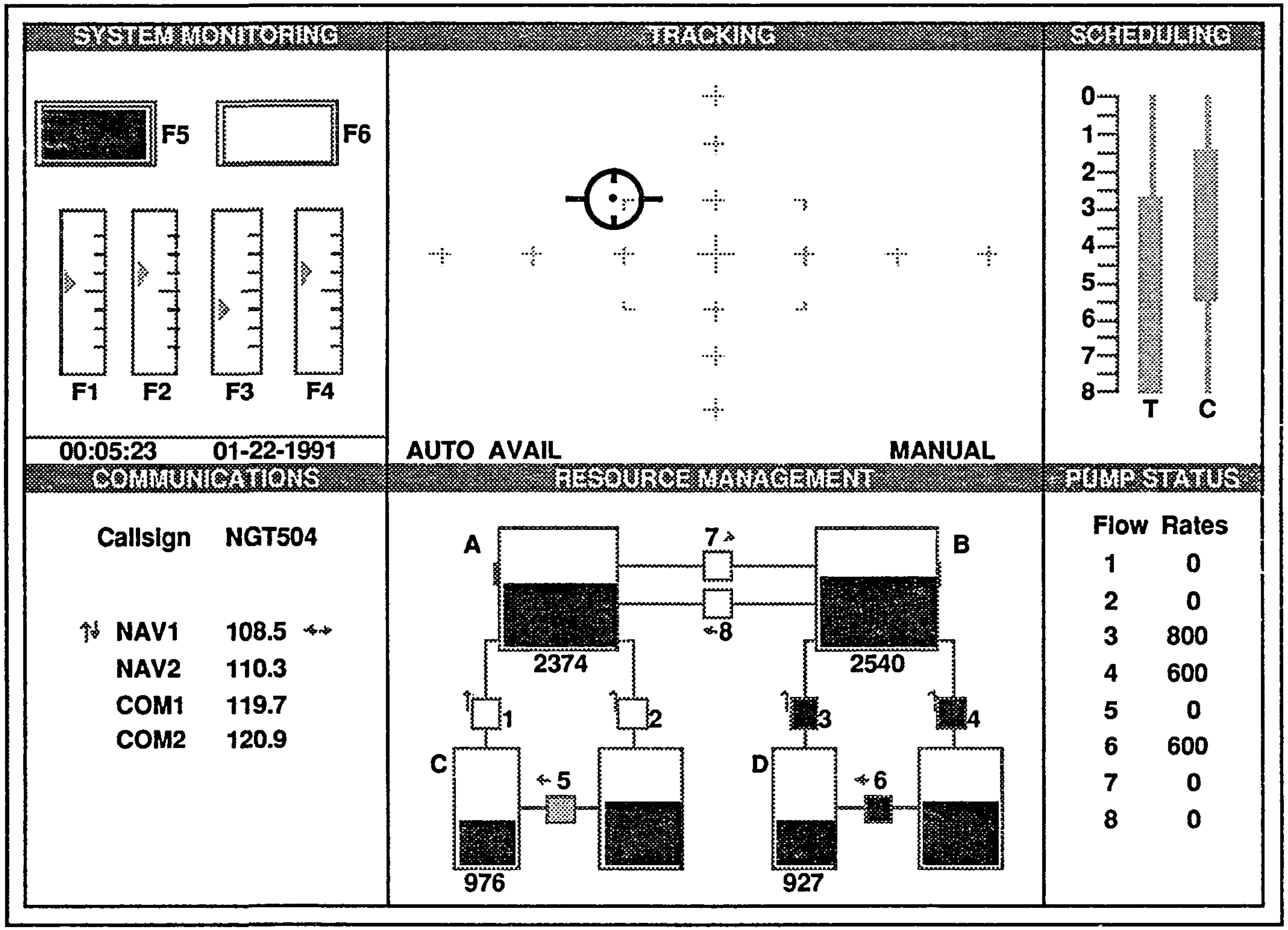


Figure 1. THE MULTI-ATTRIBUTE TASK BATTERY AS PRESENTED TO OPERATORS

scale within one unit in each direction from center. Independently and at random intervals, each display's pointer shifts its "center" position away from the middle of the vertical display. The subject is responsible for detecting this shift, regardless of direction, and responding by pressing the corresponding function key. The appropriate response key is identified below each vertical display.

When this out-of-range status is correctly identified by the subject, feedback is given in two ways. First of all, the pointer of the dial to which the subject responded moves immediately back to the center point and remains there without fluctuating for a period of 1.0 seconds. Additionally, a bar at the bottom of the dial is illuminated in yellow. If the subject fails to detect an abnormality in the probability monitoring task, the fault is automatically corrected 20 seconds from the beginning of its occurrence.

Tracking. A compensatory tracking task operated by mouse is located in the upper right window. The subject's job is to attempt to keep the target in the center of the window, within the dotted lines which form a rectangle. This task can be operated in either manual or automatic mode. The current mode is displayed as either "MANUAL" or "AUTO" in the lower right corner of the window.

Scheduling Window. The purpose of the scheduling window is to present the start and duration of the manual tracking task and the communication task. The two indicators are identified by "T" for the tracking task, and C" for the communication task. The scheduling window allows the subject to "look ahead" from 0.0 (present) to 8.0 minutes into the future. The bold lines (bars) indicate the time at which these two tasks, tracking and communication, begin. The thin lines indicate times at which either tracking or communication are not required of the subject.

Communication. The communication task consists of a series of audio messages which are presented to subjects through headphones. These messages begin with a six-digit call sign, repeated once, and a command to change the frequency of one of the channels listed on the screen. The subject must discriminate his or her call sign, "NGT504", from other three-letter, three-number combinations. The subject's call sign is always presented at the top of the Communications window. Subjects are required to change channel frequencies by the use of the arrow keys. Moving up and down changes the channel location and right/left movement increases/decreases the frequency by 0.2 Mhz.

Resource Management. The Resource Management and Pump Status windows are utilized for the resource management task. This task is presented to subjects as a fuel management task in order to make this a more meaningful task. The Resource Management window provides a diagram of this fuel management system. The six rectangular regions are tanks which hold fuel. The green levels within the tanks represent the amount of fuel in each tank, and these levels increase and decrease as the amount of fuel in a tank changes.

The lines which connect the tanks are pumps which can transfer fuel from one tank to another in the direction indicated by the corresponding arrow. The numbers underneath four of the tanks (Tanks A, B, C, and D) represent the amount of fuel in gallons for each of the tanks. This number is updated every 2 seconds as the amount of fuel in the tanks increases or decreases. The maximum capacity for either Tank A or B is 4000 gallons. Tanks C and D can contain a maximum of 2000 gallons each. Finally, the remaining two tanks have an unlimited capacity.

Subjects are instructed to maintain the level of fuel in both Tanks A and B at 2500 gallons each. This critical level is indicated graphically by a tick mark in the shaded bar on the side of these two tanks. The numbers under each of these tanks provide another means of feedback for the subject. The shaded region surrounding the tick mark represents acceptable performance. Tanks A and B are depleted of fuel at the rate of 800 gallons per minute. Therefore, in order to maintain the task objective, subjects must transfer fuel from the lower supply tanks.

The process of transferring fuel is accomplished by activating the pumps. Each pump can only transfer fuel in one direction, as indicated by the corresponding arrow. These pumps are turned on when the corresponding number key is pressed by the subject. Pressing the key a second time turns that particular pump off and so on. The pump status is indicated by the color of the square area on each pump. When that area is black, or lacking in color, the pump is off. A green light in this area indicates that the pump is actively transferring fuel.

The flow rates for each pump are presented in the "Pump Status" window. The first column of numbers represents the pump number, one through eight. When a pump is activated, its flow rate is presented next to the pump number in this window. When a pump is off, its flow rate is zero. Pumps 1 and 3 transfer fuel at the rate of 800 gallons per minute. Pumps 2, 4, 5, and 6 transfer fuel at the rate of 600 gallons per minute. Finally, Pumps 7 and 8 have a flow rate of 400 gallons per minute.

When a tank becomes full to capacity, all incoming pumps are automatically turned "off". For example, if all of the pumps were activated and Tank A reached its capacity of 4000 gallons, Pumps 1, 2, and 8 would automatically turn "off". Furthermore, if a tank were to become totally depleted of fuel, all outgoing pumps would be deactivated.

At the onset of each experiment, Tanks A and B contain approximately 2000 gallons of fuel each and Tanks C and D contain approximately 1000 gallons of fuel each. All pumps are off at the beginning of the task, leaving all strategic action to the operator's discretion.

Experiment 1 - BASELINE

The first phase of this research plan involved the collection of baseline measures for each of the four tasks of the Multi-Attribute Battery.

Subjects. Twenty male and twenty female undergraduate students at Old Dominion University were randomly selected to participate. These subjects were between the ages of eighteen and thirty-five. Each performed one of the four tasks: monitoring, tracking, communications or resource management on the basis of random group assignment.

Apparatus. A 386/20 microcomputer with EGA graphics was utilized to present the task battery to subjects in color. Data from each task was collected and stored on this computer. A second microcomputer with a voice synthesizer board was linked to the 386/20 computer in order for voice synthesized messages to be presented to the subject. These messages were transmitted through headphones which the subjects wore when participating in the communications task. The second computer was located in a room adjacent to the experimental room and, thus, was not visible to subjects.

Experimental Design. The baseline study employed a between-subjects methodology. Each subject performed only one of the four tasks. A total of 5 males and 5 females participated in each task condition. Each task condition was presented for a duration of 24 minutes. The number of stimuli presented in each task and the nature of these tasks was identical to the levels proposed for Experiment #3. In the monitoring task, subjects were presented with a total of 24 light failures and 24 dial failures which occurred randomly over the course of the task. The minimum time interval between these events, however, was 25 seconds. The communications task consisted of a total of twelve randomly occurring messages: six target and six non-target messages. The minimum time interval between the end of one signal and the beginning of another was 21 seconds. The tracking task employed the same gain and sensitivity as the following two experiments. Finally, the resource management task utilized identical flow rates and starting fuel levels (half-full in Tanks A, B, C, and D) as the final experiments also employed.

<u>Procedure</u>. Subjects were seated in a small room in front of a computer terminal which displayed the battery. The experimenter explained the general purpose of this research project and then proceeded to describe and demonstrate the task which the subject had been assigned to perform. The specific instructions to subjects for each task are recorded in Appendix A. The length of demonstration time for each task are as follows: 2.5

minutes (Tracking, if subject had used a mouse before), 3.5 minutes (Monitoring), 5 minutes (Tracking, if subject had not used a mouse), 6 minutes (Communications), and 7 minutes (Resource Management).

Following the demonstration period, the subject signed a voluntary consent form. After this, the experimental session began. As mentioned previously, each session was 24 minutes in duration. The scripts which directed the program to present events for each baseline task session are presented in Appendix B.

Experiment #2 - WORKLOAD VALIDATION

The second experiment employed the full battery of tasks. Subjects were presented with four different workload levels which simulated the levels intended for use in Experiment #3. Through subjective measures of perceived mental workload, these levels were determined to provide four different mental workload conditions to subjects. The validity of the resource management task for sensitivity to differences in strategic behavior was also examined.

Subjects. Sixteen graduate students (8 males and 8 females) participated in this study. Payment of subjects in the amount of $7.50 was necessary to ensure participation.

Apparatus. The apparatus was identical to that described in Experiment #1. In this study, however, subjects wore the headphones throughout the duration of the task, since continuous battery performance was required.

Independent Variables. Each subject participated in four proposed workload conditions: low, moderate communication, moderate tracking and high. These are referred to as low, communication, tracking and high. Low workload was composed of the monitoring and resource management tasks alone. Signals in the monitoring task consisted of 4 signal light abnormalities and 4 dial abnormalities per 4 minute block of time. The resource management task began with Tanks A, B, C, and D approximately half full and all pumps in the off position.

Moderate workload was defined as performance on monitoring, resource management and the addition of either communication or manual tracking (moderate communication and moderate tracking). Communications signals were presented at the rate of 12 signals (6 target and 6 non-target) for each 12 minute block of time.

High workload was defined as the requirement for the subject to perform all four tasks concurrently. The workload levels were presented in a partially counterbalanced manner. An example of this 64 minute script of all events presented to subjects during all four workload levels is described in Appendix C. The four different workload combination (duration = 8 minutes each) were each presented twice, for a total of 16 minutes for each task combination. Eight orders of workload presentation (presented partially counterbalanced) were available with this method. Therefore, one male and one female subject participated in each order on the partially counterbalanced list.

<u>Dependent Variables</u>. Performance on the resource management task was considered one measure of strategic behavior since different combinations of responses can lead

to either good performance or poor performance. Typically, subjects tended to respond in groups or clusters of key presses. The number of responses in different conditions provided an indication of the strategic behavior used by subjects to maintain critical levels on the resource management task. The purpose of this experiment was to determine whether or not individual variability in strategic behavior can be measured with this battery.

The second dependent variable in this experiment was the subjective measure of workload needed to validate the proposed workload levels. Subjects each provided two responses to TLX measures of mental workload for each eight minute presentation of workload: low, moderate communication, moderate tracking and high.

The NASA Task Load Index, developed by the Human Performance Group at NASA Ames Research Center, is a multi-dimensional rating scale. A weighted average of ratings on six subscales provides an overall workload rating. These subscales are: Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort and Frustration. Subjects are required to rate their perceived exertion on five of these subscales (except Own Performance) on a graded scale from "Low" to "High". The Own Performance subscale ranges from "Good" to "Poor".

The TLX rating scale can be presented to the subject at any time during the operation of the Multi-Attribute Task

Battery. A code for the onset of rating scale presentation can be added to the script which generates events for the MAT Battery. At the time of TLX presentation, specified by the script, a second screen appears in place of the MAT Battery. This screen is depicted in Figure 2. During the presentation of the second screen, all MAT Battery activity is paused until the subject either exits from the TLX screen or the maximum time for that screen is reached. Upon return to the MAT Battery screen, the timing of battery events resumes.

When the TLX screen is first presented, one pointer appears in the middle of the first subscale (i.e., 50). Each gradation represents 5 points; thus, potential scores on each subscale range from 0 to 100. The subject begins with the first subscale and uses either the mouse or the right and left arrow keys to select his or her score. After the subject has selected the score for the first subscale, he or she can press either the left mouse button, the space bar or the downward arrow key to progress to the next scale. After this, the pointer of the second subscale becomes yellow and the first pointer turns grey. The pointer of the scale which is operated is slightly larger and illuminated in yellow as opposed to the grey coloring of the other pointers. This process continues until the subject has responded to all six subscales.

After the subject has entered a response for the sixth

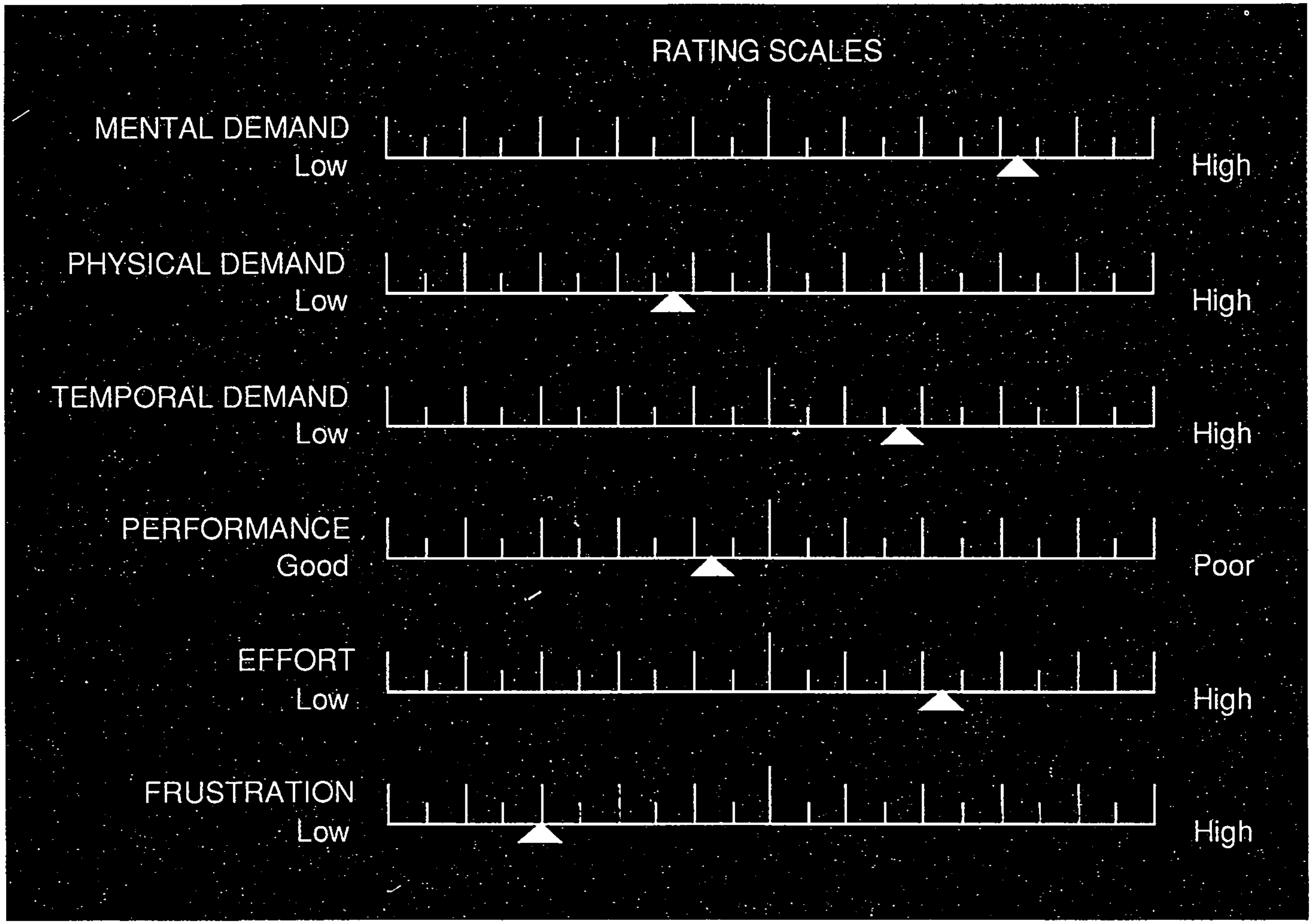


Figure 2. NASA TLX WORKLOAD RATING SCALES

subscale, the commands at the bottom of the screen change.

At this point, the subject can press either the escape or return key to exit to the MAT Battery. However, if he or she wishes to change any of the responses that have already been made, this can be done with the use of either the downward arrow key, the mouse button or the space bar. For instance, the subject can press the downward arrow key three times to return to the third subscale without changing any of the scores on the first two subscales. The pointer turns yellow when a particular subscale score can be reselected or changed. Changes are made through the same means used to initially select scores: the mouse, or the right/left arrow keys. After the subject is satisfied with all responses, the escape key or return key returns him or her to the MAT Battery.

Design. The experimental design employed in this study was a one-way within-subjects factorial design. Each level of workload was performed for sixteen minutes and subjects responded four times to Likert-type questions about perceived mental workload during each of the workload levels. For each 8 minute block of workload, the rating scale was presented at 2 minutes, 0 seconds and at 6 minutes, 0 seconds.

Procedure. Subjects were first provided with a

Procedure. Subjects were first provided with a demonstration of each task individually, followed by a training period during which the subject practiced each task separately. The length of this training period was determined by the results of Experiment #1 and was four minutes for each task, following a brief demonstration.

Following the training, subjects performed the battery as a whole for a duration of 64 minutes. Every four minutes, subjects responded to the workload measures. The entire experimental session lasted approximately 120 minutes.

Experiment #3 - OPERATOR STRATEGIES

Subjects. Fifteen undergraduate and nine graduate students (N = 24) participated as subjects in Experiment 3. Payment ($5/hour) for four hours of participation or $10.00 and two extra credit points were given as incentive for participation. The subjects ranged in age from 18 to 29, with an average age of 22.2. Ten males and fourteen females participated in this study.

Apparatus. The apparatus and experimental room were the same as described in Experiment #2.

Procedure. Subjects participated in a training program for each task individually. The length of this training program was four minutes per task following a brief demonstration of each (identical to previous studies). This training program lasted approximately 50 minutes in length. Following this, and between each hour session, a short break (less than 10 minutes) could be taken by subjects.

On Day 1, each subject performed Hours 1 and 2 of the full task battery, following the demonstration and training. The second day of experimental participation always followed two days later at the same time of day. On Day 2, subjects were allowed to ask clarification questions about any particular task before starting Hours 3 and 4. However, they were not permitted to practice any task before beginning the full battery performance. All of the monitoring, tracking, communications and resource management tasks were identical to those described in Experiment #2 in terms of event rates and signals. Appendix C presents the 64-minute script which was used for each of the hour sessions. The order of workload level presentation was partially counterbalanced between subjects.

Independent Variables. Mental workload was defined as validated in Experiment #2. Four levels were employed: (1) resource management and monitoring alone (Low); (2) resource management, monitoring and communication (Communication); (3) resource management, monitoring and tracking (Tracking); and (4) all four tasks performed simultaneously (High). Each subject participated in each workload level for a total of 64 minutes (16 minutes per hour session).

The second independent variable in this study was experience with the battery (four levels). This was defined as each of the four 64 minute blocks, containing 16 minutes of each workload level.

The experimental design employed in this study was a 4 (Workload) X 4 (Experience) within-subjects factorial design.

Dependent Variables.

1) Performance. Performance variables consisted of RMS error (tracking), deviation from criteria on the resource management task and reaction time, errors and hits/misses (communications and system monitoring).

2) Strategic Behavior. This was defined as the number of responses (pump configuration changes) on the resource management task. As shown in Experiment 2, when the number of responses to the resource management task changes while performance remains constant, a strategy for handling workload changes is implied. Another related strategy is the ability of subjects to correlate the deviations from target on the two main tanks of the resource management task. This indicates that subjects are making identical responses on the two congruent systems: (1) Tank A and its two supporting tanks; and (2) Tank B and its two supporting tanks. This simplifies the task in that the subject is performing "one task" instead of two tasks.

RESULTS

Experiment #1 - BASELINE

Various performance measures were collected for each task. These are summarized below by task.

Monitoring. For this task, reaction times to the absence of the green light, presence of the red light and abnormal dial fluctuations in the probability monitoring task were collected. Mean median reaction time and standard deviations were computed for each eight-minute block of time (total = 3 blocks), as well as percentages of hits, misses, and false alarms.

A main effect for block was found, F (2, 16) = 7.22, $p < .05$, for reaction times to the presence of the red light. Response time decreased as a function of length of time into task. Specifically, Neuman-Keuls post-hoc tests indicated that Block 3 was significantly different from Blocks 1 and 2. No learning effect was found for response times to dial deviations or to the absence of the green light. These latter tasks were found to elicit more misses than responses to the presence of a stimuli (i.e., the red light). The means for the three individual tasks by 8-minute blocks of time are presented in Figure 3.

Analyses of Variance were also performed for response times to all three tasks combined. No significant differences were found between groups as a function of blocks of time. These means are summarized in Figure 4.

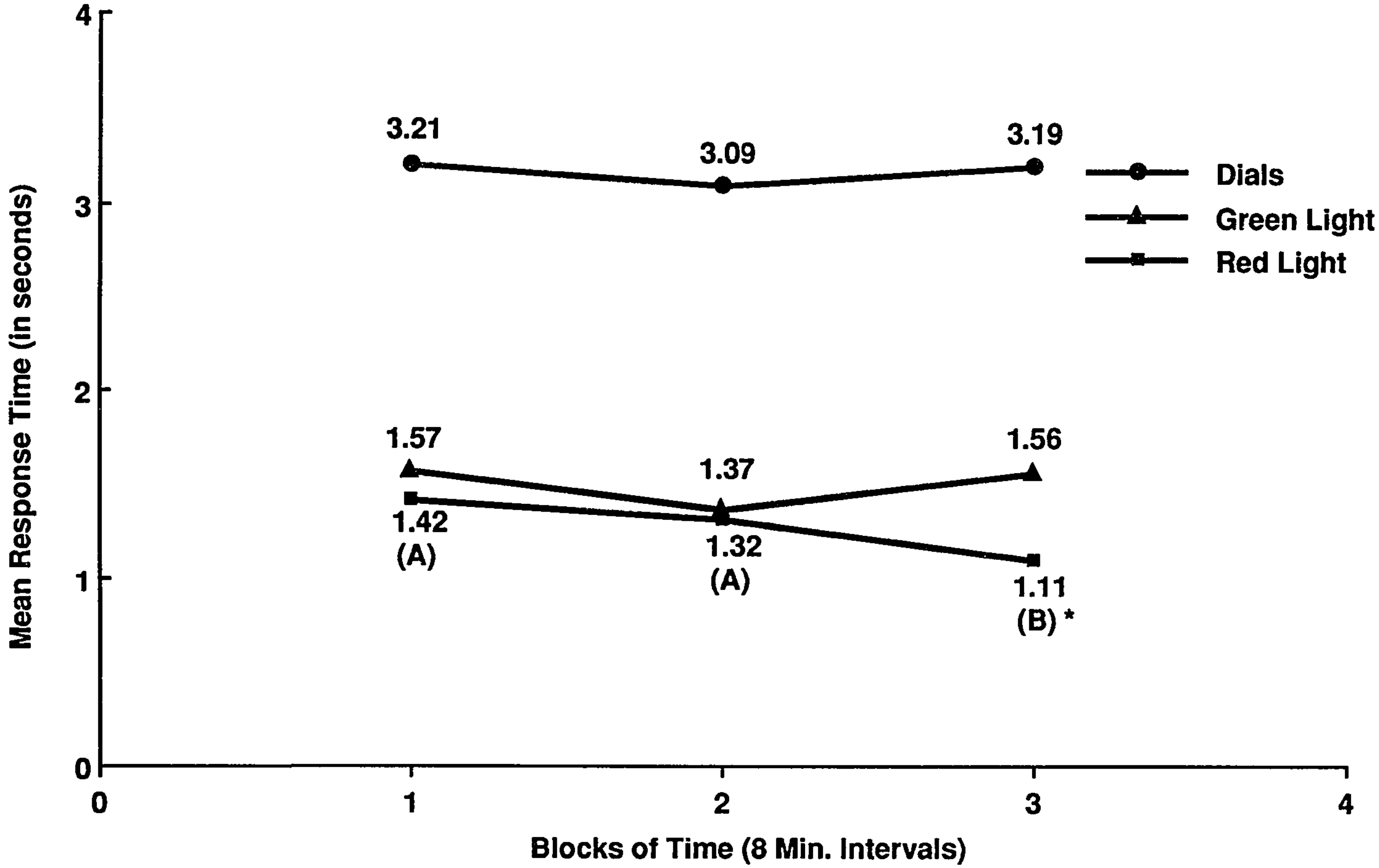


Figure 3. BASELINE MONITORING PERFORMANCE DATA (BY TASK) OVER BLOCKS OF TIME
* Different letters indicate significantly different groups, ($p < .05$)

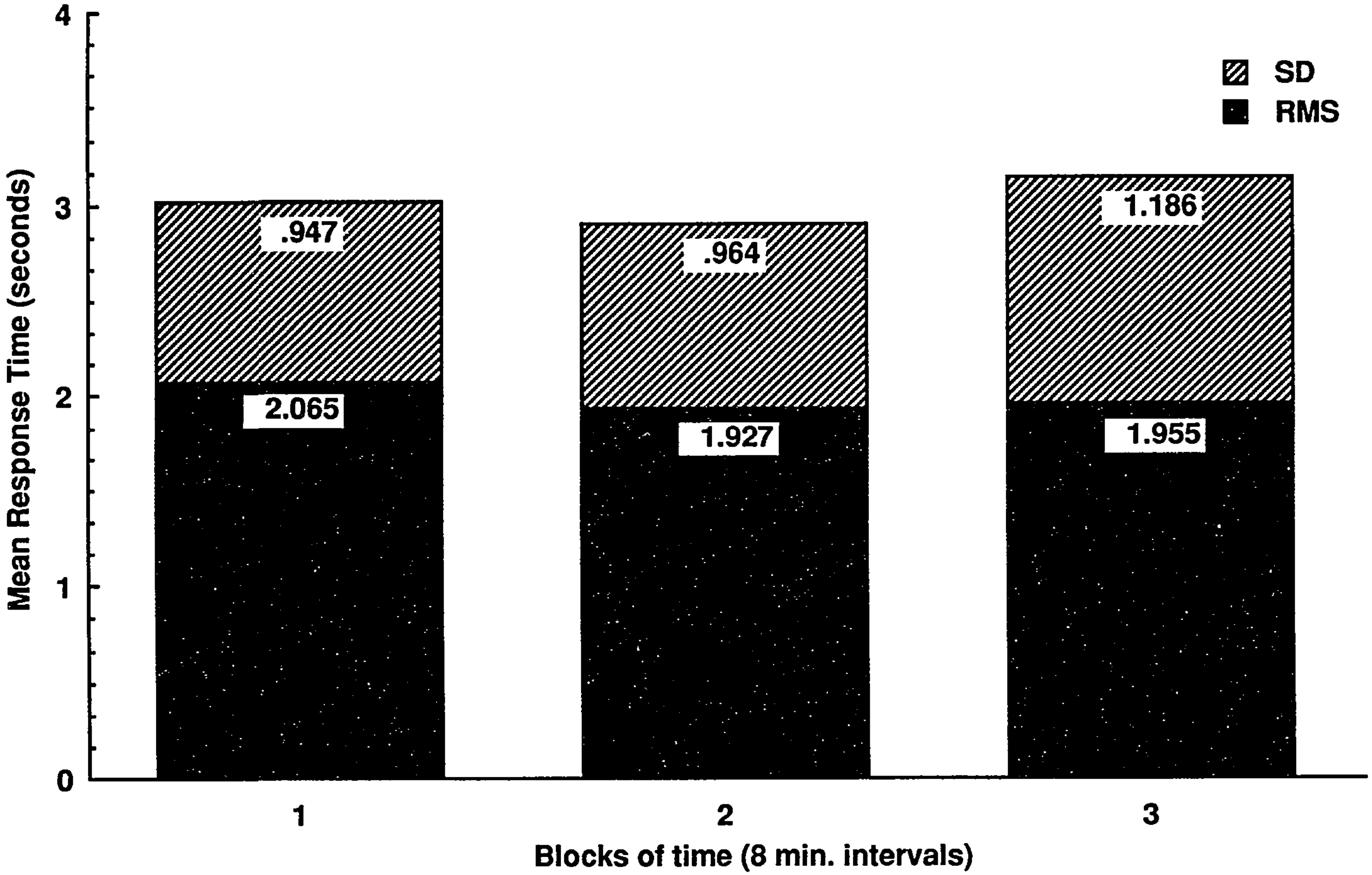


Figure 4. BASELINE MONITORING PERFORMANCE DATA (TASKS COMBINED) AS A FUNCTION OF BLOCKS OF TIME

Tracking. For the tracking task, RMS error was calculated for each 15 second interval of the task. Median RMS error was then determined for each subject across four-minute blocks (total = 6 blocks). The mean RMS scores for each four-minute block of time along with the accompanying standard deviations are shown in Figure 5.

These data were further analyzed for block and gender effects. Analyses of variance showed that RMS scores differed as a function of time into the task in four-minute blocks, $F(5,40) = 3.39$, $p < .05$. Neuman-Keuls post-hoc tests revealed that Block 1 differed significantly from Blocks 4, 5 and 6. This indicates that performance is not significantly affected by learning after approximately twelve minutes. Additionally, there was no gender difference in performance.

Communication. Data from the communication task consisted of median response time to target messages as well as accuracy of responding. These medians were each taken from groups consisting of two messages (total = 3 groups). Means and standard deviations were computed across subjects; means are presented in Figure 6 for each "ownship" message. Analyses of variance showed no significant differences for block or gender. Reaction times were constant throughout the task. Furthermore, there were no false alarm or misses.

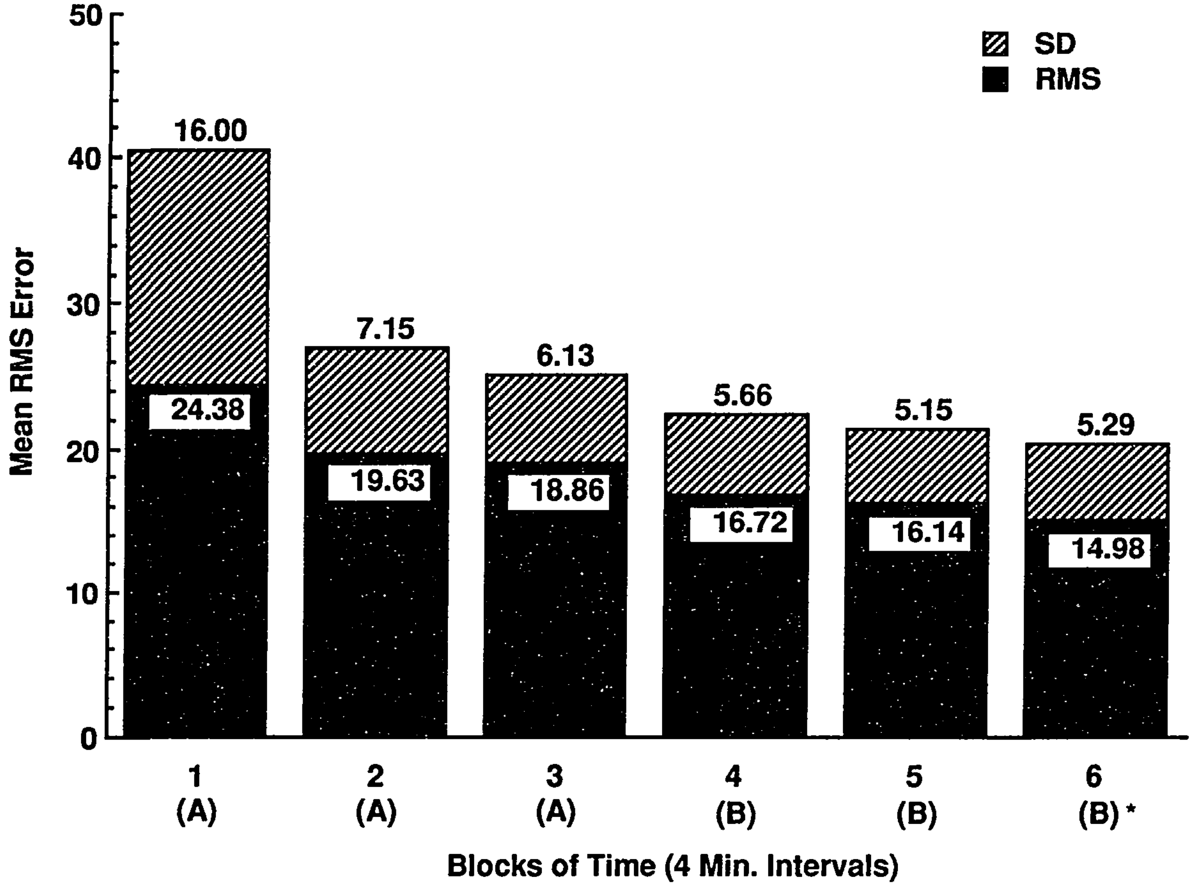


Figure 5. BASELINE TRACKING PERFORMANCE DATA AS A FUNCTION OF TIME
*** Different letters indicate significantly different groups, ($p < .05$)**

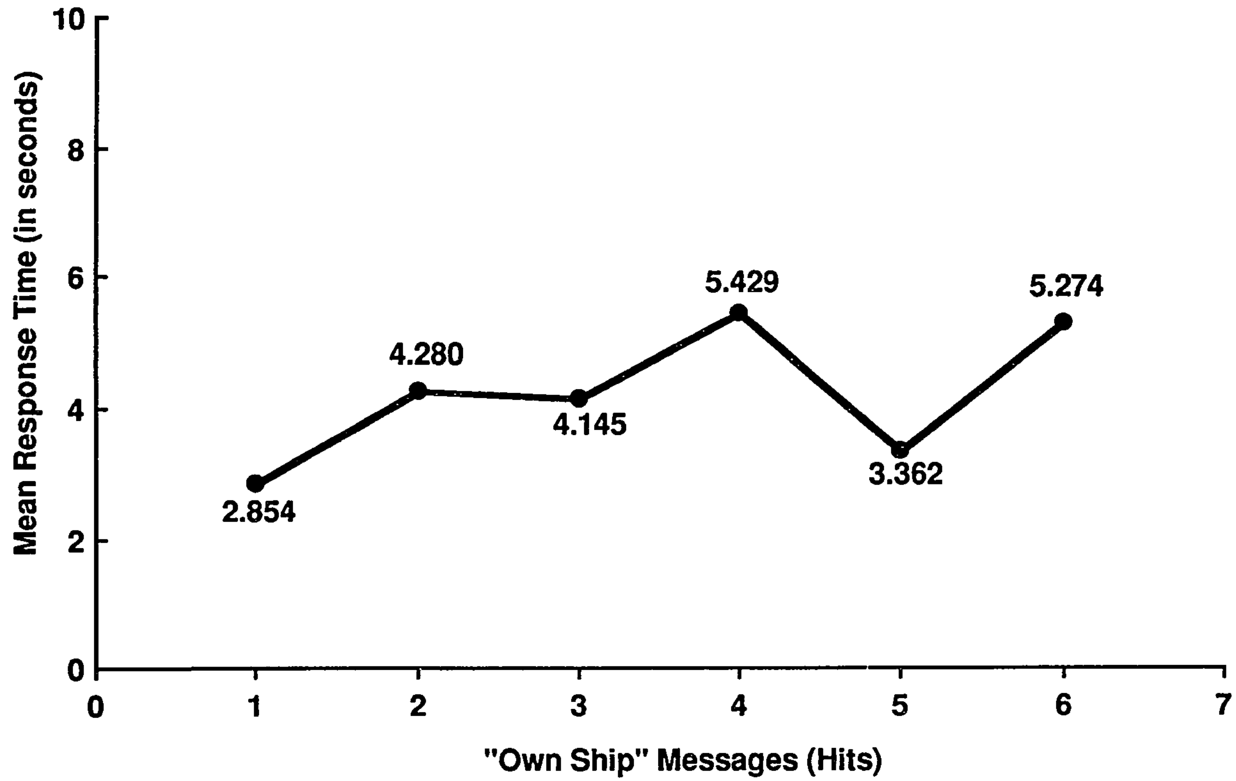


Figure 6. BASELINE COMMUNICATION PERFORMANCE DATA BY "OWNSHIP" MESSAGE (HITS)

Resource Management. From the resource management task, mean deviation from target (2500) for Tanks A and B was calculated. Those means and standard deviations are summarized in Figure 7. No significant differences between groups were found.

Summary. The results of baseline data collection yield useful information which served as a basis for comparison with full battery performance. Performance on all tasks did improve as a function of time, although this trend was not significant in every case. No gender differences were found on any of the four tasks. All source of variation tables for the baseline analyses are listed in Appendix D.

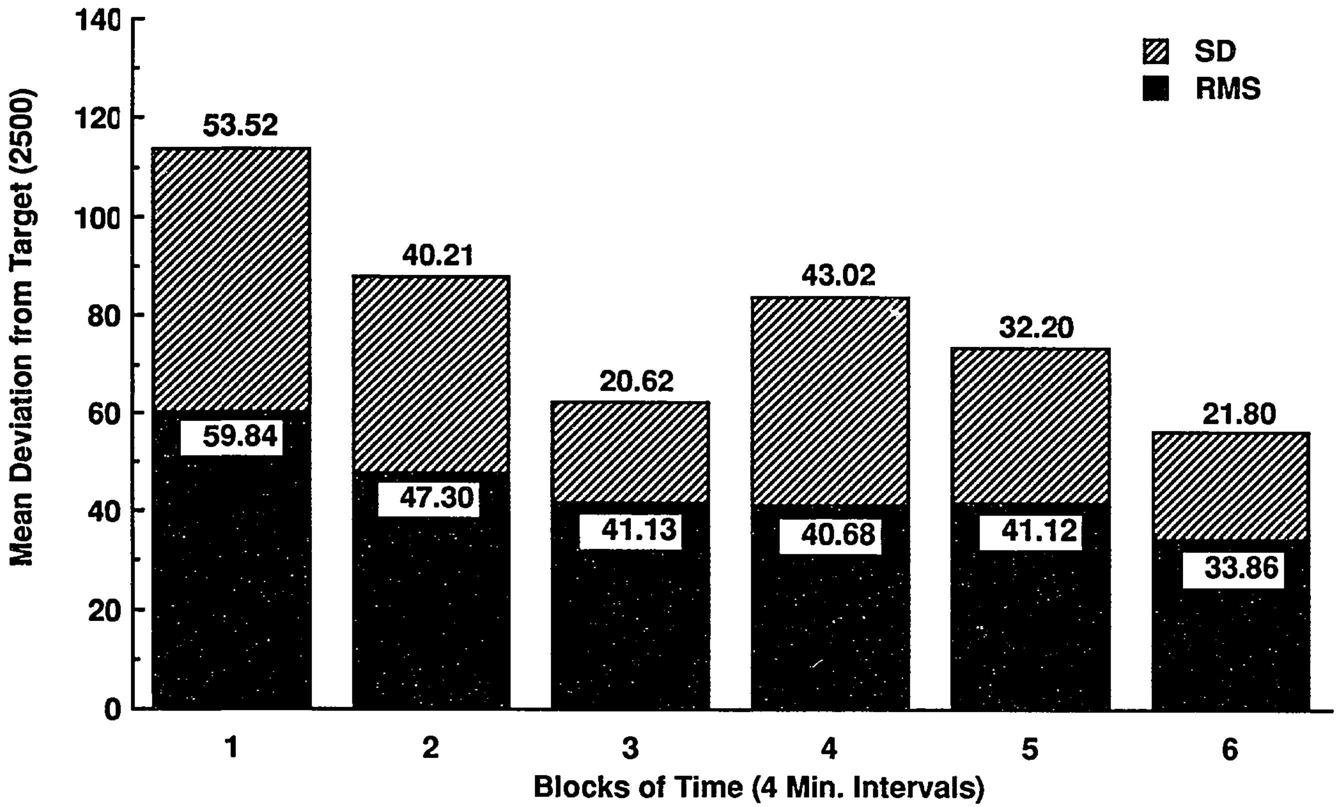


Figure 7. BASELINE RESOURCE MANAGEMENT PERFORMANCE DATA BY BLOCKS OF TIME

Experiment #2 - WORKLOAD VALIDATION

Workload Ratings.

Analyses of Variance were performed to measure subjective workload responses. A significant difference was found for responses by workload level, $F(3,189)=103.62$, $p < 0.05$. Neumann-Keuls post-hoc tests show each group to be significantly different from the others. These means are presented in Figure 8.

Strategic Behavior.

Additionally, strategic behavior responses were analyzed to determine if the resource management task was sensitive to variability in strategic behavior. When total number of responses were analyzed by workload level, significant differences were observed, $F(3, 45) = 8.99$, $p < 0.05$. Specifically, it was found that the Low and Communication workload levels were significantly different from the other two workload levels (Tracking and High). Generally, subjects made fewer pump configuration changes (key presses) as task load increased. These means are presented in Figure 9.

Analyses of variance were also performed for performance (deviation from target) as a function of workload. Although deviation from target did increase as a workload, this trend was not significant. Figure 10

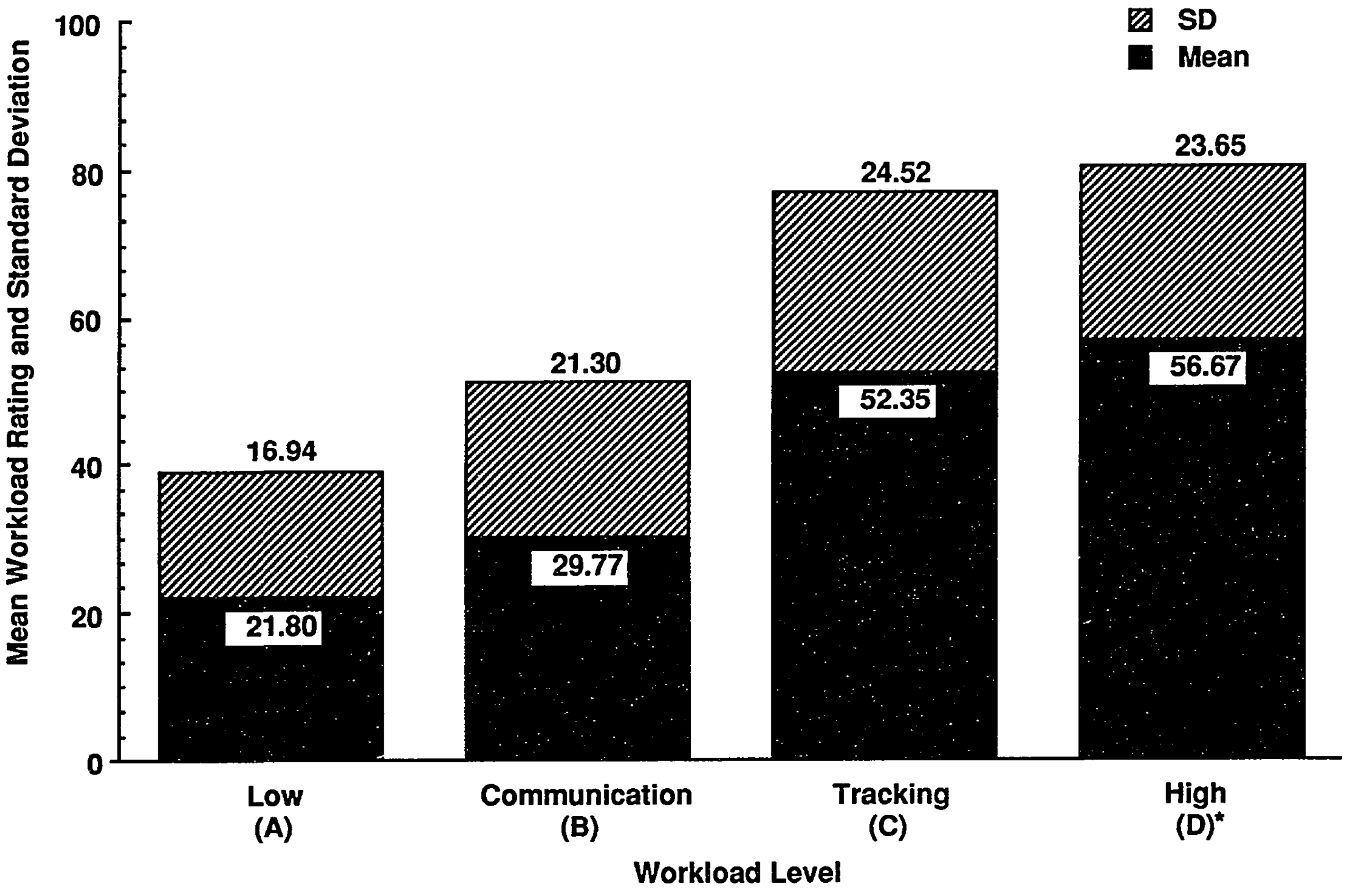


Figure 8. MEAN WORKLOAD RATING SCORES AS A FUNCTION OF WORKLOAD LEVEL
* Different letters indicate significantly different groups, ($p < .05$)

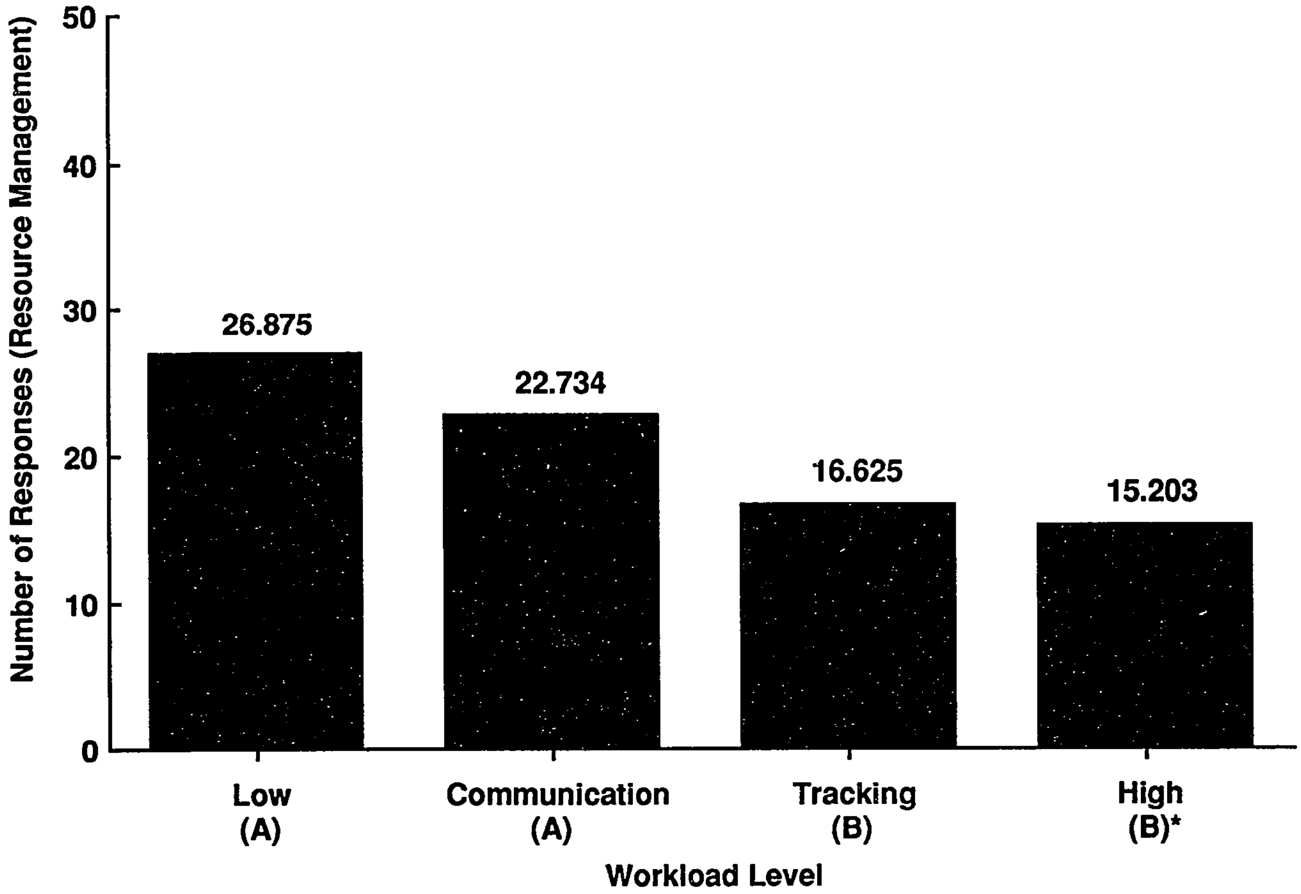


Figure 9. RESPONSE FREQUENCY TO RESOURCE MANAGEMENT AS A FUNCTION OF WORKLOAD LEVEL
* Different letters indicate significantly different groups, ($p < .05$)

presents these means. All source of variation tables for the validity analyses are listed in Appendix E.

Summary. These results showed the existence of four valid workload levels (Low, Communication, Tracking and High). Additionally, the differences in responding as a function of workload indicated variability in strategic behavior as a function of workload. Specifically, subjects made fewer changes to the pump configurations as task load increased in order to maintain performance on this task. This indicates that subjects may be developing a more efficient (i.e., fewer inputs or changes) strategy to handle increases in task load. Given the promising results summarized above, the third experiment further examined strategic behavior as a function of experience with the task (time into task) and workload (as defined above) during a longer duration of battery performance (Four 64-minute sessions).

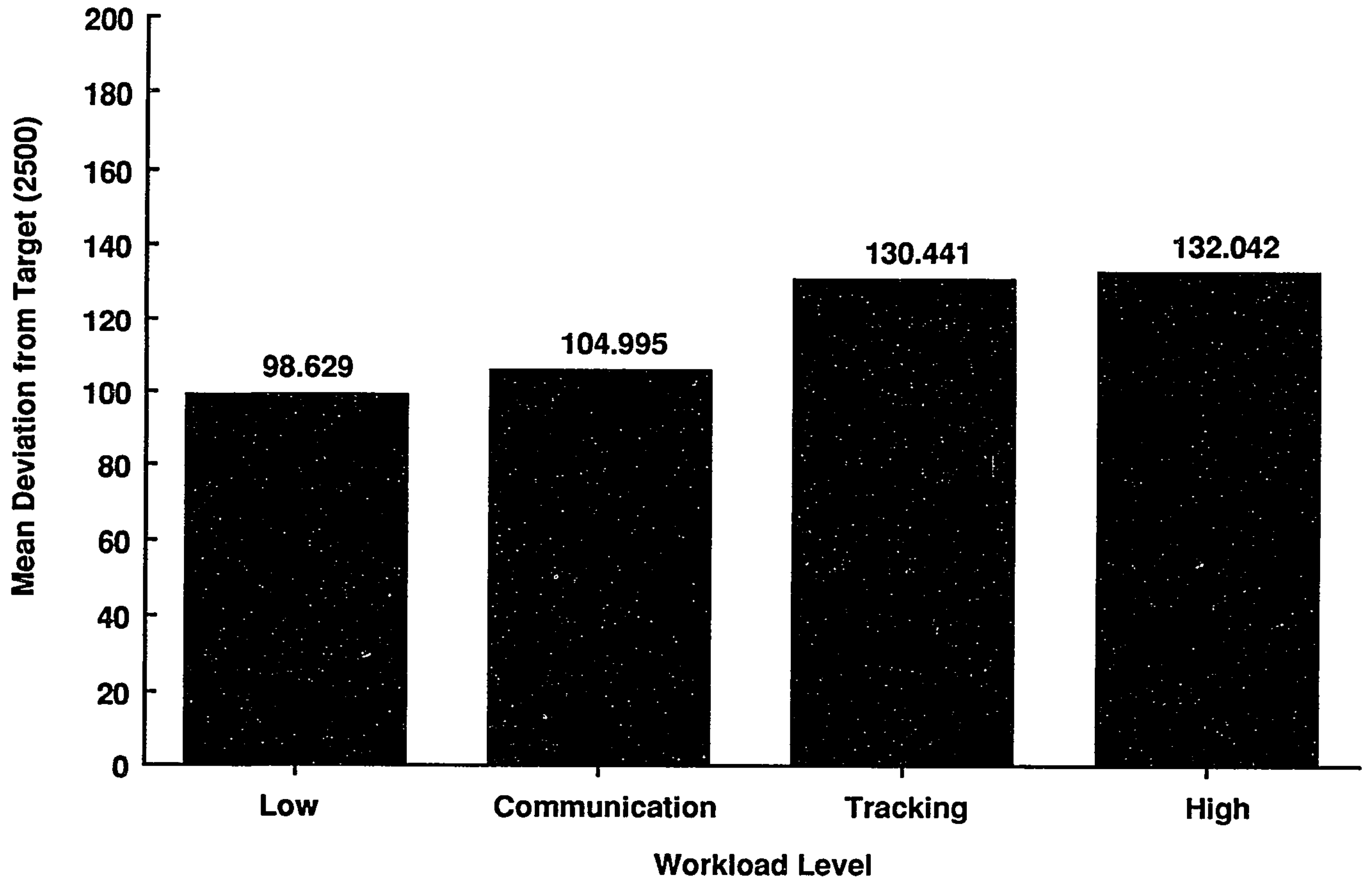


Figure 10. RESOURCE MANAGEMENT PERFORMANCE BY WORKLOAD LEVEL

Experiment #3 - OPERATOR STRATEGIES

Performance data were collected and analyzed for each task by hour and by workload level (Low, Communication, Tracking and High). Analyses of variance with Neuman-Keuls post-hoc tests were utilized for these analyses. The data are summarized below by task. The means by condition for all performance data are presented in Appendix F.

Monitoring.

The monitoring task was required of subjects throughout the entire four hours of battery performance and throughout all four workload levels. For this task, performance measures included response times and errors. These data were computed by combining all monitoring tasks (lights and dial deviations) to produce an overall reaction time and a total error frequency.

The analyses of variance indicated a significant main effect for hour into the task, $F(3,69) = 47.32$, $p<.05$. Overall, reaction time decreased both as a function of time and a function of workload. Specifically, Neuman-Keuls post-hoc tests indicated that the mean reaction times on Hours 1 and 2 were significantly different from each other and Hours 3 and 4, which did not differ from each other. These means are graphically depicted in Figure 11.

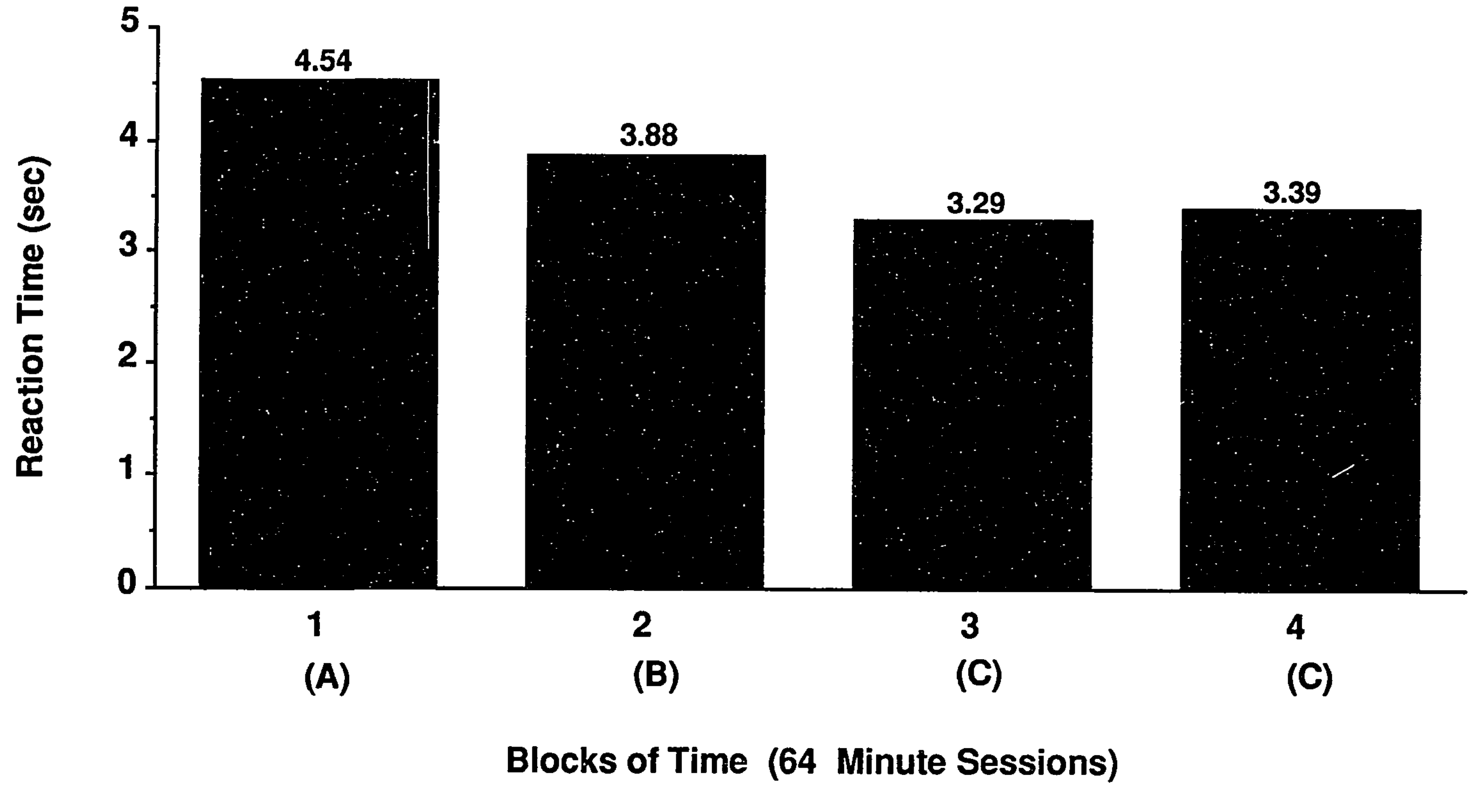


Figure 11. REACTION TIME FOR MONITORING AS A FUNCTION OF TIME

* Different Letters Indicate Significantly Different Groups, $p < 0.05$

Reaction time also differed as a function of workload, $F(3,69) = 3.81$, $p < .05$. In terms of increasing workload (determined by Study 2), the Low workload group recorded significantly higher reaction times than the Communication, Tracking and High groups. These means are presented in Figure 12. The interaction between hour and condition was not found to be significant.

Other indicators of monitoring performance included false alarms and misses to signals. A missed signal was defined as a lack of response prior to a time-out length for a particular signal occurrence (15 seconds for lights, 20 seconds for dials). A main effect for missed signals as a function of hour was revealed with analysis of variance, $F(3,69) = 22.07$, $p < .05$. As illustrated in Figure 13, Hours 1 and 2 were significantly different from each other and Hours 3 and 4. No differences in frequency of missed signals were found between Hours 3 and 4. Overall, the mean number of missed signals decreased with time into the task. There were no significant effects found for workload level.

A false alarm was defined as a response (inappropriate key press) in the absence of a signal. The analyses did not indicate any effects for false alarms as a function of time or task combination. The sources of variation for monitoring reaction times, missed signals and false alarms are summarized in Table 1.

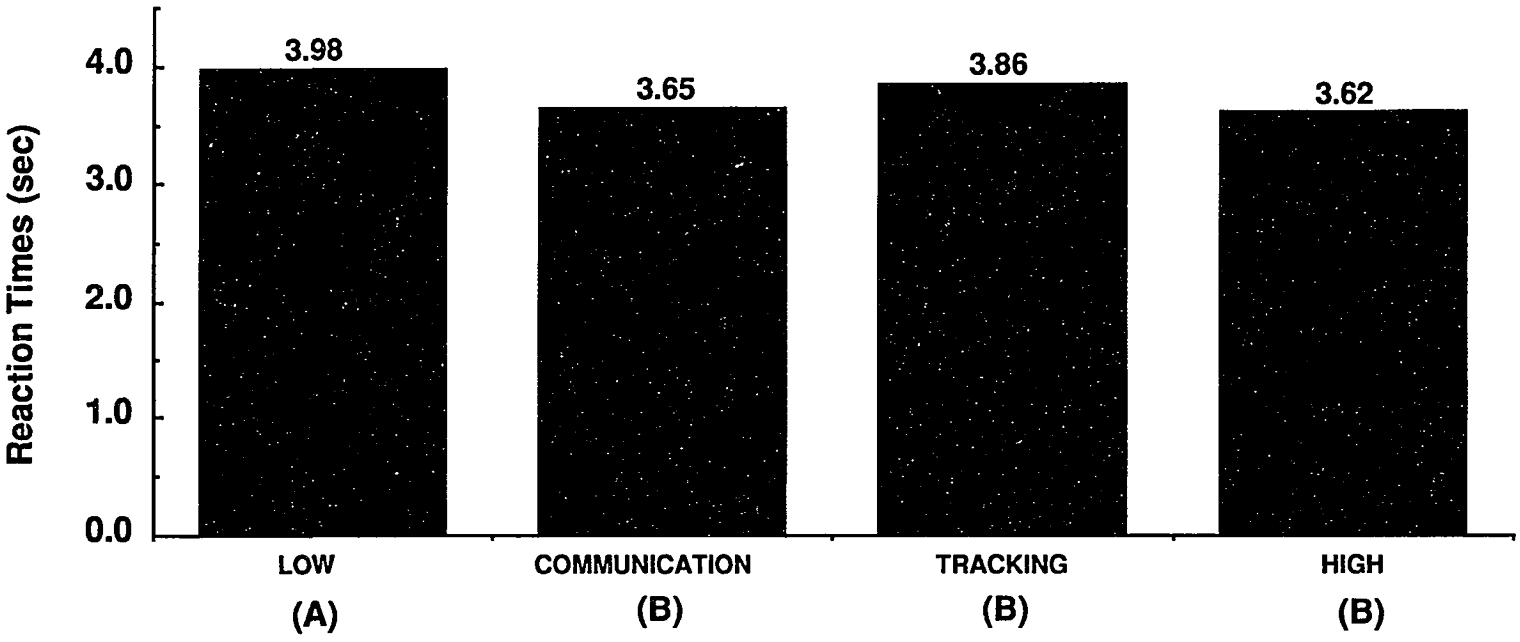


Figure 12. REACTION TIMES FOR MONITORING TASK AS A FUNCTION OF WORKLOAD LEVEL

* Different Letters Indicate Significantly Different Group, $p < 0.05$

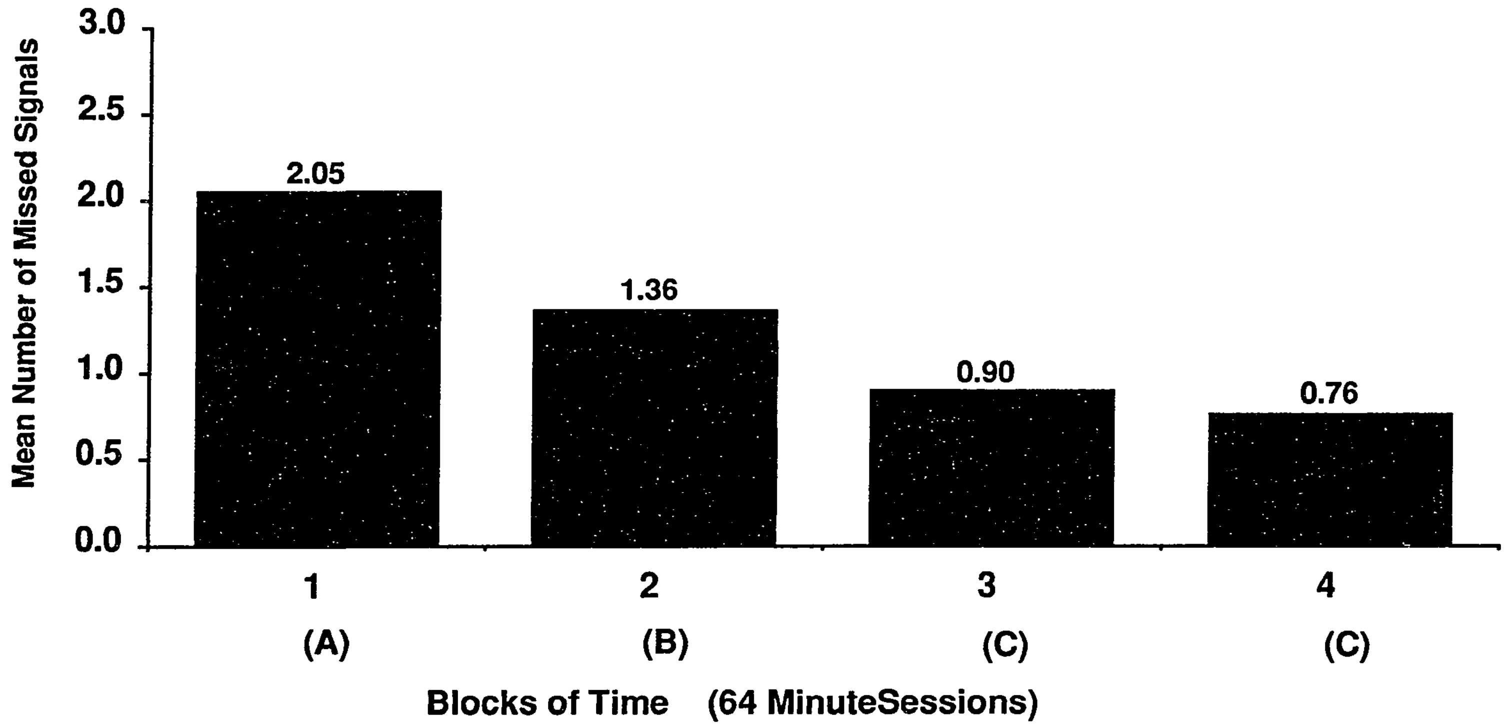


Figure 13. MISSED SIGNALS ON MONITORING TASK AS A FUNCTION OF TIME

* Different Letters Indicate Significantly Different Groups, $p < 0.05$

Table 1

Summary of the Analysis of Variance for Monitoring Data (Tasks Combined)

Source of Variation	*df*	Mean Square	*F*	Eta Square
Reaction Time				
Hour	3	62.62	47.32*	0.13
Workload Level (WL)	3	5.85	3.81*	0.01
Hour * WL	9	1.32	1.35	
Subject	23	22.65	NT	NT
Subj * Hour	69	1.32	NT	NT
Subj * WL	69	1.54	NT	NT
Subj * Hour * WL	207	0.98	NT	NT
Missed Signals				
Hour	3	65.11	22.07*	0.10
Workload Level (WL)	3	3.15	2.41	
Hour * WL	9	1.49	1.11	
Subject	23	29.60	NT	NT
Subj * Hour	69	2.95	NT	NT
Subj * WL	69	1.31	NT	NT
Subj * Hour * WL	207	1.34	NT	NT
False Alarms				
Hour	3	27.65	2.53	
Workload Level (WL)	3	16.11	2.24	
Hour * WL	9	8.03	1.45	
Subject	23	86.14	NT	NT
Subj * Hour	69	10.94	NT	NT
Subj * WL	69	7.19	NT	NT
Subj * Hour * WL	207	5.52	NT	NT

* $p < .05$, NT = No Test

In general, monitoring performance improved as subjects became more experienced with the battery. However, the finding that monitoring reaction time was highest during the low workload condition was not hypothesized. In order to examine the monitoring tasks more closely, the individual tasks were analyzed for differences in responding as a function of time and workload.

Red Light Signals. Subjects responded to the presence of the red light signals by pressing a key to turn that light "off". Analyses of variance for reaction time, misses and false alarms were performed for red light signals alone to examine differences by hour or workload level.

Mean reaction time to the presence of red light signals decreased as a function of time, F (3,69) = 5.20, $p < 0.05$. As depicted in Figure 14, mean reaction time during the Hour 1 was significantly higher than Hours 2, 3 or 4, as revealed by Neumann-Keuls post-hoc tests.

Reaction time to red light signals also varied as a function of workload level, F (3,69) = 3.30, $p < 0.05$. Again, mean reaction time was highest during the low workload condition. The communication, tracking and high workload conditions did not differ from each other. These means are graphically illustrated in Figure 15. The interaction of workload and time was not found to be significant.

Frequency of missed signals, and false alarm rates were

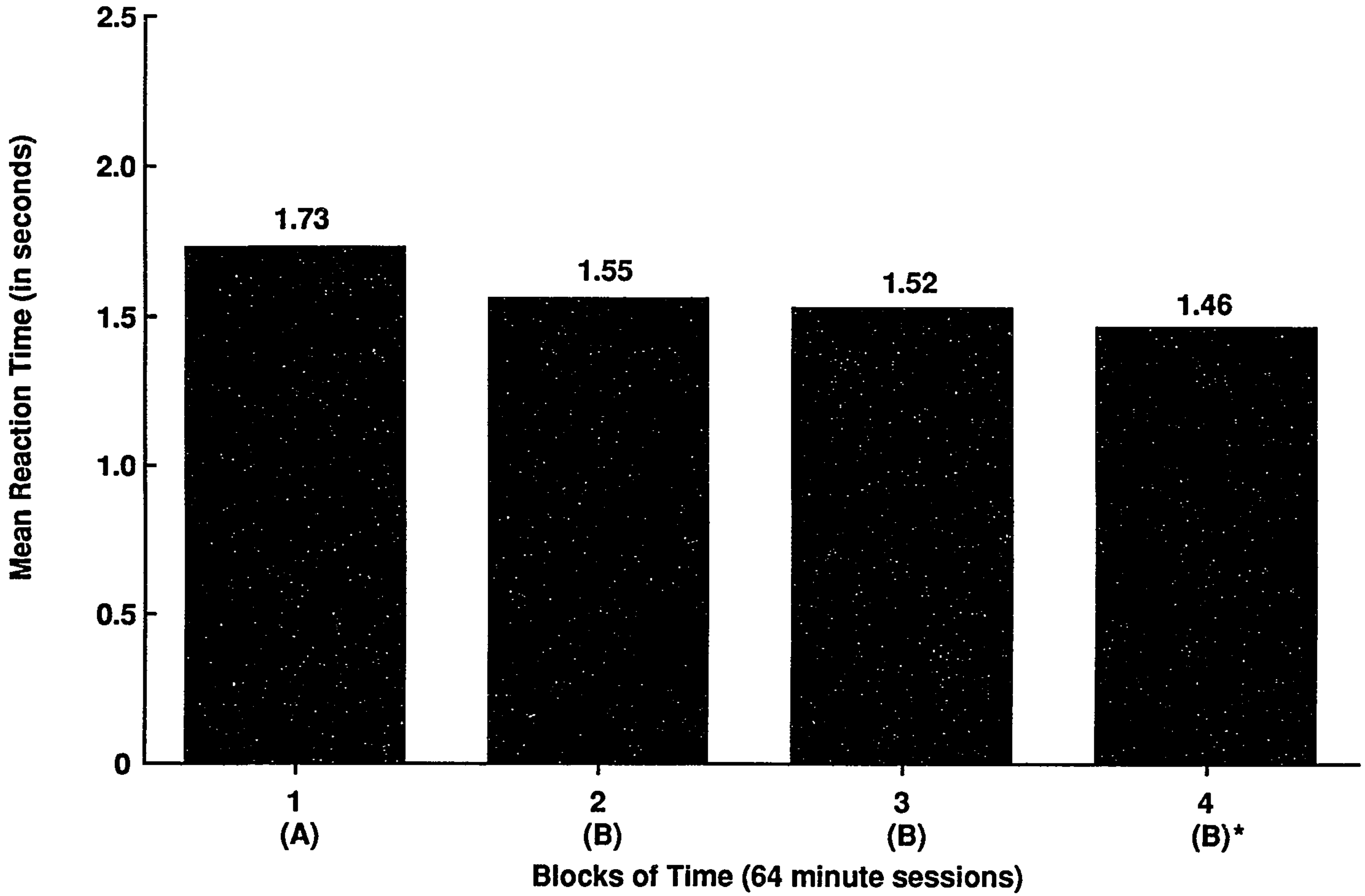


Figure 14. MEAN REACTION TIME TO RED LIGHT SIGNALS AS A FUNCTION OF TIME.
* Different letters indicate significantly different groups, ($p < .05$)

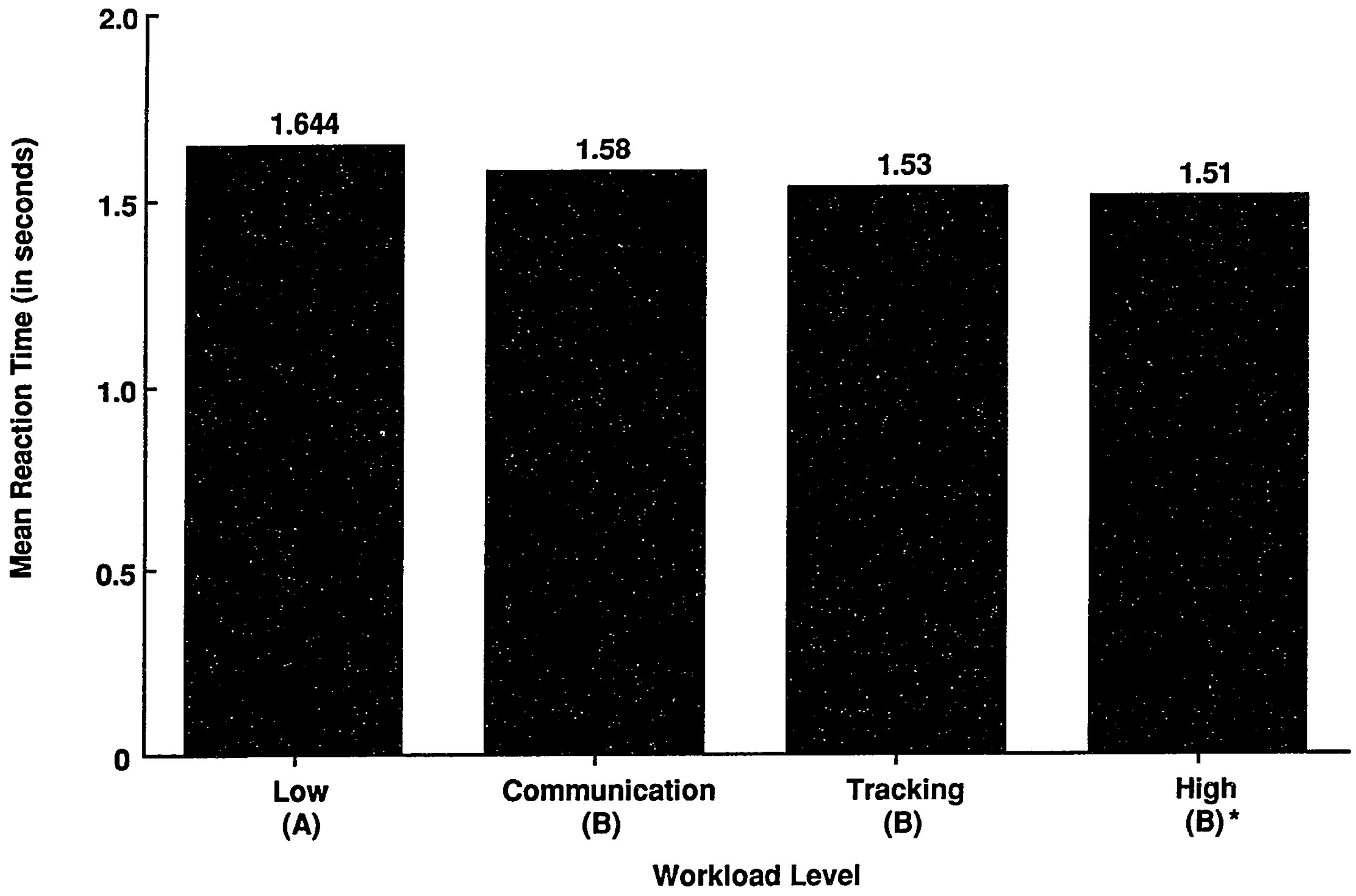


Figure 15. MEAN REACTION TIME TO RED LIGHT SIGNALS AS A FUNCTION OF WORKLOAD LEVEL
* Different letters indicate significantly different groups, ($p < .05$)

also submitted to analyses of variance. However, no differences were found for experience with the task (time) or workload level. Accuracy of responding did not vary significantly between any of the conditions. False alarm rates were typically very low (< 1.0 per condition). The sources of variation for red light signal reaction time, missed signals and false alarms are located in Table 2.

Green Light Signals. Subjects were required to respond to the absence of the green light signals by pressing a key to turn the green light back "on". These data were analyzed for differences in reaction time, misses and false alarms for hour and workload level.

Analysis of variance for mean reaction time produced a significant main effect for time, $F\ (3,69) = 13.92$, $p < 0.05$. As shown in Figure 16, mean reaction time was higher during Hours 1 and 2 than during Hours 3 and 4. No differences in reaction time were found as a function of either workload or the interaction of time and workload.

Subjects missed fewer green light signals as they became more experienced with the task battery, particularly within the low workload condition, as indicated by a significant interaction between time and workload, $F\ (9,207) = 3.35$, $p < 0.05$. Figure 17 shows that misses were higher during the low workload conditions in all four hours than any other condition. However, the mean number of misses decreased as a function of time to a greater degree during

Table 2

Summary of the Analyses of Variance for Red Light Signals

Source of Variation	df	Mean Square	F	Eta Square
Reaction Time				
Hour	3	2.51	5.20*	0.03
Workload Level (WL)	3	0.66	3.30*	0.01
Hour * WL	9	0.17	0.76	
Subject	23	4.76	NT	NT
Subj * Hour	69	0.48	NT	NT
Subj * WL	69	0.20	NT	NT
Subj * Hour * WL	207	0.23	NT	NT
Missed Signals				
Hour	3	0.01	0.57	
Workload Level (WL)	3	0.04	2.52	
Hour * WL	9	0.02	0.93	
Subject	23	0.13	NT	NT
Subj * Hour	69	0.01	NT	NT
Subj * WL	69	0.01	NT	NT
Subj * Hour * WL	207	0.02	NT	NT
False Alarms				
Hour	3	0.23	1.11	
Workload Level (WL)	3	0.16	0.92	
Hour * WL	9	0.15	1.36	
Subject	23	1.85	NT	NT
Subj * Hour	69	0.21	NT	NT
Subj * WL	69	0.18	NT	NT
Subj * Hour * WL	207	0.11	NT	NT

* $p < .05$, NT = No Test

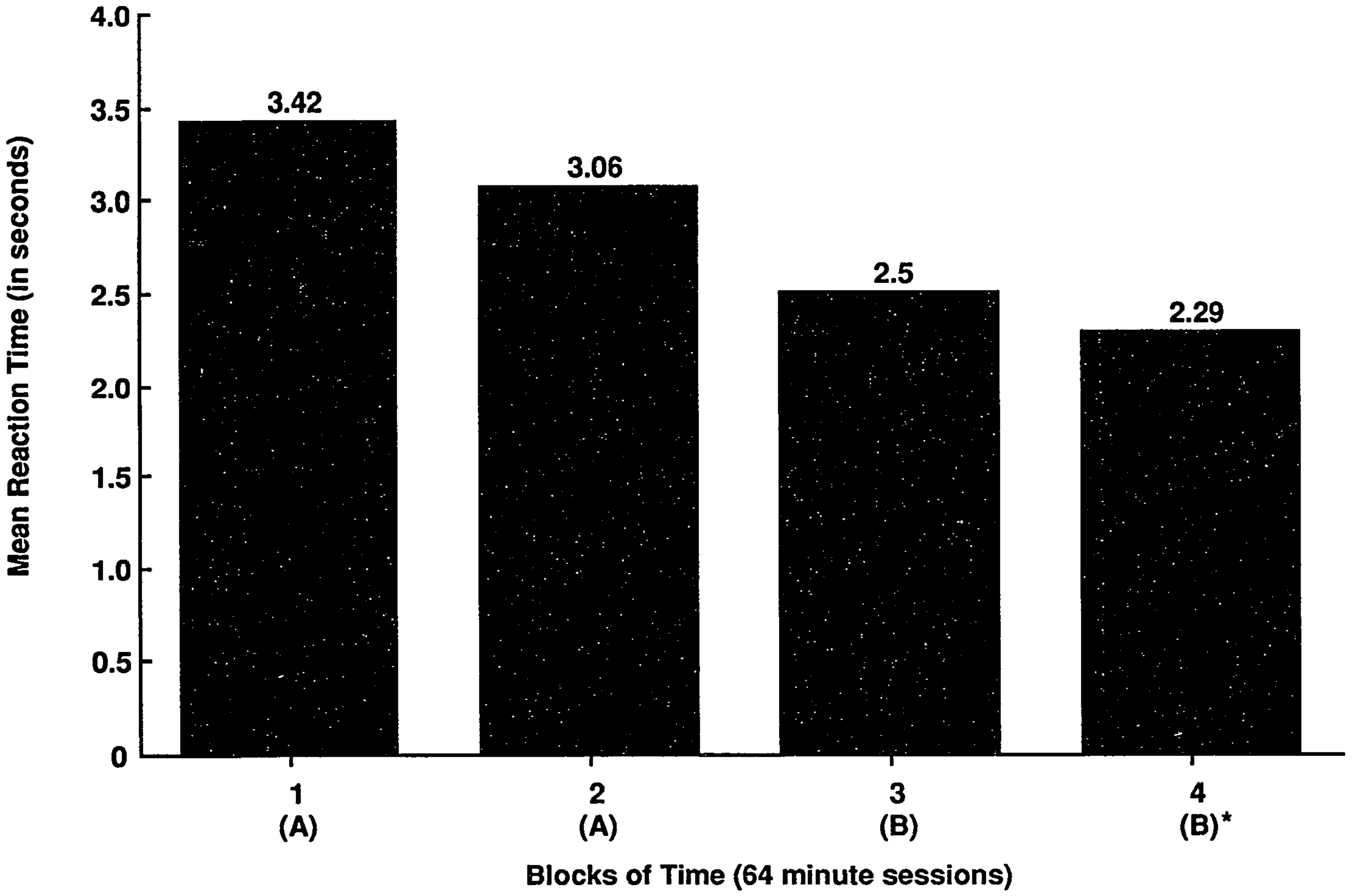


Figure 16. MEAN REACTION TIME FOR GREEN LIGHT SIGNALS AS A FUNCTION OF TIME
* Different letters indicate significantly different groups, ($p < .05$)

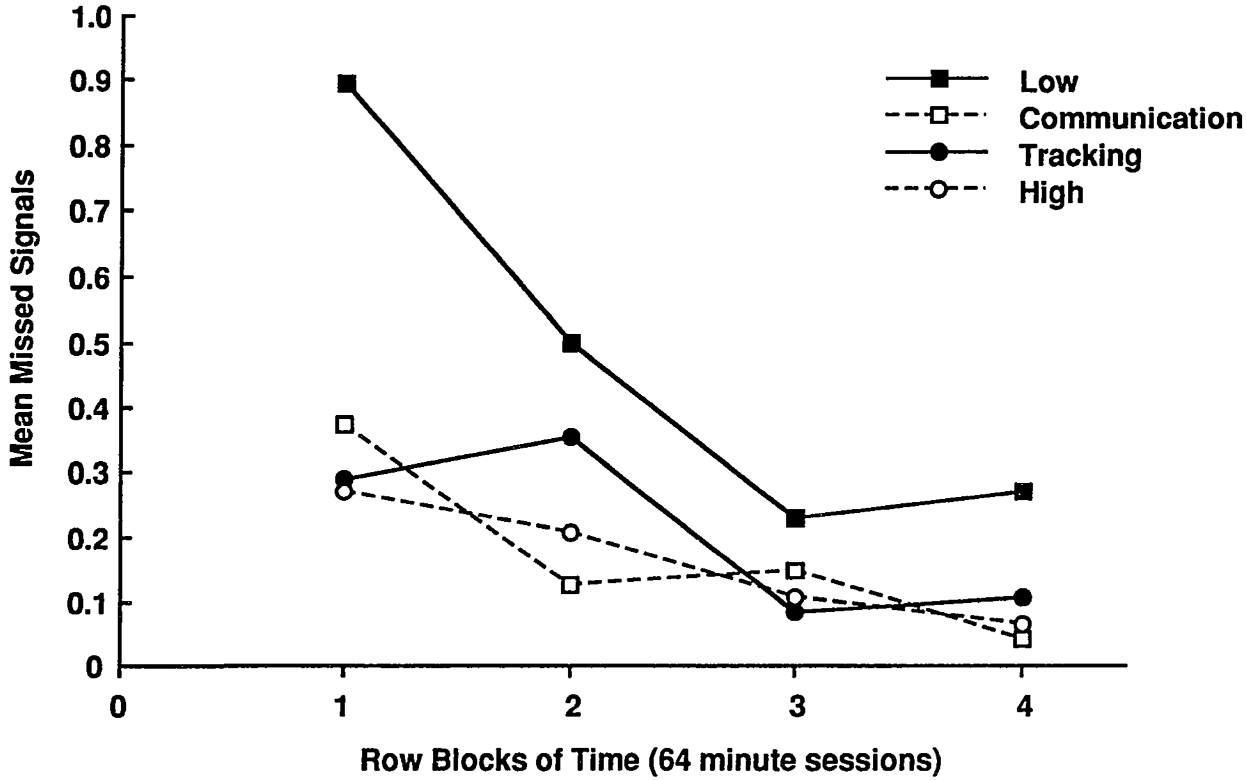


Figure 17. MEAN NUMBER OF MISSES TO GREEN LIGHTS AS A FUNCTION OF TIME AND WORKLOAD

the low workload condition than any other workload condition.

The false alarm rate did not differ significantly for either time or workload. The source of variation tables for all green light signal analyses are given in Table 3.

Dial Signals. Four moving pointer dials were presented to subjects who were instructed to respond whenever a dial deviated from midpoint averaging. These data were also analyzed for reaction time, misses and false alarms.

Mean reaction time for dials also decreased as a function of time, $F\ (3,69) = 37.27$, $p < 0.05$. As portrayed in Figure 18, subjects responded more quickly as they became more experienced with the task. Hours 3 and 4 were the only two groups which did not differ significantly from each other. Analyses of variance failed to reveal any effects for workload.

Missed dial signals varied significantly with both time, $F\ (3,69) = 17.51$, $p < 0.05$, and workload, $F\ (3,69) = 7.19$, $p < 0.05$. Subjects became increasingly more accurate as time into the task increased (See Figure 19). Hours 3 and 4 were the only two groups which did not differ significantly from each other. Additionally, as Figure 20 illustrates, subjects were more accurate in the two low workload conditions than in the two high workload conditions.

Table 3

Summary of the Analysis of Variance for Green Light Signals

Source of Variation	df	Mean Square	F	Eta Square
Reaction Time				
Hour	3	50.18	13.92*	0.05
Workload Level (WL)	3	5.19	1.26	
Hour * WL	9	3.79	1.17	
Subject	23	15.65	NT	NT
Subj * Hour	69	3.60	NT	NT
Subj * WL	69	4.13	NT	NT
Subj * Hour * WL	207	3.23	NT	NT
Missed Signals				
Hour	3	4.77	12.13*	0.06
Workload Level (WL)	3	4.21	11.96*	0.05
Hour * WL	9	0.67	3.35*	0.02
Subject	23	1.59	NT	NT
Subj * Hour	69	0.39	NT	NT
Subj * WL	69	0.35	NT	NT
Subj * Hour * WL	207	0.20	NT	NT
False Alarms				
Hour	3	0.06	0.14	
Workload Level (WL)	3	0.57	1.96	
Hour * WL	9	0.29	1.15	
Subject	23	1.70	NT	NT
Subj * Hour	69	0.40	NT	NT
Subj * WL	69	0.29	NT	NT
Subj * Hour * WL	207	0.26	NT	NT

* $p < .05$, NT = No Test

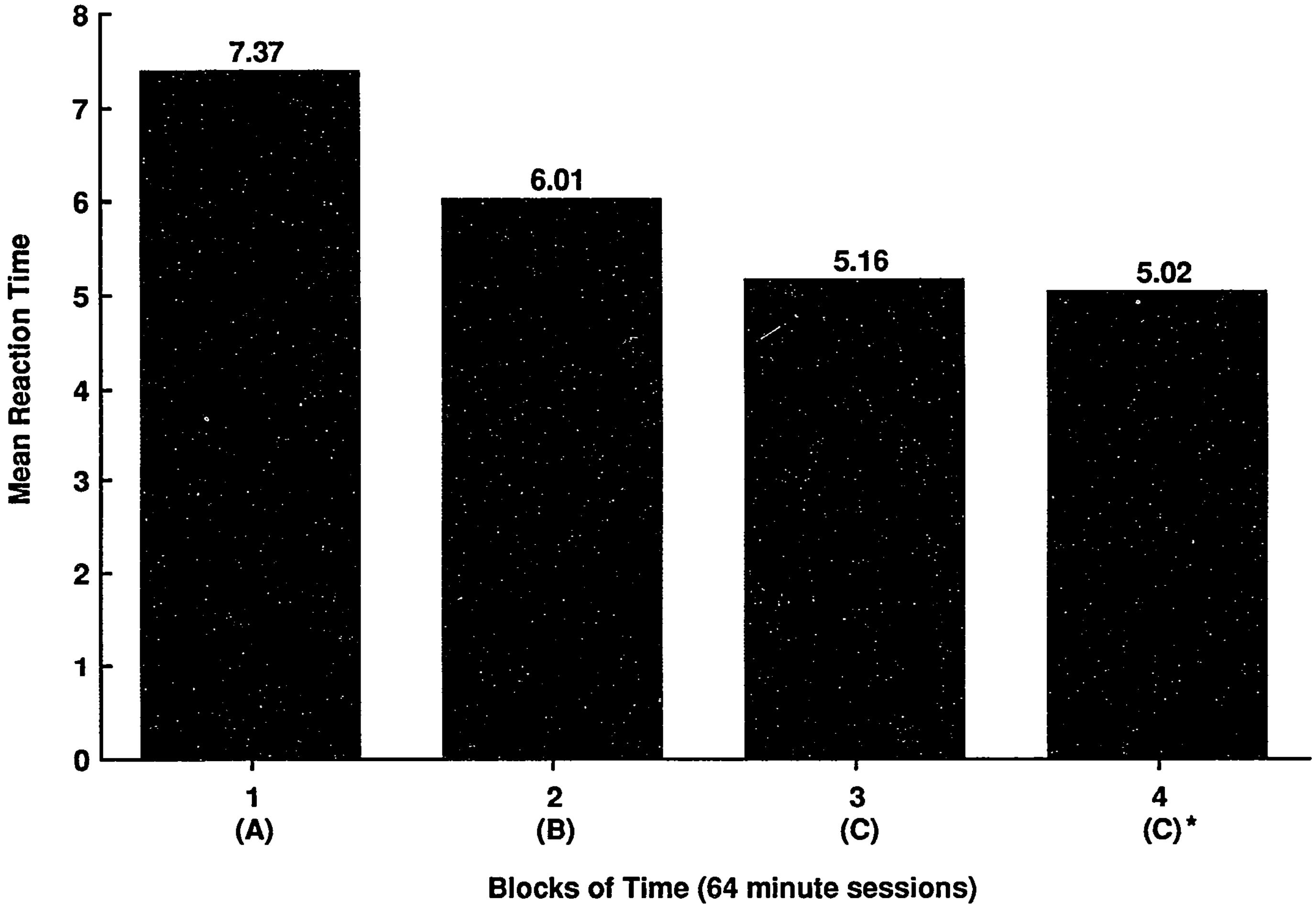


Figure 18. MEAN REACTION TIME TO DIAL SIGNALS AS A FUNCTION OF TIME
* Different letters indicate significantly different groups, ($p < .05$)

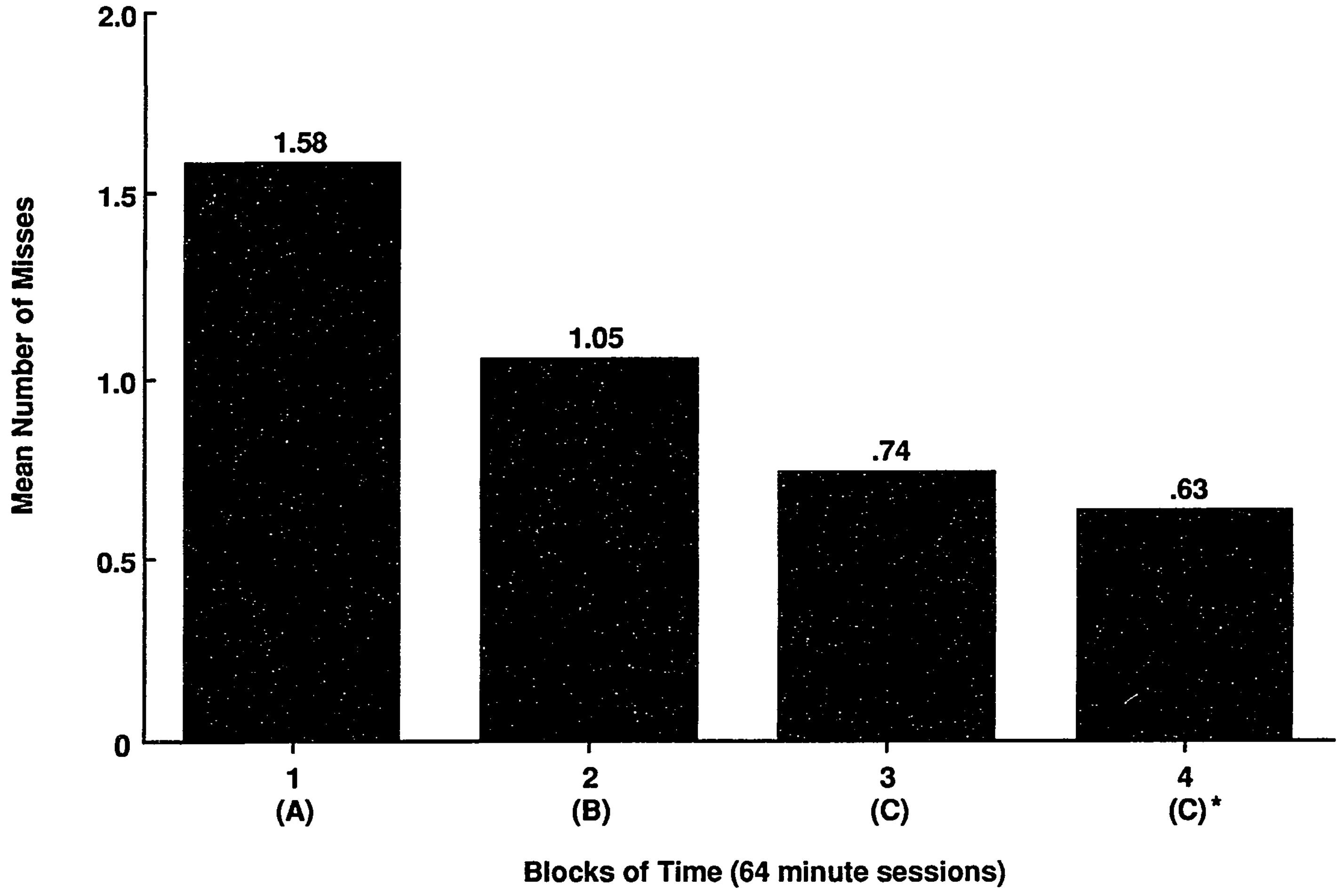


Figure 19. MEAN NUMBER OF MISSES TO DIAL SIGNALS AS A FUNCTION OF TIME
* Different letters indicate significantly different groups, ($p < .05$)

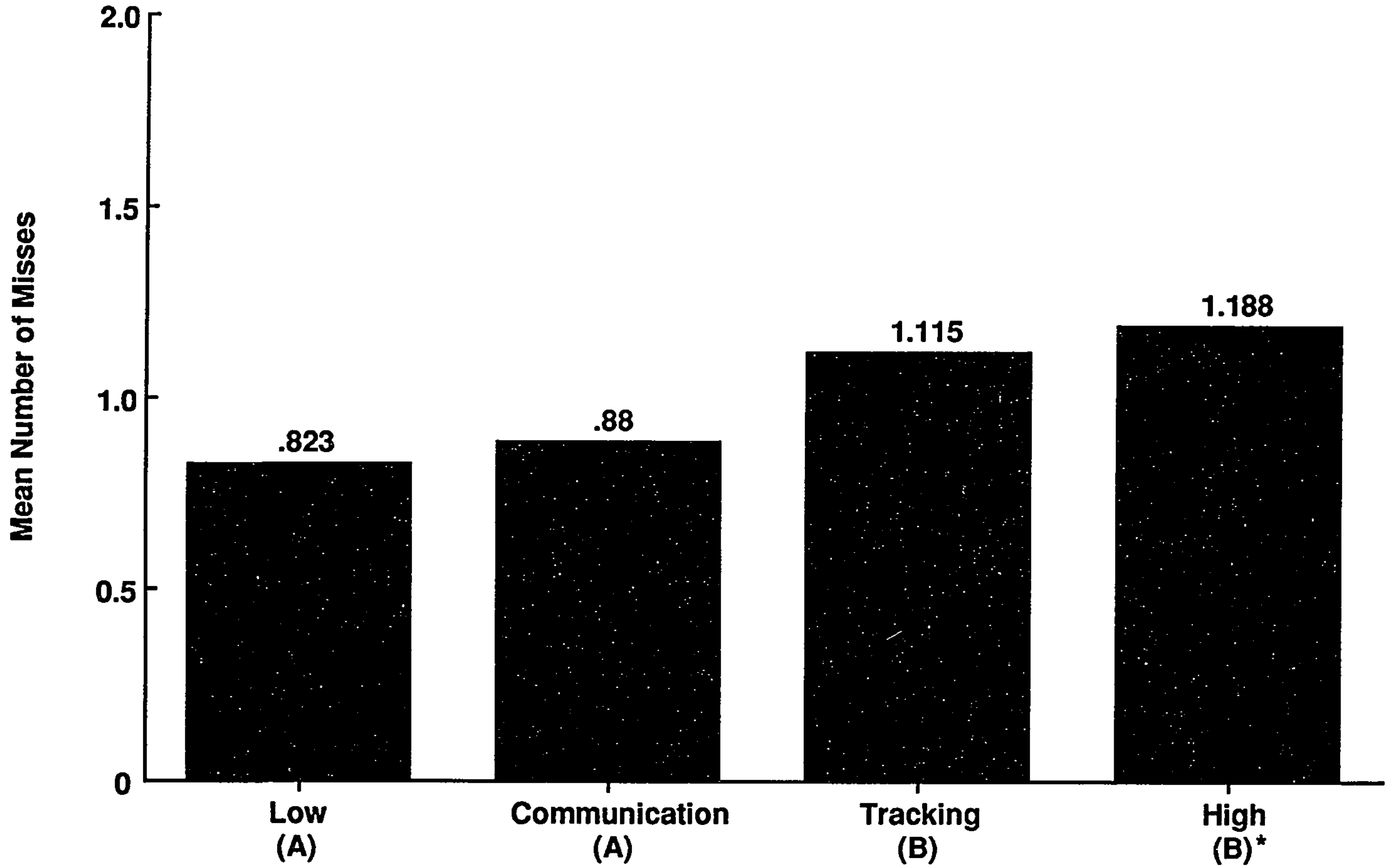


Figure 20. MEAN NUMBER OF MISSES TO DIAL SIGNALS AS A FUNCTION OF WORKLOAD
* Different letters indicate significantly different groups, ($p < .05$)

As with the red light and green light signals, the false alarm rate for dial signals did not show any significant variation by time or workload. Table 4 lists the sources of variation for dial reaction time, missed signals and false alarms.

In summary, reaction times to each of the three individual monitoring tasks became significantly faster as time into the tasks increased. The most marked decrease in reaction time occurred for responses to the dial signals and accuracy to dial signals also increased significantly as a function of time. Many subjects noted that they began to see the dial "jump out of range" peripherally as they became more experienced with this task. The green light signal was more subtle (the absence of this light) which could account for the fact that subjects were equally likely to miss this signal in the first hour as they were in the later hours of battery performance.

Subjects recorded significantly slower reaction times to the presence of red light signals during the lowest workload condition. Performance (both response time and hit rate) on the green light task was also reduced significantly during the low workload condition. However, accuracy to dial signals was greater during the low workload conditions and decreased during the higher workload conditions.

No effects for false alarm rates were found for any of the three monitoring tasks. These data were difficult to

Table 4

Summary of the Analysis of Variance for Dial Signals

Source of Variation	*df*	Mean Square	*F*	Eta Square
Reaction Time				
Hour	3	223.50	37.27*	0.13
Workload Level (WL)	3	9.86	1.95	
Hour * WL	9	4.94	1.47	
Subject	23	76.21	NT	NT
Subj * Hour	69	5.99	NT	NT
Subj * WL	69	5.05	NT	NT
Subj * Hour * WL	207	3.35	NT	NT
Missed Signals				
Hour	3	34.66	17.51*	0.08
Workload Level (WL)	3	6.10	7.19*	0.01
Hour * WL	9	0.89	0.76	
Subject	23	16.99	NT	NT
Subj * Hour	69	1.98	NT	NT
Subj * WL	69	0.85	NT	NT
Subj * Hour * WL	207	1.17	NT	NT
False Alarms				
Hour	3	357.79	1.88	
Workload Level (WL)	3	80.96	0.66	
Hour * WL	9	158.08	1.39	
Subject	23	991.83	NT	NT
Subj * Hour	69	190.60	NT	NT
Subj * WL	69	122.40	NT	NT
Subj * Hour * WL	207	114.10	NT	NT

* $p < .05$, NT = No Test

analyze because three subjects regularly pressed monitoring response keys in the absence of signals. Often, these subjects would press keys F1, F2, F3, F4, F5, and F6 in a rapid sequence as if to anticipate signals which may or may not be present. For all other subjects, however, false alarm rates were typically low (< 1.0 per condition).

Tracking

Tracking data were only collected during two task combinations, Tracking and High. Root mean square (RMS) error was utilized as the performance measure on this task. Again, the relevant independent variables were hour and workload level.

Analyses of variance yielded a main effect for hour into the task, $F(3,69) = 7.91$, $p<.05$. Performance (RMS error) improved significantly as hours into the task increased, with the exception of Hours 3 and 4, which were not significantly different from each other. These means are presented in Figure 21.

A main effect was also found for workload level, $F(1,23) = 5.85$, $p<.05$. The addition of the communication task (in the High workload condition) significantly increased RMS error (decreased performance) from the Tracking condition, as shown in Figure 22. The interaction between hour and workload was not found to be significant. All sources of variation for tracking RMS error are listed in Table 5.

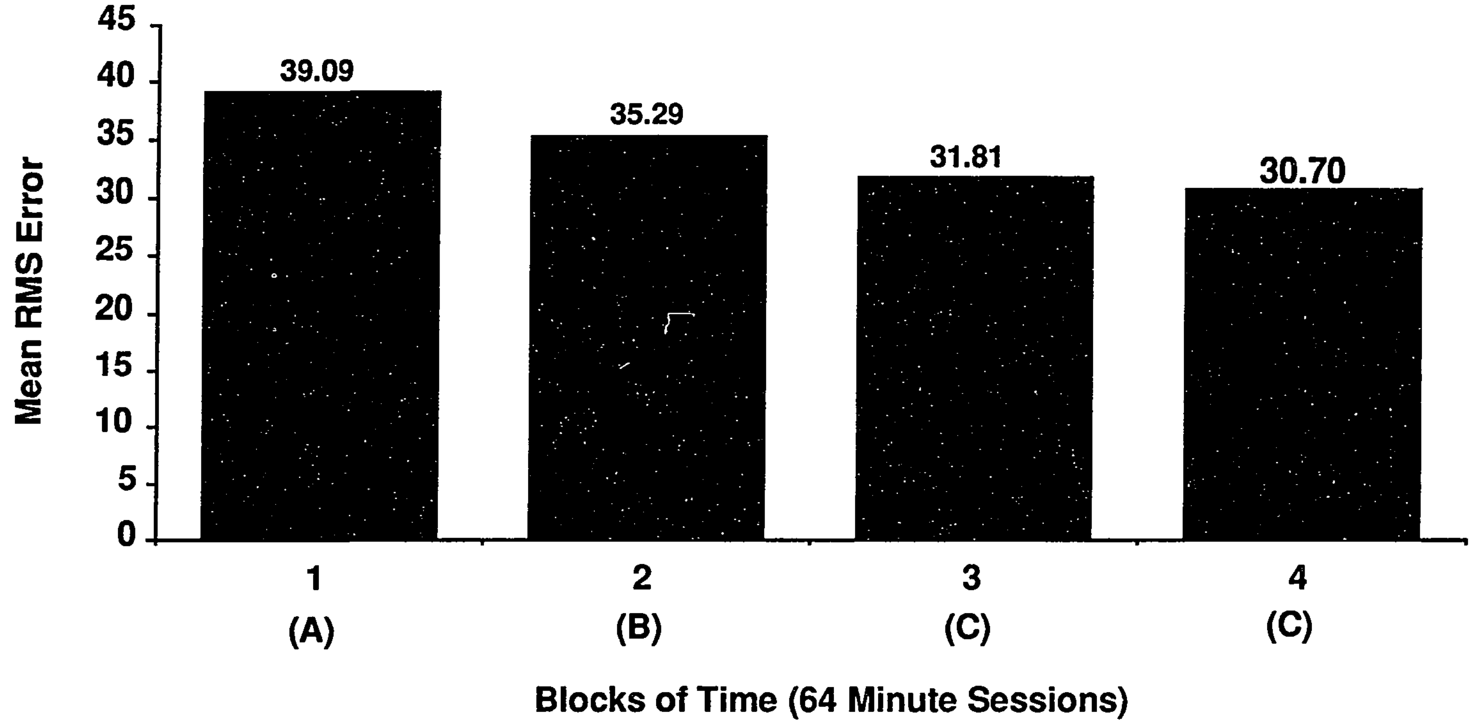


Figure 21. MEAN RMS ERROR FOR TRACKING AS A FUNCTION OF TIME

* Different Letters Indicate Significantly Different Groups, $P < 0.05$

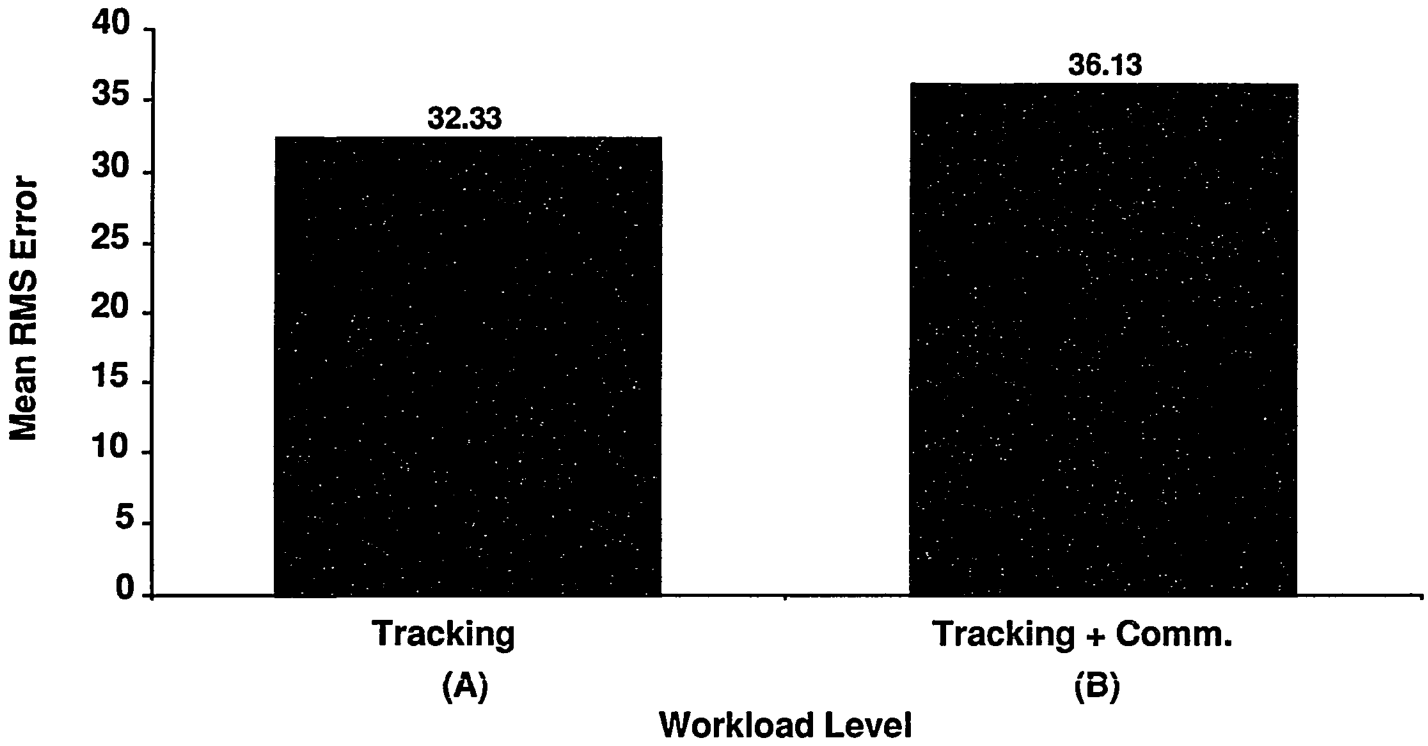


Figure 22. MEAN RMS ERROR FOR TRACKING AS A FUNCTION OF WORKLOAD LEVEL

* Different Letters Indicate Significantly Different Groups, $p < 0.05$

Table 5

Summary of the Analysis of Variance for Tracking RMS Error

Source of Variation	*df*	Mean Square	*F*	Eta Square
Hour	3	1379.27	7.91*	0.02
Workload Level (WL)	1	1386.41	5.85*	0.01
Hour * WL	3	54.34	0.96	
Subject	23	9801.08	NT	NT
Subj * Hour	69	174.45	NT	NT
Subj * WL	23	237.02	NT	NT
Subj * Hour * WL	69	56.78	NT	NT

* $p < .05$, NT = No Test

Communication

The communication task was required of subjects during only two of the four workload levels (Communication and High). The independent variables of interest were hour and task combination and the performance measures included response time and errors.

Analyses of variance for response time following the auditory message indicated main effects for both hour, $F(3,69) = 10.75$, $p<.05$, and workload level, $F(1,23) = 22.56$, $p<.05$. As illustrated in Figure 23, post-hoc tests revealed that response time decreased significantly after Hour 1 but did not differ between Hours 2, 3 and 4. Furthermore, for workload level, the addition of the tracking task (in the High workload condition) significantly increased response time, as shown in Figure 24. The interaction between hour and workload level was not statistically significant.

Communication errors were defined as incorrect channel or frequency selection. No false alarms occurred and were therefore not included. Analyses of variance for communication errors as a function of hour and workload level yielded no significant effects. In other words, communication accuracy did not change as either a function of time or workload. Table 6 summarizes the sources of variation for communication reaction time and errors.

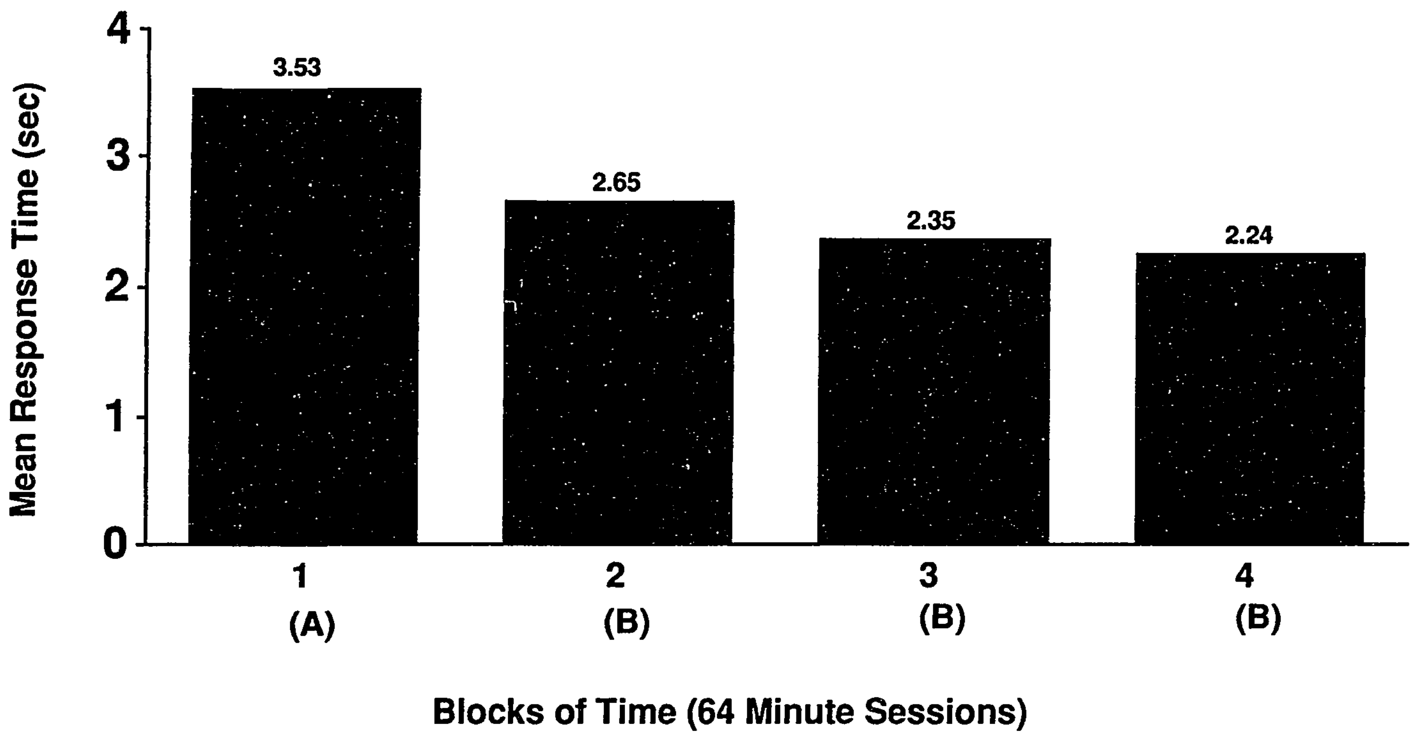


Figure 23.MEAN RESPONSE TIME FOR COMMUNICATION TASK AS A FUNCTION OF TIME

* Different Letters Indicate Significantly Different Groups, $p < 0.05$

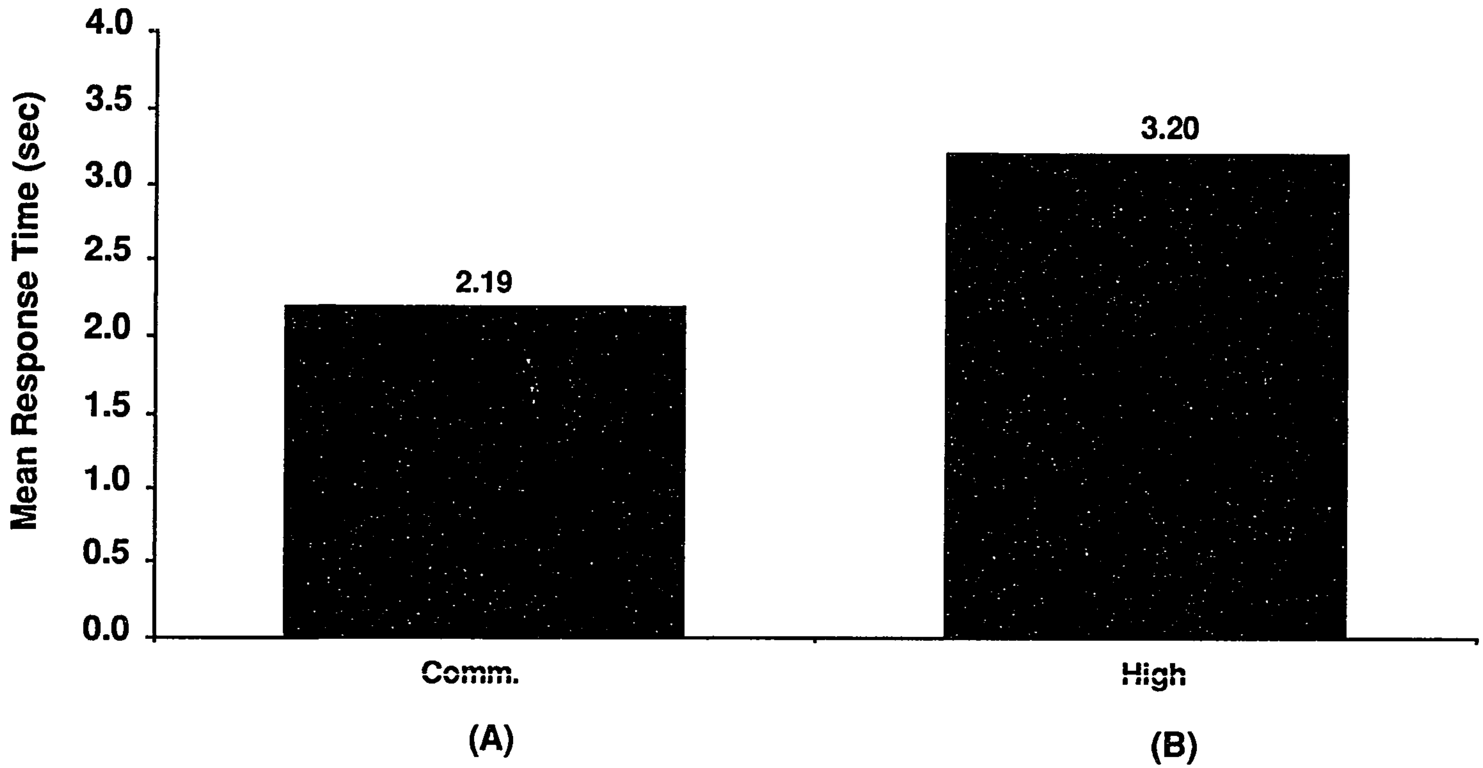


Figure 24.MEAN RESPONSE TIME FOR COMMUNICATION TASK AS A FUNCTION OF WORKLOAD LEVEL

* Different Letters Indicate Significantly Different Groups, $p < 0.05$

Table 6

Summary of the Analysis of Variance for Communication Response Time and Errors

Source of Variation	*df*	Mean Square	*F*	Eta Square
Reaction Time				
Hour	3	33.11	10.75*	0.06
Workload Level (WL)	1	97.47	22.56*	0.06
Hour * WL	3	0.35	0.16	
Subject	23	29.25	NT	NT
Subj * Hour	69	3.08	NT	NT
Subj * WL	23	4.32	NT	NT
Subj * Hour * WL	69	2.21	NT	NT
Errors				
Hour	3	0.23	1.04	
Workload Level (WL)	1	0.21	1.96	
Hour * WL	3	0.23	1.80	
Subject	23	0.26	NT	NT
Subj * Hour	69	0.22	NT	NT
Subj * WL	23	0.11	NT	NT
Subj * Hour * WL	69	0.13	NT	NT

* $p < .05$, NT = No Test

Resource Management

The mean deviation from target (2500) on the two main tanks (A and B) was used as the performance measure for the resource management task. Since subjects were instructed to keep fuel levels in these two tanks as close to 2500 gallons as possible, the mean deviation from this target was chosen as an indication of performance. Analyses of variance were calculated in order to examine how this mean deviation was affected by workload level and time into the task.

No main effect was found for hour into the task. Operators' performance on the resource management task did not change overall as a function of experience with the task. A significant main effect was observed, however, for workload level, $F(3,69) = 3.07$, $p<.05$. The first three groups (Low, Communication and Tracking) did not differ in terms of deviation from target. However, target deviations during the High task condition were significantly greater than the three lower workload conditions, as shown in Figure 25. The interaction between hour and task combination was not found to be significant. The sources of variation for resource management target deviation are summarized in Table 7.

Strategic Behavior

Response Frequency. The response frequency on the resource management task was utilized as an indication of

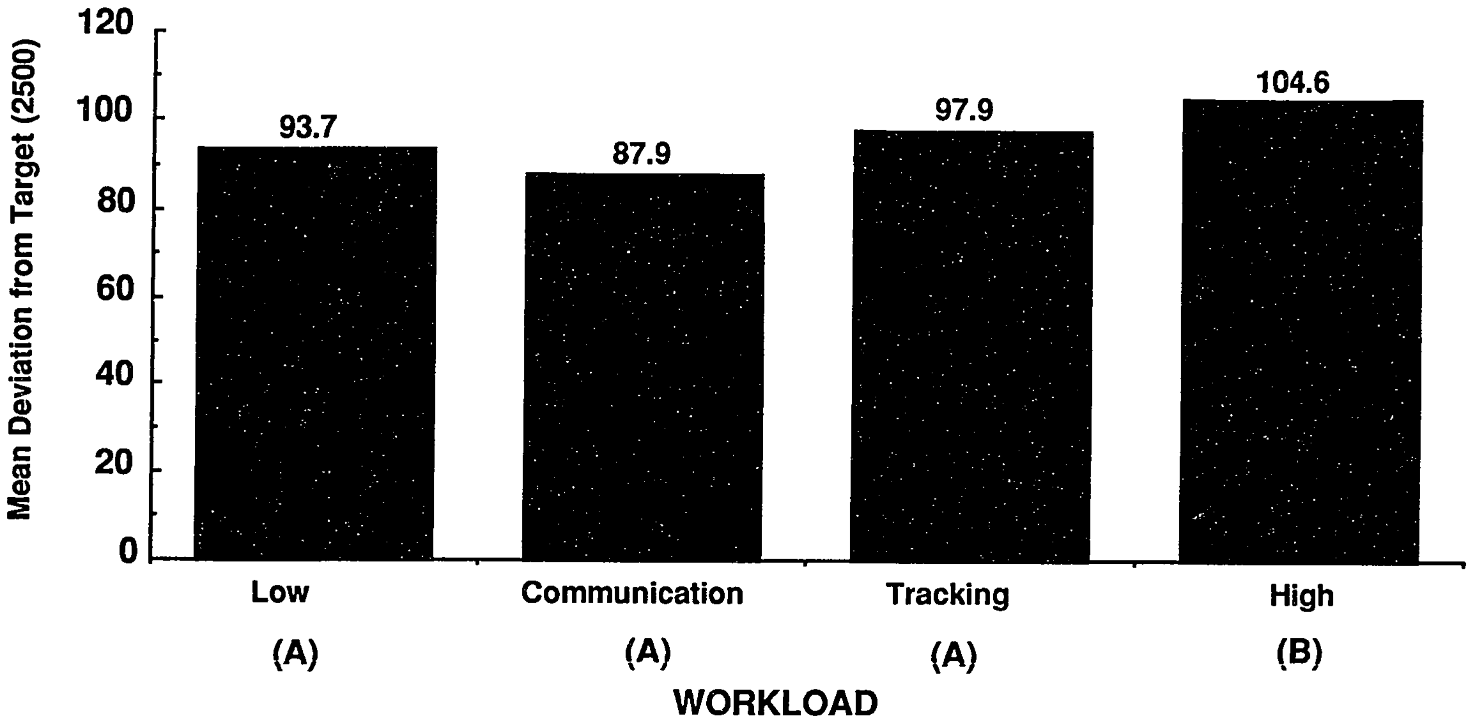


Figure 25. MEAN DEVIATION FROM TARGET FOR TANKS A & B OF THE RESOURCE MANAGEMENT TASK AS A FUNCTION OF WORKLOAD LEVEL

* Different Letters Indicate Significantly Different Groups, $p < 0.05$

Table 7

Summary of the Analysis of Variance for Resource Management

Source of Variation	df	Mean Square	F	Eta Square
Deviation From Target				
Hour	3	8059.72	0.89	
Workload Level (WL)	3	18922.31	3.07*	0.01
Hour * WL	9	2517.25	0.71	
Subject	23	159489.06	NT	NT
Subj * Hour	69	9119.91	NT	NT
Subj * WL	69	6115.49	NT	NT
Subj * Hour * WL	207	3568.69	NT	NT
Frequency of Responding				
Hour	3	123.45	0.11	
Workload Level (WL)	3	26378.26	9.38*	0.04
Hour * WL	9	635.40	1.46	
Subject	23	810154.34	NT	NT
Subj * Hour	69	1113.61	NT	NT
Subj * WL	69	2813.51	NT	NT
Subj * Hour * WL	207	435.54	NT	NT

* $p < .05$, NT = No Test

strategic behavior since it was shown with Experiment #2 that subjects decreased their frequency of responding as workload increased while maintaining performance levels. Response frequency was analyzed with the last set of data to determine whether this trend would be replicated over a longer duration of battery performance. These means are presented in Appendix G.

Analysis of Variance produced a main effect for workload level, $F(3,69) = 9.38$, $p < .05$. As illustrated in Figure 26, subjects changed pump configurations significantly more often during the first two task combinations (Low and Communication) than during the highest two task combinations (Tracking and High). This implies that during higher workload conditions, subjects developed a different method of responding which allowed them to maintain performance. This held true for the tracking condition; however, in the High workload condition, performance (mean tank deviations) did decrease significantly. This effect was found for all four 64 minute sessions and, thus, did not change as subjects became more experienced with the task. The sources of variation for resource management response frequency are presented in Table 7.

<u>Correlations between Target Deviations.</u> Looking at frequency of responding is one means of inferring a difference in responding strategy on the resource management

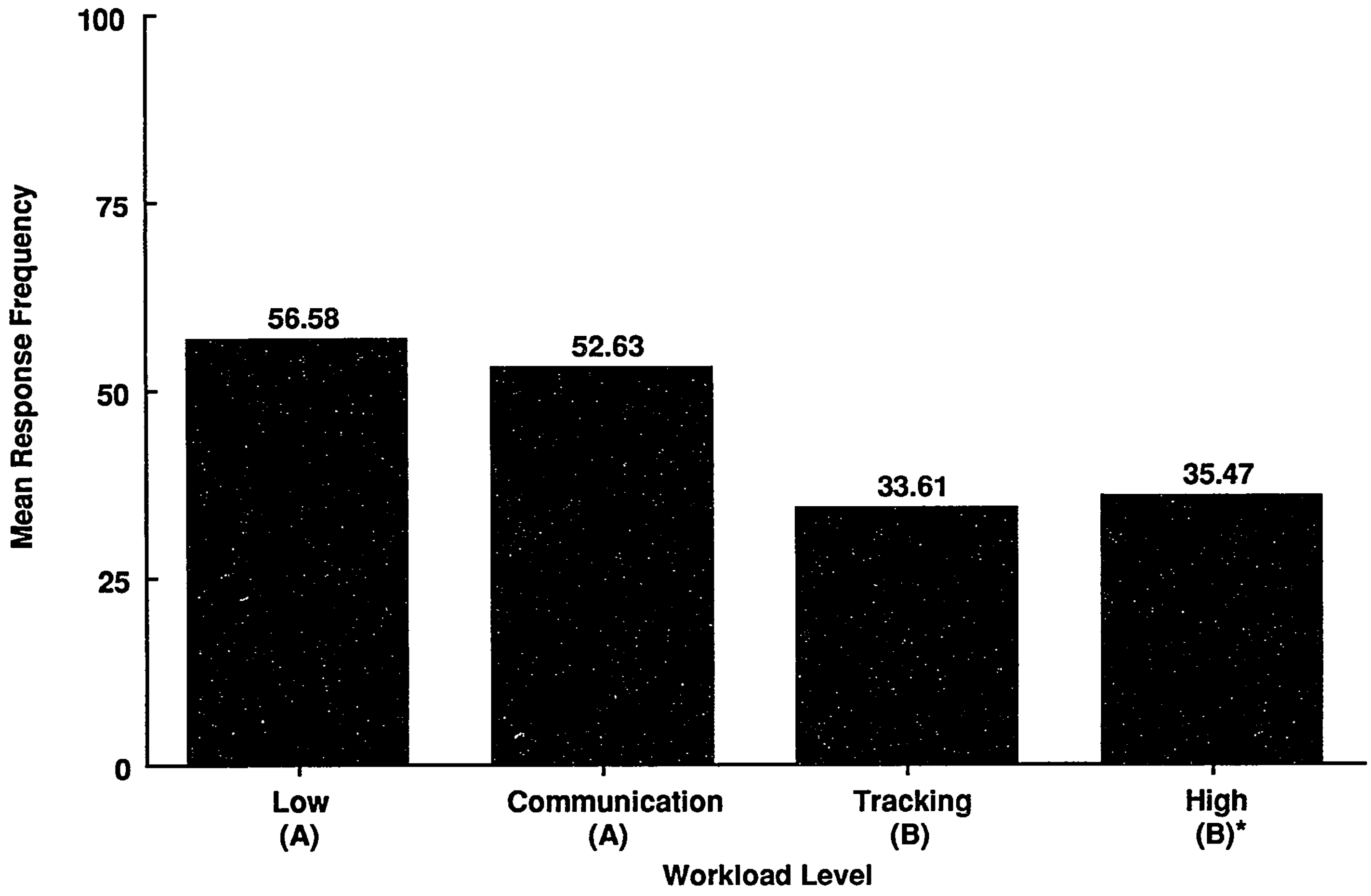


Figure 26. MEAN RESPONSE FREQUENCY ON RESOURCE MANAGEMENT AS A FUNCTION OF WORKLOAD LEVEL
* Different letters indicate significantly different groups, ($p < .05$)

task. When asked at the end of the experiment about their experience with the battery, many subjects (N=12) were able to verbalize a specific pattern of responding on the resource management task. The pattern described by the majority of subjects consisted of:

> (1) Activating pumps 1 and 3 (which are congruent) and pumps 2 and 4 (which are congruent);
>
> (2) When Tanks A and B were overfilled, but not out of acceptable range (this varied between and within subjects), pumps 1 and 3 were turned off.
>
> (3) When the levels in Tanks A and B fell to 2500 each, pumps 2 and 4 were turned off and pumps 1 and 3 turned on. This maintained the 2500 level until the supply tanks (C and D) became depleted.
>
> (4) This pump configuration was maintained until the subject determined the supply tank to be too low (this varied between and within subjects). At this time, pumps 2 and 4 were activated and pumps 1 and 3 turned off until the supply tanks were filled again to an acceptable level. Then, pumps 1 and 3 were turned back on and these four steps were repeated.

All 12 of the subjects who were able to verbalize a strategy had adopted this particular approach. However, a small number of subjects in the previous study (Experiment 2) described different patterns of responding. Additionally, it is not concluded that those subjects who could or did not verbalize this or any other strategy were not following a pattern of responding.

The strategy that is defined above consists of a fundamental assumption. This is that the levels in Tanks A and B are kept at the same level by the subject regardless of how far these levels deviate from target. When these levels deviate from target, they deviate to the same degree in the same direction (positive or negative).

Many subjects described a change in their mental representation of the resource management task over time. These subjects initially perceived this task as consisting of two separate subtasks: the maintenance of Tank A and the maintenance of Tank B. This was reflected in differences in responding between the two tank systems and, consequently, wide differences in the deviations from target in Tank A and Tank B. However, the flow rates, depletion rates and tank capacities were congruent between the three-tank system for Tank A and the three-tank system for Tank B. This information was given to subjects during training and was present in some form throughout each experimental session.

Apparently, with experience, many subjects began to perceive this task scenario as two identical systems. These subjects then made an effort to not only keep the main tanks as close to target as possible, but to also keep the level of Tank A very close to the level of Tank B. At that point, when a pump configuration change was made to the first tank system, the equivalent change was made to the second tank system (i.e., pump 1 on & pump 3 on). By doing this, subjects reduced their workload so that they were maintaining only one system.

The strategy described previously assumes congruent responses between the two systems which are made close together in time. For the subjects who utilized this strategy, the deviations between the Tank A deviation from target and the Tank B deviation from target were highly correlated.

In order to test the occurrence of this strategy statistically, correlations were made between the deviations from target of Tanks A and B for each subject for each hour of battery performance. Figure 27 illustrates the number of subjects who produced significant correlations for either 0, 1, 2, 3, or 4 hours of their experimental session. Nine of the 24 subjects maintained significant positive correlations for all four hours of the experiment. Five out of those nine subjects were among those who verbalized the strategy described previously.

Overall, analysis of variance revealed that the <u>mean</u> correlation across subjects did not differ significantly as a function of time. However, it is obvious from Figure 27 that each subject differed in terms of being able to develop or maintain significant correlations between the tank levels. Figure 28 shows how the <u>number</u> of significant correlations across subjects increased as a function of time into the experiment. During the first hour, 63% of subjects were able to maintain close levels between Tanks A and B. By the third and fourth hours, 75% of subjects showed significant correlations between the two Tanks. Again, it

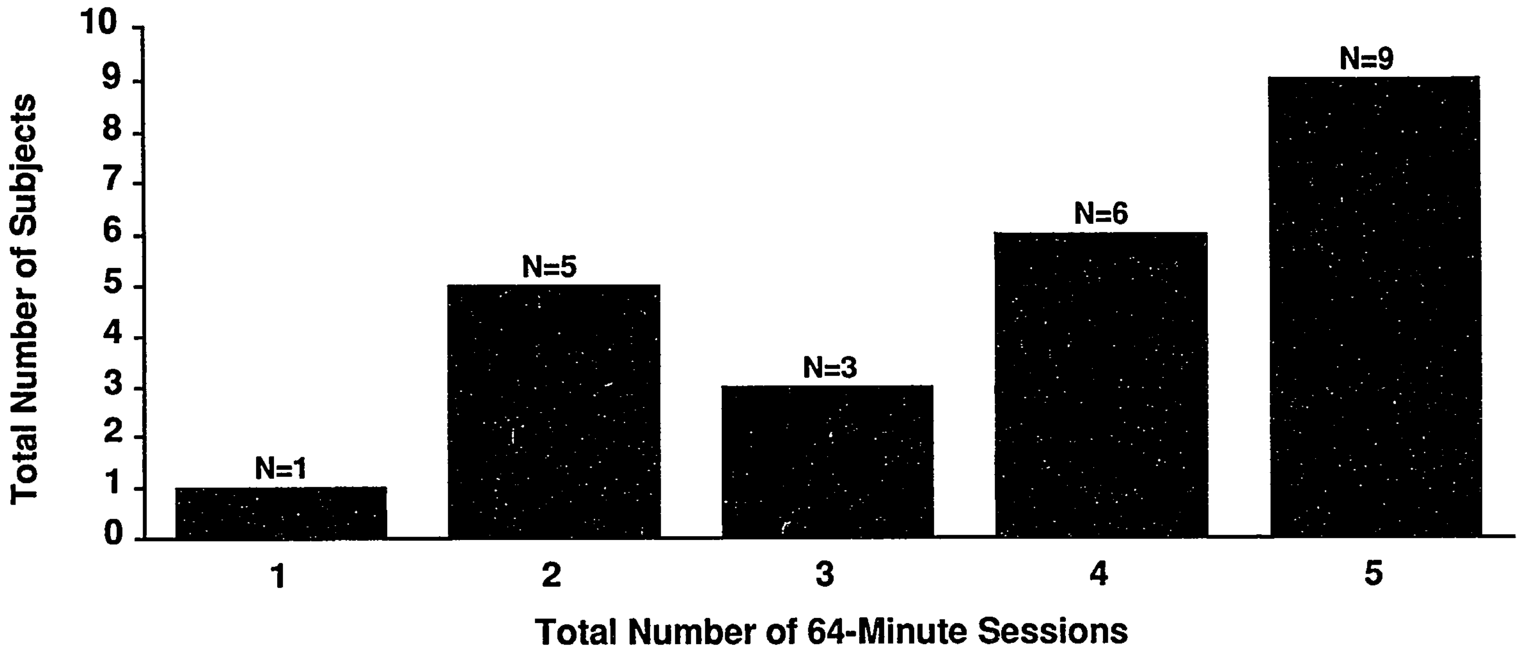


Figure 27. NUMBER OF HOURS THAT SUBJECTS MAINTAINED SIGNIFICANT CORRELATION (P<0.05) BETWEEN TANKS A & B

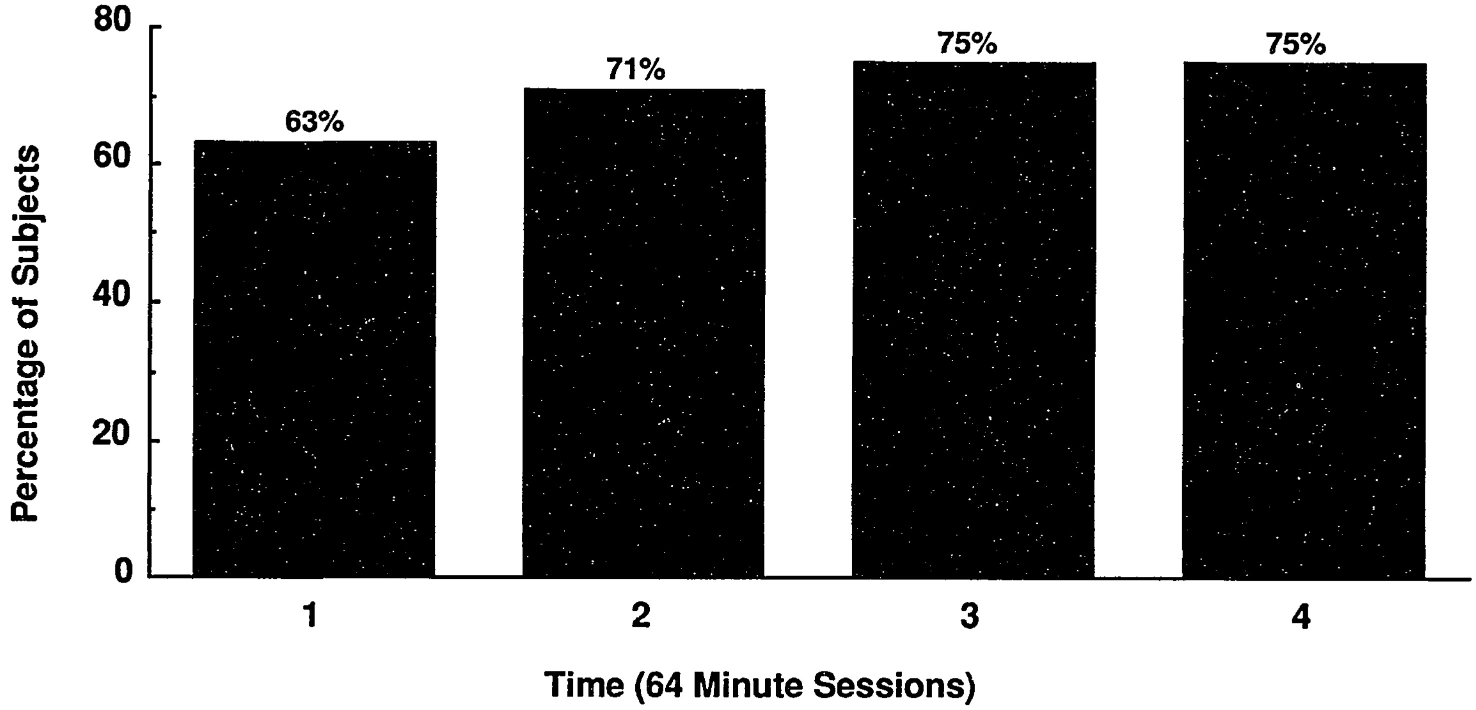


Figure 28. PERCENTAGE OF SUBJECTS WITH SIGNIFICANT CORRELATIONS BETWEEN TANKS A AND B (P<0.05) BY HOUR

is important to remember that some subjects never or rarely (one hour only) attained these significant correlations.

The strategy described here was not without drawbacks, however. Two of the subjects who used this strategy and had attained significant correlations during the first three hours, reported that during the last hour of battery performance, they had made an appropriate pump configuration change to the first system and, minutes later, realized that they had omitted the equivalent change to the second system. This type of complacency error led to a reduced correlation. Upon examining these particular subjects, it was found that the correlations were indeed very low (0.245 and 0.069) during those hours that these errors were noted by the subjects.

In summary, subject reports and correlational analyses of tank level deviations clearly indicate that strategic behavior was developed and maintained by many of the subjects. To examine the relationship between strategic behavior and performance on all four tasks, performance profiles for each subject were created and are listed in Appendix H. From this profile, each subject's performance can be examined for differences in time-sharing or attention to particular tasks as a function of strategic behavior.

<u>Performance Profile Data</u>. Subject performance profiles are referred to by subject number and can all be found in

Appendix H. One overall trend for subject performance is that if, for a given hour, performance on one task was good (better than the mean), then performance on one or more of the other tasks declined (particularly during high workload). Most often, the communication performance was the one task to suffer, which required subjects to retain and respond to an auditory message which was only presented once. Perhaps responding was retained on this task most often because as the other tasks became more automatic for subjects, short-term memory was freed to retain the message for a longer length of time before responding. The performance profile of Subject 11 demonstrates this phenomenon. Those subjects who did improve communication performance during high workload did so at the cost of monitoring, tracking or resource management performance (See Subject 14).

Few subjects followed the mean performance trend for all tasks (performance on tracking, communication and resource management declining as a function of increasing workload. Usually, performance on at least one of these tasks either did not decline or actually improved with workload, as if the subject were focusing on one task.

The majority of subjects who did follow the mean were those who showed increased correlations between the deviations of Tanks A and B on the resource management task. That is, these subjects seemed to be developing some sort of strategy similar to one described earlier in this section. For these subjects (i.e., Subjects 1, 3, 4 or 5), performance tended to be most likely a function of workload. Perhaps these subjects stayed in a strategy development mode, focusing on learning a pattern of responding rather than focusing on any particular task.

For the nine subjects who showed a significant correlation during all four hours of task performance, some interesting results were found. These subjects tended to have their best performance on the resource management task during the high workload condition. Most subjects showed better than average performance on resource management and the other tasks as well (Subjects 6 and 10), but regardless of performance level, this trend occurred (see Subject 17, Hours 1, 2, and 3 and Subject 10, Hours 2, 3, and 4). Frequently, monitoring and tracking performance were also improved during high workload, again at a cost to communication performance. These subjects may have developed their strategy early on into the experiment and could then take the time to prepare well for the high workload condition. These results and their implications are explored further in the Discussion section.

DISCUSSION

The present study involved the development of a task battery with which strategic behavior could be measured and investigated the effects of workload on operators' ability to maintain strategic behavior. The Multi-Attribute Task (MAT) Battery has been validated for use in workload and strategic behavior research. Future studies in these areas can utilize the flexibility of the MAT to manipulate aspects of a complex multiple task environment.

Strategic behavior was examined on one of the four tasks (resource management) of the MAT as a function of workload (addition of other tasks). Performance and strategic behavior measures were taken. In general, the hypotheses were supported by the data and evidence of successful strategies utilized by some of the subjects was also obtained.

The following sections discuss separately the results of the performance measures of the individual tasks, evidence of strategic behavior and the effects of workload as related to the original hypotheses, and implications for future research.

Performance Measures

Overall, performance on three of the four tasks (monitoring, tracking and communication) improved as a function of experience with the task battery. Performance on all four tasks was also dependent upon workload, although these effects were not entirely consistent with the hypotheses. Each task's performance is discussed below, followed by a summary discussing subjects' overall abilities to perform these multiple tasks.

Monitoring. In general, reaction time to all three types of monitoring signals decreased as a function of time. Mean reaction time differed between the three types of monitoring signals, with dial reaction times being the highest, followed by green light signals and red light signals, to which the lowest reaction times were reported. These differences support other studies which employed similar types of monitoring signals and found the same trends for speed of responding.

For example, in studies with the Multiple Task Performance Battery, researchers found that reaction times for a task similar to the dial monitoring task (meter monitoring) were higher than reaction times to light signals. These researchers also found that response times to the absence of a light were typically higher than response times to the presence of a light, as supported by the present study (Chiles, Jennings and Alluisi, 1978; Alluisi and Morgan, 1968).

This effect may have been intensified in the present study because the red light signal was the only piece of information on the Multi-Attribute Task Battery screen which was presented in the color red, whereas many other pieces of information (i.e., pump activation) were green. The uniqueness of the red light perhaps allowed greater ease of detection and the presence of other green signals on the screen may have made the already difficult task of identifying the absence of the green light even more difficult.

Accuracy of responding to the green light and dial signals increased with experience, but this effect was not observed for responses to the red light signals. Again, the ease of detecting this type of signal may account for this effect. The accuracy rate for red light signals was higher than accuracy for either green light or dial signals throughout all four hours of battery performance.

The lack of a vigilance decrement could be accounted for by two possible explanations. First, it is possible that the low workload condition, which consisted of monitoring and resource management, was not low enough to produce this effect. The addition of the resource management task may have added enough complexity to change any vigilance effects. Parasuraman (1986) defined a vigilance task as one involving the detection of signals presented at infrequent intervals for prolonged periods of time without rest. However, he made the distinction that multiple task situations often involve selective rather than sustained attention, thus adding new determinants of performance. The extremely low false alarm rate on all three monitoring tasks also precluded any of the standard decision theory analyses of vigilance, such as sensitivity or bias.

An alternative explanation for the lack of vigilance decrement is the length of uninterrupted task performance. Each sixty-four minute session was followed by some sort of rest period: either a five to ten minute break in the cases of Hours 1 to 2 and Hours 3 to 4, or the forty-eight hour break between Hours 2 and 3. Thackray (1990), in summarizing his research in monitoring performance, has stated that humans can be effective monitors for short periods of time, which he defined as not exceeding one hour in length with at least a five-minute rest period or change in activity.

Monitoring performance was also a function of workload. Subjects more accurately detected dial signals during low workload conditions than high workload. However, performance for red and green light signals was reduced during low workload. These results are inconsistent with Hypothesis 5 and other experiments utilizing similar monitoring tasks as part of a multiple task battery.

For example, in studies with the Multiple Task Performance Battery (MTPB) (Alluisi and Morgan, 1968), performance on these three types of monitoring tasks was best during low workload levels. However, the display layout of the Multi-Attribute Task Battery differed from the MTPB, which displayed these monitoring tasks in the top center of the display panel.

The present study also required subjects to perform the resource management task, which is located in the lower right corner of the screen (the monitoring task is in the upper left window). Subjects may have utilized the low workload time to focus on resource management performance or possibly plan for future higher workload conditions, which is discussed further in another section of this chapter (Strategic Behavior).

Therefore, although the stated hypotheses were not supported by the monitoring data, the display layout and the addition of resource management provide a possible explanation for this. By focusing on the resource management task, those tasks which were further from the visual field (red and green lights) suffered with reduced performance which was reflected in either higher reaction time or lower accuracy rates.

Tracking and Communication. Tracking and Communication performance improved with experience with the battery. This is consistent with the literature and the aforementioned hypotheses. Tracking RMS error and communication response time were also greater during the high workload condition than the tracking or communication conditions respectively. Wickens and Gopher (1977) studied dual task performance with tracking and digit-processing tasks and found that a dual task performance decrement was due to internal attention-switching behavior which manifested itself in decreased performance on both tasks.

Communication performance decreased during high workload as well. Subjects (all right-handed) operated the mouse for tracking with their right hand and therefore had to respond to the communication task with their non-dominant hand, which may have further added to the difficulty level of the high workload condition. It is also possible that this manual response to auditory input was a non-compatible combination. Some researchers have found that when a verbal response is paired with auditory input of verbal information, performance is unaffected by the addition of other tasks.

Wickens, Sandry and Vidulich (1983), for example, have tested various combinations of processing codes (verbal and spatial) and response modalities (speech or manual action) and have found that competition between tasks for manual response modalities (as with the tasks in the present study) increased task interference and disrupted time-sharing efficiency. Future studies could explore this effect further with the Multi-Attribute Task Battery. A verbal response to the auditory communication task might increase performance during all conditions involving the communication task and provide support for Wickens' theory.

Resource Management. The two performance measures taken for the resource management task were mean deviation from target and frequency of responding. Neither of these measures changed as subjects became more experienced with the task. However, as workload increased, both performance and frequency of responding decreased. Subjects decreased responding significantly during the Tracking and High workload conditions and mean deviation from target increased significantly during High workload. Perhaps subjects were attempting to plan ahead for these higher workload levels by attending more to the resource management task during the lower workload levels, which was supported by the performance decrement on the monitoring tasks during low workload levels.

The increased attention to the resource management task would be reflected in the higher frequency of responding during these lower workload periods and suggests a strategy for anticipating and managing higher workload levels. These strategies varied between and within subjects and are discussed further in the section on Strategic Behavior.

Summary. Although subjects were not trained to perform the four tasks presented with the MAT simultaneously, the general improvement in performance with experience leads to the conclusion that subjects were improving performance in the individual tasks and were developing an ability to perform multiple tasks. It is likely that certain tasks became automatic in nature, though the degree to which this occurred for any particular task probably varied between subjects.

Examination of the performance profiles shows that some subjects were able to maintain good tracking performance even with an increase in workload. Other subjects seemed to develop a pattern of responding with the resource management task that allowed that task to be processed automatically, as indicated by a stable or decreased deviation from target as workload level increased. Still others seemed to improve primarily in monitoring or communication performance.

A time-sharing ability should not be inferred from these data, however. Wickens (1984) has cautioned that improvements in time-sharing efficiency are due to the development of a time-sharing skill and not simply to increased automation of the component task. In order to test the emergence of time-sharing skill on the tasks of the Multi-Attribute Task Battery, future studies should examine whether performance for a given combination of tasks develops more rapidly when the tasks are performed together than when the same tasks are performed individually.

Further, if such a time-sharing skill exists for one task combination, it should transfer to a qualitatively different task combination (Wickens, 1984).

Strategic Behavior.

It was previously hypothesized that certain sequences of responding would correspond with better overall performance than other patterns of responding and that these strategies would develop with experience. The data collected did provide indirect evidence for most of the hypothesis. At least one strategy was identified: the majority of subjects responded in such a way as to maintain the levels of Tanks A and B at nearly equal levels. This allowed subjects to simplify the resource management task from the maintenance of two unequal systems to the maintenance of virtually one system.

In most cases, this significant correlation between tank levels was not associated with better resource management performance. However, since these significant correlations were usually associated with improvements on one or more of the monitoring, tracking or communication tasks, it is possible that these subjects developed this pattern of responding so that they could switch their attention to other tasks. Future studies might employ oculometric techniques in order to determine if eye movement data would support this hypothesis. If so, subjects would spend a higher percentage of their time fixating on the monitoring, tracking and communication tasks than the resource management task after developing this particular strategy.

<u>Planning Behavior as a Strategy</u>. Many subjects improved correlations between tank deviations as they became more experienced with the battery, which supported the hypothesis that strategies would develop over time. These subjects' performance on the individual tasks tended to reflect the overall mean fairly closely. That is, their performance on individual tasks declined as workload level increased.

Other subjects, however, showed significant correlations for all four hours, therefore developing this particular strategy very early in the experiment. The majority of these latter subjects followed a tendency to reduce resource management deviations significantly during the high workload levels. In other words, resource management performance improved during the higher workload levels.

With the given pump flow rates and tank capacities for this task, it is possible to set up the supply tanks to keep the incoming rates to the main tanks equal to these tanks' depletion rates for a maximum of ten minutes before the supply tanks are depleted. This translates into zero change in tank level for up to ten minutes. Since the scheduling window of the MAT provides information about the beginning and duration of the high workload levels before these changes occur, subjects can plan ahead and set up the tanks for this period and then attend primarily to other tasks for up to ten minutes.

It is hypothesized that those subjects who became skilled at their strategy of correlating the two tank levels (those with significant correlations all four hours) were able to use this strategy to plan for future high workload periods (which were eight minutes in duration). Those subjects who had significant correlations only during the last two hours did not show this planning strategy (reflected by a significant drop in deviation from target). However, if these subjects were to perform the battery for two more hours, it is likely that they would develop the same planning behavior as well.

Hart (1990) has stated that operators who can predict changes in advance and utilize this information achieve better workload management strategies than those who either cannot predict changes or do not utilize this information. Some subjects in the present study were able to manage high workload better when they became skilled at a particular response pattern. If workload ratings had been taken for each workload level, the subjects who were able to plan for high workload would likely rate this period lower in terms of subjective mental workload than those subjects who did not utilize this strategy.

Another line of research which should be explored given the establishment of strategic behavior is the effect of strategy disruption by an unexpected pump failure. A pump failure would be displayed as a red signal on a particular pump status box. The pump failure would indicate that a particular pump would be inactive for an unspecified length of time. Since the strategies described for the resource management task rely on activating equivalent pumps simultaneously, this strategy would be disrupted. The recovery from this disruption could be measured. Subjects may change their method of planning ahead or fall back on simply reacting to pump failures as they occur. This would probably depend on the frequency of pump failures or how "likely" the subject believed that these failures were.

Strategy errors (Slips). Fifteen out of 24 subjects (63%) either showed significant correlations during all four hours or improved correlations over time (significant correlations during the last two or three hours). Of the nine subjects who did not have a pattern of improving correlations between tanks, two described an "error" during the last hour of the experiment (the only hour without a significant correlation). These subjects (e.g., Subject 22) reported that they had been following the strategy of identical responding to both tank systems but realized that they had missed a response to the second system, resulting in an extreme deviation within Tank B. Both subjects stated that they believed that they had made the second response, but that upon identifying the deviation, realized that they "must have forgotten."

These errors provide support for research by Morris and Rouse (1985) who have found that experts often make slips, or errors based on overfamiliarity or overconfidence in system performance or their own ability as opposed to novices who make mistakes based on lack of experience or knowledge. The slips that experts made on this task occurred during the last hour of battery performance during which their pattern of responding on the resource management task was well established. Morris and Rouse (1985) have stated that these slips occur during largely automatic execution of a well-known action sequence and that slips flourish in an environment where there is little novelty. Operators then rely on the belief that since nothing has ever gone wrong before, nothing will go wrong in the future.

Overconfidence in Automation. Slips occurred in the present study because subjects became overconfident in their own performance. Many researchers are also interested in the effects of overconfidence in automated systems which are often found in the flight environment (Morris and Rouse, 1985; Ternham, 1978).

The Aviation Safety Reporting System has published many air crew reports describing over-reliance on automated systems which have led to accidents or near accidents (Billings, Lauber, Funkhouser, Lyman & Huff, 1976). Wiener (1988) has also noted that one result of automation may be excessive reliance upon those automated systems, whether these automated systems are found in the flight environment or other complex environments such as nuclear plants.

Another possible result of automation is that the operator loses a clear internal representation of the system since he or she is depending on automation. This lack of a clear mental representation would then be a detriment to the operator during an abnormal situation where he or she is called upon to perform the tasks manually. Braune and Trollip (1982) have concluded that expectations about resolving a variety of unexpected situations cannot be made without a good internal representation about the system of operation.

The Multi-Attribute Task Battery can be modified to allow subjects control over automation of the tracking task. Studies examining the effects of control of automation in a multiple task environment may be a step in sorting through these issues. Further experiments could examine the use of automation as a workload management strategy.

As mentioned previously, failures in the resource management task would lead to a decreased confidence in the system performance of this task. These pump failures would disrupt the largely automated response pattern that many subjects develop for performing this task. However, the development of an automatic response pattern indicates the existence of a good mental representation of the system on the part of these subjects. Therefore, these subjects would probably recover from system failure more quickly than subjects who had not developed a pattern of responding on this task.

Strategies and Performance. Overall, those subjects who adopted the strategy of aligning the two tank systems did not show better performance overall on battery performance. It is possible that the other subjects may have adopted strategies that were not captured with the techniques utilized in this study. Future research might also utilize verbal protocols to determine what other types of strategies are being utilized.

Hart (1990) is one of very few researchers who have proposed a framework for types of strategies as a function of workload. She has found that during low workload periods, operators may perform unrelated tasks to keep alertness up. Several subjects in the present study may have been trying to maintain alertness by making irrelevant entries or key presses to the communication task.

According to Hart (1990), moderate workload yields reactive strategies, where subjects are merely responding to stimuli as they occur without the resources to plan for future events. High workload, therefore, should cause subjects to defer or shed some tasks by delaying a response or delegating tasks to another individual or to automation.

Hart (1990) has suggested that tasks can be ordered by priority, difficulty, duration or by the resources required of them. Some subjects may have focused on tasks that were difficult for them with the goal of improving performance on that task. An alternative to this is that subjects may have deferred monitoring and resource management tasks when communication or tracking were required with the knowledge that they would be able to return to them whereas there was a limited period of time to perform the communication and tracking tasks.

Because the tasks were not assigned any priority during the instructions, it was difficult to assess whether or not subjects assigned their own priorities to the tasks. Assigning priorities to the individual tasks in future studies, either directly, or through consequences for poor performance, may allow the emergence of different strategies for allocating time to the individual tasks.

It is important to assess whether different strategies are associated with better performance than other strategies, because if this is true for a given task or job, there are important training implications. For example, techniques for scanning radar screens can be taught to air traffic controllers if it is shown that certain techniques produce better performance than others. However, equally successful air traffic controllers may have qualitatively different scanning strategies. Research which addresses the issue of training an individual in a performance strategy which might be different from the strategy that he or she would naturally adopt without specific training would provide some much needed answers in this area.

Problems may still arise when two controllers with different strategies are communicating with one another. The effects of individual differences in strategic behavior on team performance is yet another area which is open for research.

Summary

Strategic behavior had been defined previously in this paper as the following:

> The action (or inaction) that an operator takes in order to change the task structure, sequence of responding, or allocation of mental resources with the purpose of achieving a more manageable workload, dealing with unexpected change in the environment or achieving one's goal safely and efficiently.

In the present study, there is evidence that subjects developed a sequence of responding on the resource management task which allowed them to manage the high workload periods. During the low workload condition, subjects deferred responding to monitoring so that they could allocate mental resources to the resource management task and potentially use the low workload condition as a strategy development period.

The data from this study also provide support that skilled operators use different strategies than novice operators but that skilled operators are still prone to errors. These errors, based on overconfidence or complacency, were found to occur in the present study when subjects showed evidence of developing their strategic behavior into a highly automatic pattern of responding during which these errors were not immediately noticed.

Many current work environments are highly complex and require operators to perform multiple tasks simultaneously. Standard measures such as reaction time and errors, while important measures of performance, may not be enough to capture the characteristics which define a good operator. In order to train individuals to become the best possible operators of these highly complex environments or to design the work environment to provide a compatible system for safe and efficient performance, other measures of performance are needed.

Strategic behavior is one of such measures which also attempts to capture the planning, resource allocation and prioritization of tasks which all occur during complex task situations. It is important that the concept can be operationally defined and measured before further applications are made. The present study provided a good first step in this direction.

Given the existence of strategic behavior within a certain task situation, one important question becomes that of training. Should we train operators to utilize a particular strategy or allow them to develop the strategy that they choose as long as overall performance is not affected? Other questions are also raised. These include the effects of an abnormal condition that disrupts a pre-existing strategy and the understanding of errors that can be made even when an individual has become an expert at a particular strategy. Future research which addresses these issues will lead to a more thorough understanding of complex multi-task performance.

REFERENCES

Alluisi, E.A. and Morgan, B.B. (1968). Effects of practice and work load on the performance of a code transformation task (COTRAN). NASA CR-1261. Washington, D.C.: National Aeronautics and Space Administration.

Berg, S.L. and Sheridan, T.B. (1984). Measuring workload differences between short-term memory and long-term memory scenarios in a simulated flight environment. Twentieth Annual Conference on Manual Control. Moffett Field, CA: NASA-Ames Research Center.

Billings, C.E., Lauber, J.K., Funkhouser, H., Lyman, G., and Huff, E.M. (1976). NASA Aviation Safety Reporting System Report No. 76-1. Moffett Field, CA: NASA Technical Memorandum TM-X-3445.

Bisseret, A. (1981). Application of signal detection theory to decision making in supervisory control. Ergonomics, 24, 81-94.

Braune, R.J. and Trollip, S.R. (1982). Towards an internal model in pilot training. Aviation, Space and Environmental Medicine, 53(10), 996-999.

Chiles, W.D., Jennings, A.E., and Alluisi, E.A. (1978). The measurement and scaling of workload in complex performance. U.S. Department of Transportation: FAA-AM-78-34.

Damos, D.L., Smist, T.E. and Bittner, A.C. (1983). Individual differences in multiple-task performance as a function of response strategy. Human Factors, 25(2), 215-226.

Foushee, H.C. and Helmreich, R.L. (1988). Group interaction and flight crew performance. In E.L. Wiener and D.C. Nagel (Eds.), Human Factors in Aviation. New York: Academic Press, Inc.

Fuld, R.B. (1987). Computer monitoring vs. self-monitoring: The impact of automation on error detection. Unpublished master's thesis, University of Illinois: Urbana-Champaign.

Giffin, W.C. and Rockwell, T.H. (1984). Computer-aided testing of pilot response to critical in-flight events. Human Factors, 26(5), 573-581.

Green, D.M. and Swets, J.A. (1966). Signal Detection Theory and Psychophysics. New York: Wiley.

Hart, S.G. (1990). Pilots' workload coping strategies. Proceedings of Challenges in Aviation Human Factors: The National Plan. Washington D.C.: American Institute of Aeronautics and Astronautics.

Hart, S.G. (August, 1987). Research Papers and Publications (1981-1987): Workload Research Program. NASA Technical Memorandum 100016.

Hayes-Roth, B. and Hayes-Roth, F. (1979). A cognitive model of planning. Cognitive Science, 3, 275-310.

Jennings, A.E. and Chiles, W.D. (1977). An investigation of time-sharing ability as a factor in complex performance. Human Factors, 19, 535-547.

Jensen, R.S. and Benel, R.A. (1977). Judgment evaluation and instruction in civil pilot training (Final Report FAA-RD-78-24). Springfield, VA: National Technical Information Service.

Johannsen, G. and Rouse, W.B. (1983). Studies of planning behavior in normal, abnormal and emergency situations. IEEE Transactions on Systems, Man, and Cybernetics, SMC-13(3), 267-278.

Keele, S.W., Neill, W.T., and DeLamos, S.M. (1977). Individual differences in attentional flexibility (Technical Report). Eugene, OR: University of Oregon, Center for Cognitive and Perceptual Research.

Knaeper, A. and Rouse, W.B. (1983). A model of human problem solving in dynamic environments. Proceedings of the Human Factors Society - 27th Annual Meeting. Santa Monica, CA: Human Factors Society.

Mackie, R.R. (1987). Vigliance research - Are we ready for countermeasures? Human Factors, 29(6), 707-723.

Morris, N.M. and Rouse, W.B. (1986). Human operator response to error-likely situations in complex engineering systems. NASA-CR-177484. Moffett Field, CA: NASA-Ames Research Center.

Morris, N.M. and Rouse, W.B. (1985). An experimental approach to validating a theory of human error in complex systems. Proceedings of the Human Factors Society - 29th Annual Meeting. Santa Monica, CA: Human Factors Society.

Morris, N.M. and Rouse, W.B. (1983). The effects of level of knowledge upon human problem solving in a process control task. Proceedings of the Human Factors Society - 27th Annual Meeting. Santa Monica, CA: Human Factors Society.

Parasuraman, R. (1986). Vigilance, monitoring and search. nI Boff, K.R., Kaufman, L., & Thomas, J.P. (Eds), Handbook of Perception and Human Performance. Volume II: Cognitive Processes and Performance. New York: John Wiley and Sons.

Reason, J. (1983). On the nature of mistakes. In N. Moray and J.W. Senders (Eds.), Preprints of the NATO Conference on Human Error, Bellagio, Italy.

Schiffrin, R.M. and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 84(2), 127-190.

Sheridan, T.B. and Ferrell, W.A. (1974). Man-machine Systems: Information, Control, and Decision Models of Human Performance. Cambridge, MA: MIT Press.

Slovic, P., Fischoff, B., and Lichtenstein, S. (1982). Facts versus fears: Understanding perceived risk. In D. Kahneman, P. Slovic and A. Tversky (Eds.), Judgment under Uncertainty: Heuristics and Biases. Cambridge: Cambridge University Press.

Stokes, A.F. and Wickens, C.D. (1988). Aviation displays. In E.L. Wiener & D.C. Nagel (Eds.), Human Factors in Aviation. New York: Academic Press, Inc.

Svenson, O. (1985). Cognitive strategies in a complex judgment task: Analyses of concurrent verbal reports and judgments of cumulated risk over different exposure times. Organizational Behavior and Human Decision Processes, 36(1), 1-15.

Ternhem, K.E. (1978, April). The automatic complacency. Aerospace Safety, 12-13.

Thackray, R.I. (1990). Controller vigilance and monitoring performance. Proceedings of Challenges in Aviation Human Factors: The National Plan. Washington, D.C.: American Institute of Aeronautics and Astronautics.

Thorndyke, P.W. (1980). Performance models for spatial and locational cognition (Technical Report R-2676-ONR). Washington, DC: Rand Corporation.

Welford, A.T. (1978). Mental work-load as a function of demand, capacity, strategy and skill. Ergonomics, 21(3), 151-167.

Wickens, C.D. (1984). Engineering Psychology and Human Performance. Glenview, ILL: Scott, Foresman & Co.

Wickens, C.D. and Gopher, D. (1977). Control theory measures of tracking as indices of attention allocation strategies. Human Factors, 19(4), 349-365.

Wickens, C.D., Mountford, S.J., and Schreiner, W. (1981). Multiple resources, task-hemispheric integrity, and individual differences in time-sharing. Human Factors, 23(2), 211-230.

Wickens, C.D., Sandry, D.L., and Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing and output. Human Factors, 25(2), 227-248.

Wiener, E.L. (1989). Human factors of advanced technology ("Glass Cockpit") transport aircraft. NASA CR-177528. Moffett Field, CA: NASA-Ames Research Center.

Wiener, E.L. (1988). Cockpit automation. In E.L. Wiener & D.C. Nagel (Eds.) Human Factors in Aviation. New York: Academic Press.

Williges, R.C., Williges, B.H., and Fainter, R.G. (1988). Software interfaces for aviation systems. In E.L. Wiener & D.C. Nagel (Eds.), Human Factors in Aviation. New York: Academic Press, Inc.

Appendix A

Instructions to Subjects

PROJECT MAT

The overall purpose of this research is to understand pilot performance during different phases of flight. The task which is displayed before you is a computerized simulation of the kinds of tasks that pilots perform. Each window of the screen represents a different kind of task, as indicated by each heading: monitoring, tracking, communications and resource management, for example. Eventually, subjects will be asked to perform all of these tasks together. However, the purpose of today's study is to examine performance on these tasks individually. That is, you will only be asked to perform one of these tasks.

SYSTEM MONITORING

Specifically, the task which you will be performing today is the system monitoring task. The only information that will be relevant to you will be found in the monitoring window. You will not need to attend to any other task.

This task consists of two parts: lights and dials. You will be monitoring the two lights at the top of this task for any changes. You will also be monitoring the four dials beneath them for any directional changes in the fluctuation of the pointer. Let me demonstrate how this task will appear in the "normal condition." (Begin DEMOT.DTB) (Pause).

Let me first explain the changes that can occur with the lights. As you can see, during normal conditions, the left light is on in green. (Pause off). But occasionally, this green light will go out. When this happens you must press the "F5" key as indicated next to that light. You will receive feedback in that the light will immediately turn back on. (Pause on). Any questions?

The second light is normally off (pause off), but occasionally, a red light will turn on in this position. To respond to this, you must press the "F6" key, also indicated next to that light. Again, as soon as you respond correctly, the red light will disappear. To summarize, you will be monitoring the lights for the absence of the green light, and for the presence of the red light on the right.

When these events occur, you must respond to them. Now, I'll let you practice for a few seconds. (One occurrence of each light condition is presented to the subject). Any questions? (pause on)

The second part of this task consists of monitoring the four dials below the lights. Normally, the yellow pointers fluctuate from one unit below to one unit above the center line. (pause off) Is that obvious to you? Your task is to monitor these four dials and detect any change in the fluctuation of the pointer. In other words, if the pointer of one of these dials fluctuates either above or below the normal range, you must respond. The correct response is the key that is indicated below the dial which is out of range. (one example each of below and above range fluctuations will be presented - experimenter responds).

You'll notice that feedback to a correct response is given by the presence of a yellow bar at the bottom of the dial that was out of range. Again, the abnormal fluctuation can occur in either direction - above or below - but there is only one response per dial. Any questions? Now, I'll let you practice with the dials for one minute. (Subject is presented with two signals to respond to).

During the experimental task, you will be monitoring both lights and dials and looking for changes in any of them for a sustained period of 24 minutes. Respond as quickly and as accurately as possible.

TRACKING

The task that you will be performing today is the tracking task. All of the information that you need to perform this task is contained in the section titled "Tracking". Are you right-handed? (Mouse and pad are placed on the side of handedness).

Have you ever used a mouse before?

IF YES...

What have you used the mouse for? (This is noted). Go to (***).

IF NO...

The mouse is one way of controlling your position on the screen. Typically, you would use the arrow keys to move up and down and right and left. The mouse is another way to move around. The mouse pad correlates roughly to the area of the screen so that if you move the mouse up on the pad, your position on the screen is moved up accordingly. The same is true for moving right, left, diagonally, etc. Basically, if you wish to move to a different area of the screen, you must move your mouse in that identical direction on the mouse pad.

One important point to remember about the mouse is that you may have to move the mouse farther with your hand then the distance on the screen indicates (experimenter demonstrates this). Since there is not much room on the pad, you may have to pick up the mouse and set it back down to continue your movement. As long as the mouse is not touching the pad or any other surface, you will not affect your position on the screen.

***(DEMOT.DTB On and paused).

The overall purpose of this task is to keep the plane, represented by the green circle, within the dotted rectangular area in the center of this task.

(Pause off). If you do not control the plane with the mouse, the plane will drift away from the center. You must control the plane with movements of the mouse. Basically, you must compensate for this random drifting by pulling the plane back to center with corresponding movements with the mouse. For example, if the plane is drifting to the right, moving the mouse to the left will return the ship to center. Most of the time, however, you will be working in two dimensions: horizontal and vertical. So you will be making many diagonal movements. Let me demonstrate (experimenter controls mouse). Watch both the screen and the mouse. You'll notice that, if the plane is away from the center, you must make rather large movements to return it. If the plane is already in the center, smaller movements will be required. Now, you practice for a few minutes before we start the experiment. (Those who have not used a mouse require two demot.dtb runs).

Remember, the overall purpose of this task is to keep the plane in the center rectangular area. Try to maintain this at all times. If the plane leaves the rectangular area, try to return the plane to center as quickly as possible.

COMMUNICATIONS

The task that you will be performing today is the communication task. All of the information that you need from the screen is contained in the lower left corner area under the title "Communications". The overall purpose of this task is to discriminate between audio signals which will be presented through headphones and to respond as indicated.

The messages that you will hear begin with a six-digit call sign followed by a command. These call signs consist of three letters followed by three numbers. You must respond only when you hear your personal call sign which is NGT504. This number will remain on the screen at all times as a reminder to you. Any other call sign is not meaningful to you. Your call sign will sound like this. Please listen to the first part of the following message: Ctl F1. The call sign will be presented twice so that you can identify it precisely. Other call signs will be presented in the same way. The following is an example of a different call sign: Ctl F2. Do you feel comfortable discriminating between your call sign and other call signs? (If no, repeat other examples).

The second part of the message involves a command and you must respond to those that follow your call sign only. Do not respond if the message begins with a different call sign. The second part of the message is a command requiring you to change one of the frequency numbers listed on the screen. Let me explain this part of the task. There are four channels listed in the left column on the screen: NAV1, NAV2, COM1, COM2. These will be referred to in the audio message as: First Navigation, Second Navigation, First Radio and Second Radio. Notice that COM1 and COM2 are referred to as radio channels. In the right column are the frequency numbers that correspond to each channel.

The following is another example of one of these messages: Ctl F1. In this message, you were directed to change second navigation to 111.6. That channel is the second channel listed on the screen, or NAV2. You must change the frequency of this channel from what it is now to 111.6. In order to do this, you must use the arrow keys on the right part of the keyboard. The up and down keys change channels for you. Try using these keys. You can see that you are moving up and down through the different channels. Now go to second navigation.

In order to change the frequency numbers, you must use the left and right arrow keys. The right arrow key increases the number by intervals of 0.2 and the left arrow key decreases the number by intervals of 0.2. Now change this frequency to 111.6. Let's try a few more examples. (Alternate between Ctl F1 and Ctl F2). Do you feel ready to begin? Let's first adjust the headphones for you. Put the headphones on and let me know if the volume level is adequate. CTL F2.

Remember, the overall goal of this task is to correctly distinguish messages beginning with your call sign and respond to those commands. Please respond as quickly and accurately as you can.

RESOURCE MANAGEMENT

The task that you will be performing is the resource management task. All of the information that you will need is contained within the two lower right windows with the headings "Resource Management" and "Pump Status".

This task is considered a fuel management task. The rectangles are tanks which hold fuel, the green levels within the tanks that increase and decrease are fuel, and the lines which connect the tanks are pumps which transfer fuel from one tank to another in the direction that is indicated by the arrow. The numbers underneath four of the tanks represent the amount of fuel in gallons for each of these tanks. This number will be increasing and decreasing as these levels change. The capacity for the main tanks, A & B, is 4000 gallons each. The supply tanks, C & D, contain a maximum of 2000 gallons each. The tanks on the right of each three-tank system have an unlimited capacity - they never run out.

Your overall goal with this task is to maintain the levels of fuel in tanks A & B at 2500 gallons for as long as possible. This critical level is indicated by the tic mark in the shaded area on the side of each of these tanks. You are to keep the level of fuel from dropping below this shaded area during any interval in which the level deviates from 2500 gallons. You must transfer fuel to tanks A & B in order to meet this criteria because tanks A & B lose fuel at the rate of 800 gallons per minute. So you can see that with their present levels of approximately 2400 gallons each, these tanks would become empty in slightly more than 3 minutes without the transfer of additional fuel. Tanks C & D only lose fuel if they are transferring fuel to another tank.

Let me now demonstrate the process of transferring fuel. Notice that every pump has a number, a square box and an arrow next to it. The arrow indicates the direction through which fuel can be transferred with that pump. Each pump can only transfer fuel in one direction. The pumps are activated by pressing the key corresponding to the pump that you wish to activate. Use the number keys across the top of the keyboard rather than those on the right hand side of the keyboard. I'll demonstrate by turning all of the pumps on.

When I turned the pumps on, two things occurred. First, the square on each pump turned green. That means that the pump is actively transferring fuel. When the pump is off, the square is black. The second change on the screen is the numbers that appeared in the "Pump Status" window. Let's focus on that now.

Under "Pump Status", two columns of numbers are present. The first column, numbers one through eight, indicate the pump numbers and these correspond directly to the pumps in the diagram. The second column of numbers indicates the flow rates in gallons per minute of each pump when that pump is on. For example, Pump 1 transfers 800 gallons of fuel per minute from Tank C to Tank A. The flow rate for any given pump is only presented if that pump is on and actively transferring fuel. Pumps 1 and 3 transfer at a rate of 800 gallons per minute, Pumps 7 & 8 transfer at 400 gallons per minute and the other pumps have flow rates of 600 gallons per minute. Are these flow rates clear to you?

So far, you've seen two conditions for the pumps: on and off. Pressing the pump number key once turns the pump on; pressing the key again turns that pump off, and so on.

If a tank fills up to its capacity, all incoming pump lines will be turned off automatically. This is because a full tank cannot receive any more fuel. You will have to turn those pumps back on at a later time, if that is what you wish. Conversely, if a tank becomes empty, all outgoing pumps will automatically be turned off. This is because an empty tank can no longer transfer fuel. (These two conditions are demonstrated in the DEMORM.DTB script). Again, you will have to turn these pumps on again if that is what you wish to do. Any questions?

Your overall goal is to prevent the fuel level in Tanks A & B as close to 2500 gallons each for as long as you possibly can. There may be more than one way to achieve this goal; you may use the method that works the best for you. If the fuel level in these tanks should deviate from this level, however, please return the fuel level back to this point as soon as possible.

APPENDIX B

Baseline Task Scripts

BASELINE MONITORING SCRIPT

TIME			EVENT
0	0	57	RED
0	1	3	SCALE 2 DOWN
0	1	11	SCALE 3 DOWN
0	1	42	GREEN
0	2	32	SCALE 4 UP
0	2	43	SCALE 3 UP
0	2	58	GREEN
0	3	24	SCALE 2 UP
0	4	7	RED
0	5	47	SCALE 1 UP
0	6	2	RED
0	6	16	SCALE 2 DOWN
0	7	1	GREEN
0	7	15	SCALE 2 DOWN
0	7	33	SCALE 4 DOWN
0	7	45	GREEN
0	8	6	SCALE 1 UP
0	9	1	GREEN
0	9	40	SCALE 2 UP
0	11	4	RED
0	11	5	SCALE 4 DOWN
0	11	31	SCALE 4 UP
0	11	34	RED
0	11	55	SCALE 2 UP
0	12	49	GREEN
0	12	53	SCALE 4 DOWN
0	13	9	SCALE 3 DOWN
0	13	27	SCALE 2 DOWN
0	14	23	RED
0	14	26	SCALE 3 UP
0	15	1	RED
0	15	33	GREEN
0	16	12	SCALE 1 UP
0	16	48	SCALE 1 UP
0	16	52	GREEN
0	17	24	RED
0	17	34	SCALE 4 DOWN
0	17	50	RED
0	18	8	SCALE 3 DOWN
0	19	45	RED
0	20	31	GREEN
0	20	35	SCALE 4 UP
0	21	6	GREEN
0	21	37	SCALE 2 UP
0	22	8	SCALE 1 DOWN
0	22	32	GREEN
0	22	57	SCALE 4 UP
0	23	6	RED
0	23	59	END

BASELINE TRACKING SCRIPT

TIME			EVENT
0	0	0	MANUAL
0	23	59	END

BASELINE COMMUNICATIONS SCRIPT

TIME			EVENT
0	0	0	COMM TASK START
0	1	42	COMM TASK OWN MESSAGE
0	3	24	COMM TASK OTHER MESSAGE
0	6	2	COMM TASK OWN MESSAGE
0	7	15	COMM TASK OTHER MESSAGE
0	9	1	COMM TASK OTHER MESSAGE
0	11	31	COMM TASK OTHER MESSAGE
0	12	49	COMM TASK OWN MESSAGE
0	14	26	COMM TASK OTHER MESSAGE
0	16	12	COMM TASK OTHER MESSAGE
0	17	34	COMM TASK OTHER MESSAGE
0	21	37	COMM TASK OWN MESSAGE
0	22	37	COMM TASK OWN MESSAGE
0	23	58	COMM TASK END
0	23	59	END

BASELINE RESOURCE MANAGEMENT SCRIPT

TIME	EVENT
0 0 0	BEGIN
0 23 59	END

** No pump failures were presented during the baseline run.

Appendix C

64-Minute Script

Hour	Minutes	Seconds	Event	
0	0	1	GREEN	
0	1	32	SCALE 4 DOWN	
0	1	44	GREEN	
0	2	5	SCALE 3 DOWN	
0	2	10	RED	
0	2	22	SCALE 1 UP	
0	2	58	SCALE 3 UP	
0	3	29	RED	
0	4	19	SCALE 3 DOWN	LOW
0	4	21	RED	
0	5	42	SCALE 1 UP	
0	6	23	SCALE 1 UP	
0	6	32	RED	
0	6	54	GREEN	
0	7	21	GREEN	
0	7	29	SCALE 2 DOWN	
0	8	1	GREEN	
0	9	32	SCALE 4 DOWN	
0	9	44	GREEN	
0	10	5	SCALE 3 DOWN	
0	10	10	RED	
0	10	22	SCALE 1 UP	
0	10	58	SCALE 3 UP	
0	11	29	RED	
0	12	19	SCALE 3 DOWN	LOW
0	12	21	RED	
0	13	42	SCALE 1 UP	
0	14	23	SCALE 1 UP	
0	14	32	RED	
0	14	54	GREEN	
0	15	21	GREEN	
0	15	29	SCALE 2 DOWN	
0	15	50	AUTO END	
0	16	0	MANUAL	
0	16	0	COMM TASK START	
0	16	15	COMM TASK OTHER	
0	16	54	RED	
0	17	14	COMM TASK OWN	
0	17	26	GREEN	
0	17	36	SCALE 1 UP	
0	18	11	GREEN	
0	18	27	COMM TASK OTHER	

0	18	28	RED	
0	18	40	SCALE 4 UP	
0	18	49	COMM TASK OWN	
0	19	4	SCALE 1 DOWN	
0	19	28	SCALE 2 UP	
0	20	0	RED	HIGH
0	20	21	COMM TASK OWN	
0	21	1	RED	
0	21	4	SCALE 3 DOWN	
0	21	17	COMM TASK OTHER	
0	21	25	SCALE 1 UP	
0	21	34	RED	
0	21	47	SCALE 4 UP	
0	22	13	COMM TASK OTHER	
0	22	29	RED	
0	22	59	SCALE 4 DOWN	
0	23	7	COMM TASK OWN	
0	23	59	COMM TASK END	

0	24	27	SCALE 3 UP	
0	25	32	RED	
0	26	29	SCALE 4 UP	
0	26	39	SCALE 2 DOWN	
0	26	49	GREEN	
0	27	3	SCALE 2 DOWN	
0	27	54	RED	
0	27	59	GREEN	
0	28	25	RED	
0	28	26	SCALE 3 UP	TRACKING
0	28	54	RED	
0	29	30	GREEN	
0	30	32	SCALE 3 DOWN	
0	31	14	RED	
0	31	17	SCALE 1 DOWN	
0	31	31	SCALE 2 DOWN	

0	32	0	COMM TASK START	
0	32	15	COMM TASK OTHER	
0	32	54	RED	
0	33	14	COMM TASK OWN	
0	33	26	GREEN	
0	33	36	SCALE 1 UP	
0	34	11	GREEN	
0	34	27	COMM TASK OTHER	
0	34	28	RED	
0	34	40	SCALE 4 UP	
0	34	49	COMM TASK OWN	
0	35	4	SCALE 1 DOWN	
0	35	28	SCALE 2 UP	
0	36	0	RED	
0	36	21	COMM TASK OWN	HIGH
0	37	1	RED	

```
0 37  4   SCALE 3 DOWN
0 37 17   COMM TASK OTHER
0 37 25   SCALE 1 UP
0 37 34   RED
0 37 47   SCALE 4 UP
0 38 13   COMM TASK OTHER
0 38 29   RED
0 38 59   SCALE 4 DOWN
0 39  7   COMM TASK OWN
0 39 59   COMM TASK END

0 40 27   SCALE 3 UP
0 41 32   RED
0 42 29   SCALE 4 UP
0 42 39   SCALE 2 DOWN
0 42 49   GREEN
0 43  3   SCALE 2 DOWN
0 43 54   RED
0 43 59   GREEN
0 44 25   RED
0 44 26   SCALE 3 UP                    TRACKING
0 44 54   RED
0 45 30   GREEN
0 46 32   SCALE 3 DOWN
0 47 14   RED
0 47 17   SCALE 1 DOWN
0 47 31   SCALE 2 DOWN

0 48  0   AUTO
0 48  0   COMM TASK START
0 48 13   GREEN
0 48 33   COMM TASK OWN
0 49 16   RED
0 49 32   GREEN
0 50  2   COMM TASK OWN
0 50 32   COMM TASK OTHER
0 50 56   SCALE 4 DOWN
0 51  5   COMM TASK OTHER
0 51 20   SCALE 4 UP
0 51 23   SCALE 2 DOWN
0 52  5   RED
0 52 20   COMM TASK OTHER               COMMUNICATION
0 52 34   GREEN
0 52 41   SCALE 1 UP
0 53  2   COMM TASK OWN
0 53 38   COMM TASK OTHER
0 53 51   SCALE 2 DOWN
0 54  2   GREEN
0 54 38   SCALE 2 UP
0 54 43   SCALE 3 UP
0 55  8   RED
0 55 18   COMM TASK OWN
```

0	56	13	GREEN
0	56	33	COMM TASK OWN
0	57	16	RED
0	57	32	GREEN
0	58	2	COMM TASK OWN
0	58	32	COMM TASK OTHER
0	58	56	SCALE 4 DOWN
0	59	5	COMM TASK OTHER
0	59	20	SCALE 4 UP
0	59	23	SCALE 2 DOWN
1	0	5	RED
1	0	20	COMM TASK OTHER
1	0	34	GREEN
1	0	41	SCALE 1 UP
1	1	2	COMM TASK OWN
1	1	38	COMM TASK OTHER
1	1	51	SCALE 2 DOWN
1	2	2	GREEN
1	2	38	SCALE 2 UP
1	2	43	SCALE 3 UP
1	3	8	RED
1	3	18	COMM TASK OWN
1	3	57	COMM TASK END
1	3	58	RATING
1	3	59	END

COMMUNICATION

Appendix D

Sources of Variation Tables for Baseline Data

Summary of Analysis of Variance for Baseline Monitoring Response Time to Red Light Signals

Source of Variation	df	Mean Square	F	Eta Square
Block of Time (3)	2	0.25	7.22*	0.01
Gender	1	1.29	6.32*	0.03
Block * Gender	2	0.00	0.12	

$* p < 0.05$

Summary of Analysis of Variance for Baseline Monitoring Response Time to Green Light Signals

Source of Variation	df	Mean Square	F	Eta Square
Block of Time (3)	2	0.12	0.35	
Gender	1	3.86	6.67*	0.06
Block * Gender	2	0.27	0.78	

$* p < 0.05$

Summary of Analysis of Variance for Baseline Monitoring Response Time to Probability Monitoring Signals

Source of Variation	df	Mean Square	F	Eta Square
Block of Time (3)	2	0.04	0.18	
Gender	1	2.65	2.61	
Block * Gender	2	0.39	1.77	

$* p < 0.05$

Summary of Analysis of Variance for Baseline Tracking RMS Error as a Function of Gender and Block of Time

Source of Variation	df	Mean Square	F	Eta Square
Block of Time (6)	5	114.37	3.39*	0.13
Gender	1	526.19	2.12	
Block * Gender	5	4.29	0.13	
Subject(Gender)	8	247.88	NT	NT
Subj * Block(Gender)	40	33.72	NT	NT

* $p < 0.05$

Summary of Analysis of Variance for Baseline Tracking RMS Error as a Function of Mouse Experience and Block of Time

Source of Variation	df	Mean Square	F	Eta Square
Block of Time (6)	5	114.37	3.39*	0.13
Mouse Experience	1	8.41	0.03	
Block * Experience	5	3.94	0.12	
Subject(Exp)	8	312.61	NT	NT
Subj * Block(Exp)	40	33.76	NT	NT

* $p < 0.05$

Summary of Analysis of Variance for Baseline Communication Response Time as a Function of Gender and "Own Ship" Message

Source of Variation	df	Mean Square	F	Eta Square
Message (6)	5	350.80	3.37*	0.22
Gender	1	485.47	3.94	
Message * Gender	5	150.20	1.44	
Subject(Gender)	8	123.24	NT	NT
Subj * Block(Gender)	40	104.18	NT	NT

* $p < 0.05$

Summary of Analysis of Variance for Baseline Communication Response Time as a Function of Gender and Block of Time

Source of Variation	df	Mean Square	F	Eta Square
Block of Time (3)	2	130.50	2.21	
Gender	1	185.46	2.91	
Block * Gender	2	38.61	0.65	
Subject(Gender)	8	63.66	NT	NT
Subj * Block(Gender)	16	58.94	NT	NT

* $p < 0.05$

Appendix E

Sources of Variation Tables for Validity Data

Summary of Analysis of Variance for Validity TLX Scores as a Function of Workload Level and Order of Ratings

Source of Variation	df	Mean Square	F	Eta Square
Presentation Order	3	113.14	0.63	
Workload Level	3	18483.11	103.62*	0.32
Subject	15	5512.35	NT	NT
Subj * Order	45	67.27	NT	NT

* p < 0.05, NT = No Test

Summary of Analysis of Variance for Validity Resource Management Performance (Deviation from Target)

Source of Variation	df	Mean Square	F	Eta Square
Workload Level	3	9470.50	3.32*	0.05
Subject	15	29574.11	NT	NT
Subject * Workload	45	2849.73	NT	NT

* p < 0.05, NT = No Test

Summary of Analysis of Variance for Frequency of Responding on the Resource Management Task

Source of Variation	df	Mean Square	F	Eta Square
Tank (2)	1	28.89	0.46	
Workload Level (4)	3	1890.70	8.99*	0.09
Tank * Workload	3	181.36	1.87	
Subject	15	1854.35	NT	NT
Subject * Tank	15	62.30	NT	NT
Subject * Workload	45	210.31	NT	NT
Subj * Tank * Workload	45	97.09	NT	NT

* p < 0.05

Appendix F

Performance Data Means

RT = Reaction Time
FA = False Alarms
Condition 1 = Low Workload
Condition 2 = Communication
Condition 3 = Tracking
Condition 4 = High Workload

Overall Monitoring Means

Hour	Cond.	N	RT	Misses	FA
			Means		
1	1	48	4.8338	2.2500	0.9792
1	2	48	4.4300	1.7708	0.8542
1	3	48	4.6123	2.1250	1.1875
1	4	48	4.2913	2.0417	0.8542
2	1	48	4.0727	1.2708	1.8750
2	2	48	3.7829	1.0417	1.0000
2	3	48	3.6998	1.7292	2.9375
2	4	48	3.9920	1.4167	1.5208
3	1	48	3.5472	0.8750	1.2500
3	2	48	3.1569	0.8750	1.3542
3	3	48	3.4988	0.7292	1.8333
3	4	48	2.9499	1.1042	1.6250
4	1	48	3.4967	0.7708	1.5208
4	2	48	3.2123	0.6250	0.9375
4	3	48	3.6231	0.7500	1.0000
4	4	48	3.2659	0.8750	1.2708

Red Light Means

Hour	Cond.	N	RT	Misses	FA
			Means		
1	1	48	1.9058	0.0000	0.0625
1	2	48	1.7077	0.0625	0.1875
1	3	48	1.6954	0.0000	0.1458
1	4	48	1.6065	0.0208	0.1042
2	1	48	1.5948	0.0000	0.2292
2	2	48	1.6235	0.0208	0.1667
2	3	48	1.4829	0.0208	0.1458
2	4	48	1.5077	0.0417	0.1042
3	1	48	1.4908	0.0208	0.1042
3	2	48	1.4452	0.0208	0.2083
3	3	48	1.4254	0.0208	0.1042
3	4	48	1.4883	0.0000	0.1875
4	1	48	1.5846	0.0000	0.0417
4	2	48	1.5413	0.0417	0.1042
4	3	48	1.4529	0.0000	0.0208
4	4	48	1.5142	0.0000	0.1667

Green Light Means

Hour	Cond.	N	Means RT	Misses	FA
1	1	48	3.7385	0.8958	0.1667
1	2	48	3.4738	0.3750	0.1042
1	3	48	3.0410	0.2917	0.1458
1	4	48	3.4413	0.2708	0.0625
2	1	48	3.0840	0.5000	0.2083
2	2	48	2.8310	0.1250	0.0208
2	3	48	3.1883	0.3542	0.2917
2	4	48	3.1196	0.2083	0.1250
3	1	48	2.7540	0.2292	0.1458
3	2	48	2.8888	0.1458	0.1250
3	3	48	2.2729	0.0833	0.2083
3	4	48	2.1752	0.1042	0.0833
4	1	48	2.6994	0.2708	0.2917
4	2	48	1.8835	0.0417	0.1667
4	3	48	2.3723	0.1042	0.0417
4	4	48	2.2244	0.0625	0.0833

Dial Signal Means

Hour	Cond.	N	Means RT	Misses	FA
1	1	48	7.3158	1.3750	1.0625
1	2	48	7.0042	1.3542	0.5625
1	3	48	7.8242	1.8333	1.0625
1	4	48	7.3206	1.7500	0.0583
2	1	48	5.9850	0.7708	3.5417
2	2	48	5.8421	0.8958	1.6042
2	3	48	5.7050	1.3542	7.5417
2	4	48	6.5171	1.1875	1.7500
3	1	48	5.2633	0.6250	1.8333
3	2	48	4.6163	0.7083	6.1875
3	3	48	5.4692	0.6250	1.8333
3	4	48	4.7123	1.0000	2.8333
4	1	48	4.9717	0.5208	1.9167
4	2	48	4.9433	0.5417	1.1458
4	3	48	5.6223	0.6458	1.6250
4	4	48	5.1027	0.8125	2.5417

Tracking Means

Hour	Cond.	N	Means RMS Error
1	1	48	37.3245
1	2	48	40.8635
2	1	48	32.4556
2	2	48	38.1330
3	1	48	29.8316
3	2	48	33.7971
4	1	48	29.6884
4	2	48	31.7074

Communication Means

Hour	Cond.	N	Means RT	Misses
1	1	48	3.0869	0.1042
1	2	48	3.9786	0.2917
2	1	48	2.1465	0.1042
2	2	48	3.1651	0.1250
3	1	48	1.7618	0.1042
3	2	48	2.7100	0.0625
4	1	48	1.7617	0.1042
4	2	48	2.9337	0.1250

TLX Score Means

Hour	Means TLX Score
1	55.312
2	56.425
3	47.233
4	45.721

Resource Management Means

Hour	Cond.	N	Means Deviation from Target
1	1	48	98.0256
1	2	48	89.2432
1	3	48	102.5977
1	4	48	118.2408
2	1	48	89.3628
2	2	48	89.9772
2	3	48	105.4742
2	4	48	101.7366
3	1	48	94.2551
3	2	48	86.3607
3	3	48	93.7495
3	4	48	102.2864
4	1	48	93.1979
4	2	48	85.8031
4	3	48	89.7749
4	4	48	95.8853

Appendix G

Strategic Behavior Means

Resource Management Response Frequency

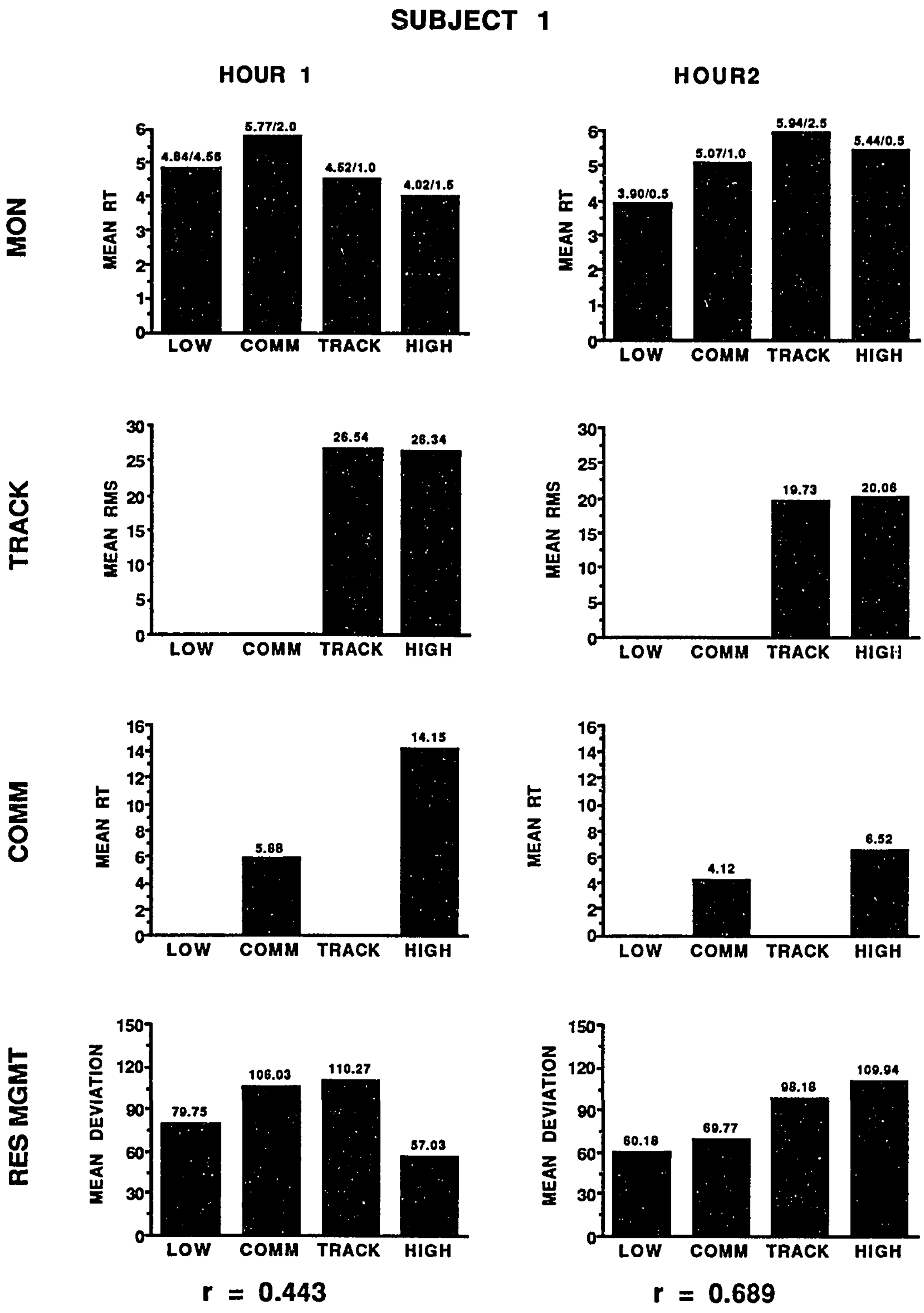
Means

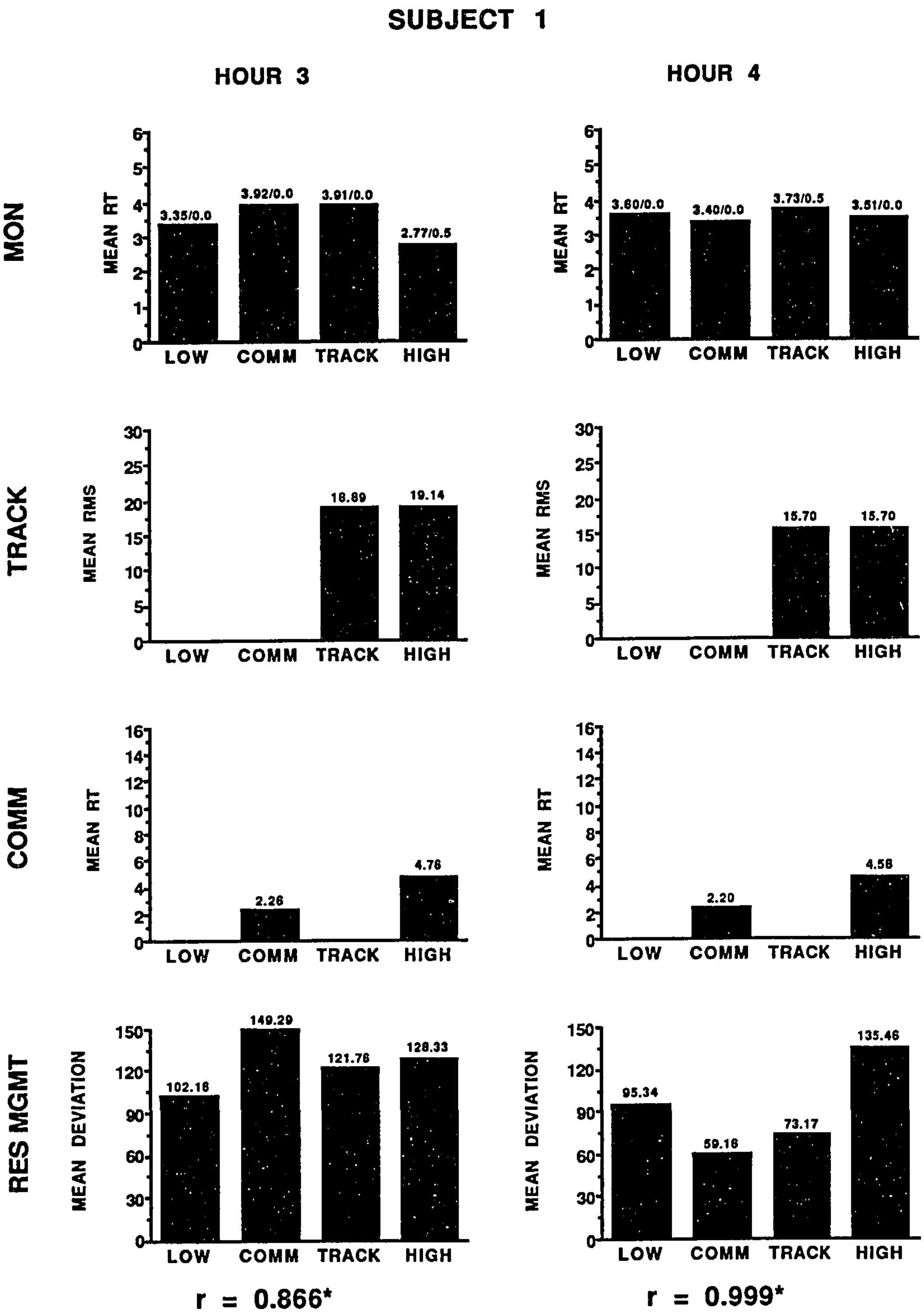
Hour	Condition	N	Number of Responses
1	1	48	53.3333
1	2	48	57.4792
1	3	48	34.7917
1	4	48	31.4792
2	1	48	59.7292
2	2	48	52.3542
2	3	48	33.3542
2	4	48	37.4375
3	1	48	53.5208
3	2	48	54.1042
3	3	48	33.0208
3	4	48	37.1458
4	1	48	59.7500
4	2	48	46.5833
4	3	48	33.2708
4	4	48	35.8333

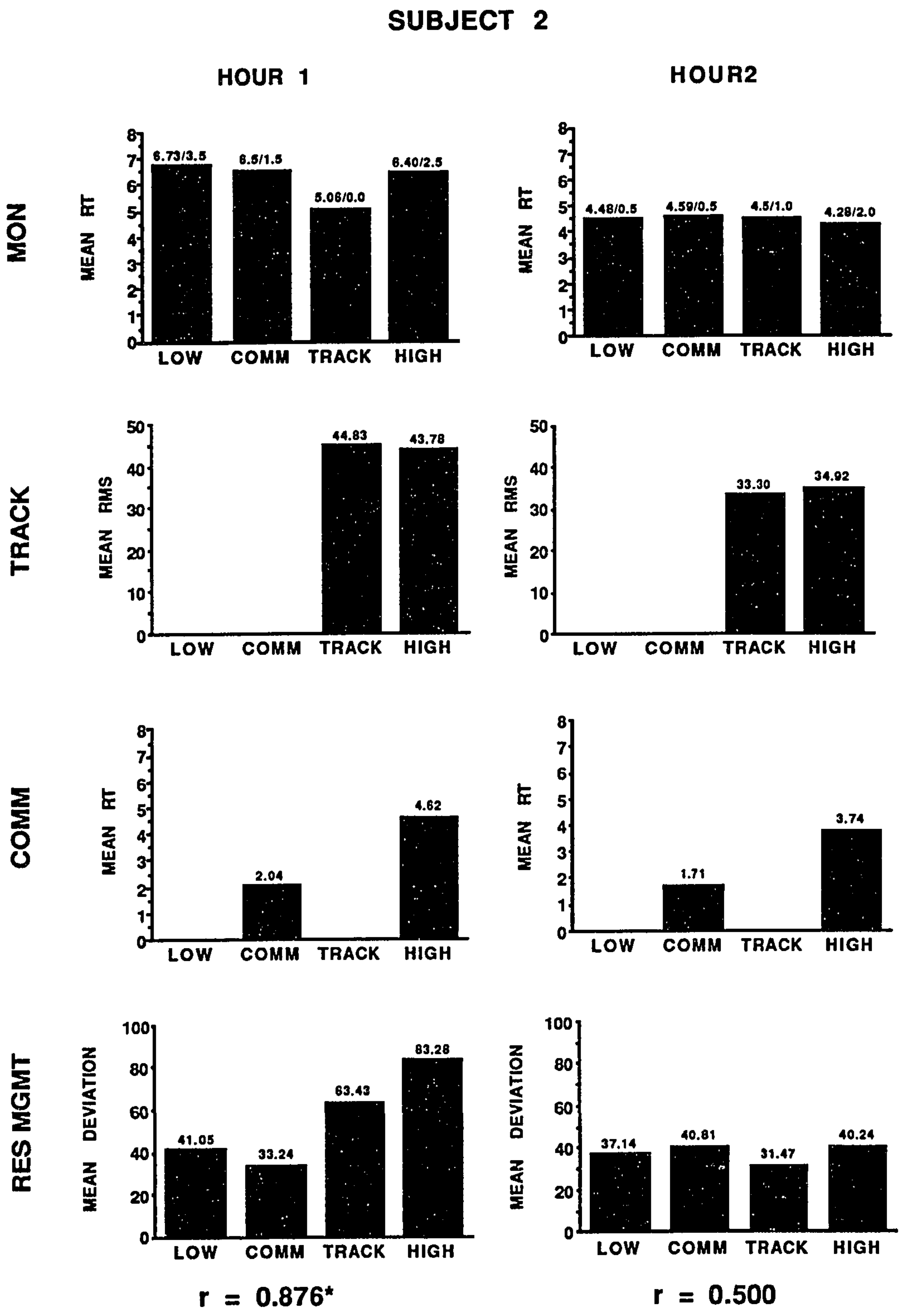
Appendix H

Performance Profiles for Subjects

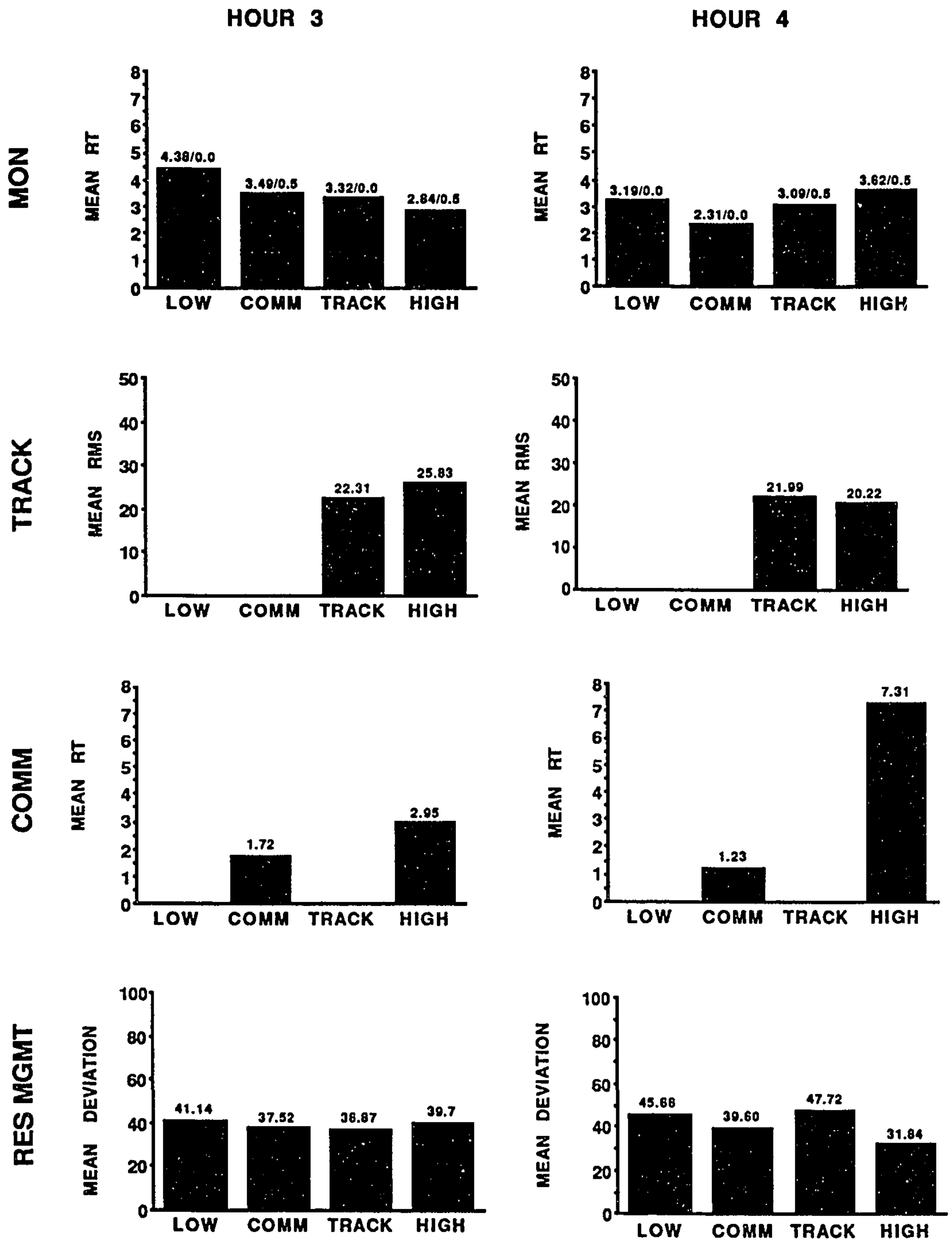
Note: Axes may be different for different subjects. The correlation between Tanks A and B is listed by hour. An asterisk (*) indicates a significant correlation ($p < 0.05$).


SUBJECT 1
HOUR 1
HOUR2
MON
MEAN RT
4.84/4.56
5.77/2.0
4.52/1.0
4.02/1.5
3.90/0.5
5.07/1.0
5.94/2.5
5.44/0.5
LOW
COMM
TRACK
HIGH
TRACK
MEAN RMS
26.54
26.34
19.73
20.06
COMM
MEAN RT
5.88
14.15
4.12
6.52
RES MGMT
MEAN DEVIATION
79.75
106.03
110.27
57.03
60.18
69.77
98.18
109.94
r = 0.443
r = 0.689


SUBJECT 1
HOUR 3
HOUR 4
MON
MEAN RT
3.35/0.0
3.92/0.0
3.91/0.0
2.77/0.5
3.60/0.0
3.40/0.0
3.73/0.5
3.51/0.0
TRACK
MEAN RMS
18.89
19.14
15.70
15.70
COMM
MEAN RT
2.26
4.76
2.20
4.58
RES MGMT
MEAN DEVIATION
102.16
149.29
121.76
128.33
95.34
59.16
73.17
135.46
LOW
COMM
TRACK
HIGH
r = 0.866*
r = 0.999*

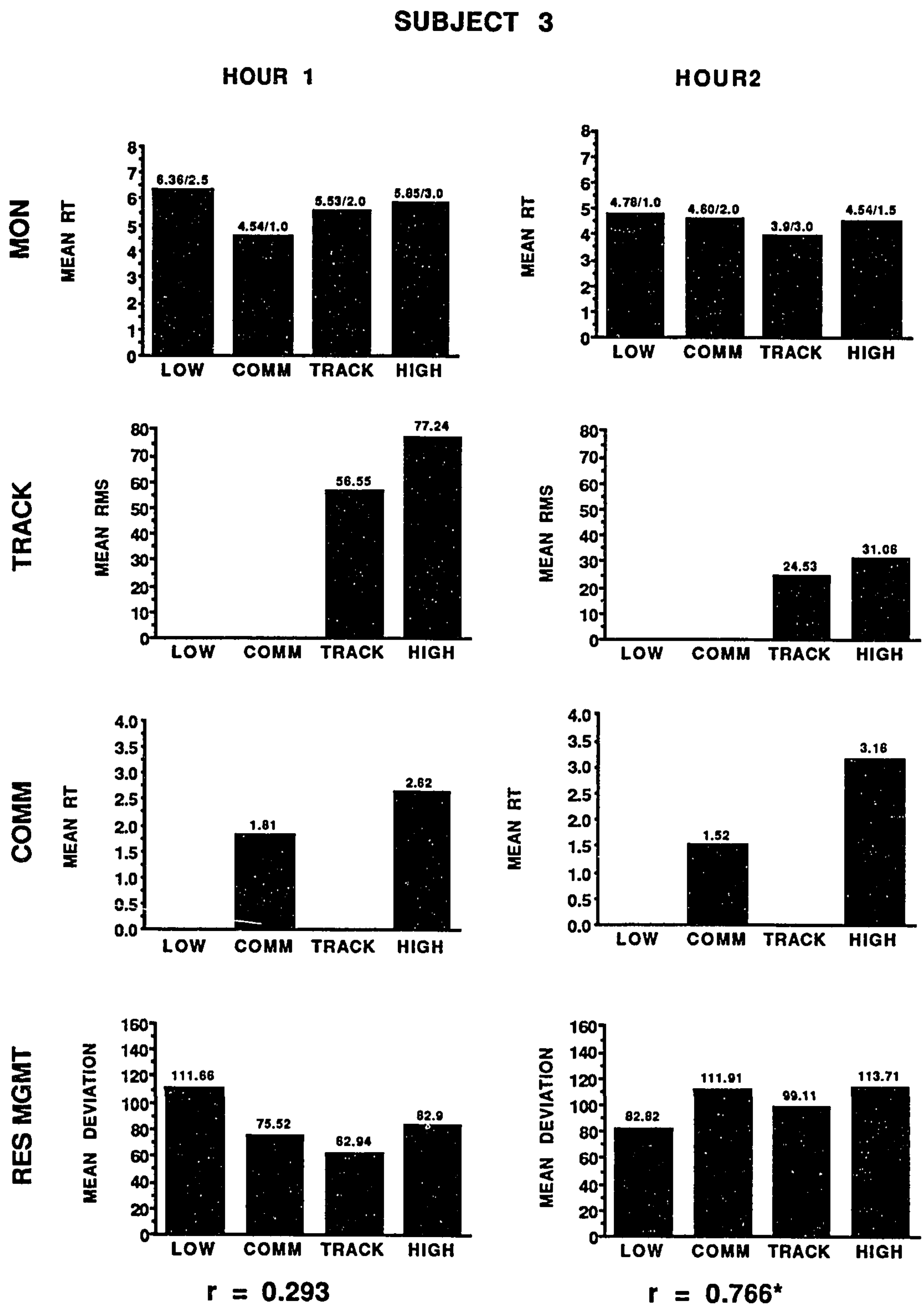

SUBJECT 2
HOUR 1
HOUR2
MON
MEAN RT
6.73/3.5
6.5/1.5
5.06/0.0
6.40/2.5
4.48/0.5
4.59/0.5
4.5/1.0
4.28/2.0
TRACK
MEAN RMS
44.83
43.78
33.30
34.92
COMM
2.04
4.62
1.71
3.74
RES MGMT
MEAN DEVIATION
41.05
33.24
63.43
83.28
37.14
40.81
31.47
40.24
LOW
COMM
TRACK
HIGH
r = 0.876*
r = 0.500

SUBJECT 2

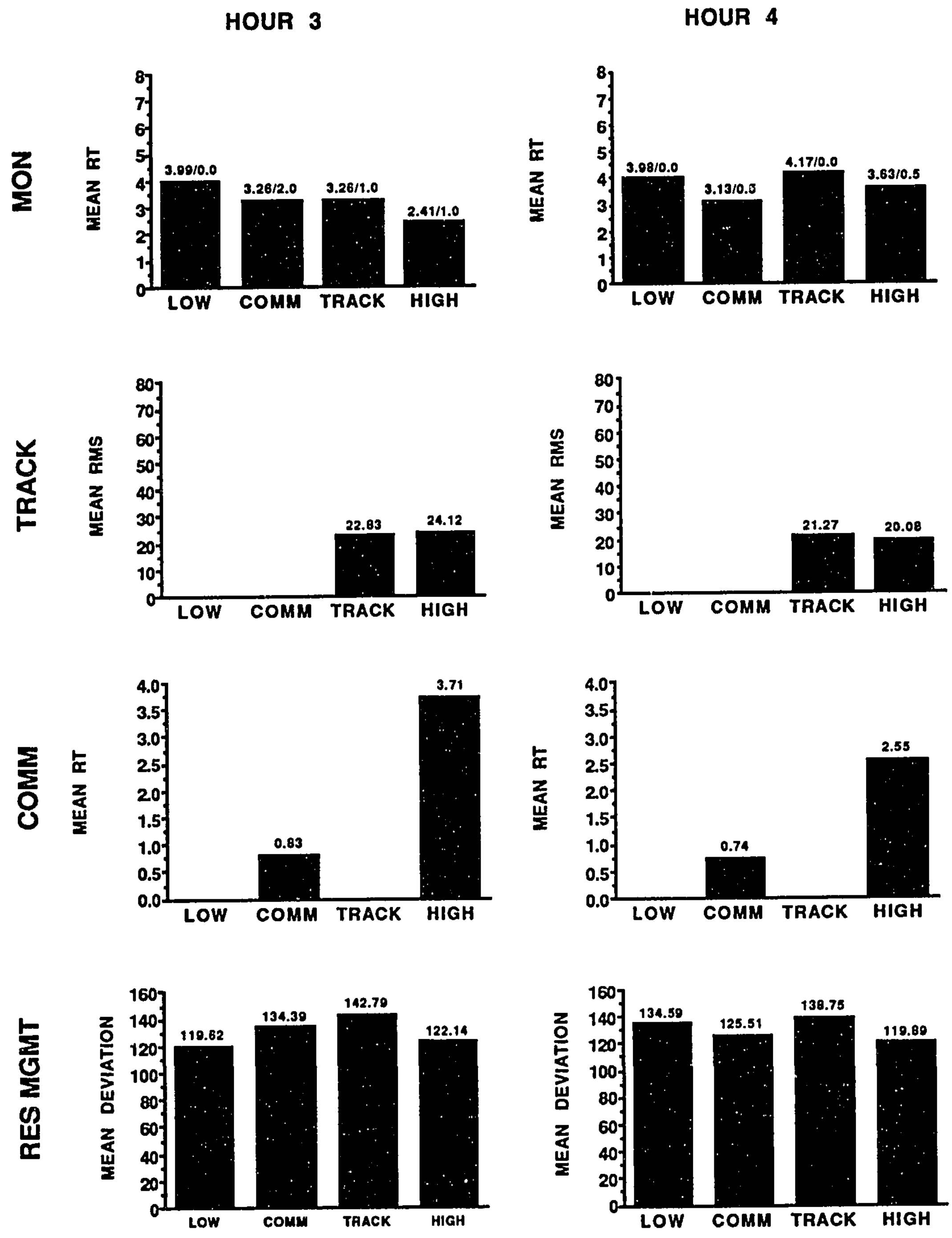


r = 0.583

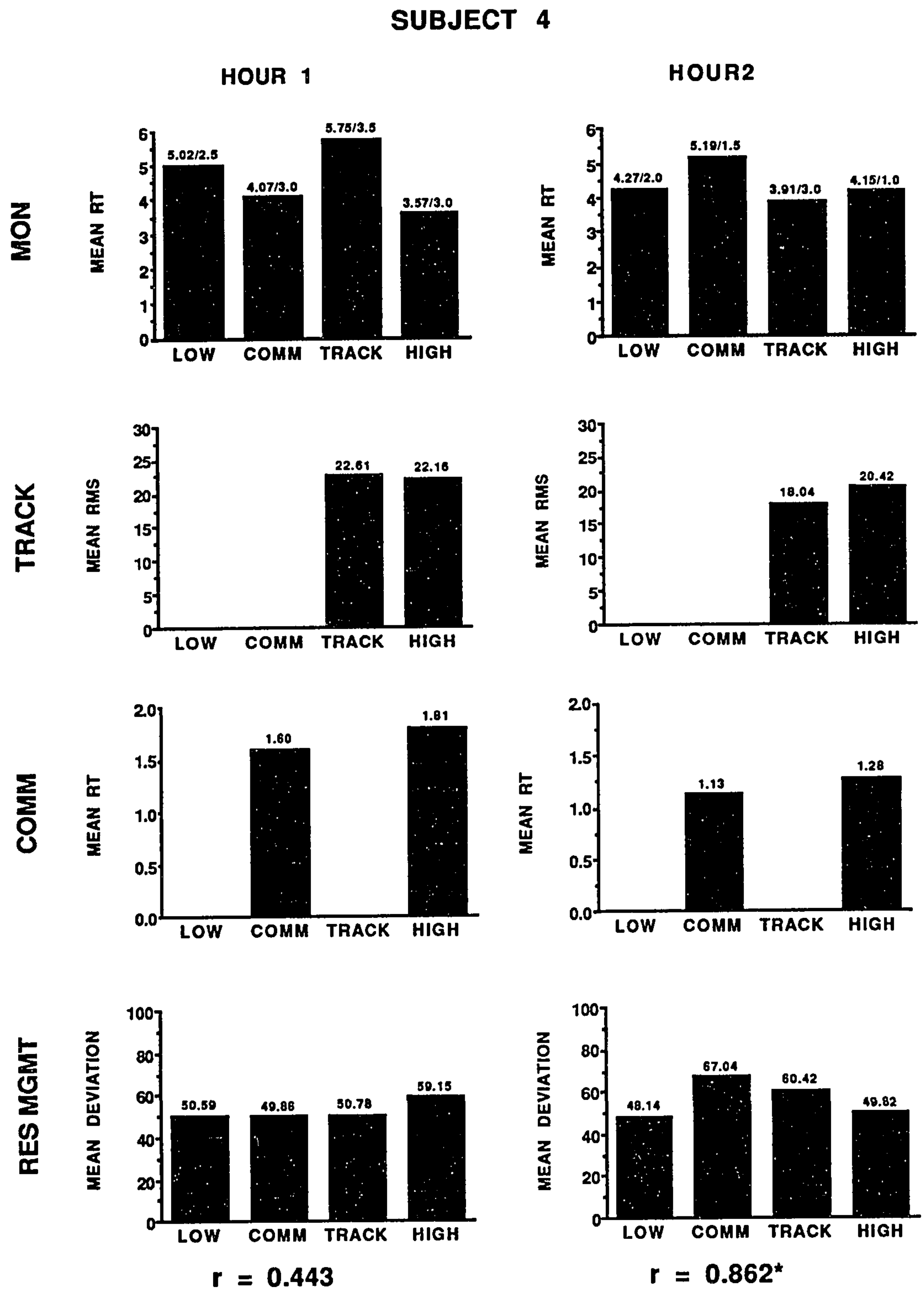
r = 0.899*

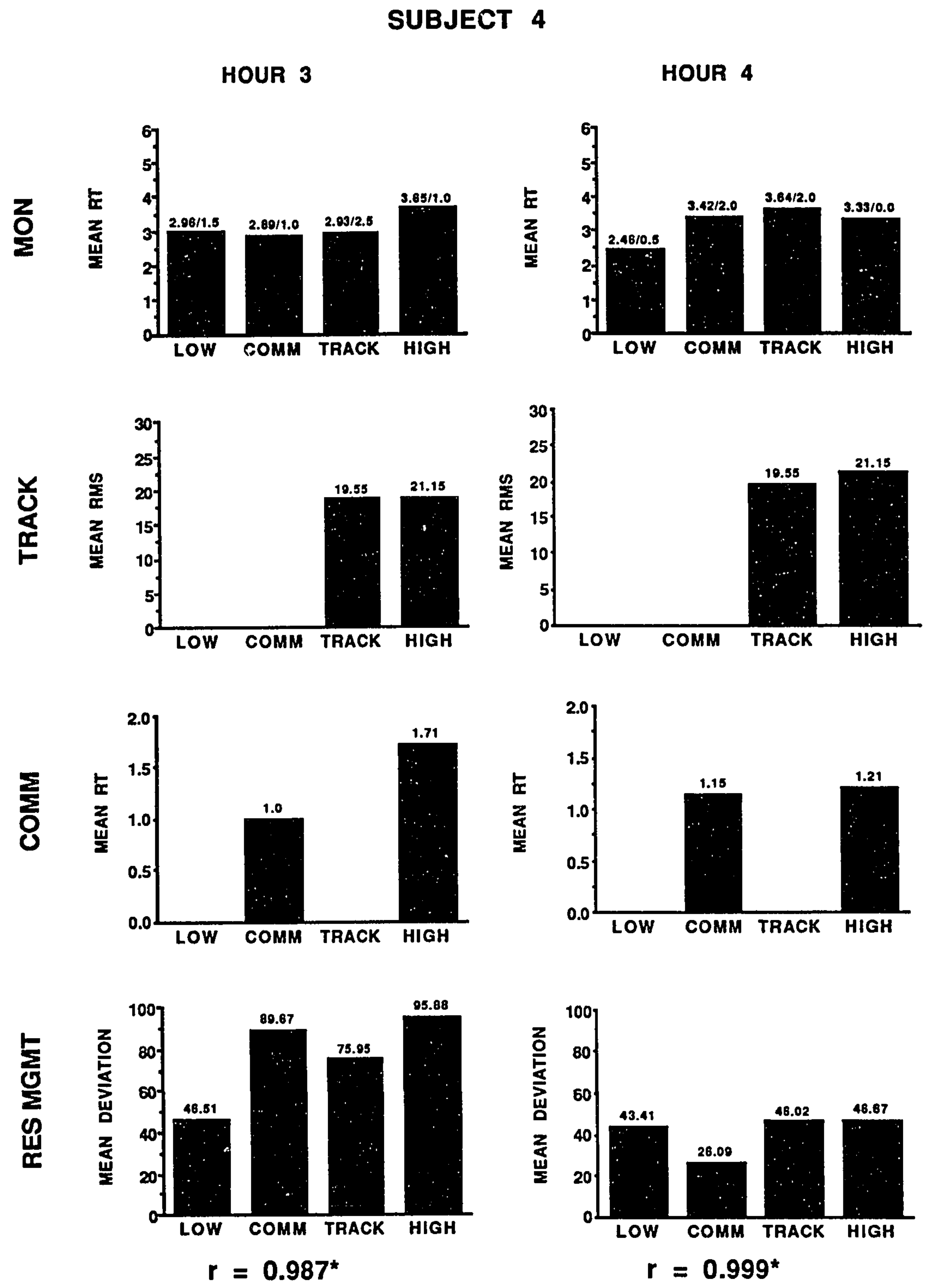

SUBJECT 3
HOUR 1
HOUR2
MON
MEAN RT
6.36/2.5
4.54/1.0
5.53/2.0
5.85/3.0
4.78/1.0
4.60/2.0
3.9/3.0
4.54/1.5
LOW
COMM
TRACK
HIGH
TRACK
MEAN RMS
56.55
77.24
24.53
31.06
COMM
MEAN RT
1.81
2.62
1.52
3.16
RES MGMT
MEAN DEVIATION
111.66
75.52
62.94
82.9
82.82
111.91
99.11
113.71
r = 0.293
r = 0.766*

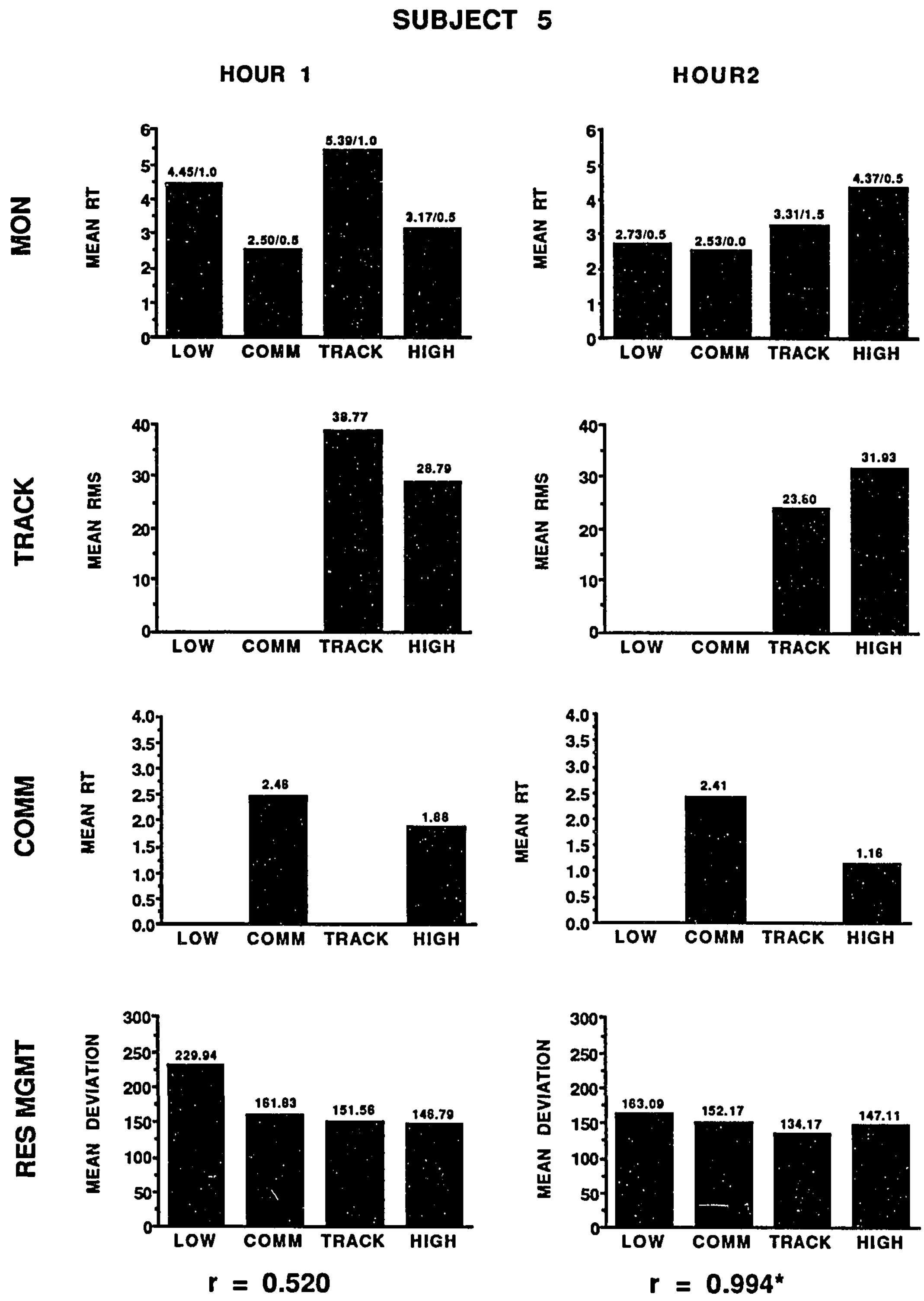
SUBJECT 3

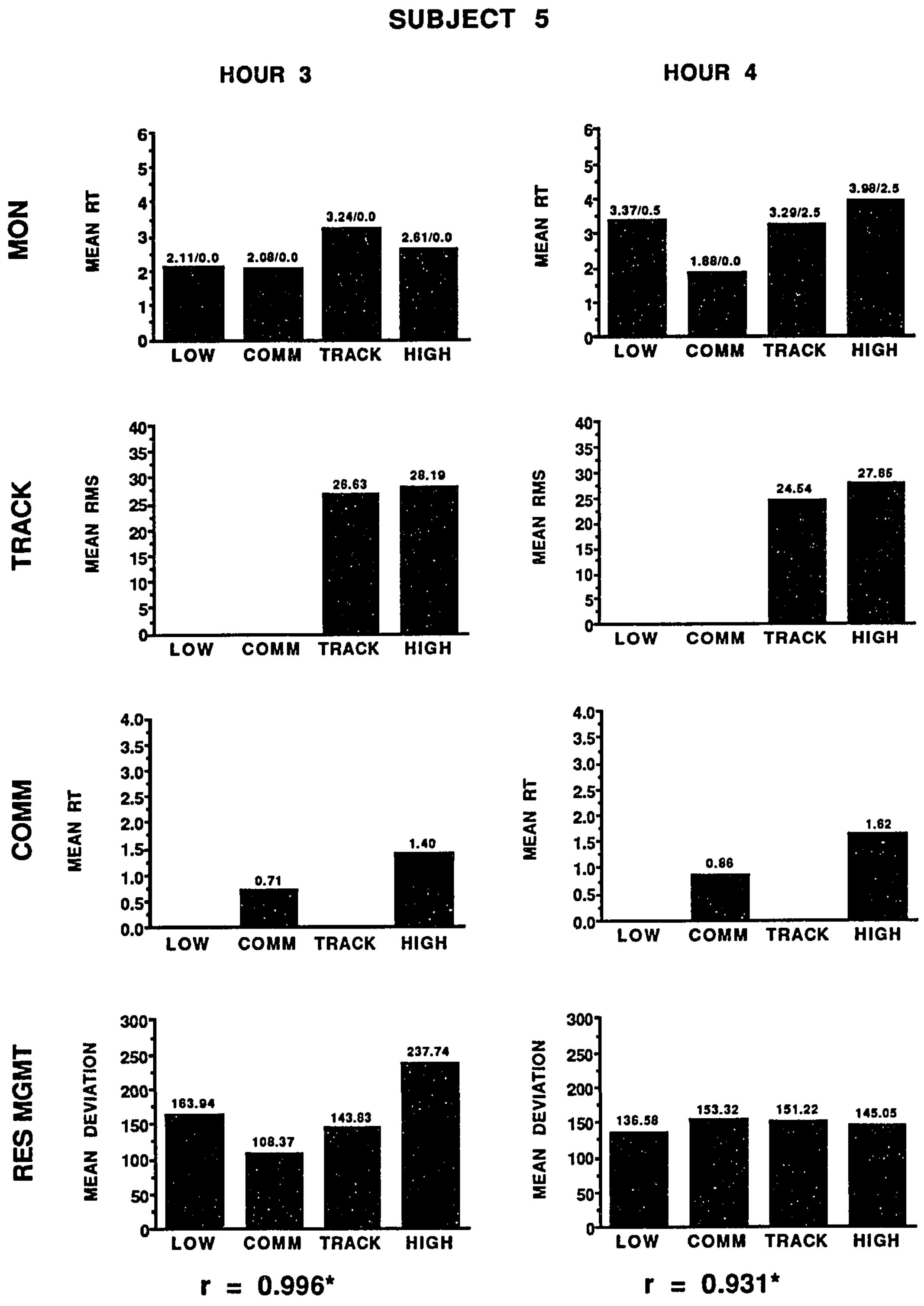


r = 0.974* r = 0.973*

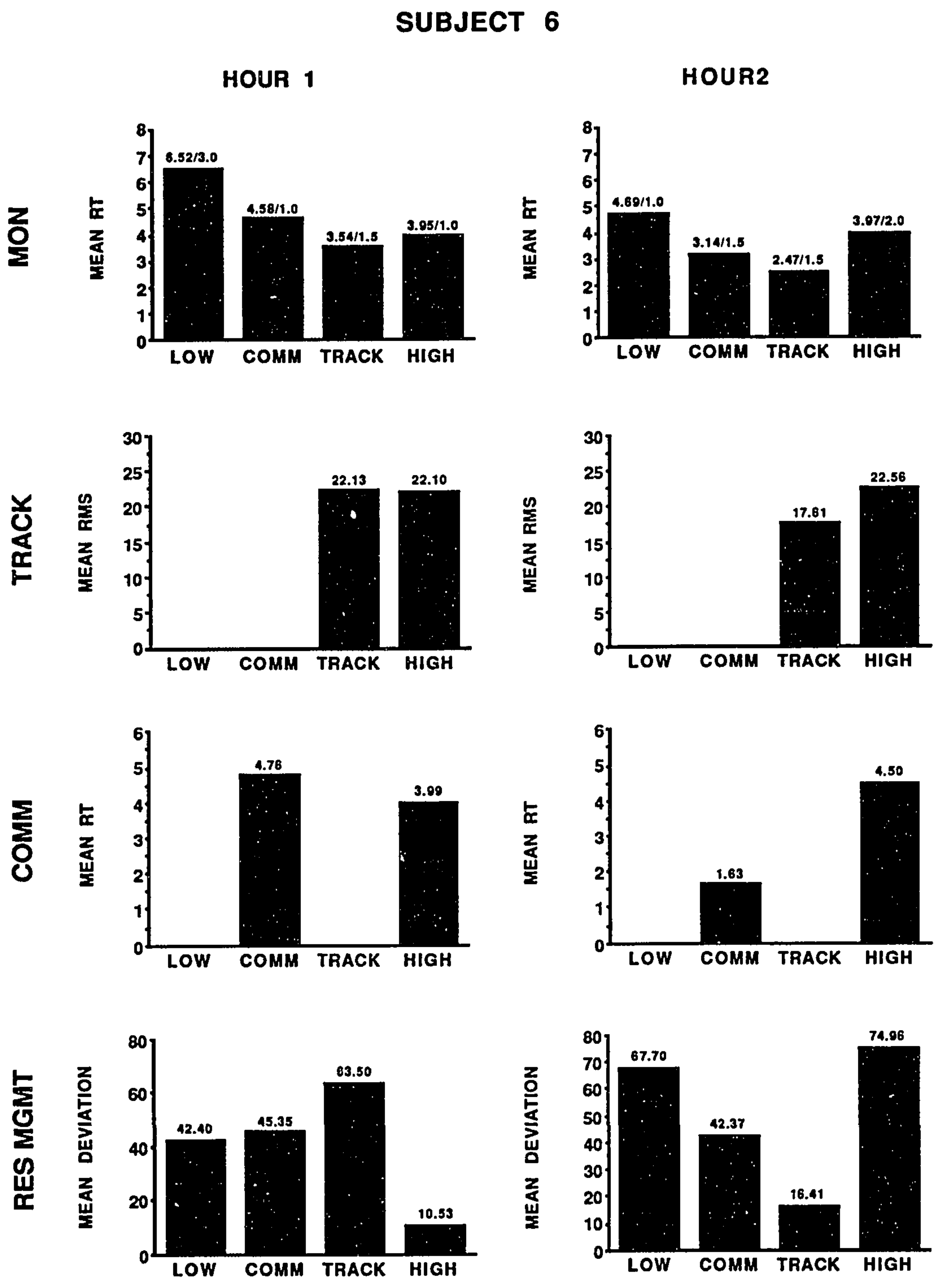

SUBJECT 4
HOUR 1
HOUR2
MON
TRACK
COMM
RES MGMT
MEAN RT
MEAN RMS
MEAN DEVIATION
5.02/2.5
4.07/3.0
5.75/3.5
3.57/3.0
4.27/2.0
5.19/1.5
3.91/3.0
4.15/1.0
22.61
22.16
18.04
20.42
1.60
1.81
1.13
1.28
50.59
49.86
50.78
59.15
48.14
67.04
60.42
49.82
LOW
COMM
TRACK
HIGH
r = 0.443
r = 0.862*


SUBJECT 4
HOUR 3
HOUR 4
MON
MEAN RT
2.96/1.5
2.89/1.0
2.93/2.5
3.65/1.0
2.46/0.5
3.42/2.0
3.64/2.0
3.33/0.0
LOW
COMM
TRACK
HIGH
TRACK
MEAN RMS
19.55
21.15
COMM
MEAN RT
1.0
1.71
1.15
1.21
RES MGMT
MEAN DEVIATION
46.51
89.67
75.95
95.88
43.41
26.09
46.02
46.67
r = 0.987*
r = 0.999*


SUBJECT 5
HOUR 1
HOUR2
MON
MEAN RT
4.45/1.0
2.50/0.5
5.39/1.0
3.17/0.5
2.73/0.5
2.53/0.0
3.31/1.5
4.37/0.5
LOW
COMM
TRACK
HIGH
TRACK
MEAN RMS
38.77
28.79
23.60
31.93
COMM
MEAN RT
2.48
1.88
2.41
1.16
RES MGMT
MEAN DEVIATION
229.94
161.83
151.56
148.79
163.09
152.17
134.17
147.11
r = 0.520
r = 0.994*



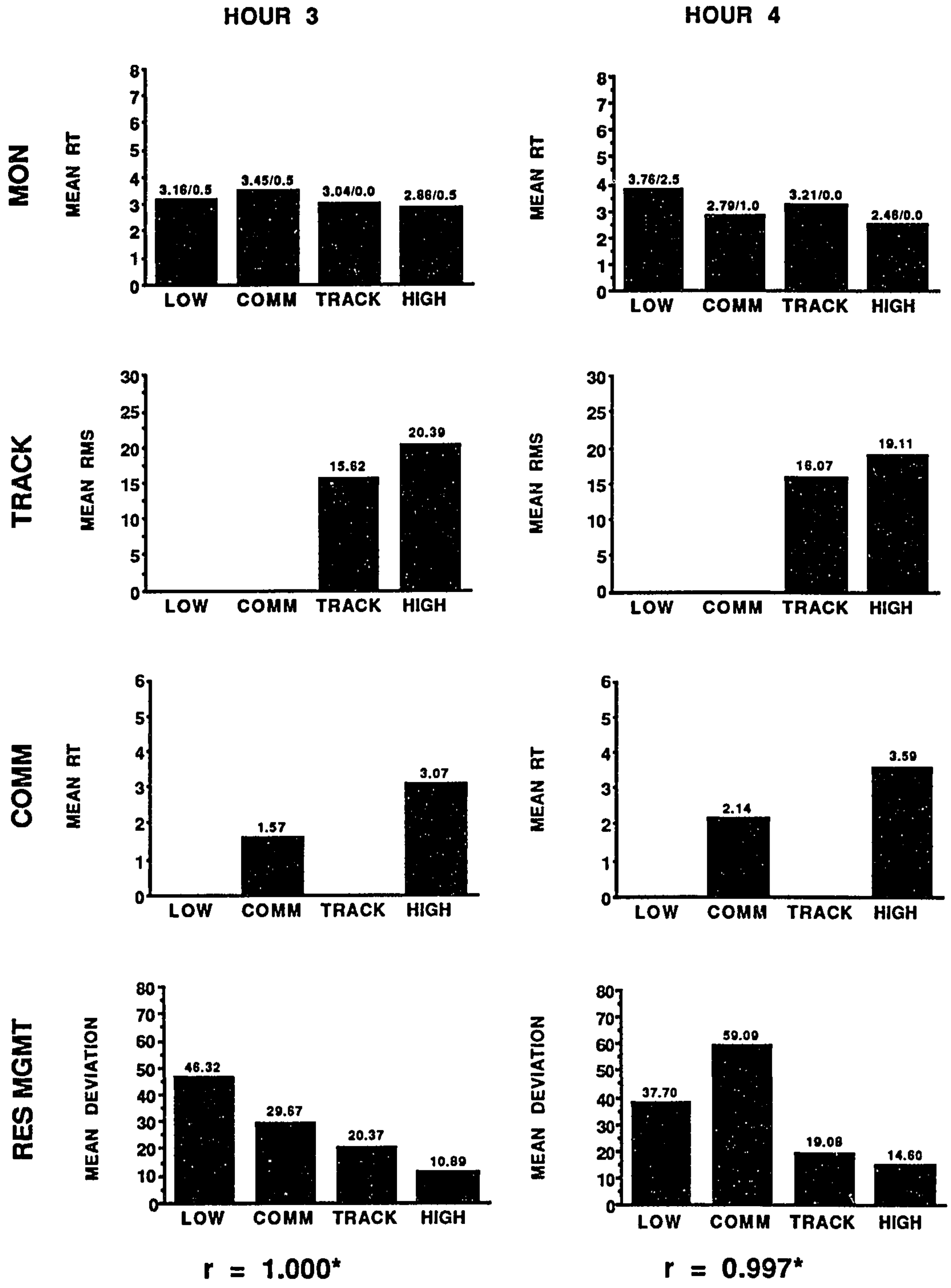
SUBJECT 5
HOUR 3
HOUR 4
MON
MEAN RT
2.11/0.0
2.08/0.0
3.24/0.0
2.61/0.0
3.37/0.5
1.88/0.0
3.29/2.5
3.98/2.5
LOW
COMM
TRACK
HIGH
TRACK
MEAN RMS
26.63
28.19
24.54
27.85
COMM
0.71
1.40
0.86
1.62
RES MGMT
MEAN DEVIATION
163.94
108.37
143.83
237.74
136.58
153.32
151.22
145.05
r = 0.996*
r = 0.931*

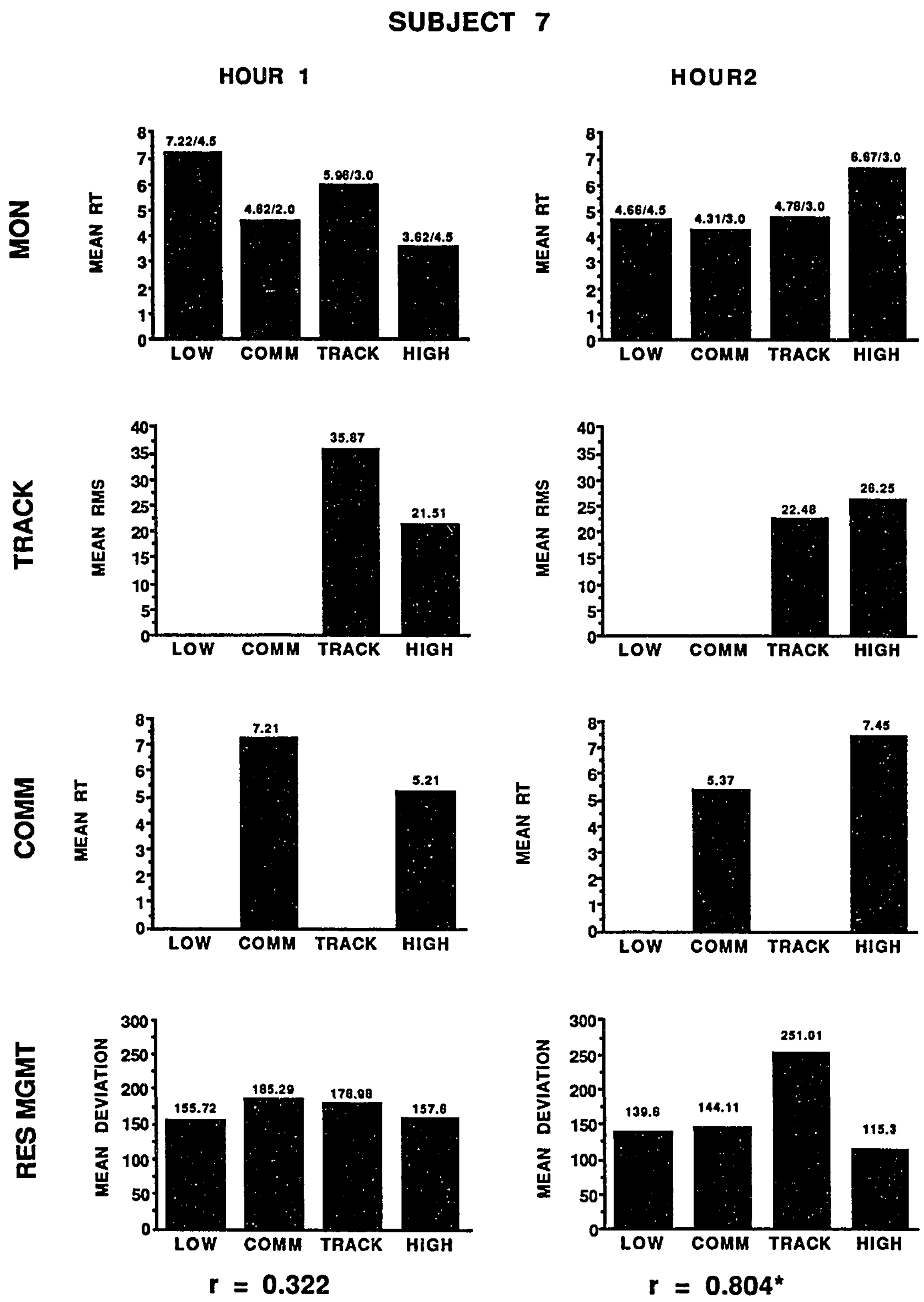

SUBJECT 6
HOUR 1
HOUR2
MON
MEAN RT
6.52/3.0
4.58/1.0
3.54/1.5
3.95/1.0
4.69/1.0
3.14/1.5
2.47/1.5
3.97/2.0
TRACK
MEAN RMS
22.13
22.10
17.81
22.56
COMM
MEAN RT
4.76
3.99
1.63
4.50
RES MGMT
MEAN DEVIATION
42.40
45.35
63.50
10.53
67.70
42.37
16.41
74.96
LOW
COMM
TRACK
HIGH

r = 0.996*

r = 0.993*

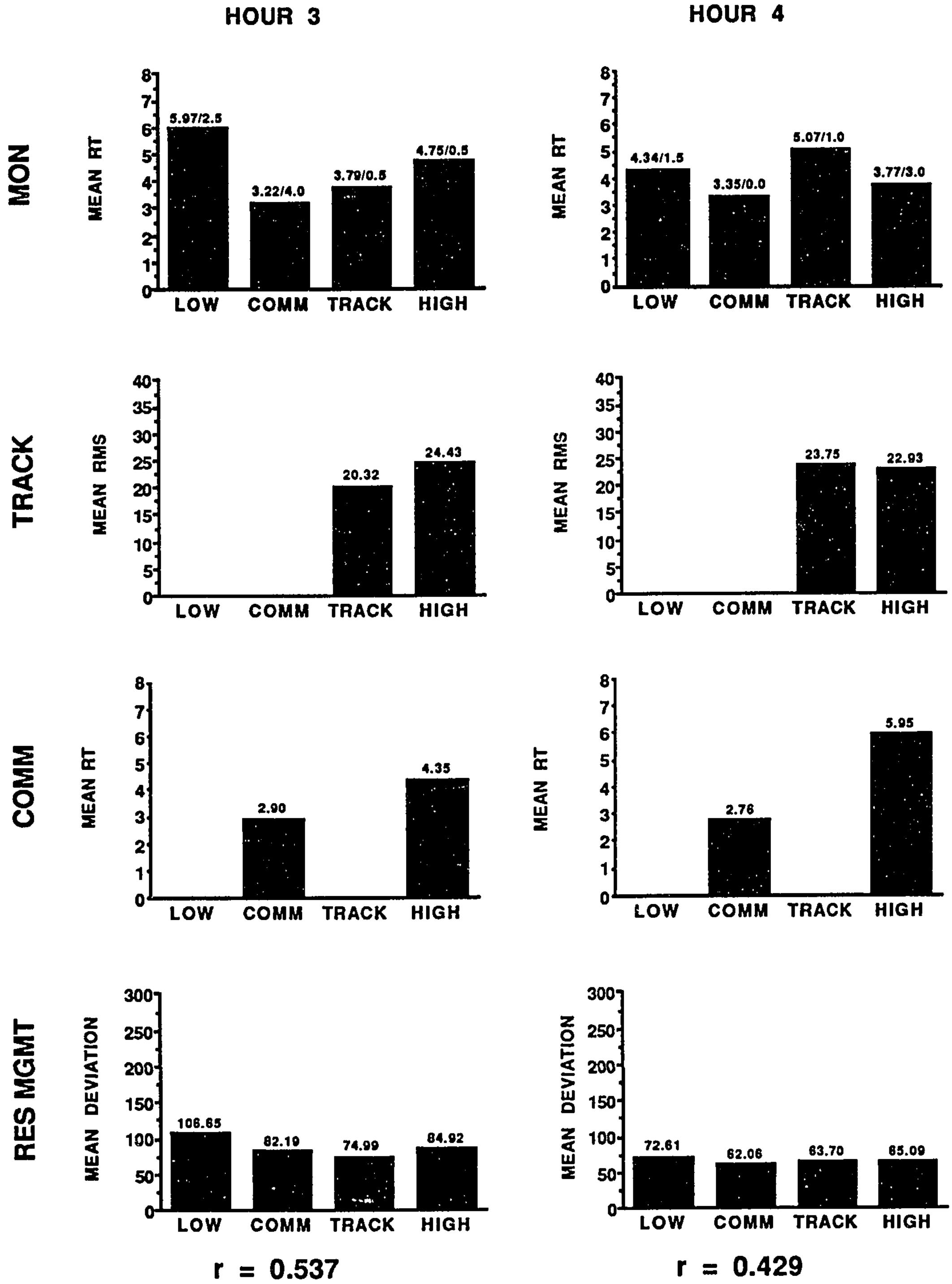
SUBJECT 6

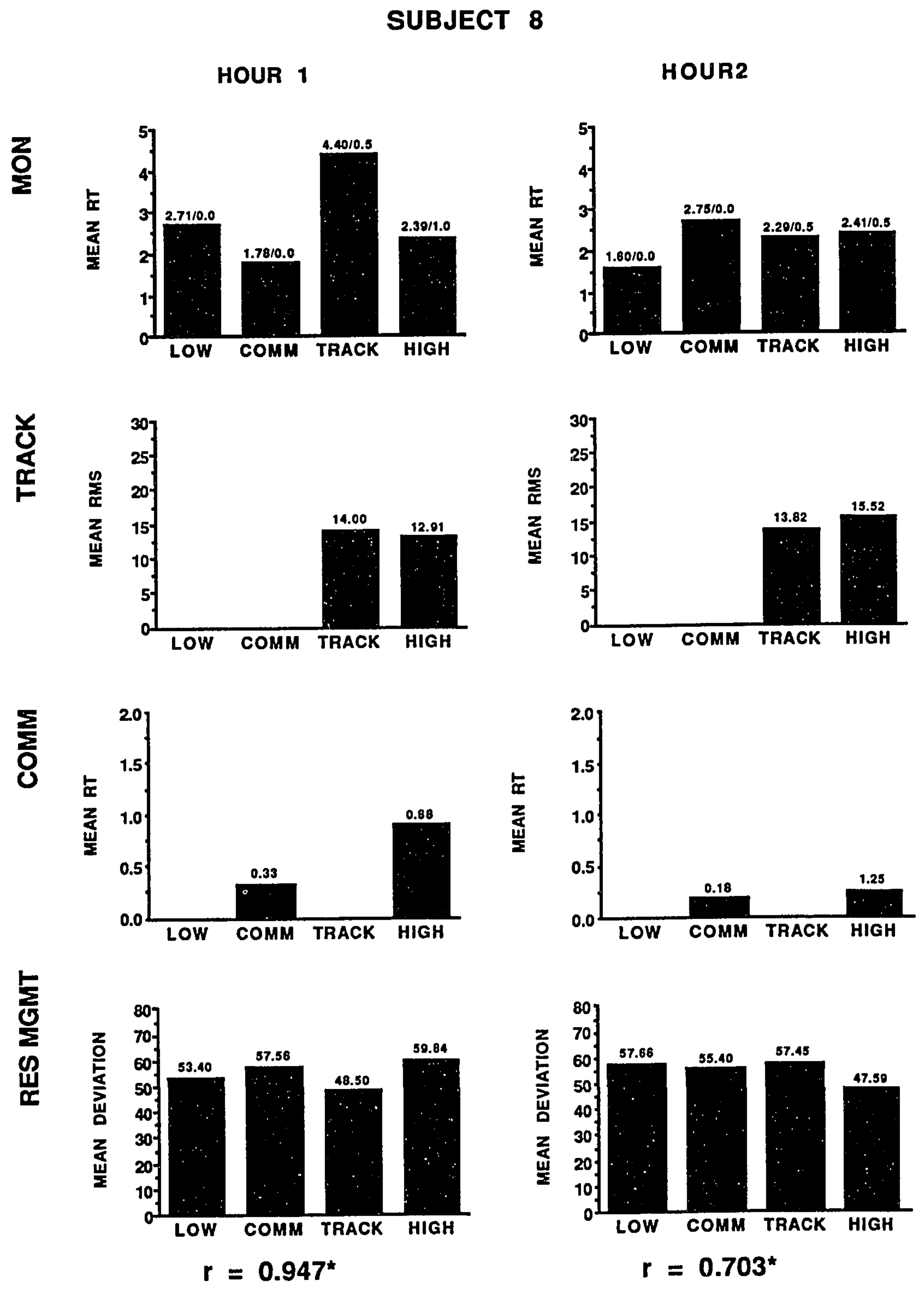

HOUR 3
HOUR 4
MON
MEAN RT
3.16/0.5
3.45/0.5
3.04/0.0
2.86/0.5
3.76/2.5
2.79/1.0
3.21/0.0
2.46/0.0
LOW
COMM
TRACK
HIGH
TRACK
MEAN RMS
15.62
20.39
16.07
19.11
COMM
MEAN RT
1.57
3.07
2.14
3.59
RES MGMT
MEAN DEVIATION
46.32
29.67
20.37
10.89
37.70
59.09
19.08
14.60
r = 1.000*
r = 0.997*



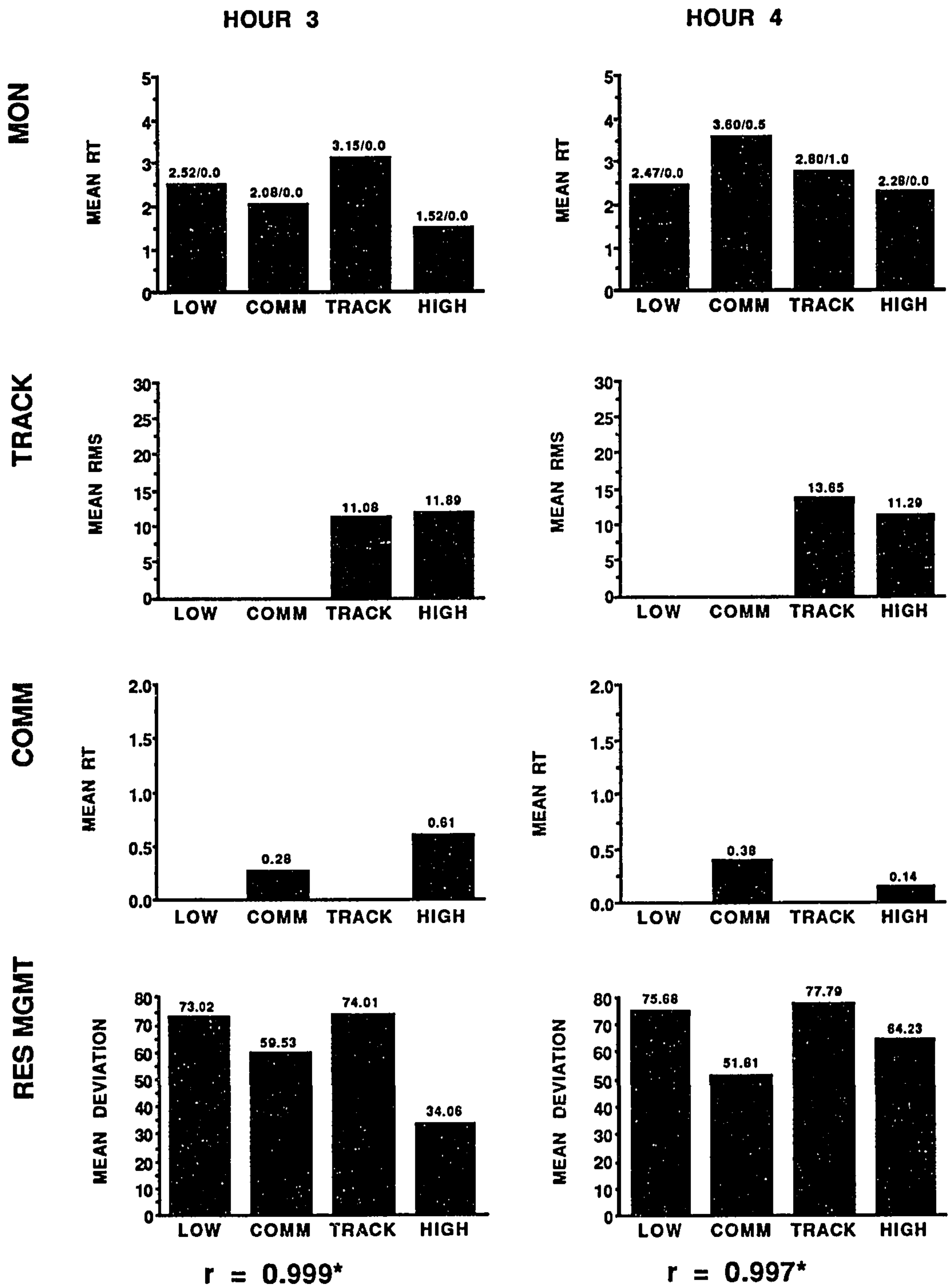
SUBJECT 7
HOUR 1
HOUR2
MON
MEAN RT
7.22/4.5
4.62/2.0
5.96/3.0
3.62/4.5
4.66/4.5
4.31/3.0
4.78/3.0
6.67/3.0
LOW
COMM
TRACK
HIGH
TRACK
MEAN RMS
35.87
21.51
22.48
26.25
COMM
MEAN RT
7.21
5.21
5.37
7.45
RES MGMT
MEAN DEVIATION
155.72
185.29
178.98
157.6
139.8
144.11
251.01
115.3
r = 0.322
r = 0.804*

SUBJECT 7

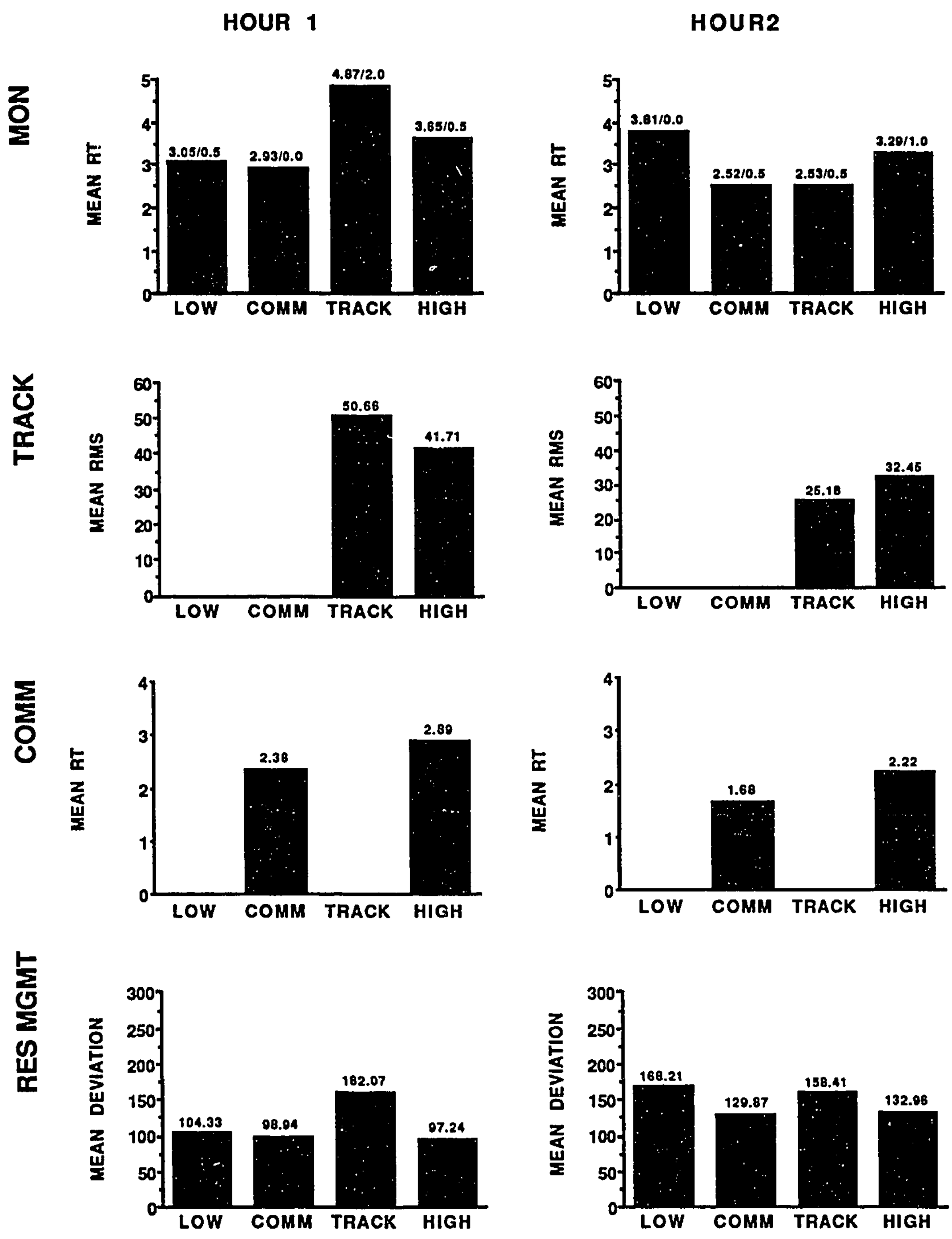

HOUR 3
HOUR 4
MON
MEAN RT
5.97/2.5
3.22/4.0
3.79/0.5
4.75/0.5
4.34/1.5
3.35/0.0
5.07/1.0
3.77/3.0
TRACK
MEAN RMS
20.32
24.43
23.75
22.93
COMM
MEAN RT
2.90
4.35
2.76
5.95
RES MGMT
MEAN DEVIATION
106.65
82.19
74.99
84.92
72.61
62.06
63.70
65.09
LOW
COMM
TRACK
HIGH
r = 0.537
r = 0.429


SUBJECT 8
HOUR 1
HOUR2
MON
TRACK
COMM
RES MGMT
MEAN RT
MEAN RMS
MEAN DEVIATION
2.71/0.0
1.78/0.0
4.40/0.5
2.39/1.0
1.60/0.0
2.75/0.0
2.29/0.5
2.41/0.5
14.00
12.91
13.82
15.52
0.33
0.88
0.18
1.25
53.40
57.56
48.50
59.84
57.66
55.40
57.45
47.59
LOW
COMM
TRACK
HIGH
r = 0.947*
r = 0.703*

SUBJECT 8


HOUR 3
HOUR 4
MON
TRACK
COMM
RES MGMT
MEAN RT
MEAN RMS
MEAN DEVIATION
LOW
COMM
TRACK
HIGH
2.52/0.0
2.08/0.0
3.15/0.0
1.52/0.0
2.47/0.0
3.60/0.5
2.80/1.0
2.28/0.0
11.08
11.89
13.65
11.29
0.28
0.61
0.38
0.14
73.02
59.53
74.01
34.06
75.68
51.61
77.79
64.23
r = 0.999*
r = 0.997*

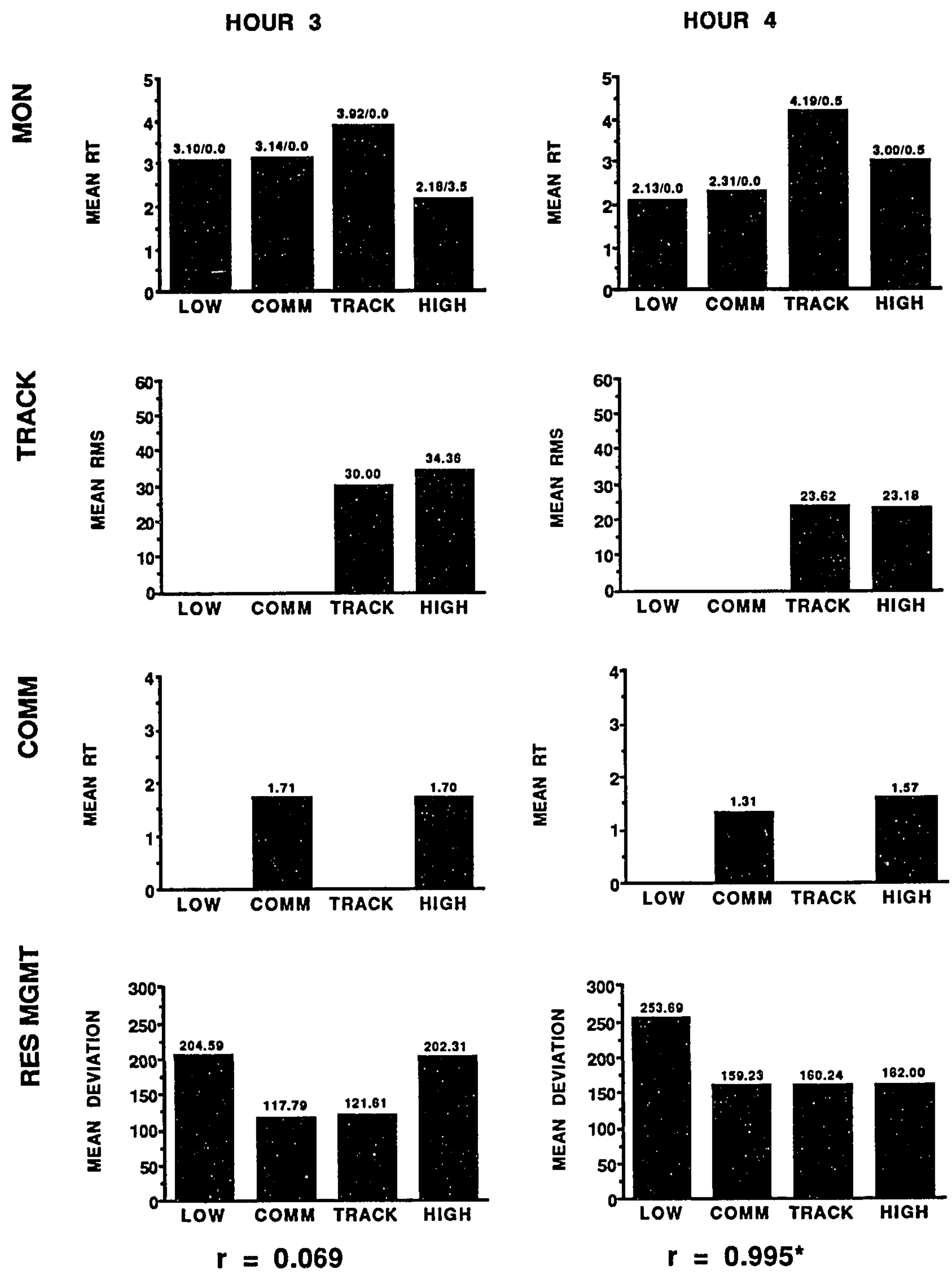
SUBJECT 9

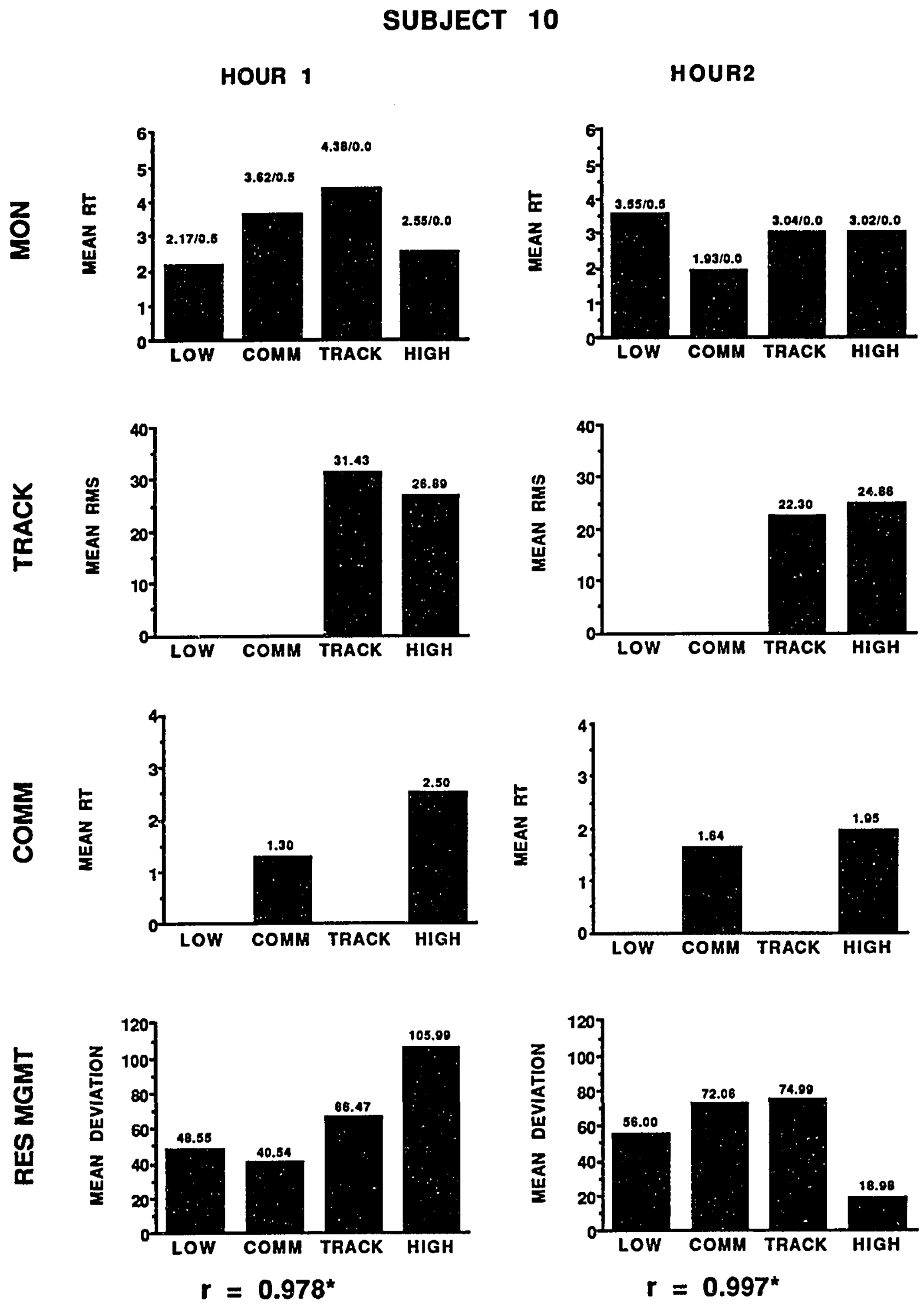


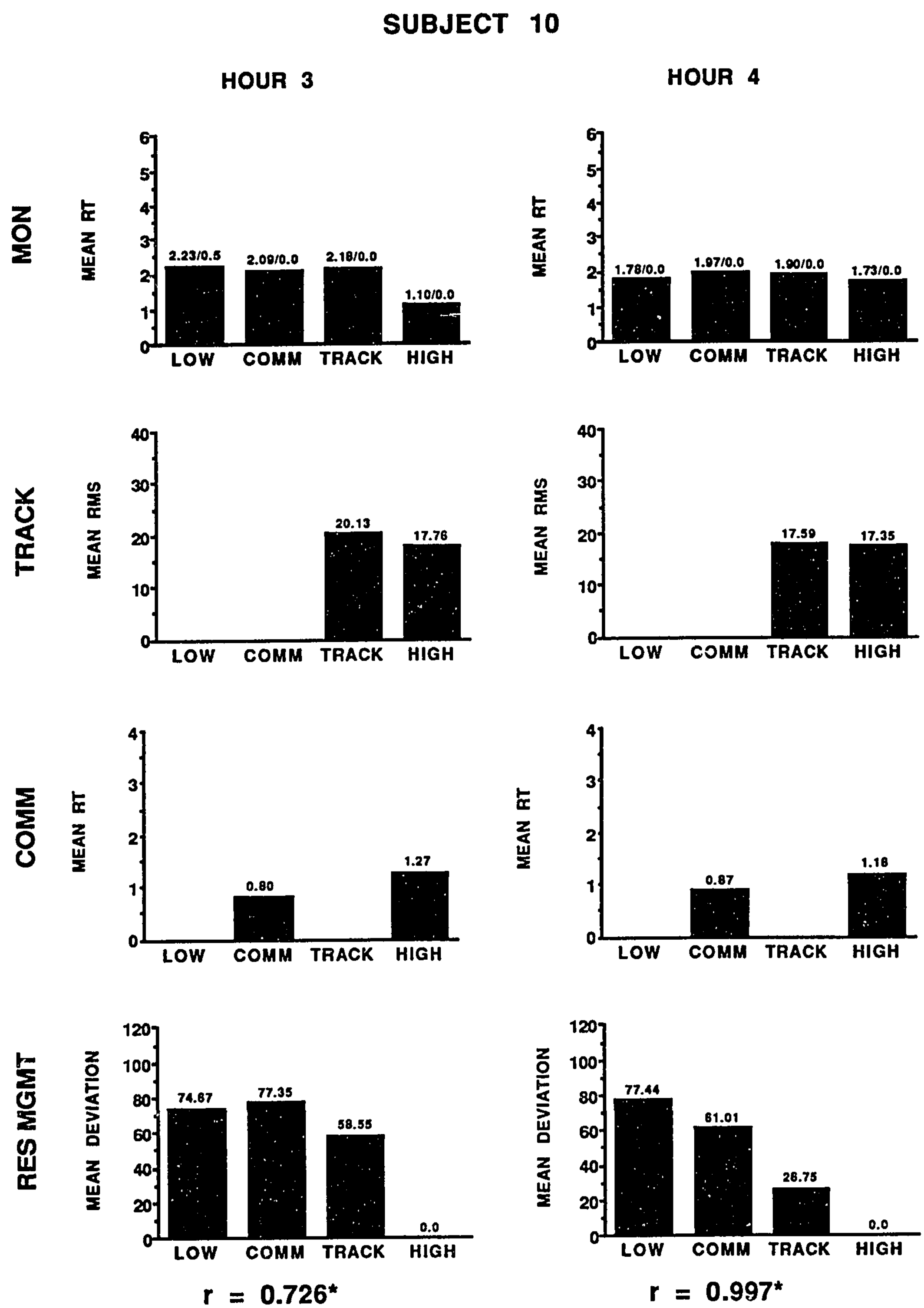
r = 0.994*

r = 0.589

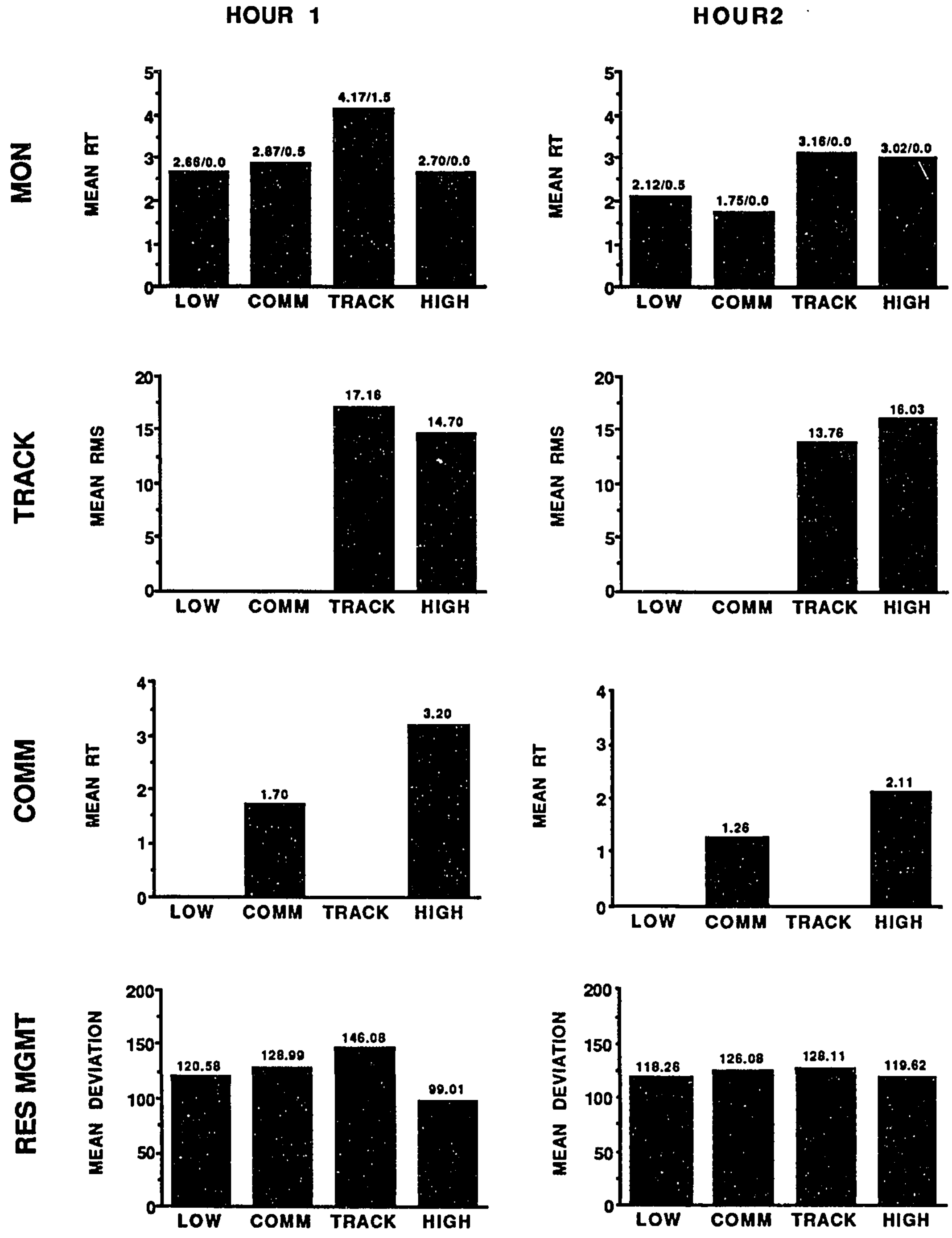
SUBJECT 9


HOUR 3
HOUR 4
MON
TRACK
COMM
RES MGMT
MEAN RT
MEAN RMS
MEAN DEVIATION
LOW
COMM
TRACK
HIGH
3.10/0.0
3.14/0.0
3.92/0.0
2.18/3.5
2.13/0.0
2.31/0.0
4.19/0.5
3.00/0.5
30.00
34.36
23.62
23.18
1.71
1.70
1.31
1.57
204.59
117.79
121.61
202.31
253.69
159.23
160.24
162.00
r = 0.069
r = 0.995*


SUBJECT 10
HOUR 1
HOUR2
MON
MEAN RT
2.17/0.5
3.62/0.5
4.38/0.0
2.55/0.0
3.55/0.5
1.93/0.0
3.04/0.0
3.02/0.0
TRACK
MEAN RMS
31.43
26.89
22.30
24.86
COMM
1.30
2.50
1.64
1.95
RES MGMT
MEAN DEVIATION
48.55
40.54
66.47
105.99
56.00
72.06
74.99
18.98
LOW
COMM
TRACK
HIGH
r = 0.978*
r = 0.997*

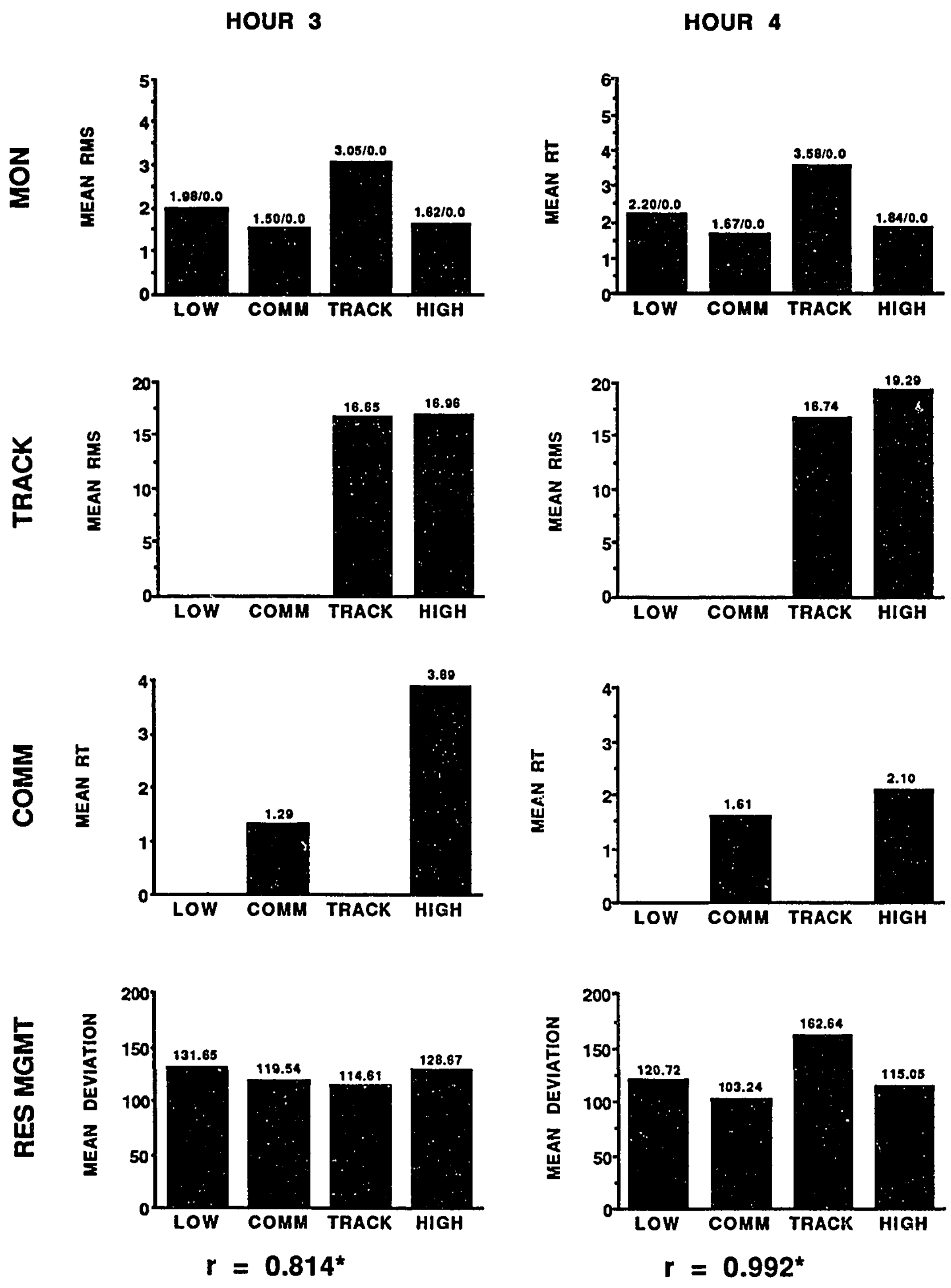

SUBJECT 10
HOUR 3
HOUR 4
MON
MEAN RT
2.23/0.5
2.09/0.0
2.18/0.0
1.10/0.0
1.78/0.0
1.97/0.0
1.90/0.0
1.73/0.0
LOW
COMM
TRACK
HIGH
TRACK
MEAN RMS
20.13
17.76
17.59
17.35
COMM
MEAN RT
0.80
1.27
0.87
1.18
RES MGMT
MEAN DEVIATION
74.67
77.35
58.55
0.0
77.44
61.01
26.75
0.0
r = 0.726*
r = 0.997*

SUBJECT 11

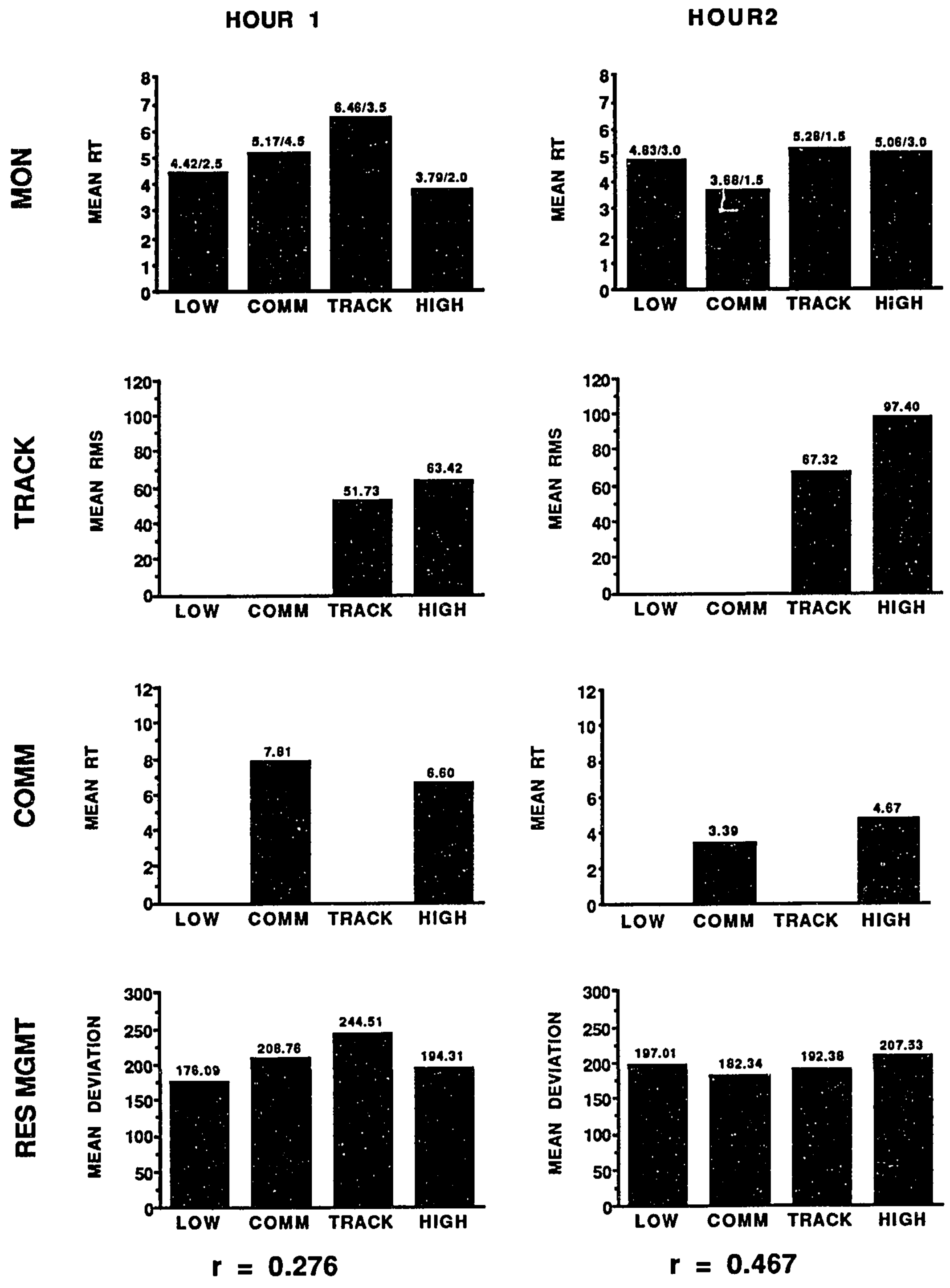


r = 0.812*   r = 0.797*

SUBJECT 11

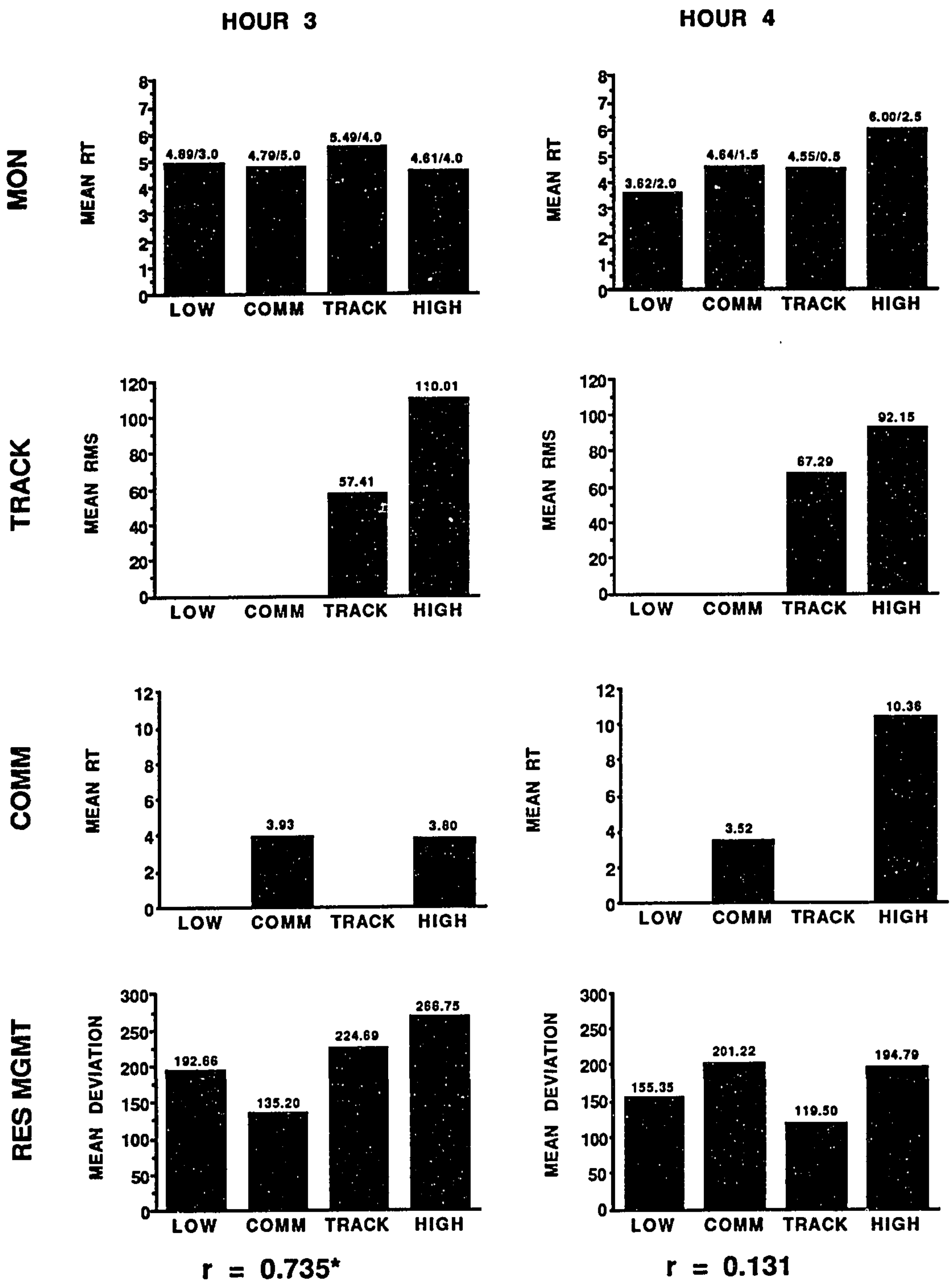

HOUR 3
HOUR 4
MON
MEAN RMS
MEAN RT
1.98/0.0
1.50/0.0
3.05/0.0
1.62/0.0
2.20/0.0
1.67/0.0
3.58/0.0
1.84/0.0
TRACK
MEAN RMS
16.65
16.96
16.74
19.29
COMM
MEAN RT
1.29
3.89
1.61
2.10
RES MGMT
MEAN DEVIATION
131.65
119.54
114.61
128.67
120.72
103.24
162.64
115.05
LOW
COMM
TRACK
HIGH
r = 0.814*
r = 0.992*

SUBJECT 12

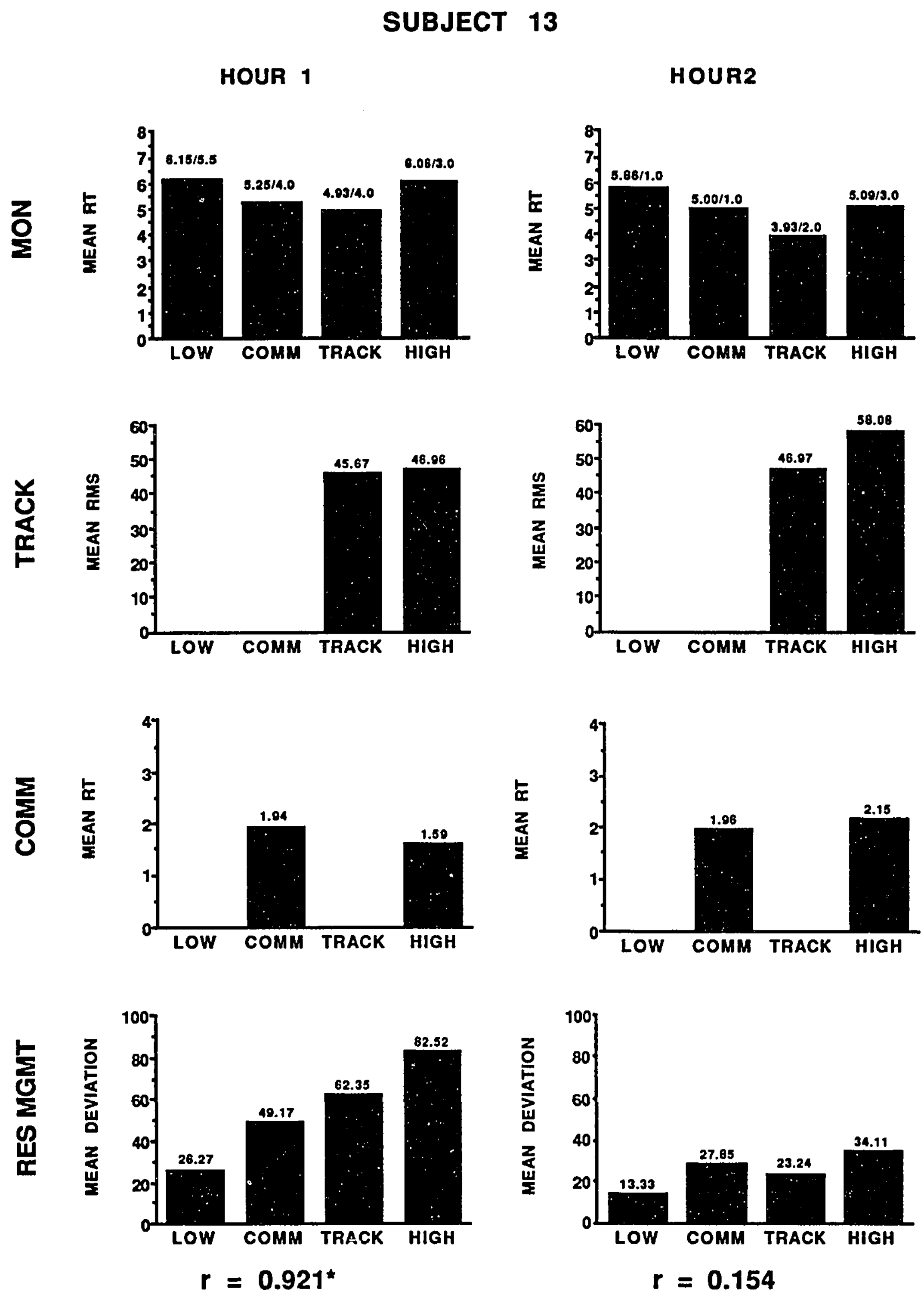


HOUR 1
HOUR2
MON
MEAN RT
4.42/2.5
5.17/4.5
6.46/3.5
3.79/2.0
4.83/3.0
3.68/1.5
5.28/1.5
5.06/3.0
TRACK
MEAN RMS
51.73
63.42
67.32
97.40
COMM
MEAN RT
7.81
6.60
3.39
4.67
RES MGMT
MEAN DEVIATION
176.09
208.76
244.51
194.31
197.01
182.34
192.38
207.33
LOW
COMM
TRACK
HIGH
r = 0.276
r = 0.467

SUBJECT 12

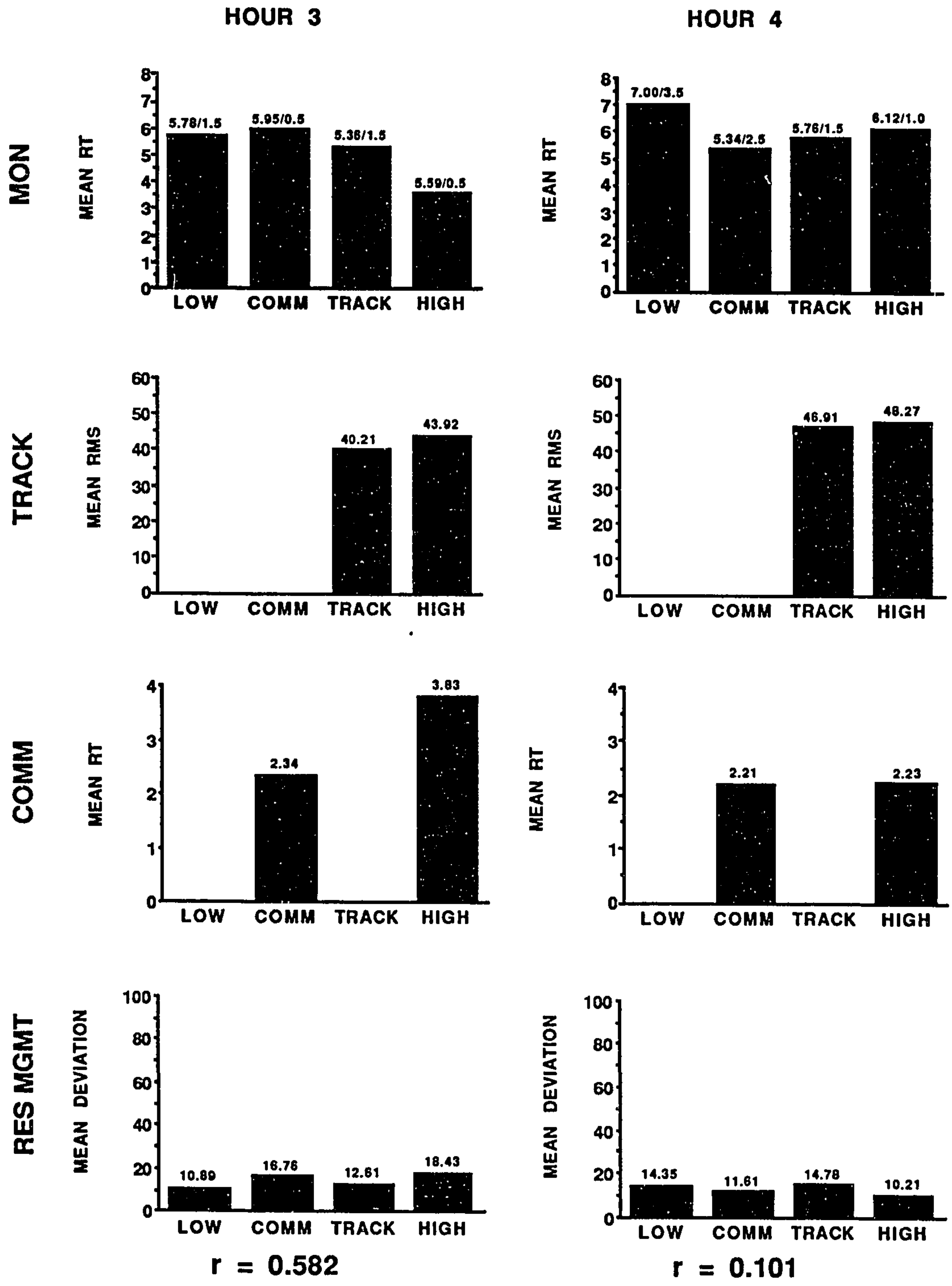


HOUR 3
HOUR 4
MON
TRACK
COMM
RES MGMT
MEAN RT
MEAN RMS
MEAN DEVIATION
LOW
COMM
TRACK
HIGH
4.89/3.0
4.79/5.0
5.49/4.0
4.61/4.0
3.62/2.0
4.64/1.5
4.55/0.5
6.00/2.5
57.41
110.01
67.29
92.15
3.93
3.80
3.52
10.36
192.66
135.20
224.69
266.75
155.35
201.22
119.50
194.79
r = 0.735*
r = 0.131

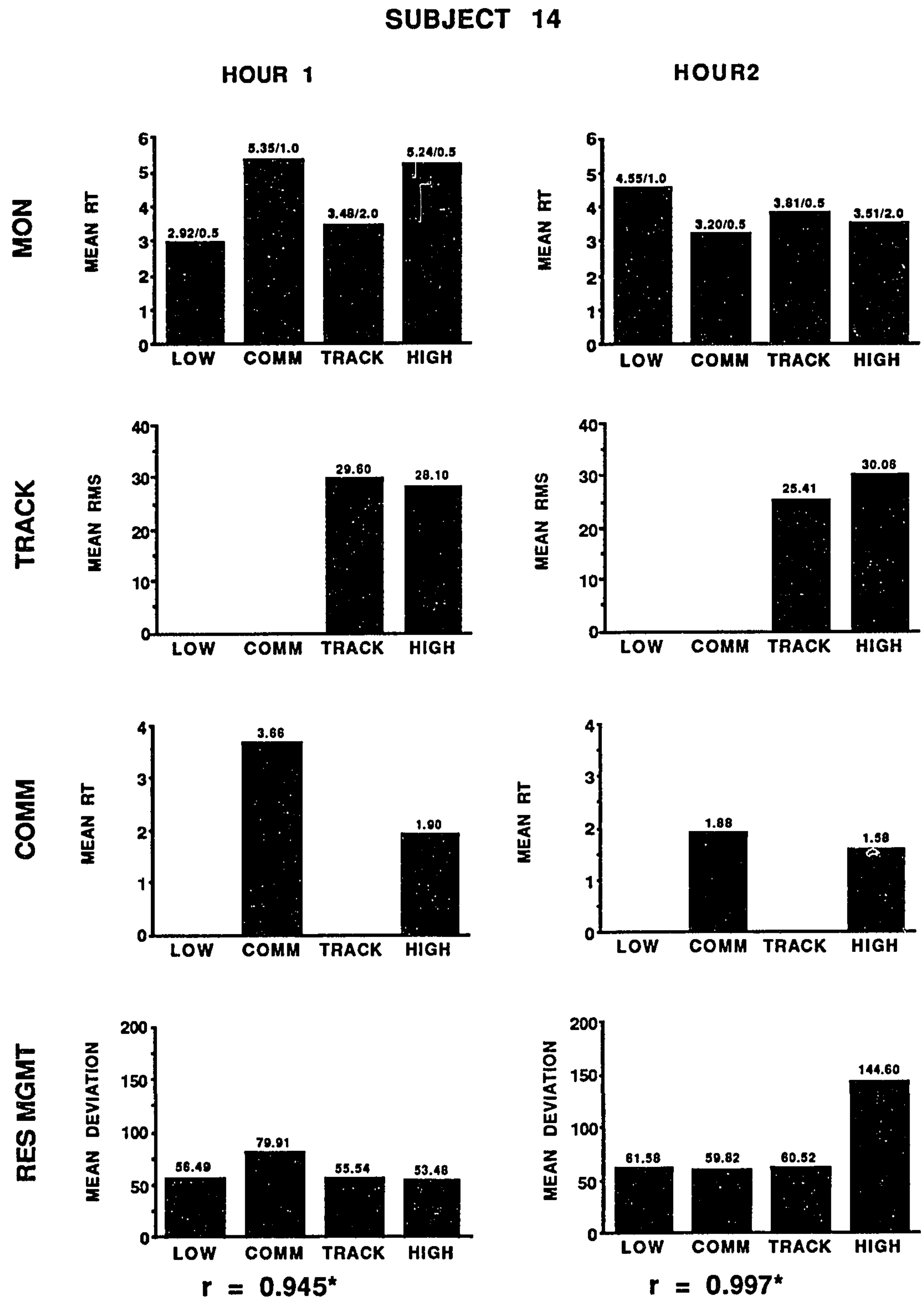


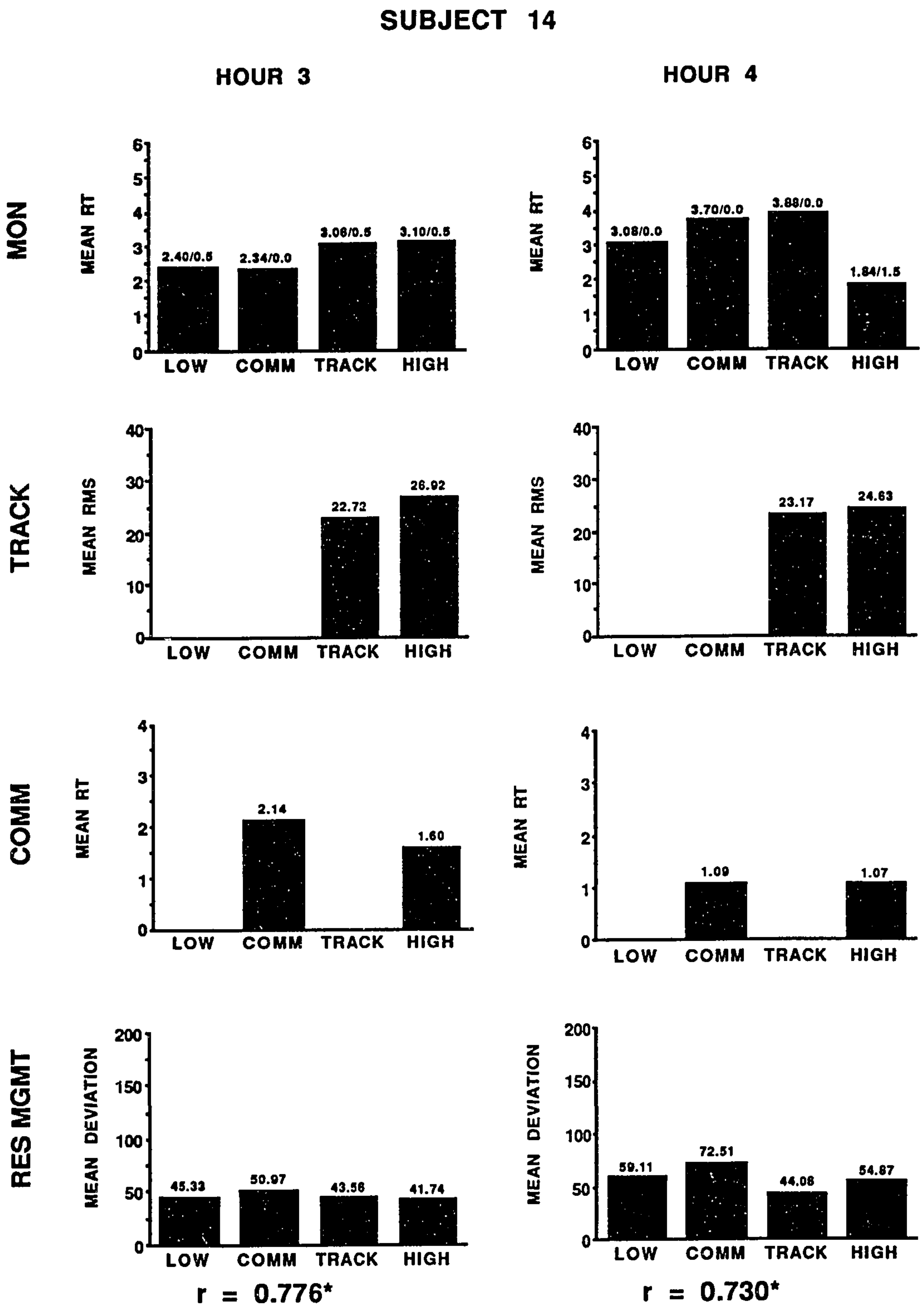
SUBJECT 13
HOUR 1
HOUR2
MON
TRACK
COMM
RES MGMT
MEAN RT
MEAN RMS
MEAN DEVIATION
LOW
COMM
TRACK
HIGH
6.15/5.5
5.25/4.0
4.93/4.0
6.06/3.0
5.86/1.0
5.00/1.0
3.93/2.0
5.09/3.0
45.67
46.96
46.97
58.08
1.94
1.59
1.96
2.15
26.27
49.17
62.35
82.52
13.33
27.85
23.24
34.11
r = 0.921*
r = 0.154

SUBJECT 13

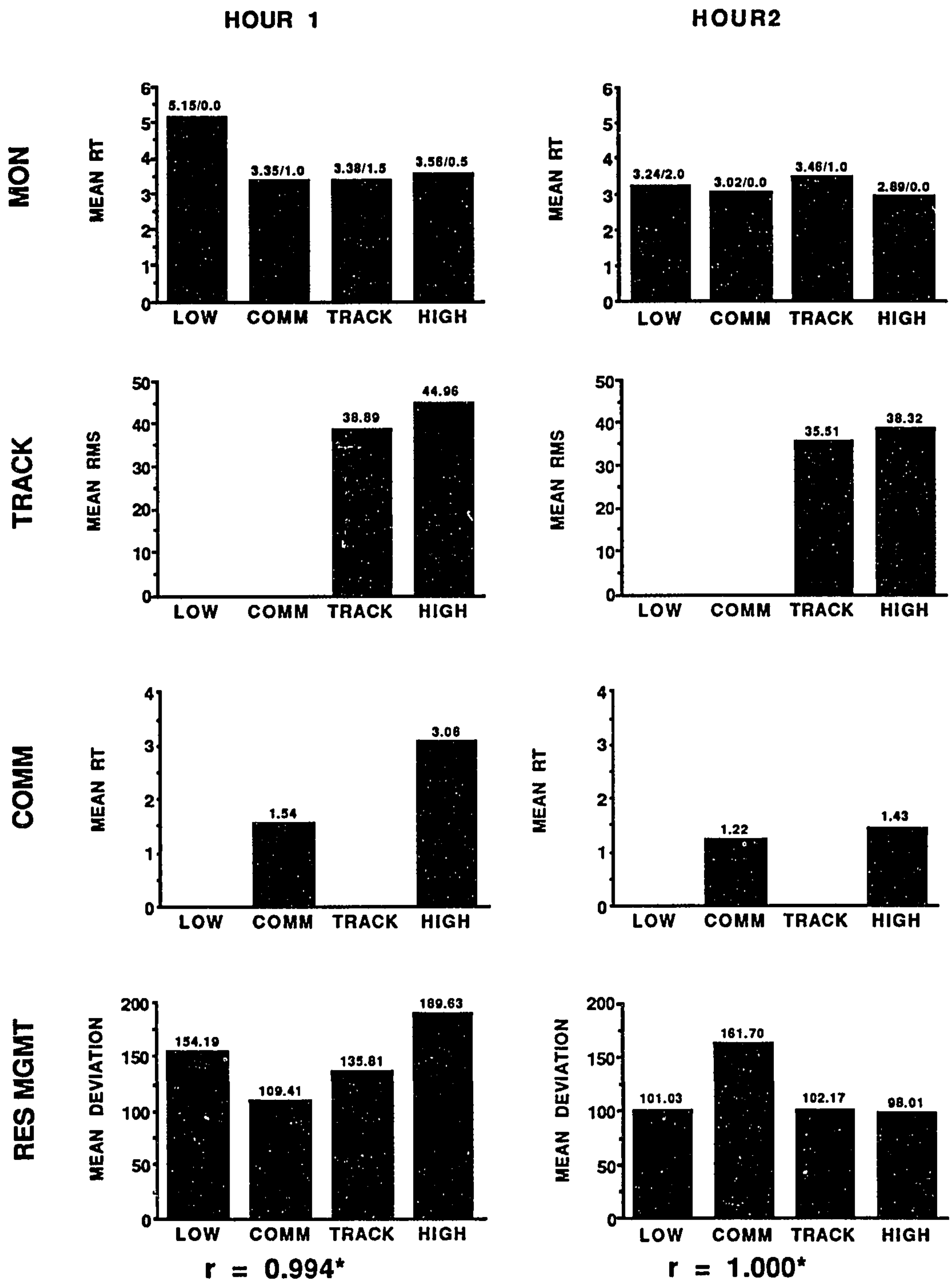


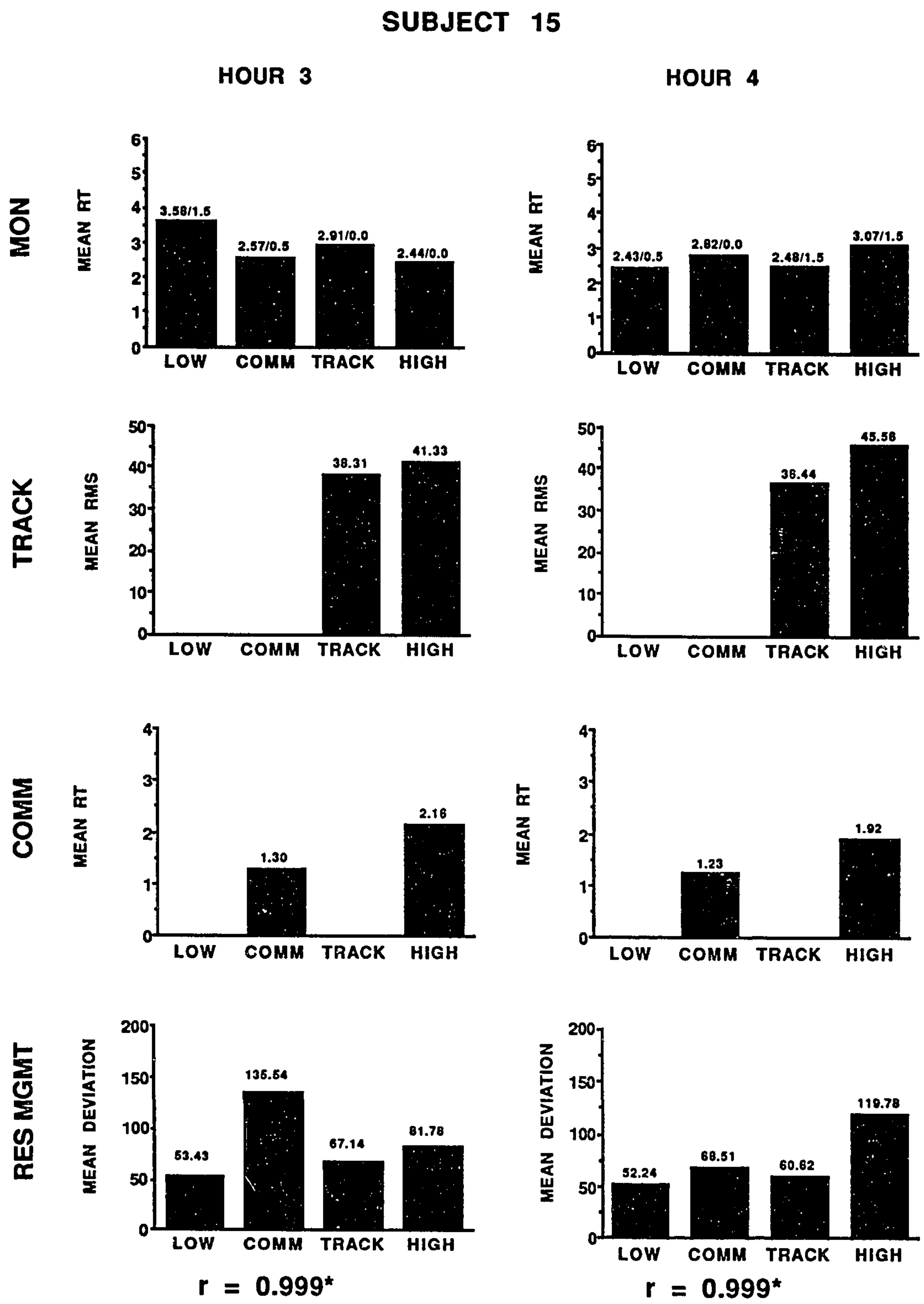
HOUR 3
HOUR 4
MON
MEAN RT
5.78/1.5
5.95/0.5
5.36/1.5
5.59/0.5
7.00/3.5
5.34/2.5
5.76/1.5
6.12/1.0
LOW COMM TRACK HIGH
TRACK
MEAN RMS
40.21
43.92
46.91
48.27
COMM
MEAN RT
2.34
3.83
2.21
2.23
RES MGMT
MEAN DEVIATION
10.89
16.76
12.61
18.43
14.35
11.61
14.78
10.21
r = 0.582
r = 0.101

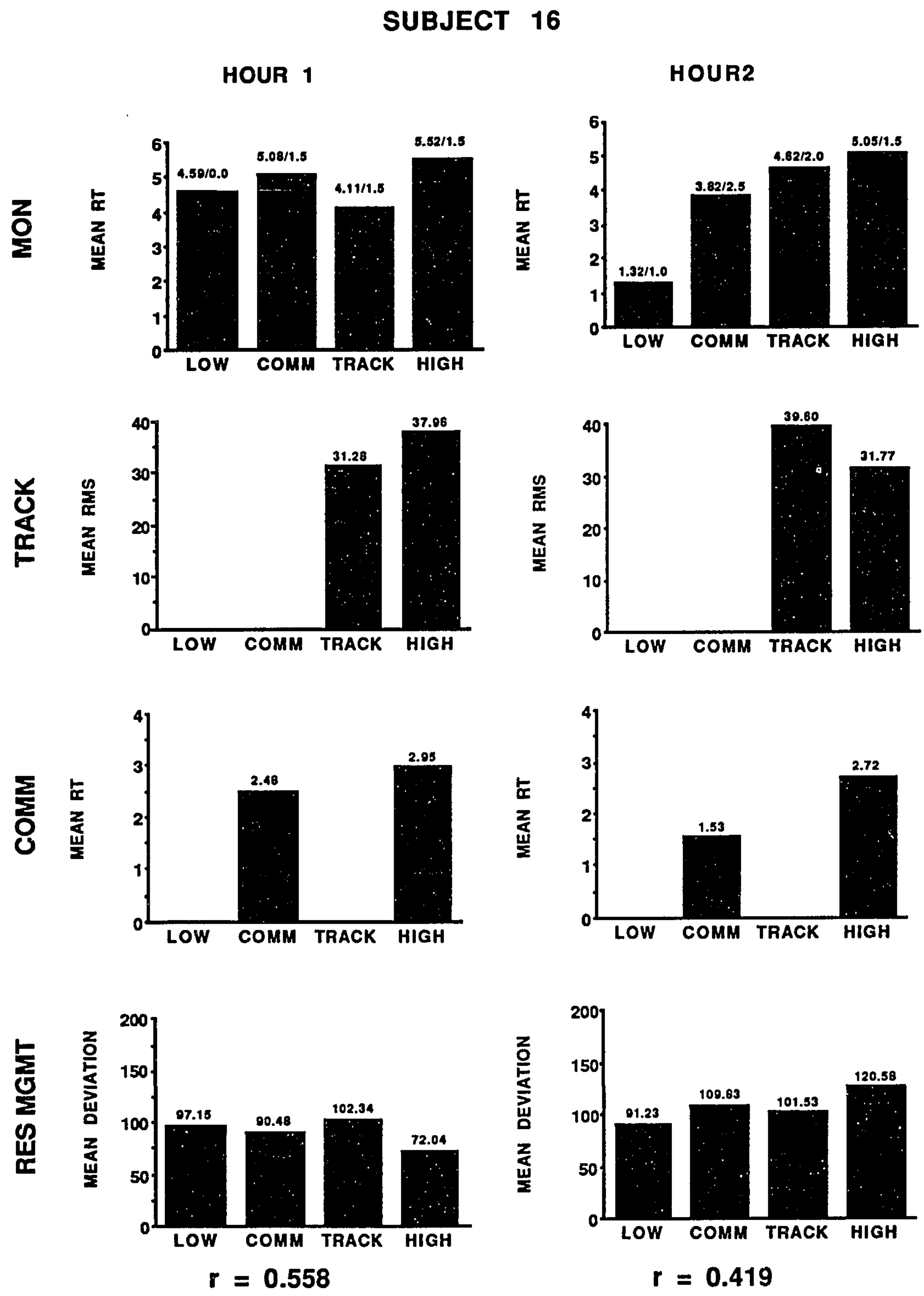

SUBJECT 14
HOUR 1
HOUR2
MON
MEAN RT
2.92/0.5
5.35/1.0
3.48/2.0
5.24/0.5
4.55/1.0
3.20/0.5
3.81/0.5
3.51/2.0
LOW
COMM
TRACK
HIGH
TRACK
MEAN RMS
29.60
28.10
25.41
30.06
COMM
MEAN RT
3.66
1.90
1.88
1.58
RES MGMT
MEAN DEVIATION
56.49
79.91
55.54
53.48
61.58
59.82
60.52
144.60
r = 0.945*
r = 0.997*


SUBJECT 14
HOUR 3
HOUR 4
MON
MEAN RT
2.40/0.5
2.34/0.0
3.06/0.5
3.10/0.5
3.08/0.0
3.70/0.0
3.88/0.0
1.84/1.5
LOW
COMM
TRACK
HIGH
TRACK
MEAN RMS
22.72
26.92
23.17
24.63
COMM
1.09
1.07
2.14
1.60
RES MGMT
MEAN DEVIATION
45.33
50.97
43.56
41.74
59.11
72.51
44.08
54.87
r = 0.776*
r = 0.730*

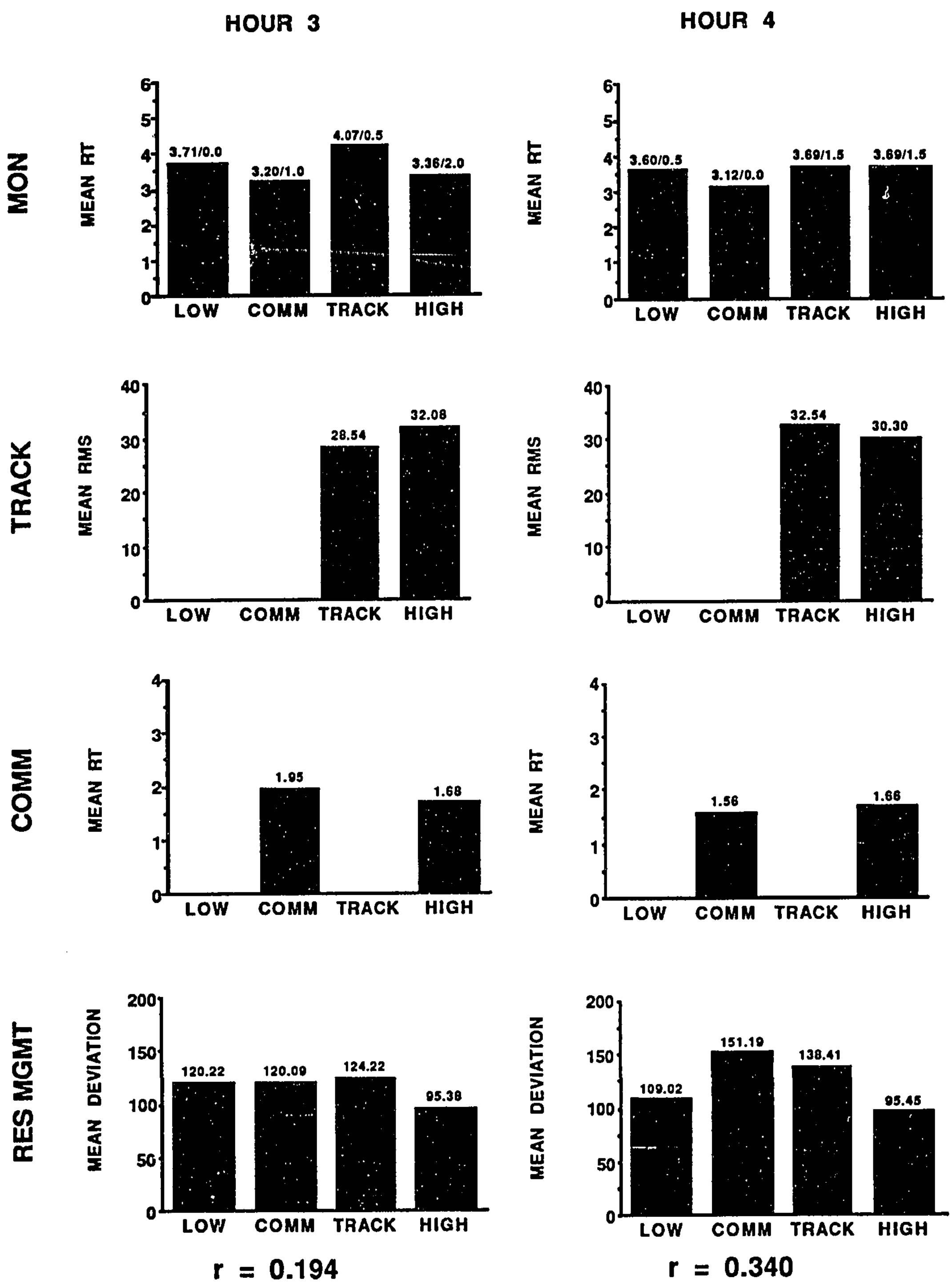
SUBJECT 15


HOUR 1
HOUR2
MON
MEAN RT
5.15/0.0
3.35/1.0
3.38/1.5
3.56/0.5
3.24/2.0
3.02/0.0
3.46/1.0
2.89/0.0
LOW COMM TRACK HIGH
TRACK
MEAN RMS
38.89
44.96
35.51
38.32
COMM
MEAN RT
1.54
3.08
1.22
1.43
RES MGMT
MEAN DEVIATION
154.19
109.41
135.81
189.63
101.03
161.70
102.17
98.01
r = 0.994*
r = 1.000*

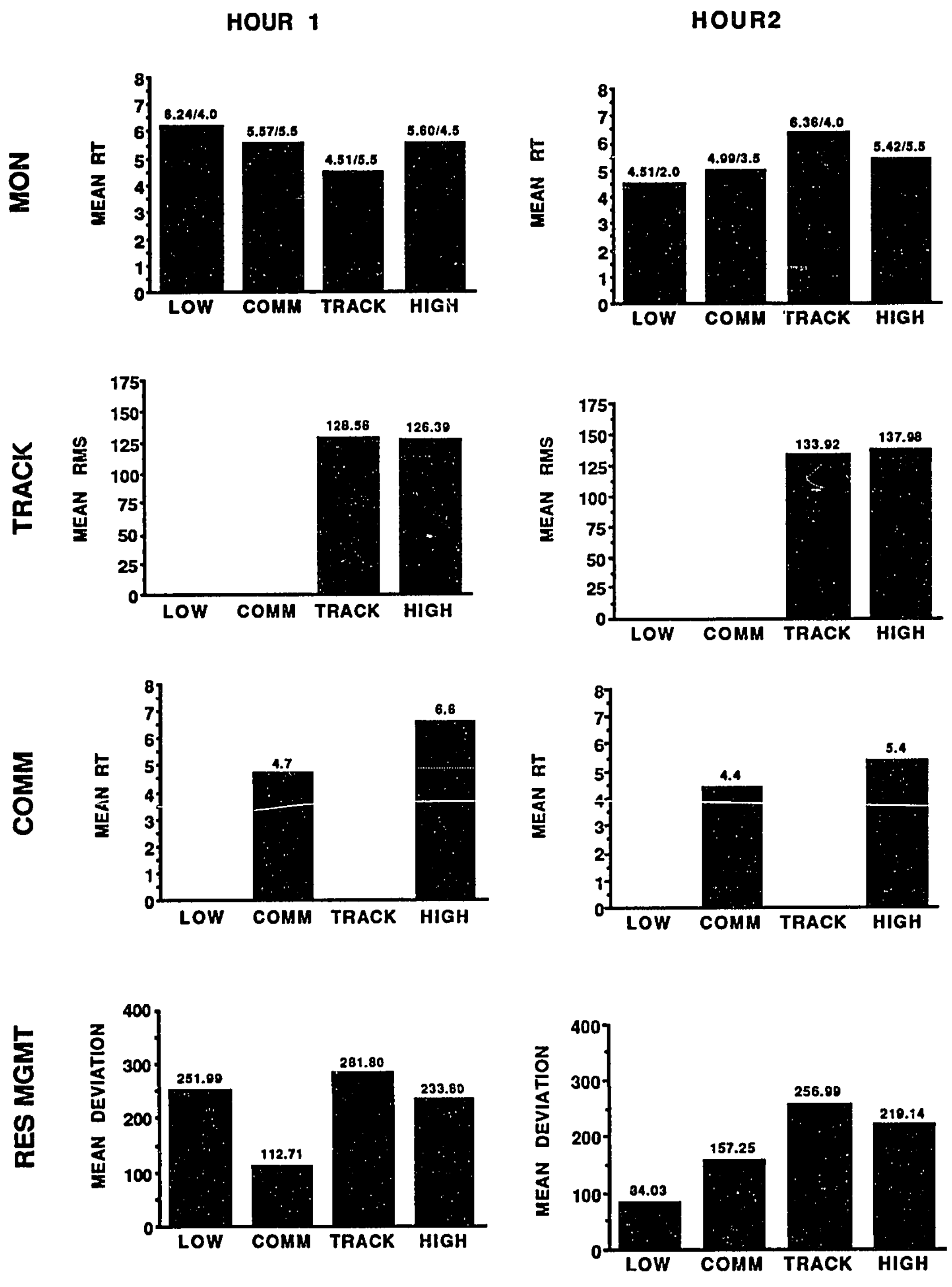

SUBJECT 15
HOUR 3
HOUR 4
MON
MEAN RT
3.58/1.5
2.57/0.5
2.91/0.0
2.44/0.0
2.43/0.5
2.82/0.0
2.48/1.5
3.07/1.5
LOW
COMM
TRACK
HIGH
TRACK
MEAN RMS
38.31
41.33
36.44
45.58
COMM
MEAN RT
1.30
2.16
1.23
1.92
RES MGMT
MEAN DEVIATION
53.43
135.54
67.14
81.78
52.24
68.51
60.62
119.78
r = 0.999*
r = 0.999*


SUBJECT 16
HOUR 1
HOUR2
MON
MEAN RT
4.59/0.0
5.08/1.5
4.11/1.5
5.52/1.5
1.32/1.0
3.82/2.5
4.62/2.0
5.05/1.5
TRACK
MEAN RMS
31.28
37.96
39.60
31.77
COMM
2.48
2.95
1.53
2.72
RES MGMT
MEAN DEVIATION
97.15
90.48
102.34
72.04
91.23
109.83
101.53
120.58
LOW
COMM
TRACK
HIGH
r = 0.558
r = 0.419

SUBJECT 16


HOUR 3
HOUR 4
MON
MEAN RT
3.71/0.0
3.20/1.0
4.07/0.5
3.36/2.0
3.60/0.5
3.12/0.0
3.69/1.5
3.69/1.5
LOW
COMM
TRACK
HIGH
TRACK
MEAN RMS
28.54
32.08
32.54
30.30
COMM
MEAN RT
1.95
1.68
1.56
1.66
RES MGMT
MEAN DEVIATION
120.22
120.09
124.22
95.38
109.02
151.19
138.41
95.45
r = 0.194
r = 0.340

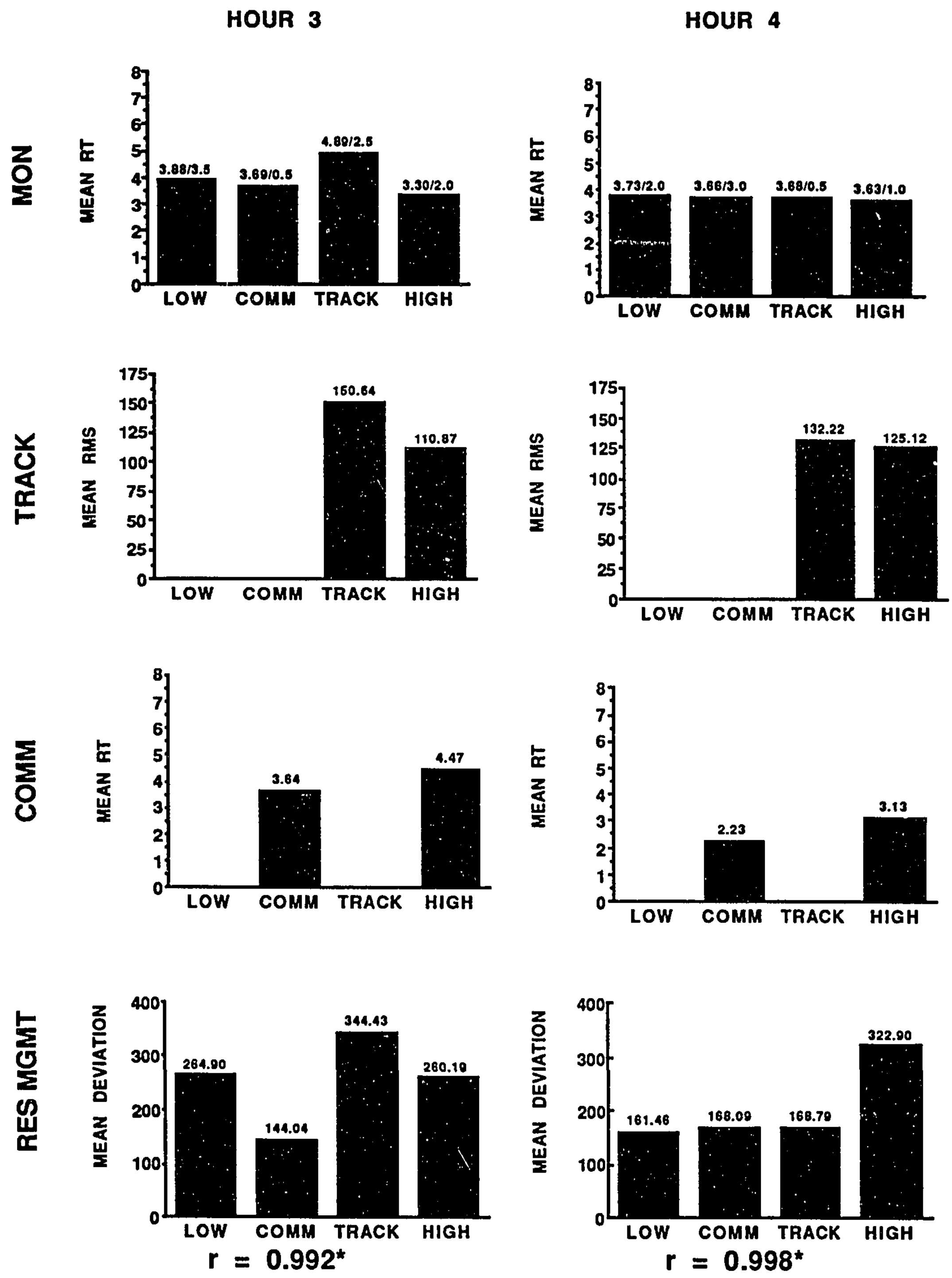
SUBJECT 17



r = 0.851*

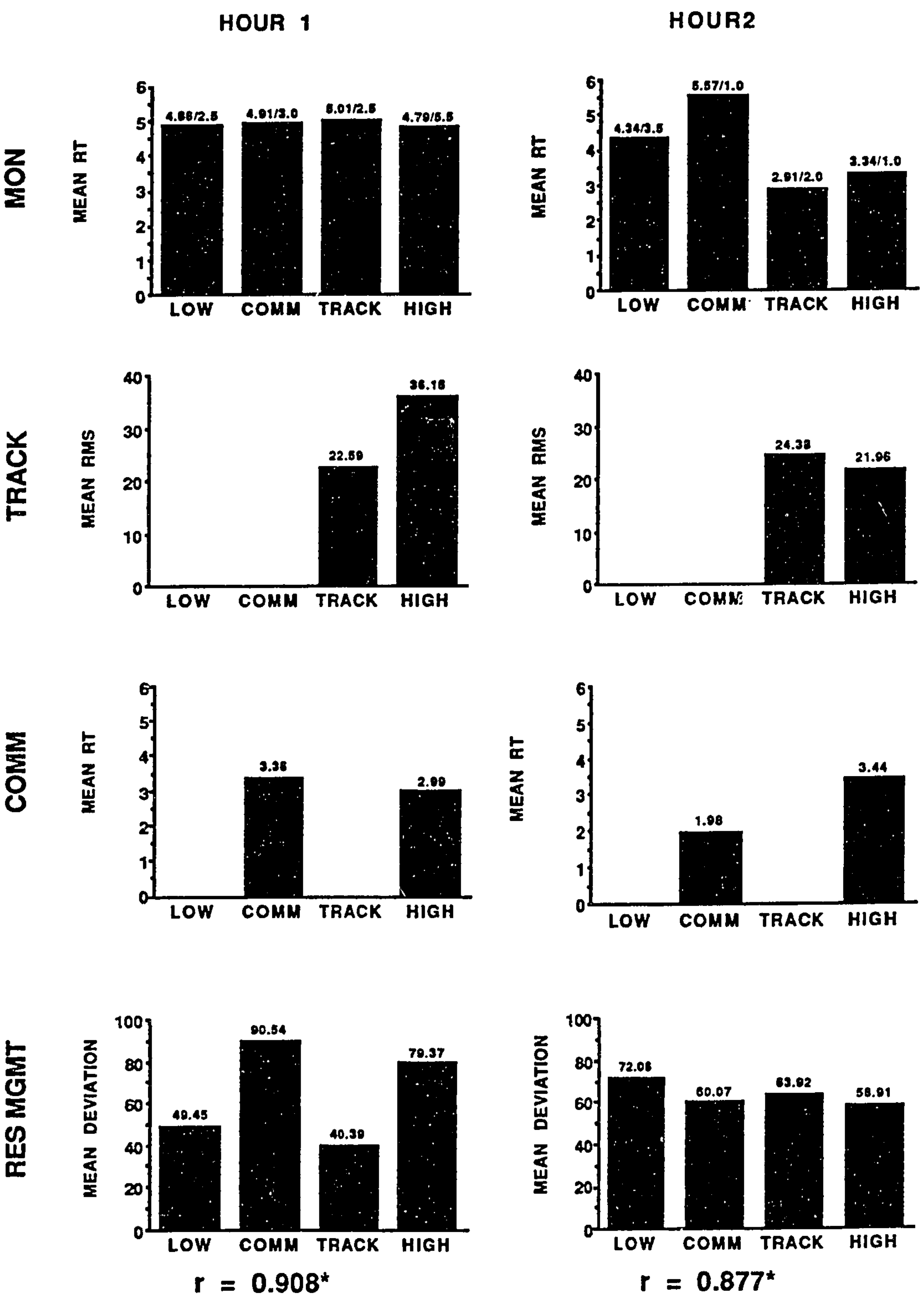
r = 0.968*

SUBJECT 17



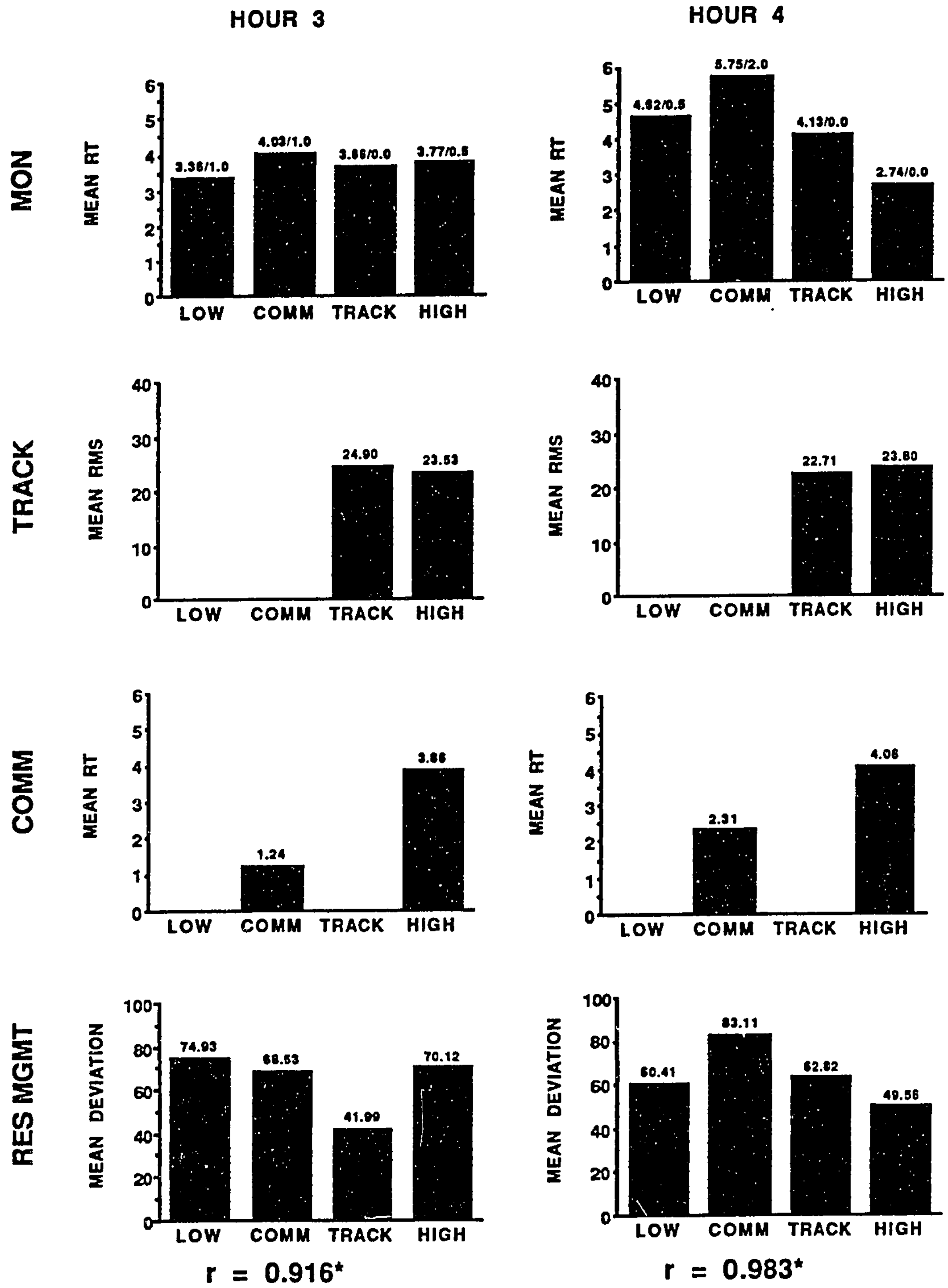
HOUR 3
HOUR 4
MON
MEAN RT
3.88/3.5
3.69/0.5
4.89/2.5
3.30/2.0
3.73/2.0
3.66/3.0
3.68/0.5
3.63/1.0
TRACK
MEAN RMS
150.64
110.87
132.22
125.12
COMM
MEAN RT
3.64
4.47
2.23
3.13
RES MGMT
MEAN DEVIATION
264.90
144.04
344.43
260.19
161.46
168.09
168.79
322.90
LOW
COMM
TRACK
HIGH
r = 0.992*
r = 0.998*

SUBJECT 18

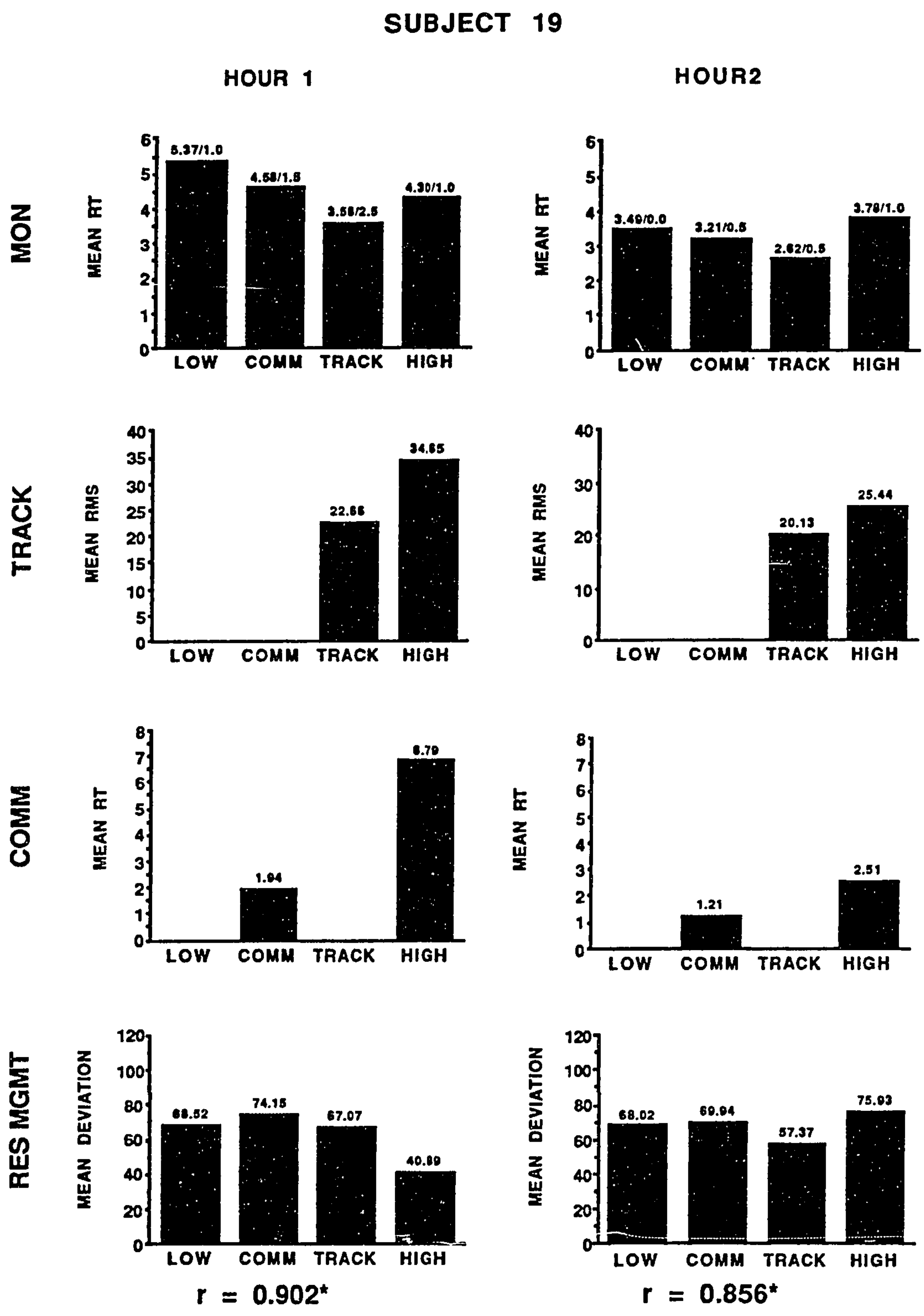


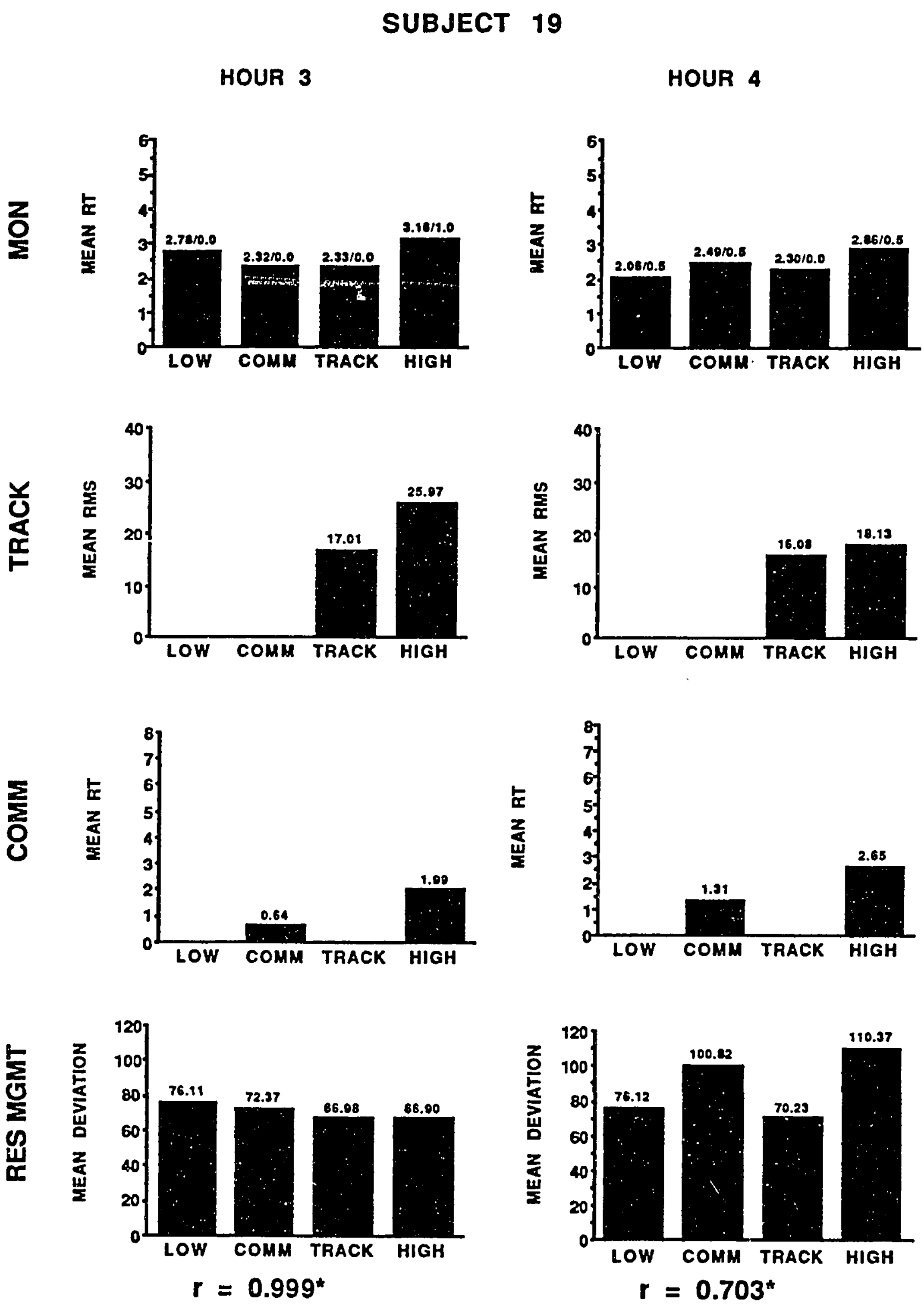
HOUR 1
HOUR2
MON
TRACK
COMM
RES MGMT
MEAN RT
MEAN RMS
MEAN DEVIATION
LOW
COMM
TRACK
HIGH
4.88/2.5
4.91/3.0
5.01/2.5
4.79/5.5
4.34/3.5
5.57/1.0
2.91/2.0
3.34/1.0
22.59
36.18
24.38
21.96
3.36
2.99
1.98
3.44
49.45
90.54
40.39
79.37
72.08
60.07
63.92
58.91
r = 0.908*
r = 0.877*

SUBJECT 18

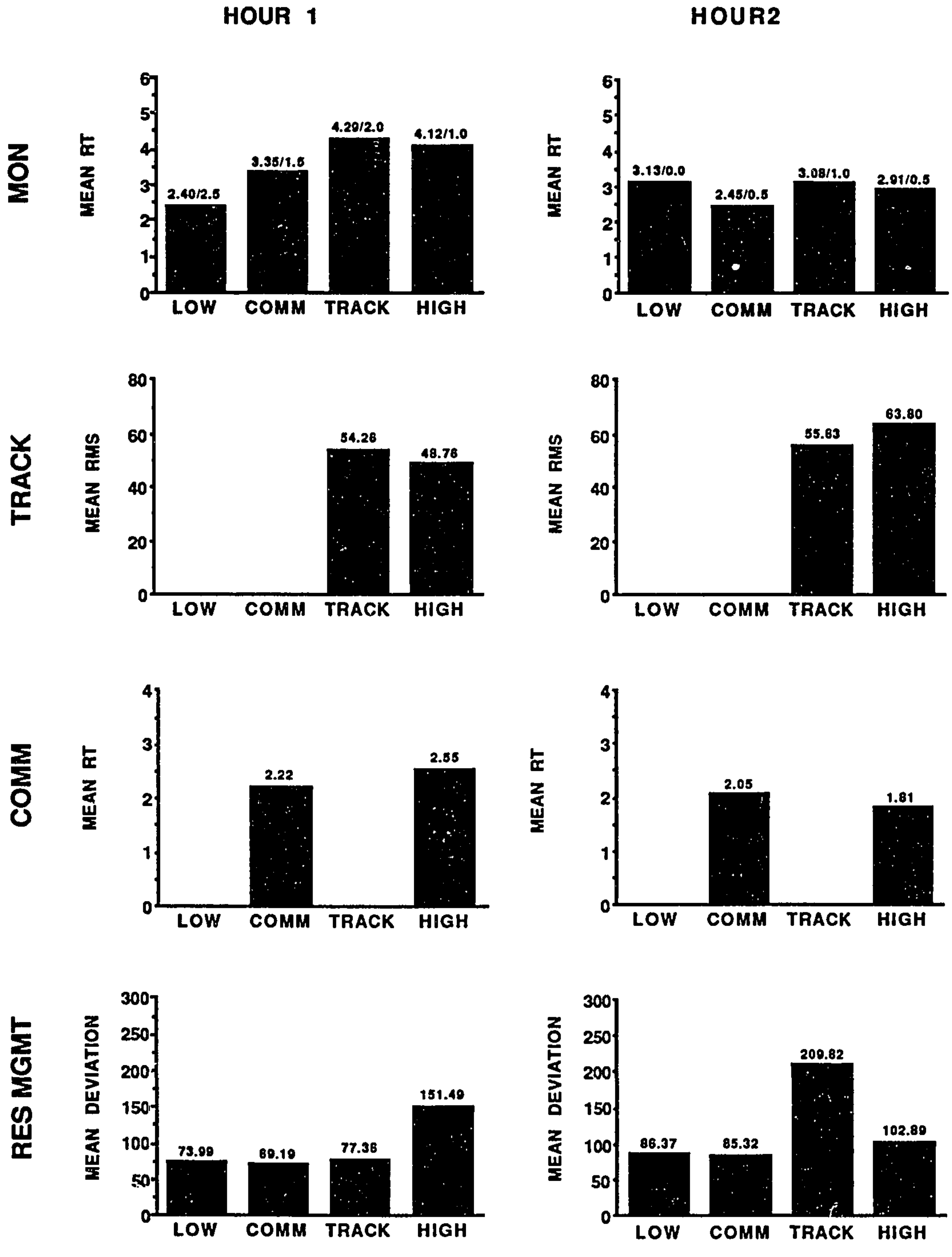


HOUR 3
HOUR 4
MON
TRACK
COMM
RES MGMT
MEAN RT
MEAN RMS
MEAN DEVIATION
LOW
COMM
TRACK
HIGH
3.36/1.0
4.03/1.0
3.66/0.0
3.77/0.5
4.62/0.5
5.75/2.0
4.13/0.0
2.74/0.0
24.90
23.53
22.71
23.80
1.24
3.86
2.31
4.06
74.93
68.53
41.99
70.12
60.41
83.11
62.82
49.56
r = 0.916*
r = 0.983*


SUBJECT 19
HOUR 1
HOUR2
MON
TRACK
COMM
RES MGMT
MEAN RT
MEAN RMS
MEAN DEVIATION
LOW
COMM
TRACK
HIGH
5.37/1.0
4.58/1.5
3.58/2.5
4.30/1.0
3.49/0.0
3.21/0.5
2.62/0.5
3.78/1.0
22.66
34.65
20.13
25.44
1.94
6.79
1.21
2.51
68.52
74.15
67.07
40.89
68.02
69.94
57.37
75.93
r = 0.902*
r = 0.856*


SUBJECT 19
HOUR 3
HOUR 4
MON
MEAN RT
2.78/0.0
2.32/0.0
2.33/0.0
3.16/1.0
2.06/0.5
2.49/0.5
2.30/0.0
2.86/0.5
LOW
COMM
TRACK
HIGH
TRACK
MEAN RMS
17.01
25.97
16.08
18.13
COMM
MEAN RT
0.64
1.99
1.31
2.65
RES MGMT
MEAN DEVIATION
76.11
72.37
66.98
66.90
76.12
100.82
70.23
110.37
r = 0.999*
r = 0.703*

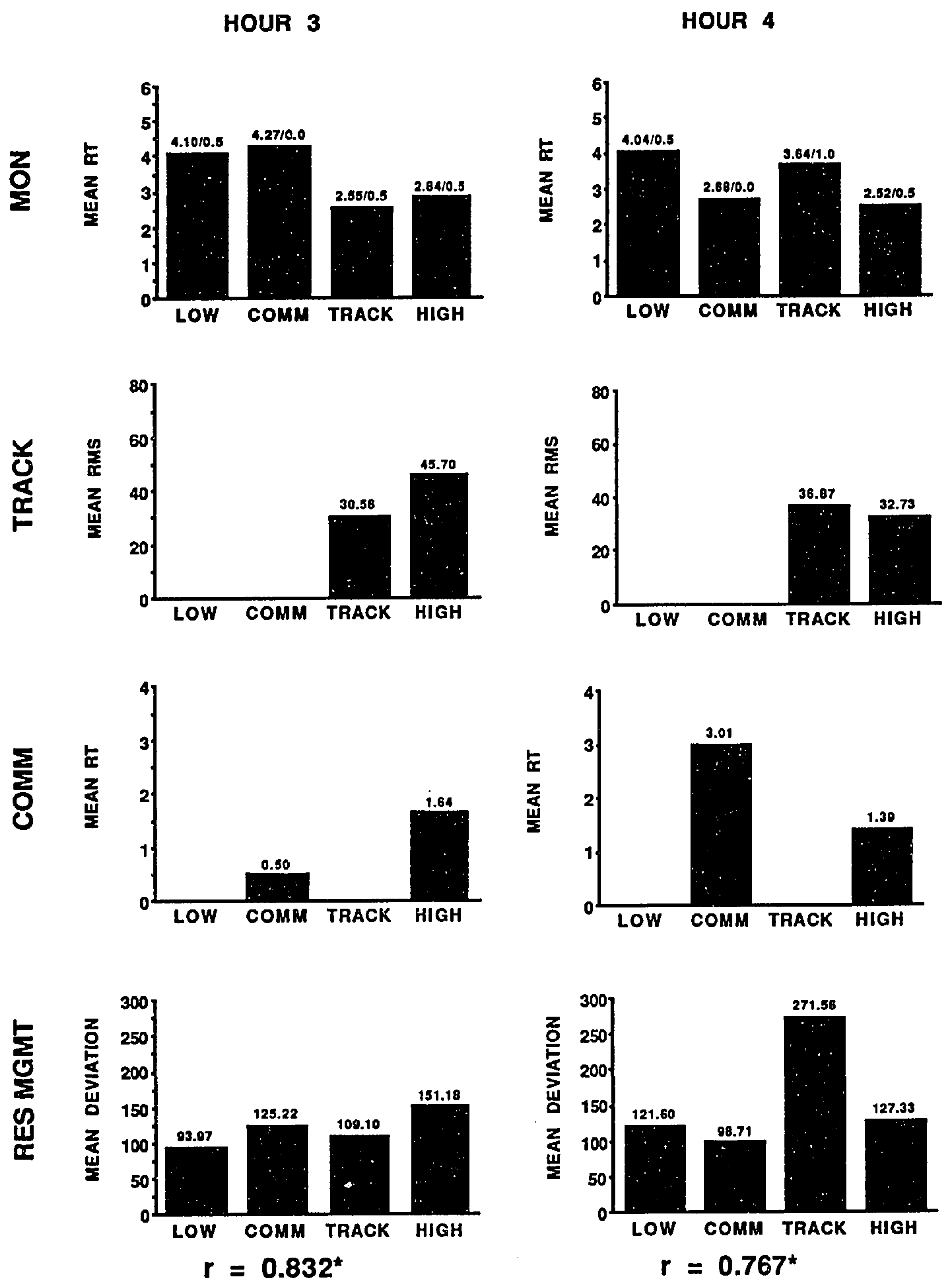
SUBJECT 20



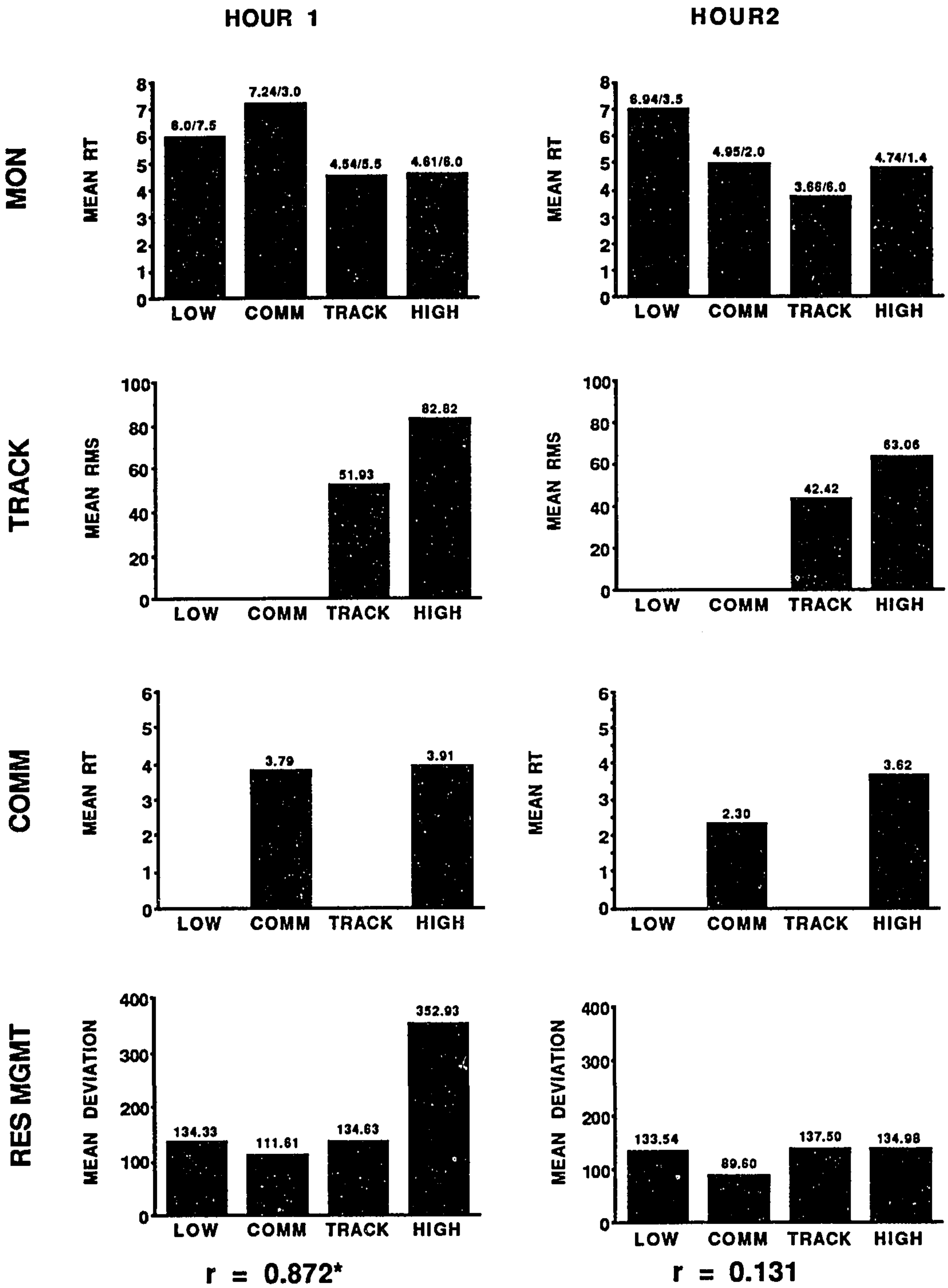
r = 0.895*

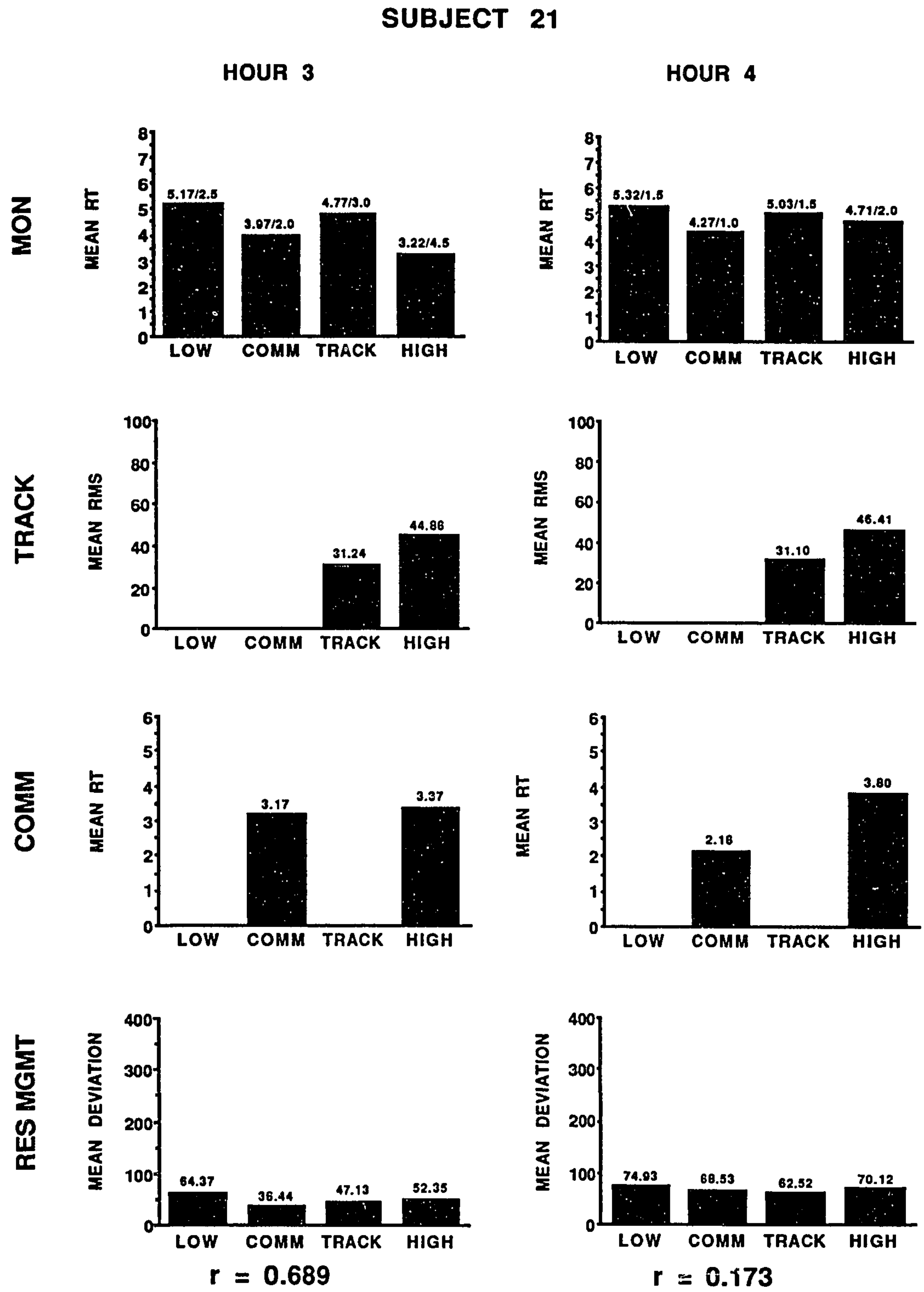
r = 0.131

SUBJECT 20

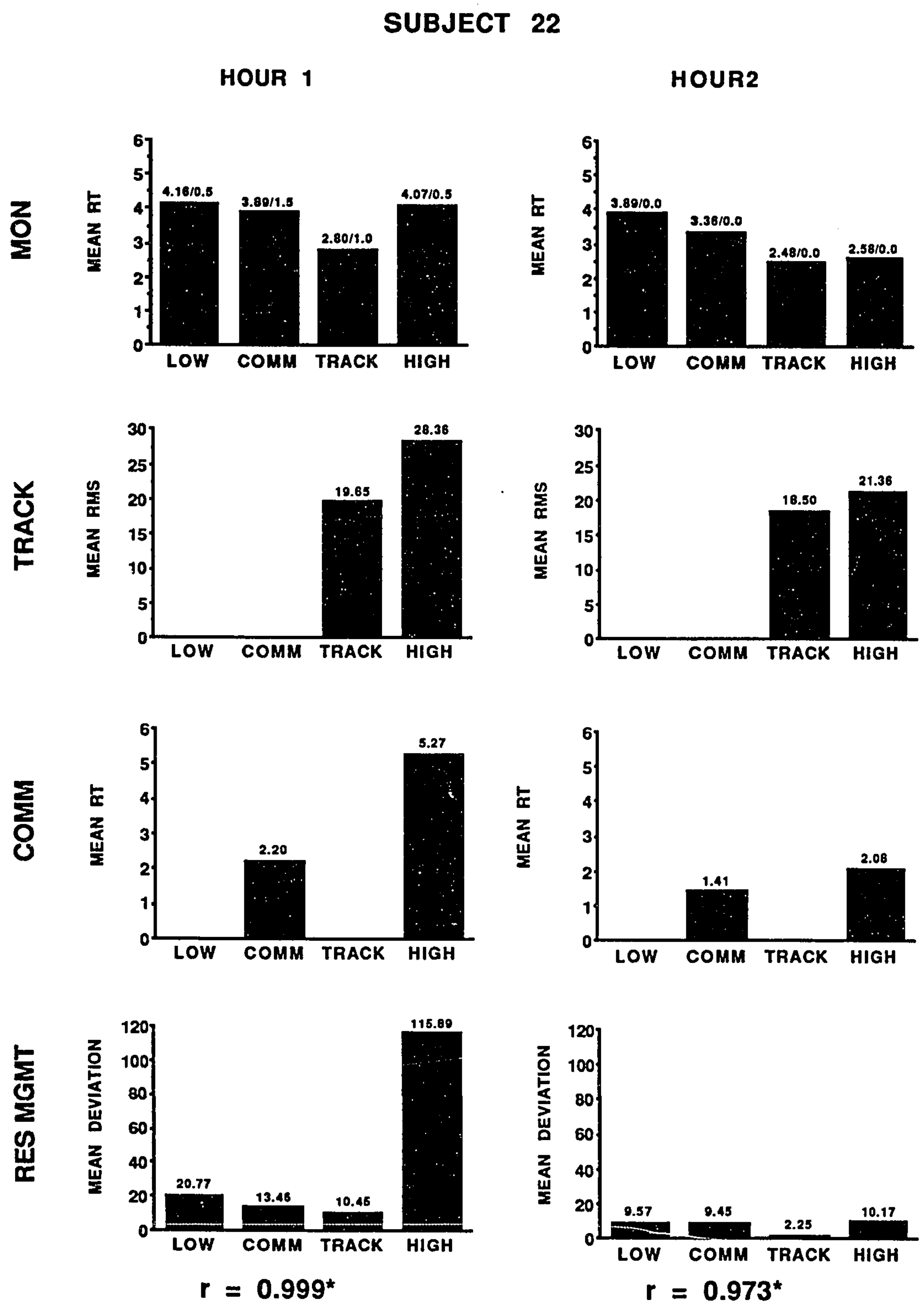

HOUR 3
HOUR 4
MON
MEAN RT
4.10/0.5
4.27/0.0
2.55/0.5
2.84/0.5
4.04/0.5
2.68/0.0
3.64/1.0
2.52/0.5
TRACK
MEAN RMS
30.56
45.70
36.87
32.73
COMM
MEAN RT
0.50
1.64
3.01
1.39
RES MGMT
MEAN DEVIATION
93.97
125.22
109.10
151.18
121.60
98.71
271.56
127.33
LOW
COMM
TRACK
HIGH
r = 0.832*
r = 0.767*

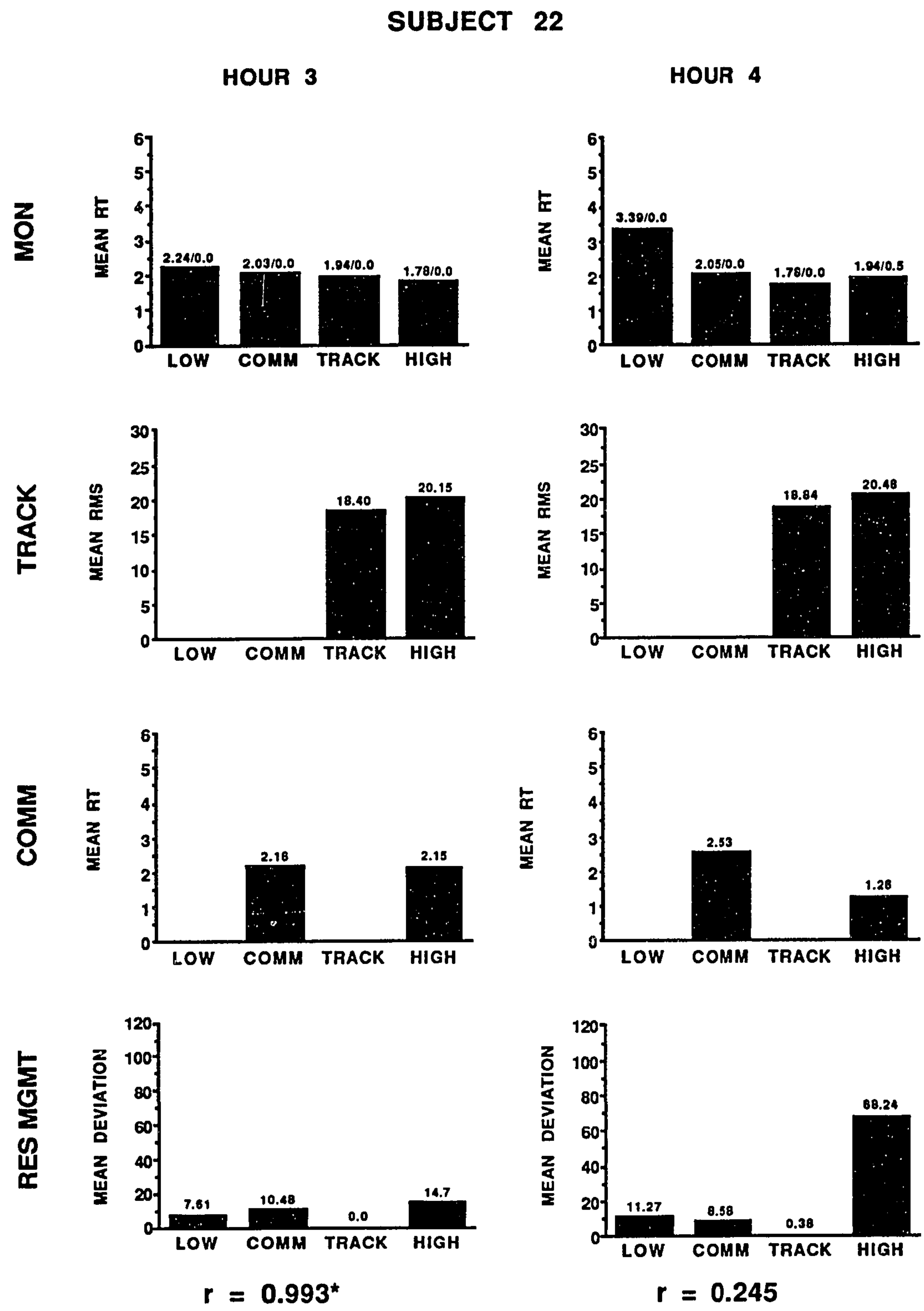
SUBJECT 21

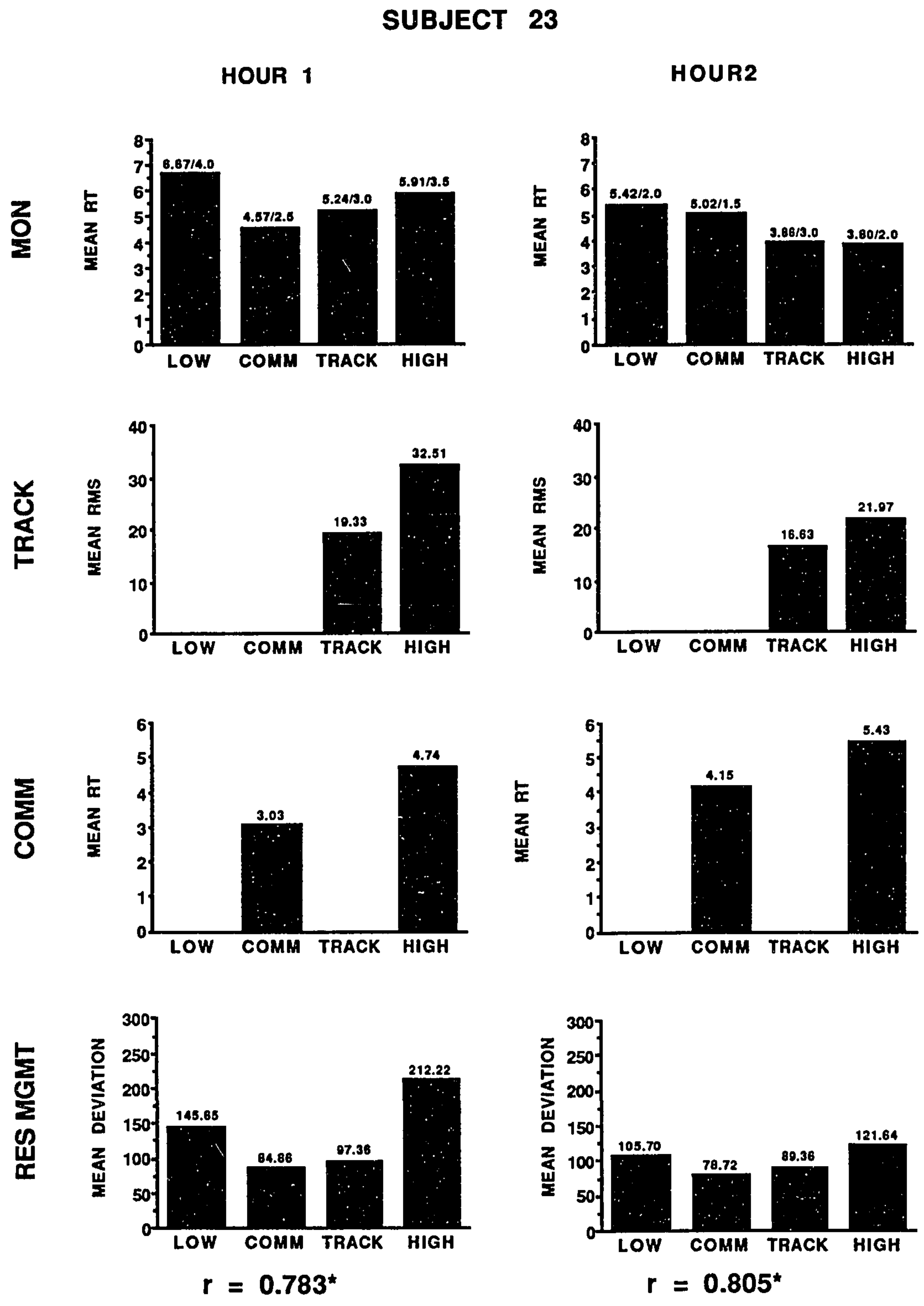

HOUR 1
HOUR2
MON
TRACK
COMM
RES MGMT
MEAN RT
MEAN RMS
MEAN DEVIATION
LOW
COMM
TRACK
HIGH
6.0/7.5
7.24/3.0
4.54/5.5
4.61/6.0
6.94/3.5
4.95/2.0
3.66/6.0
4.74/1.4
51.93
82.82
42.42
63.06
3.79
3.91
2.30
3.62
134.33
111.61
134.63
352.93
133.54
89.60
137.50
134.98
r = 0.872*
r = 0.131



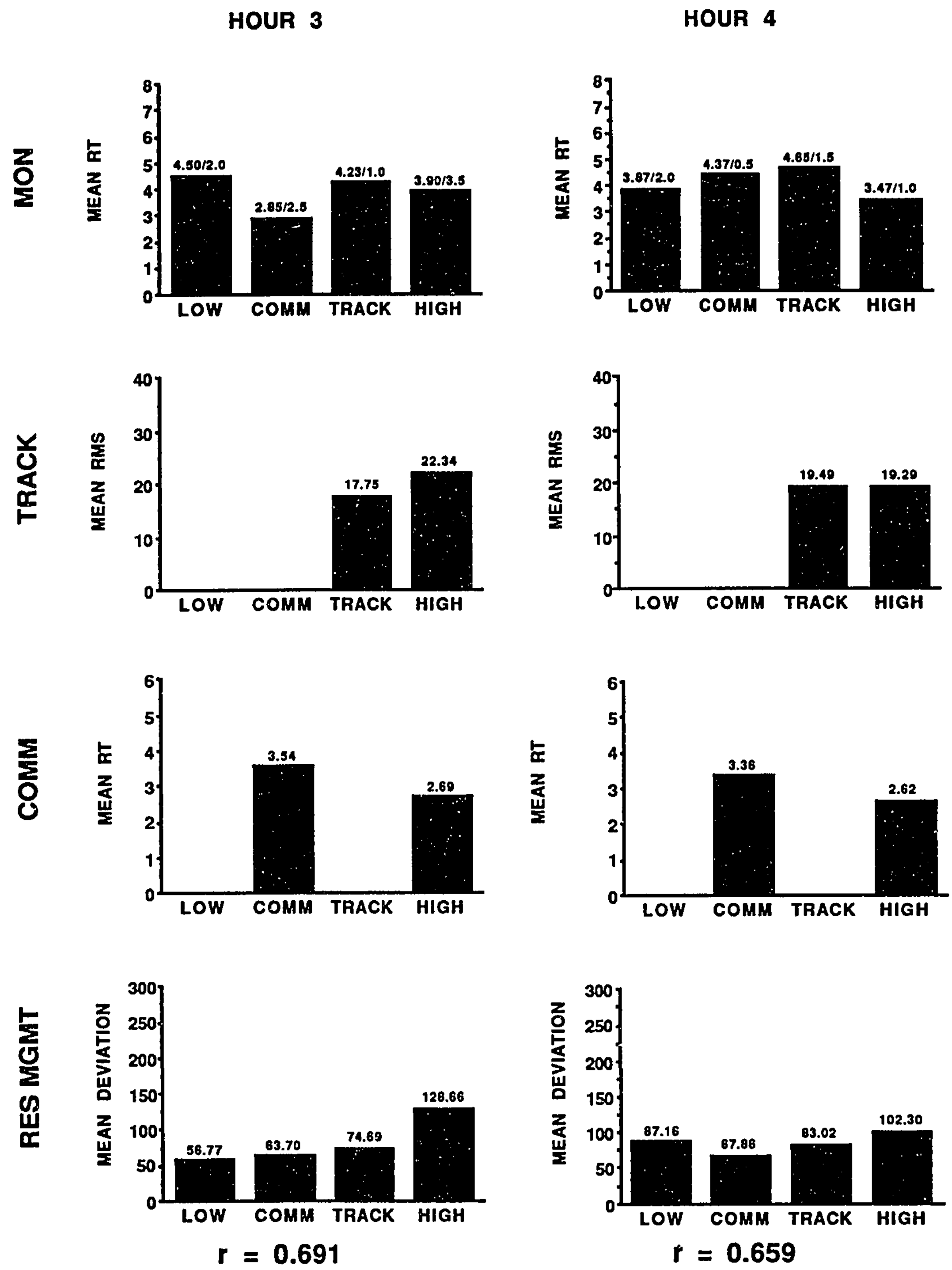
SUBJECT 21
HOUR 3
HOUR 4
MON
MEAN RT
5.17/2.5
3.97/2.0
4.77/3.0
3.22/4.5
LOW
COMM
TRACK
HIGH
5.32/1.5
4.27/1.0
5.03/1.5
4.71/2.0
TRACK
MEAN RMS
31.24
44.86
31.10
46.41
COMM
MEAN RT
3.17
3.37
2.18
3.80
RES MGMT
MEAN DEVIATION
64.37
36.44
47.13
52.35
74.93
68.53
62.52
70.12
r = 0.689
r = 0.173


SUBJECT 22
HOUR 1
HOUR2
MON
TRACK
COMM
RES MGMT
MEAN RT
MEAN RMS
MEAN DEVIATION
4.16/0.5
3.89/1.5
2.80/1.0
4.07/0.5
3.89/0.0
3.36/0.0
2.48/0.0
2.58/0.0
19.65
28.36
18.50
21.36
2.20
5.27
1.41
2.06
20.77
13.46
10.45
115.89
9.57
9.45
2.25
10.17
LOW
COMM
TRACK
HIGH
r = 0.999*
r = 0.973*

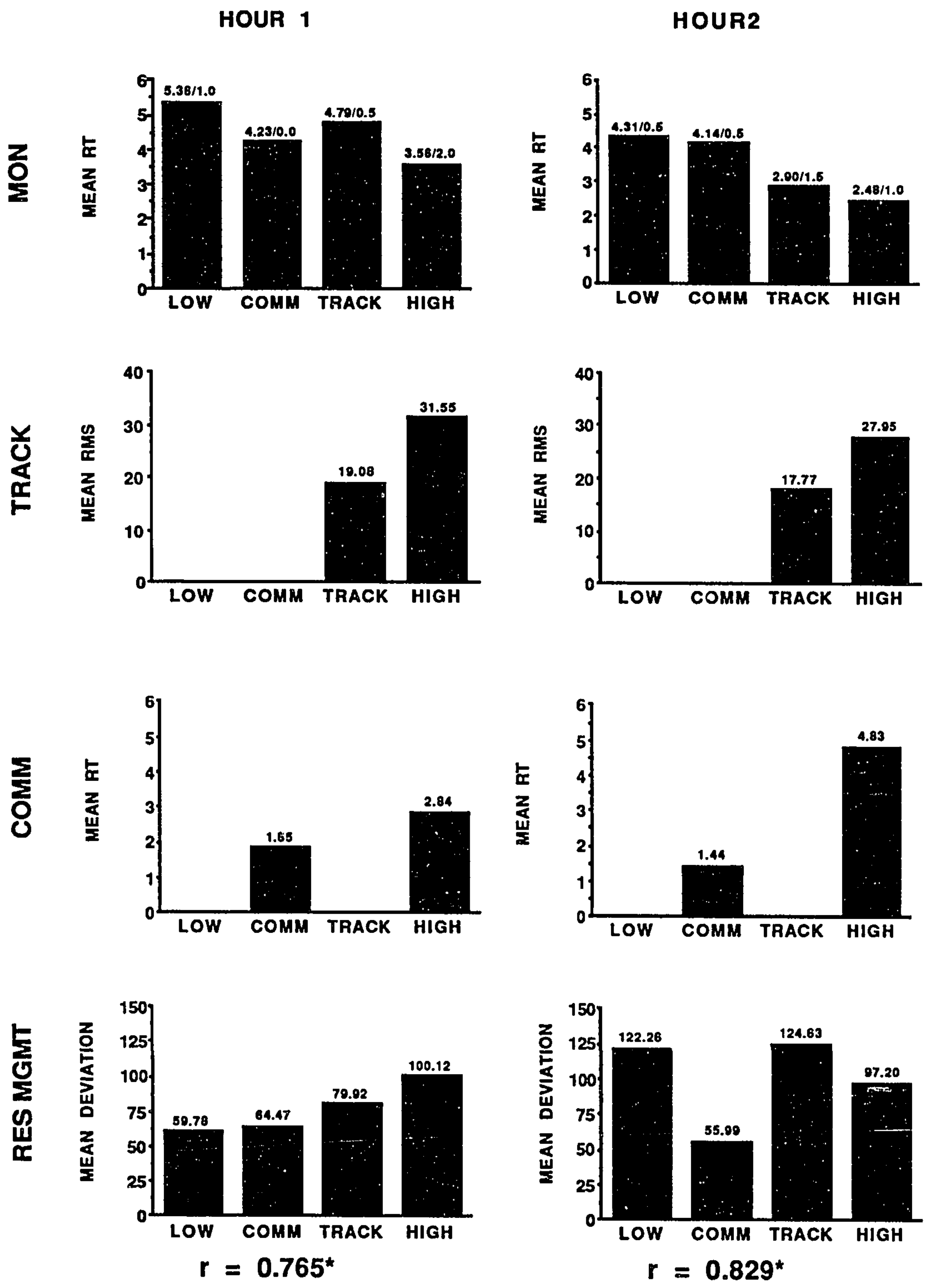

SUBJECT 22
HOUR 3
HOUR 4
MON
TRACK
COMM
RES MGMT
MEAN RT
MEAN RMS
MEAN DEVIATION
LOW
COMM
TRACK
HIGH
2.24/0.0
2.03/0.0
1.94/0.0
1.78/0.0
3.39/0.0
2.05/0.0
1.78/0.0
1.94/0.5
18.40
20.15
18.84
20.48
2.16
2.15
2.53
1.28
7.61
10.48
0.0
14.7
11.27
8.58
0.38
68.24
r = 0.993*
r = 0.245


SUBJECT 23
HOUR 1
HOUR2
MON
MEAN RT
6.67/4.0
4.57/2.5
5.24/3.0
5.91/3.5
5.42/2.0
5.02/1.5
3.86/3.0
3.80/2.0
LOW
COMM
TRACK
HIGH
TRACK
MEAN RMS
19.33
32.51
16.63
21.97
COMM
3.03
4.74
4.15
5.43
RES MGMT
MEAN DEVIATION
145.65
84.86
97.36
212.22
105.70
78.72
89.36
121.64
r = 0.783*
r = 0.805*

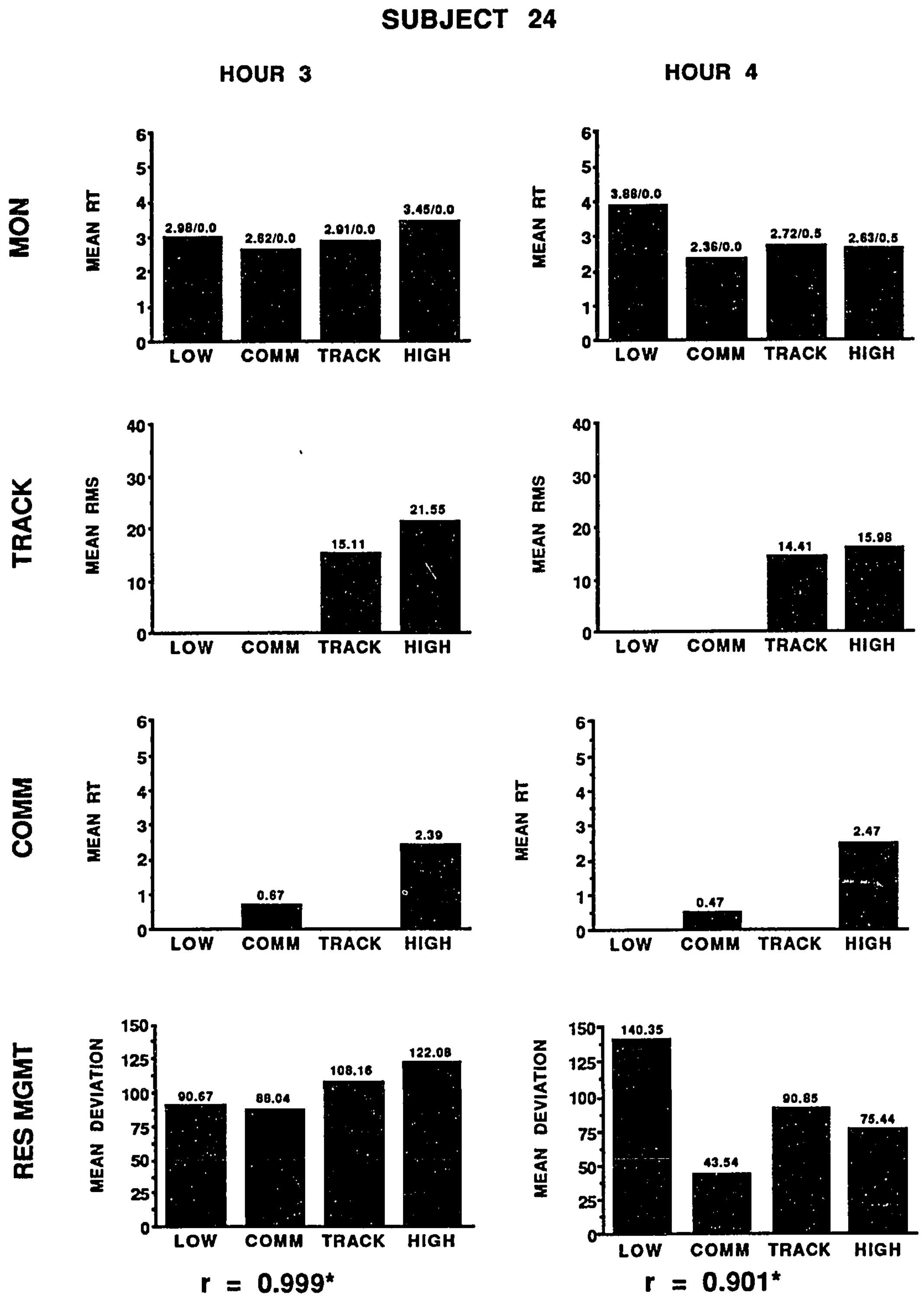
SUBJECT 23


HOUR 3
HOUR 4
MON
TRACK
COMM
RES MGMT
MEAN RT
MEAN RMS
MEAN DEVIATION
LOW
COMM
TRACK
HIGH
4.50/2.0
2.85/2.5
4.23/1.0
3.90/3.5
3.87/2.0
4.37/0.5
4.65/1.5
3.47/1.0
17.75
22.34
19.49
19.29
3.54
2.69
3.36
2.62
56.77
63.70
74.69
128.66
87.16
67.86
83.02
102.30
r = 0.691
r = 0.659

SUBJECT 24



HOUR 1
HOUR2
MON
MEAN RT
5.38/1.0
4.23/0.0
4.79/0.5
3.56/2.0
4.31/0.5
4.14/0.5
2.90/1.5
2.48/1.0
TRACK
MEAN RMS
19.08
31.55
17.77
27.95
COMM
MEAN RT
1.65
2.84
1.44
4.83
RES MGMT
MEAN DEVIATION
59.78
64.47
79.92
100.12
122.26
55.99
124.63
97.20
LOW
COMM
TRACK
HIGH
r = 0.765*
r = 0.829*


SUBJECT 24
HOUR 3
HOUR 4
MON
TRACK
COMM
RES MGMT
MEAN RT
MEAN RMS
MEAN DEVIATION
LOW
COMM
TRACK
HIGH
2.98/0.0
2.62/0.0
2.91/0.0
3.45/0.0
3.88/0.0
2.36/0.0
2.72/0.5
2.63/0.5
15.11
21.55
14.41
15.98
0.67
2.39
0.47
2.47
90.67
88.04
108.16
122.08
140.35
43.54
90.85
75.44
r = 0.999*
r = 0.901*