1992-09

DDG 51 operational evaluation: measures of workload from Combat Information Center communication patterns

Conlon, Thomas M.

Monterey, California. Naval Postgraduate School

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Naval Postgraduate School 6b. OFFICE SYMBOL OR	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5000	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Including Security Classification)
DDG 51 Operational Evaluation: Measures of Workload from Combat Information Center Communication Patterns

12 PERSONAL AUTHOR(S)
CONLON, Thomas Mark

13 TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. Page Count
Master's thesis	FROM TO	1992, SEPTEMBER	67

16. SUPPLEMENTAL NOTATION
The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	DDG-51, Arleigh Burke, Stress, Communication, Workload, OPEVAL

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

This thesis analyzes 2,700 verbal transmissions collected from an audio tap on DDG 51's internal communication network during the ship's OPEVAL. The frequency and duration of these voice transmissions are analyzed to explore for systematic changes. These changes are induced by eight simulated combat scenarios. The data shows that CIC team member communication patterns varied as a function of workload. The use of verbal communication patterns as unobtrusive, noninvasive measures of workload in operational settings is discussed and recommendations are made to further develop these measures.

20 DISTRIBUTION/AVAILABILTIY OF ABSTRACT [X] UNCLASSIFIED/UNLIMITED [ ] SAME AS RPT. [ ] DTIC	1a. REPORT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Frank Petho	22b. TELEPHONE (Include Area Code) (408)646-2157 22c. OFFICE SYMBOL OR/Pe

DD Form 1473, JUN 86 — Previous editions are obselete. — S/N 0102-LF-014-6603

SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

Approved for public release; distribution is unlimited.

DDG 51 OPERATIONAL EVALUATION: MEASURES OF WORKLOAD FROM COMBAT INFORMATION CENTER COMMUNICATION PATTERNS

by

Thomas M. Conlon
Lieutenant, United States Navy
B.S., United States Naval Academy, 1986

Submitted in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 1992

ABSTRACT

This thesis analyzes 2,700 verbal transmissions collected from an audio tap on DDG 51's CIC internal communication network during the ship's OPEVAL. The frequency and duration of these voice transmissions are analyzed to explore for systematic changes. These changes are associated with different workload levels and the levels of stress induced by eight simulated combat scenarios. The data shows that CIC team member communication patterns varied as a function of workload. The use of verbal communication patterns as unobtrusive, noninvasive measures of workload in operational settings is discussed and recommendations are made to further develop these measures.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

Communication is the act of sharing information. This thesis is about human communication, but human communication in a very narrow, restricted, and unique sense. It is about how members of a Combat Information Center team aboard the lead ship of a new class of United States Navy guided-missile destroyers, the ARLEIGH BURKE (DDG 51), shared information during their ship's Congressionally mandated operational evaluation. It is about information shared by crewmen as they attempted to thwart raids by very realistic and challenging threats to demonstrate their ship's ability to fight. It is about communication - how information is shared - in a simulated combat environment; an environment characterized by high costs, high stakes, high drama, and high workload. From a methodological perspective, this thesis is about using natural human communication patterns as unobtrusive, noninvasive indices of workload. This introduction will consider five broad themes.

- **First**. Modern surface combat systems have become increasingly lethal and at the same time, increasingly complex. Complexity usually increases operator workload, increases which typically require distributing that workload across many skilled operators. It is those operators who ultimately must communicate with each other to accomplish a mission.

- **Second**. The AEGIS combat system exemplifies such a system. It is highly complex, very lethal, and manned by a diversified crew whose training, background, and rank all differ. However, the crew needs to share information - communicate - during critical, and in some cases life-threatening, high workload evolutions.

- **Third**. There may be substantive changes in this shared behavior - communication - among team members during high-workload evolutions. These changes may be systematic alterations in the frequency and duration of verbal transmissions. If these changes, or communication patterns, can be quantified and set within a proper theoretical context, they could be used as unobtrusive, noninvasive measures of workload and stress.

- **Fourth**. The Operational Test and Evaluation (OPEVAL) of USS ARLEIGH BURKE (DDG 51) presented an opportunity to examine human communication patterns under different levels of workload.

- **Fifth**. If high workload systematically alters a team's communication patterns, then those patterns should be accounted for by current models and findings from the study of human information processing.

A. INCREASING SYSTEM COMPLEXITY

> The effort to stay in control of technology becomes more difficult all the time. No one, no thing, has ever been perfect, but the price of error is higher than ever before. For hundreds of millennia in the prehistoric past, individuals defended their own land and built their own shelters. Settlements were far apart; accidents affected relatively few people. Today, living close together in complex social networks, we may become victims of other people's mistakes - on the positive side, we owe our survival to reliable strangers. We depend more and more on a rare breed of specialists trained to hold the line against chaos. They make their share of mistakes. But they strive to develop ways of catching error early and preventing it from blossoming into catastrophe. (Pfeiffer, 1989, p. 39)

Simply stated, system complexity has narrowed the margin for error, has made the need for good design more crucial, and has made the consequences of error potentially more catastrophic.

Military research and development has undoubtedly advanced American commercial and industrial technologies (Binkin, 1986, p. 3). In 1986 alone, for example, the Department of Defense expended fifteen times the research and development (R&D)

funds than did France, Germany, or Britain and 80 times that of Japan (Packard Commission, 1989, p. 3). These R&D expenditures have produced increasingly complex, albeit lethal, military systems, which in turn, have increased the need to consider human factors during system design. In fact, the interface between the operator and the machine in modern combat systems can be a critical and limiting factor in system performance.

Increased complexity not only substantiates the need for proper weapon system design, it has driven the need to consider other aspects of human factors; that is, the accompanying manpower, personnel, and training systems needed to accommodate more complicated and sophisticated human performance requirements. A quick historical survey of the number of specialized skills required to fight a ship reflects this trend. In 1805, Nelson's Fleet at Trafalgar, had only four ratings: able-bodied seaman, less than able-bodied seaman, carpenter, and marine. In 1916, Britain's WWI Navy at Jutland had twelve ratings (Keegan, 1987, pp. 65-66). Today, the U.S. Navy has 112 ratings and 1409 Naval Enlisted Classifications or subspecialties. The evolutionary increases in combat system complexity are clearly reflected in the distribution of labor needed to operate and maintain them.

B. AEGIS: AN EXAMPLE OF INCREASED COMPLEXITY

The most recent example of a complex system in today's Navy is the AEGIS weapon system aboard the billion dollar TICONDEROGA class missile cruisers and the new ARLEIGH BURKE class missile destroyers. The AEGIS concept is, from a technical perspective, a quantum leap from older systems. The AEGIS weapon system affords its ships the world's most capable anti-air warfare (AAW) capability. The system was designed and developed to provide carrier battlegroups defense against aircraft and anti-ship missiles (Polmar, 1987, p. 113).

1. The AEGIS System

The AEGIS weapon system is a sophisticated computer-aided data processing, analysis, and display system, designed to handle coordinated Soviet air and missile saturation attacks. Its centerpiece is the AN/SPY-1A phased arrayed radar; a radar that can simultaneously search and track hundreds of air and surface targets. AEGIS is to the Navy as the Airborne Early Warning Aircraft (AWACs) is to the Air Force: an all seeing eye. There is, however, a significant difference between the two systems. While the sole function of AWACs is to acquire and transmit data to ground control stations, AEGIS is an independent **weapons** platform as well. It not only detects and classifies hostile targets, it can destroy them. (Allard, 1990, p. 163)

AEGIS equipped cruisers and destroyers protect carrier battle groups (CVBGs) by detecting, classifying, and tracking hundreds of targets simultaneously; in the air, on the surface and under the sea. They also bring additional **offensive** power to the CVBG in missiles and guns. Vessels equipped with the AEGIS system destroy attackers by using a variety of weapons including ship and air-launched torpedoes, anti-submarine rockets, deck guns, surface-to-surface and surface-to-air missiles, and rapid fire PHALANX close-in weapon systems, all aided by electronic jammers and decoys. The variety of missions includes anti-air, anti-submarine, anti-ship, and strike warfare, including bombardment of shore positions. (CG 47 Class Services, Naval Sea Systems Command, 1987, p. 1-13)

In 1985, Vice Admiral H.C. Mustin, then commander of the Second Fleet, summarized the importance of the AEGIS weapon system. He stated:

> AEGIS has brought clarity to the air battle. . . . the importance of our new ability to put the surface-to-air-missile ships in the outer defense zone, where they can shoot approaching bombers *before* they reach missile launch range, cannot be overstated. ... with AEGIS, we can win the air battle against all comers. Without AEGIS, we cannot win. (Allard, 1990, p. 163)

2. The VINCENNES Incident

The AEGIS weapon system achieved notoriety during the 1988 Iran/Iraq war, when the USS VINCENNES shot down an Iranian Airbus A300, Iran Air Flight 655, with 290 passengers aboard. It took seven minutes to shoot down Flight 655, but subsequent investigations by the Navy, Congress, and international organizations lasted six months (Hill, 1989, p. 108). Investigations continue. Recent accusations by *Newsweek* and *Nightline*, for example, have renewed interest in the VINCENNES incident and Congressional hearings are being held to determine the validity of the original reports (Newsweek, July, 13 1992, pp. 28-39).

Independent psychologists who reviewed the VINCENNES incident testified before the House Armed Services Committee in October 1988. They testified that the stress of combat, heightened workload due to information overload, and a communications breakdown in the ship's Combat Information Center (CIC) contributed to the tragedy

(Squires, 1988, p. A3). The Navy's investigative team, headed by Rear Admiral William Fogarty, drew similar conclusions:

> Since it appears that combat induced stress on personnel may have played a significant role in this incident, it is recommended the CNO direct further study be undertaken into the stress factors impacting on personnel in modern warships with highly sophisticated command, control, communications, and intelligence systems, such as AEGIS. This study should also address the possibility of establishing a psychological profile for personnel who must function in this environment. It is also suggested that the CNO consider instituting a program for Command, Control, Communication, and Intelligence (C3I) stress management to test and evaluate the impact of human stress on C3I operations in complex warships such as the AEGIS cruiser. Integral to this program would be the incorporation of measures of human effectiveness into battle simulation techniques to assess the effect of peak overloads and stress on human players. (CNO Memorandum Ser 11B1/14-89, 1989, p. 3)

The system hardware was vindicated by the Navy investigation as working exactly as designed and the investigation concluded that human error had caused the tragic loss of life. Questions then surfaced as to the cause of the human error; something not so easily explained or understood.

C. WORKLOAD AND STRESS

1. Stress

For the purpose of this paper, stress is defined as:

> A loading, a burden, a pressure on the individual, which may come from physical or psychological sources. For practical purposes, a stressor can be considered any condition that taxes a person's resources or threatens his well-being (McGrath, 1989, pp. 1-2).

The dangers of combat are well known, but more subtle impacts induced by high ambient noise, crowding, heat, fatigue, lack of sleep, high workload, anxiety, and competition are more obscure. Figure 1 shows a conceptual breakdown of general stress as induced

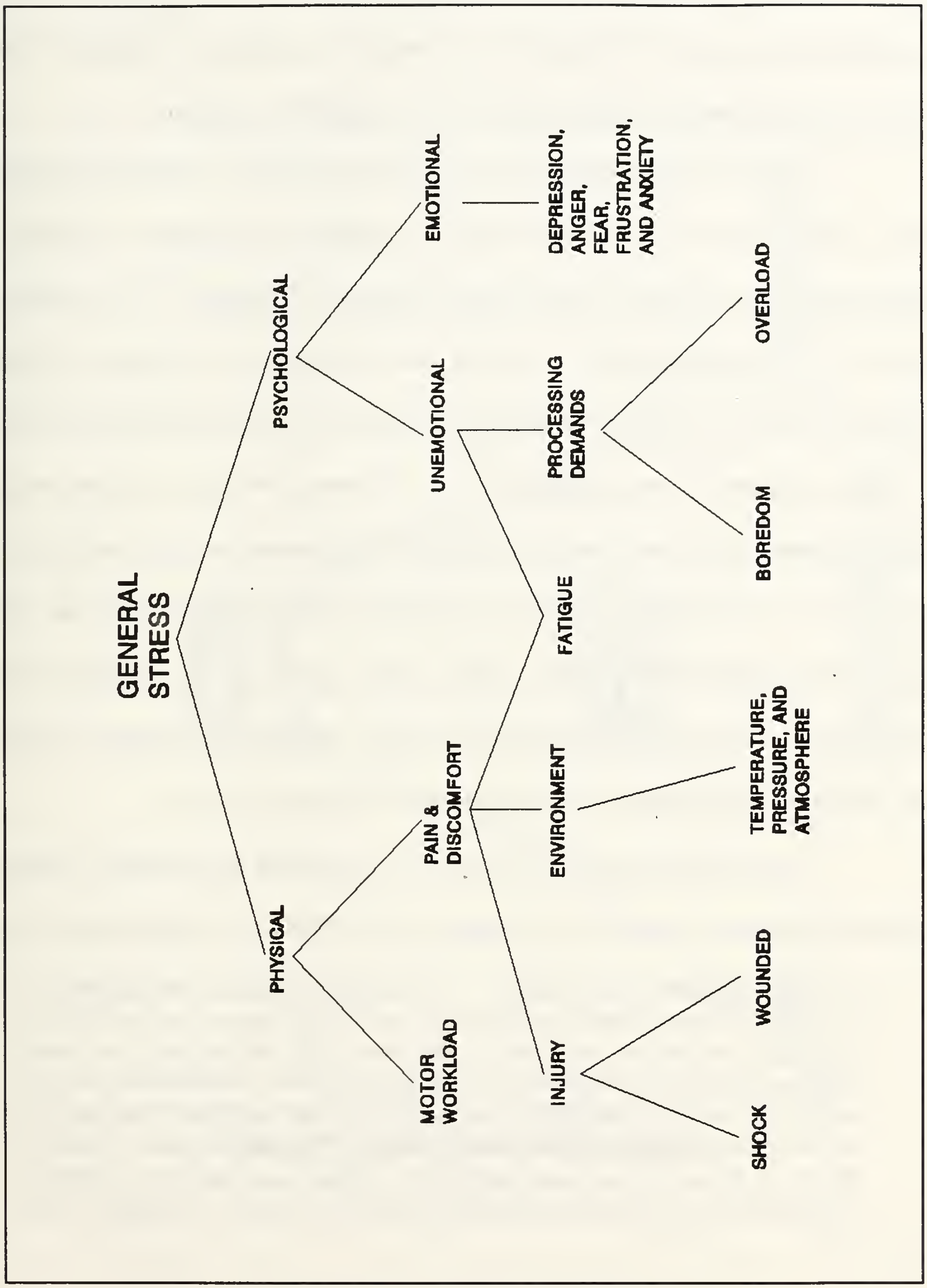


Figure 1. Stressors

by a broad range of stressors. (Poock and Martin, 1984, p. 1-3). It is clear that stress, as defined above, can be induced by a variety of conditions, including increased information processing wrought by high levels of cognitive workload.

The relationship between levels of stress and human performance measures, such as accuracy of response and speed of response, is not linear. Instead, the relationship takes the form of the inverted-U depicted in Figure 2. This function is called the Yerkes-Dodson Law. It holds that performance is not always adversely impacted by stress. In fact, as reflected by the shape of the curve, optimal performance is actually achieved in the presence of stress. However, both too much and too little stress adversely impact performance, especially at the extreme regions, the tails, of the inverted-U. Performance is assumed to be affected by the extent to which the stressor activates the central nervous system. High central nervous system activation induces high arousal. Low activation induces low arousal. Different performance decrements are associated with different levels of arousal. (McGrath, 1989, p. 7)

The Yerkes-Dodson Law bears on two important considerations concerning human performance, especially as it relates to task difficulty and high arousal stress.

> First, the optimum level of stress . . . is inversely related to the difficulty of the task. In other words, the optimum level shifts downward for difficult tasks and upward for easy tasks. The more difficult the task, the more sensitive it will be to the effects of high arousal stress. Performance in CIC can get the worst of this effect, because under the conditions that produce high arousal, the tasks become more difficult. ...The second point, is that the effects of stress are not necessarily bad. Stress can, and does, improve performance on tasks where the arousal levels are too low. (McGrath, 1989, p. 7)

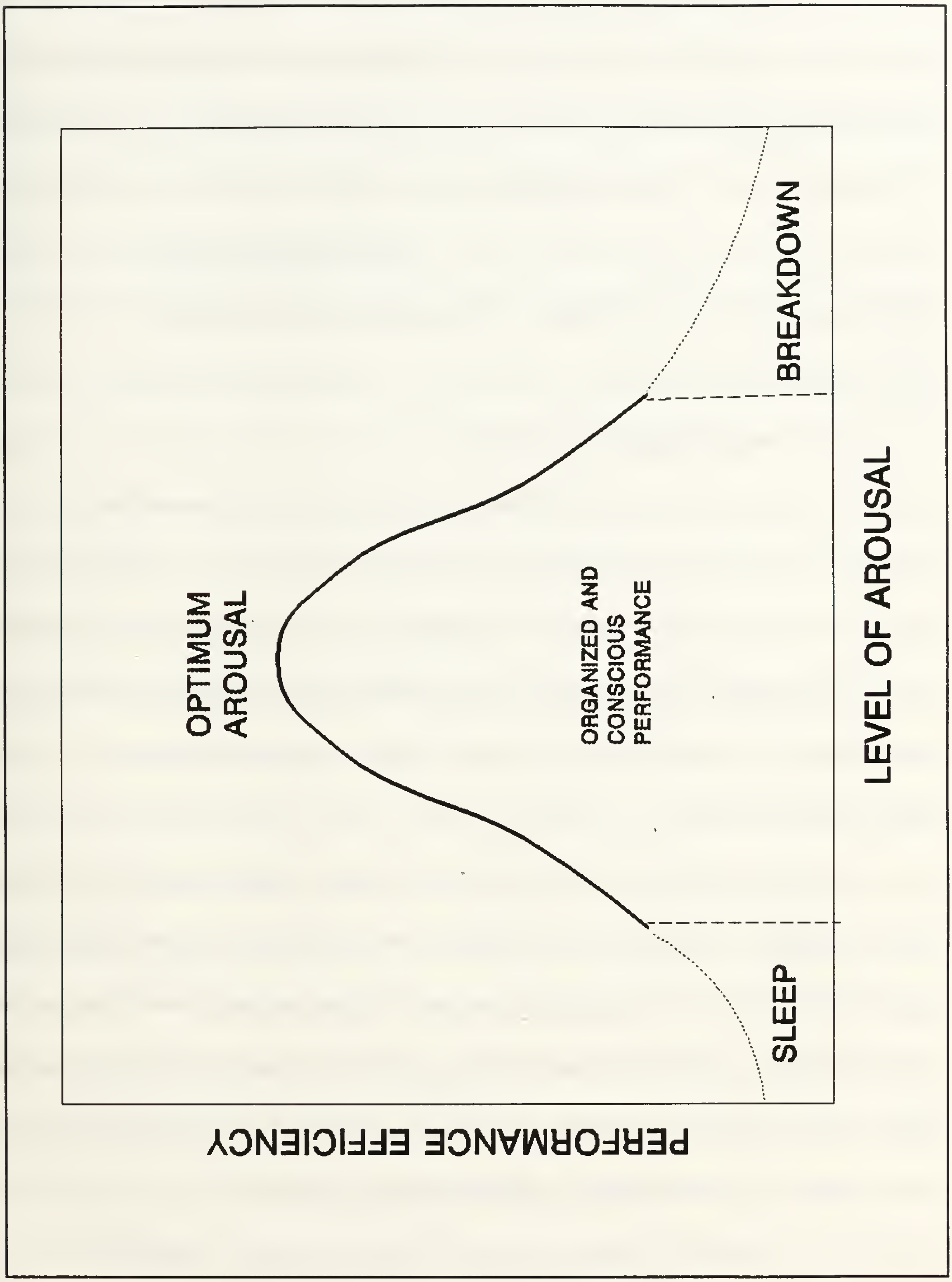


Figure 2. Yerkes-Dodson Law

Naval operations typically occur at the low end of the arousal spectrum, with occasional periods of extremely high arousal levels induced by abrupt changes in the environment. This condition is accurately captured by a popular anonymous aphorism; "Standing a watch in CIC is hours of boredom punctuated by moments of shear terror." Because this thesis focused on CIC team members performing tasks in a simulated combat environment, the remainder of this section discusses the causes and effects of high arousal stress.

2. Causes of High Arousal Stress

Figure 1 showed a list of stressors which span the spectrum from low to high arousal stress. The most typical causes of high arousal stress are high anxiety, high workload, and adverse environmental conditions. Each of these stressors can have its own unique effects on performance, but as a group, they all increase the arousal level; that is, they induce relatively high levels of central nervous system activation. (McGrath, 1989, pp. 7-11)

High arousal stressors typically manifest themselves in three ways. First, they induce a feeling of frustration or a distinct sense of arousal. Second, they stimulate physiological changes; for example, an increase in heart rate, heightened blood pressure, faster respiration, and higher core body temperature. Third, and a particularly important effect from this paper's perspective, high arousal stressors affect the efficiency with which people process information. (Wickens, 1992, p. 412)

3. Effects of High Arousal Stress

The most significant cognitive effects associated with high arousal stress are (a) attentional narrowing, (b) short term memory loss, (c) activation, and (d) communication degradation. These effects are discussed below.

a. Attentional Narrowing

Attentional narrowing refers to a sharp constriction of a person's field or range of attention under conditions of high central nervous system activation; that is, during states of high arousal. High arousal stress increases alertness and musters attentional resources, but at the same time, attention becomes very narrowly focused (McGrath, 1989, p. 12). Attention tends to be focused centrally at the expense of "paying attention" to events at the periphery of the problem space. This narrowing is analogous to viewing the contents of a room through the keyhole of its door.

Attentional narrowing has important implications for a person who must perform more than one task at a time. There is evidence which indicates that if a person has to simultaneously perform more than one task, for example, tracking a cursor on a display while communicating, then performance on the secondary or subsidiary task (in this example, communicating) may deteriorate in high workload situations. Deterioration can be expected in manual dexterity, sensory-motor tasks, and performance of the secondary tasks in general. (Hockey, 1983, p. 137) This "tunnel vision," which essentially reflects a highly focused attentional field, is not manifest at the sensory level. Its impact is central. It affects the central cognitive processes. (McGrath, 1989, pp. 13-14)

b. Short Term Memory Loss

The human memory system is typically conceptualized as having three stages. The first stage temporarily stores information at the sensory level; for example, information can be stored at the visual or auditory level. Sensory storage lasts for about a quarter of a second and requires no effort to retain it. Short term memory, also called "working memory," is the second memory stage and occurs between information stored at the sensory level and information stored in long term memory, the third stage in the process. Information cannot pass from short term memory into long term memory without applying considerable effort; that is, a person must "pay attention" to the information in short term memory or "rehearse" it, if it is to be retained in long term memory. Without rehearsing the information in short term memory, it fades and is quickly lost, usually in under twenty seconds (Wickens, 1992, p. 220). Short term memory is adversely affected by high arousal stress. This impact probably stems from environmental conditions, such as the rate with which information is flowing or its shear volume. Information overload can effectively block a person's ability to rehearse information temporarily held in the short term memory register. In the present case, high workload situations in CIC during combat situations prevents rehearsal. Unless written down, such as with a grease pencil on a display screen, discrete point estimations of tactical data quickly fade from memory.

During high arousal situations, information is usually not committed to long term memory. Most of it fades from short term memory in roughly twenty seconds. Given that, operators must frequently refresh their short term memory stores.

Two examples in which instrumentation is used to compensate for these memory deficits are (a) instrumented ranges that record performance during high workload air combat maneuvering exercises and (b) inport training evolutions that systematically record all team performance.

c. Activation

Activation is the tendency of high arousal stress to rapidly instigate action with little or no consideration for the consequences of the action. Under conditions of high arousal, operators have the desire to "do something" quickly, even though it may not be prudent. Responsiveness or reaction times will be quicker, but more mistakes will be made. (Wickens, 1992, p. 419)

d. Communication Degradation

Individuals usually become less communicative and are less willing to pass detailed information in high arousal conditions. Misunderstandings between team members tend to occur more frequently due to attentional narrowing and the demands placed on short term memory stores. (McGrath, 1989, p. 14) Studies have also shown that there are quantitative changes in verbal communications patterns produced under stress compared to normal, non-stressed communication (Hicks, 1979, pp. 124-125). Communication patterns refer to changes in the duration and frequency of verbal transmissions, not their content. These changes, which will be discussed in greater detail later, occur if the level of arousal exceeds a certain threshold.

4. Measurement of Stress: Issues

Three major issues often arise in non-laboratory attempts to measure the impact of stress on human operators in the Navy. They are (a) the acceptability of certain measurement techniques to the operational chain of command, (b) the generalizability of findings produced by laboratory based experimental procedures, and (c) the obtrusiveness of the data collection itself.

a. Acceptability to the Chain of Command

Evaluating the impact of operator stress on mission performance is a controversial issue within the operational chain of command. Many of today's military leaders recognize the potentially catastrophic consequences of stress. Some, however, reasonably believe that combat stress cannot be simulated without the threat of mortal danger and yet remain within acceptable safety and budgetary limits. This position is exemplified in the following Congressional testimony by the Director of the Department of Defense Office of Operational Test and Evaluation.

> Operational tests are run in the most realistic combat conditions possible, consistent with safety and available test resources. It is unlikely that an operational test can be devised that can put operators under stresses identical to combat and still meet the requirements of safety. It would be of little value and would probably be unsafe to run an operational test of a weapons system so as to cause operators to "distort data" or suffer from "task fixation." Tests when an operator is not acting rationally, will not provide pertinent information on which to judge the effectiveness of the system. (Congressional Record, H.A.S.C. No. 100-94, September 14, 1988, p. 157)

Despite limitations imposed by safety and budget constraints, testing today's billion dollar surface combatants requires a level of realism necessary to ensure the tests' findings are indeed valid. This realism is part of the test and evaluation procedure, and it is precisely this realism that induces high workloads on the crew. Given that workload has become an important aspect of the OPEVAL, it must somehow be measured and made part of the comprehensive system evaluation and its subsequent report.

b. *Generalizability of Findings*

The results from studies of stress which incorporate laboratory induced stress do not fully generalize to real world situations (Hicks, 1979, p. 110). When dealing with a CIC team at sea, evaluating the impact of increased workload places extraordinary methodological demands on the researcher and irregular scheduling demands on the operational unit. These demands are extraordinary because it is difficult, if not impossible, to control all the intervening variables in an operational environment. It is also difficult to completely control the schedules of the operators. Results from such experiments, typically produce findings that are either unreliable or cannot readily generalize beyond the immediate test session or test environment.

c. *Obtrusiveness of Data Collection*

Physiological methods used to measure stress are typically obtrusive and occasionally invasive; for example, rectal thermometers were recently used to measure core body temperature of sailors aboard ships in the Persian Gulf. The interruption, discomfort, or the simple presence of data collection equipment may bias the data by simply altering routine behavior. Alternatively, psychological techniques used to measure workload usually cannot be administered during actual operations; hence, data, especially survey data, is usually collected after the fact. Unobtrusive, noninvasive measures of workload would satisfy an operator's desire to remain unencumbered by extraneous and alien data collection requirements and a researchers desire to collect accurate and reliable data.

D. OPERATIONAL TEST AND EVALUATION OF DDG 51

During the October 1988 House Armed Services Committee inquiry into the VINCENNES incident, the Life Science Director of the Cognitive and Neural Sciences Division in the Office of Naval Research reported that the Navy spends about $30 million a year on human performance research. He also reported that his office had actively been investigating this area for forty years (Squires, 1988, p. A3). Despite the considerable investment in time and money, the Office of Naval Technology started an exploratory development program called Tactical Decision Making Under Stress (TADMUS) in 1989.

> The objective of the TADMUS program is to apply recent developments in decision theory, individual and team training, and information display to the problem of enhancing tactical decision quality under conditions of stress. This will be accomplished by a cooperative program in human factors and training technology. (Office of Naval Technology, 1991, p. 1)

During the same time period, the Navy was faced with criticism on its test and evaluation procedures of AEGIS. The criticism centered on reports that the quantity, realism, and difficulty of scenarios used to test AEGIS were inadequate. (Allard, 1990, p. 163)

The Chief of Naval Operations (CNO) replied to the criticism by citing that "more testing had been done on the AEGIS weapon system than on any other system to date" (Trost, 1988, p. A21). He further defended AEGIS by explaining the testing procedures used:

> In the Navy, the command primarily charged with the responsibility for testing our weapons systems is the Operational Test and Evaluation Force (OPTEVFOR). This command is headed by a two-star admiral who reports directly to me and the Secretary of the Navy. In fact, many of the systems it tests do not qualify for placement in the fleet. OPTEVFOR tries to defeat new systems by challenging them with known threats - and anticipated future threats - in order to ensure that fleet operators make the systems perform properly (Trost, 1988, p. A21).

In 1991, CNO tasked Commander OPTEVFOR (COMOPTEVFOR) to incorporate TADMUS testing into the operational evaluation (OPEVAL) of ARLEIGH BURKE (DDG 51) (Kren, 7 July 1992).

In November 1991, OPTEVFOR enlisted the assistance of Naval Command, Control, Ocean Surveillance Center's RDT&E Division (NRaD) to identify methods to assess the stress experienced by DDG 51 crew members and to do so with as little interference on operations as possible. The measures were to be unobtrusive and noninvasive (COMOPTEVFOR Memorandum of Agreement, 1991, p. 3). The following three methods were chosen by NRaD and agreed upon by COMOPTEVFOR (NRaD, 1992. p. 1).

- Subjective workload assessments from CIC watchstanders.
- Subjective assessments of performance pressure by experts observing video and audio tape recordings of the CIC team.

- Objective measurements of workload using the War Diaries; that is, data reconstructed from AEGIS's computers.

These three data collection methods were used during DDG 51 OPEVAL to provide a basis for NRaD to report on the levels of stress present during the simulated combat scenarios. The data to support the present study's focus on communication patterns, a dimension considered but not implemented by NRaD, was extracted from the audio tape recordings of the CIC team.

E. COMMUNICATION PATTERNS AS INDICES OF WORKLOAD

Communication patterns; again, changes in the frequency and duration of verbal transmissions, may be used as indices of workload pending empirical validation. Workload, stress, and ineffective communication have been implicated as causative factors in many accidents involving complex systems. Very few studies, however, have focused on communication patterns during periods of increased workload. The analysis presented in this thesis centers on exploring for distinct changes in communication patterns among CIC team members during various levels of workload imposed by realistic operational scenarios. This analysis will search for quantitative differences in frequency and duration of communications as a function of increasing workload.

The importance of these measures is that the data collection method is completely unobtrusive and noninvasive. The method described in this thesis requires little more than a line tap on the internal communications circuit, which when undisclosed to its users, eliminates performance biases from operators. If these methods are validated, they will provide the Navy with an unobtrusive, inexpensive, uncomplicated, and rapid means of evaluating the impact of workload on a CIC team. Moreover, temporal analysis of CIC team communication patterns can serve as a basis for the development of more realistic team trainers to study workload effects on team performance.

1. Predicted Findings

The following predictions are based on the findings associated with high arousal stress's impact on information processing.

- The average time of a verbal transmission by a CIC team member will decrease as levels of workload increase.
- The frequency of verbal transmissions will increase as levels of workload increase.
- The magnitude of the dependent variables identified above - average time and frequency of transmission - will covary with changes in the level of workload.

The following chapter describes the method by which the communication data were collected and analyzed to test the validity of these predictions.

II. METHOD

This analysis sought to compare and contrast speech patterns produced by a CIC team while performing their duties under different levels of workload. This section describes the method by which the analysis was conducted. There are four parts. The first part describes how the CIC team was exposed to different levels of workload induced by different simulated combat scenarios. The second part discusses the quantitative indices used to analyze human communication patterns. The third part describes the techniques used to collect voice data from CIC during simulated combat. The fourth and final part describes the statistical approach used to treat the data and test hypotheses.

A. RAIDS

The USS ARLEIGH BURKE (DDG 51) underwent OPEVAL during the January-February 1992 time frame. There were three levels of workload imposed by three different simulated combat scenarios or "raids" launched against DDG 51 during its OPEVAL. The three raid levels were named NO-RAID, MANNED-RAID, and MISSILE-RAID. These three raids levels, which will be referred to as "Composite Raids" throughout this thesis, were comprised of eight different exercises. The eight exercises that comprised the three Composite Raids will be referred to as "Component Raids."

COMOPTEVFOR's schedule of events (SOE) was developed to support testing the ship's systems by presenting varying threat profile densities and formats to exercise the full capacity of shipboard combat systems. The SOE also determined which test events could be recorded for subsequent workload analysis. Following pre-trial examination of OPTEVFOR's test plan, NRaD, the agency responsible to OPTEVFOR for stress analysis, identified segments of the test events which were classified as high activity anti-air warfare (AAW) scenarios. NRaD also identified relatively low activity periods for comparative baseline assessments. (NRaD, 1992, p. 4)

The high activity exercises selected included two broad categories of assaults; that is, manned aerial raids and anti-ship missile raids. Three manned aerial raids (MR-3, MR-11, and MR-12) were launched against DDG 51 on 17 January 1992. A fourth manned aerial raid of significantly greater proportions than the three preceding raids was launched 3 February 1992. This large stream raid, dubbed MR-MAX, tested DDG 51's ability to handle anticipated aerial saturation attacks. The final two high activity exercises involved live missile firings against simulated anti-ship missile drones. These two missile raids, designated MF-4E and MF-7, were executed 31 January and 2 February 1992, respectively. (NRaD, 1992, p. 4)

The relatively low periods of activity used for baseline comparisons were two periods, both on 17 January 1992. These two periods, termed NO-RAID-1 and NO-RAID-2, were selected because they were similar to normal underway operations. (NRaD, 1992, p. 5)

During early OPEVAL events (MR-3, MR-11, and MR-12), weapon engagements were simulated; that is, the CIC team would rehearse the firing sequence but not actually release a missile. The SOE indicated there would be medium to high density multi-warfare threats, with the possibility of commercial and friendly forces mixed with the threat. This combination of factors produced the potential for high track density and increased workload. (NRaD, 1992, pp. 4-5)

Events in the later stages of the OPEVAL included live firing events (MF-4E and MR-7), together with a simulated maximum density manned raid (MR-MAX) of approximately 50 aircraft (NRaD, 1992, pp. 4-5). The Composite Raids, which were comprised of these Component Raids, are briefly described below.

1. NO-RAID

Two relatively low activity periods were considered baseline workload levels. These periods were called NO-RAID-1 and NO-RAID-2. They were free from any scheduled air activity and considered transit time for the ship by the SOE. These conditions provided data from what was considered a "normal watch" while the ship steamed independently. As such, they represent baseline activity levels.

2. MANNED-RAID

The second workload category consisted of four manned aircraft raids of varying intensity; MR-3, MR-11, MR-12, AND MR-MAX. All engagements were simulated and typified scenarios presented during inport training exercises in team trainers. There were, however, differences during the OPEVAL that would add to the amount of workload and stress experienced by the crew. These differences included the presence of heavy electronic jamming, the anxiety of performance pressure induced by the OPEVAL, and other variables such as fatigue, motion sickness, real and simulated equipment failures, and in the case of MR-MAX, the relative size of the incoming raid.

3. MISSILE-RAID

The third workload category was induced by live missile firings. The two events of this category, MF-4E and MF-7, were live fire exercises at multiple air targets. Both scenarios presented the CIC team with challenging and realistic engagement geometries. Workload would almost certainly be greater than that of the normal watch period (NO-RAID-1 and NO-RAID-2) and probably greater than that of the MANNED-RAID category. Although most, if not all, of the same factors that contributed to high levels of workload and stress in the MANNED-RAID scenarios were present in the live fire exercises, there were at least two factors that could limit the level of stress compared to actual combat. They were (a) the constraints imposed by range safety considerations and (b) the ability of the CIC team to deduce the threat axis by knowing the physical limits of the missile test range.

B. INDICES OF WORKLOAD

The analysis of communication patterns during these raids included three quantitative measures. These measures, which will be discussed below, were Mean Transmission Time (MTT), Speech to Pause Ratio (SPR), and Speech Time to Total Time Ratio (ST/TT). These three measures were derived from a simple observation which was defined as the duration of a verbal utterance on the communication network, measured in seconds, by any CIC team member.

1. Mean Transmission Time (MTT)

Mean Transmission Time is the average duration of voice transmissions from the CIC team during each simulated raid. It provided a convenient measure of the typical length of vocal transmissions by CIC team members under varying levels of workload.

2. Speech-to-Pause Ratio (SPR)

The Speech-to-Pause Ratio (SPR) is the ratio of the total Speaking Time (ST) to the total Pause Time (PT). Speaking Time is the sum of all transmission times on the network over the raid. Since simultaneous transmissions by more than one team member cannot occur on the network, ST cannot exceed the duration of the raid. Pause Time is the sum of the times during which no transmissions or keying of a transmitter was detected.

Total Time (TT) was measured in minutes, hence, the need to divide the product of the number of transmissions and MTT by 60 to achieve a like unit of measurement for ST. The relationship between Speaking Time and Pause Time and their basis in the SPR is formulated below.

$$SPR = \frac{ST}{PT}, \quad \text{where}$$

$$ST = \frac{(NUMBER\,OF\,TRANSMISSIONS \times MTT)}{60},$$

$$MTT = MEAN\,TRANSMISSION\,TIME,$$

$$PT = (TT - ST),$$

and

$$TT = TOTAL\ TIME\ OF\ THE\ RAID\ (MINUTES).$$

3. Speech-Time-to-Total-Time (ST/TT) Ratio

The Speech-Time-to-Total-Time Ratio (ST/TT) is the ratio of the amount of time in which speech occurred to the length of the entire raid measured in minutes. This measure highlighted the total speech time during a given raid against the total elapsed time of the raid. This ratio is related to SPR, but, in fact, is different. SPR compares the Speech Time to the Pause Time, while ST/TT compares the Speech Time to the Total Time of the raid.

C. DATA COLLECTION TECHNIQUE

COMOPTEVFOR directed that only unobtrusive, noninvasive methods could be used to collect data. Therefore, audio taps were installed on Internal Communications Net 15 and used to record voice transmissions between CIC team members during the two NORAID episodes and six subsequent Component Raid scenarios. Internal Communications Net 15 aboard AEGIS combatants is the primary means by which members of CIC coordinate tactical employment of the ship. Naval Warfare Analysis Center (NWAC) time stamped the data collected to enable synchronized post-event reconstruction of the CIC team's voice communications. (NRaD, 1992, pp. 9-10)

D. STATISTICAL ANALYSIS

This section on statistical analysis is presented in two parts. The first part discusses the actual data and the adjustments made to it to account for unplanned events during various raids. The second part discusses the scaling technique employed to provide a basis to rank order the various raids in terms of increasing workload.

1. Data

NRaD compiled written transcripts of voice communications from the audio tapes taken during DDG 51 OPEVAL. The length of each transmission (in seconds) was recorded during the transcription process. These individual transmission times were entered into a commercial computer statistical software package, *STATGRAPHICS*, for exploratory data analysis. They were entered two ways. The first format treated data from each of the eight Component Raids individually; for example MR-3, MR-MAX, MF-7, etc. The second format collapsed across these specific raids to produce the three Composite Raid categories: NO-RAID, MANNED-RAID, and MISSILE-RAID. The positions on the CIC watch teams were occupied by the same operators for all raids. An exception to this was the Commanding Officer's presence during all of the missile firing events.

There were two other irregularities in the data. First, two raids in the MANNED-RAID data (MR-3 and MR-MAX) each had an unusually long transmission burst; 30 and 23 seconds, respectively. These transmissions were deleted because their content was atypical: they contained miscellaneous discussions that did not pertain to the current operational environment.

Likewise, the longest transmission was deleted from two raids in the MISSILE-RAID data (MF-4E and MF-7). In these two transmissions, a 24 and 33 second burst respectively, the content dealt with range safety procedures, an unavoidable artificiality made for safety considerations.

2. Tests

a. Tests of Significant Differences

Preliminary screening to determine if there was a statistical difference between the distributions of transmission times across the Composite Raids was accomplished using a Chi-Squared test for homogeneity. The null hypothesis was stated as follows:

H_0: The lengths of verbal utterances during the Composite Raids (NO-RAID, MANNED-RAID, and MISSILE-RAID) are the same.

The k-sample problem for a categorical variable is the problem of testing whether the distributions of the variable are the same for k populations, based on independent random samples from each population. As stated previously, the chi-square test for homogeneity is the appropriate statistical procedure for this purpose and has the following form:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}},$$

where the summation is over all cells in the two-way table, O_{ij} represents the observed frequency for the *ij*th category of the variable, and E_{ij} represents the expected frequency of the variable. E_{ij} is the row total multiplied by the column total for the same row and column divided by the grand total. The chi-square test will be based on the rule,

$$\text{Reject } H_0 : \chi^2 \geq c \ ,$$

where the cut-off value *c* is to be determined to control the type I error probability to the value specified by the preassigned significance level α. The chi-square distribution is indexed by an integer-valued parameter called the degrees of freedom. The degrees of freedom equal the number of rows in the table minus one times the number of columns minus one. (Koopmans, 1987, pp. 412-415)

The Kolmogorov - Smirnov two-sample test was used to verify the results of the chi-squared test for a pair-wise comparison of the Composite Raids. The

Kolmogorov-Smirnov test evaluates overall goodness-of-fit to determine whether two samples may reasonably have come from the same distribution. The procedure requires calculating the maximum vertical distance between the cumulative distribution functions (CDF's) of two samples. If the distance is large enough, the hypothesis that two samples come from the same distribution is rejected.

b. Subjective Workload Level

Before the OPEVAL, operational subject matter experts and behavioral researchers from OPTEVFOR and NRaD determined that in terms of workload, the raids, from lowest to highest, ranked as follows: NO-RAID, MANNED-RAID, and MISSILE-RAID. Except for MR-MAX, there was no attempt to predict how the Component Raids which comprised these three Composite Raids would rank within each category. For MR-MAX, the judges held that workload levels should rank closer to the missile firing events simply because of the size of the raid. After compiling summary statistics and deriving temporal measures for each Component Raid, the Component Raids were ranked according to the temporal measures. The rationale to rank the Component Raids by these criteria was straightforward: if each of the measures of communication patterns actually tapped into workload, then the rank order of the Component Raids made on the basis of these measures should be the same.

The Component Raids were ranked three ways. First, they were ranked by increasing magnitudes of the ST/TT ratio and the SPR; and second, they were ranked by decreasing magnitude of MTT. However, since these two criteria produced different rankings of workload[1], a third ranking was done to determine which criterion most accurately reflected workload. A subjective scaling method, scaling by magnitude estimation, was selected to produce the third set of workload rankings.

This scaling method required subject matter experts to mark a point on a line that corresponded to the subjective magnitude of the dimension being rated, in this case, workload. The subject matter experts were ten Surface Warfare qualified Lieutenants. Respondents were given two reference points on which to rank the seven Component Raids. The low workload reference point was based on independent operation of a ship during peacetime. The high workload reference point was based on carrier battle group operations during a wartime footing. The reference points were events not included in the Component Raids being rated in order to produce clear agreement as to the rank of those references. The scale's unit of measurement is totally arbitrary, but by providing two reference points, an interval scale is implied because both a scale and an origin are determined. (Zatkin, 1983, pp. 1-6) APPENDIX A contains the test questionnaire.

---

[1] The differences in Component Raids will be discussed in Chapter IV.

III. RESULTS

The results of the analysis on communication patterns will be presented in two parts. The first part considers the three temporal measures of frequency and duration of verbal transmissions (MTT, SPR, and ST/TT). The second part submits these measures to statistical analyses and rates the relative workload attributed to each scenario on a subjective basis. Both sections treat the data on the three Composite Raid scenarios (NO-RAID, MANNED-RAID, and MISSILE-RAID) first, then breaks those scenarios into their Component Raids; that is, NO-RAID-1, NO-RAID-2, MR-3, MR-11, MR-12, MR-MAX, MF-7, and MF-4E.

A. TEMPORAL MEASUREMENTS

1. Composite Raids

TABLE 1 shows the results of the temporal measures taken when DDG 51 communication data was grouped by NO-RAID, MANNED-RAID, and MISSILE-RAID; that is, when Component Raids (MR-3, MR-11, and so forth) were collapsed to form the three Composite Raid scenarios. The last column of TABLE 1, labeled COMBINED, shows the temporal measures collapsed across all Composite Raids. Figure 3 shows the trends in the temporal measures across the three Composite Raid scenarios as a function of increasing workload.

TABLE 1. TEMPORAL MEASURES FROM COMPOSITE RAIDS

	NORAID	MANRAID	MISSILE	COMBINED
n	305	1808	651	2764
MTT (SEC)	2.92	2.73	27.6	2.71
σ^2	7.37	4.54	3.85	4.70
σ	2.71	2.13	1.96	2.17
ST (MIN)	14.8	82.3	27.6	124.8
TT (MIN)	75.0	265.6	81.6	422.2
ST/TT	.20	.31	.34	.30
SPR	.25	.45	.51	.42

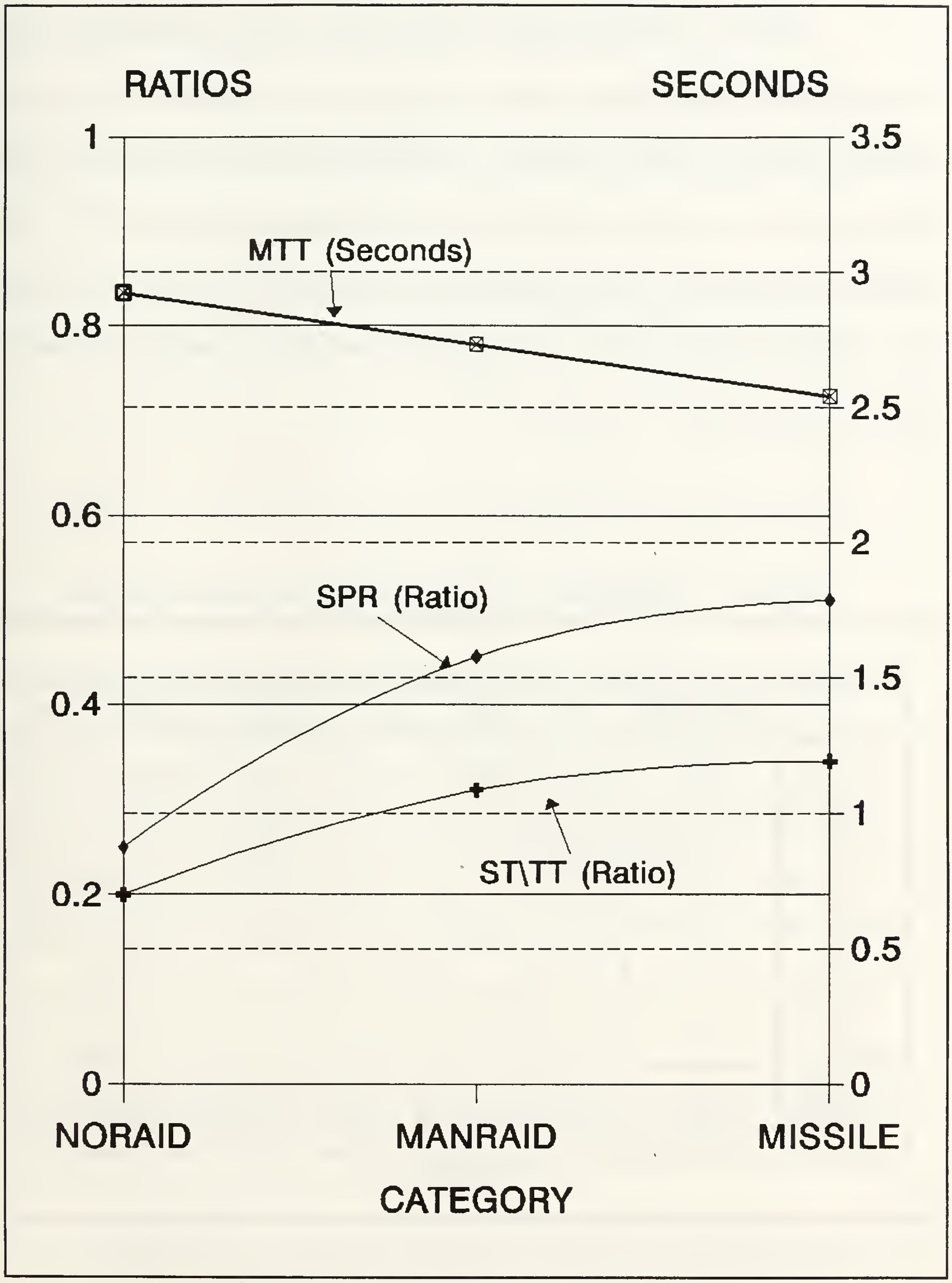


Figure 3. Composite Raid Trends

2. Component Raids

TABLE 2 shows the temporal measures from each Component Raid. Figure 4 decomposes the three Composite Raid scenarios into their Component Raids and rank orders the Component Raids according to increasing magnitudes of SPR and ST/TT ratio. Figure 5 rank orders these same scenarios by decreasing magnitude of MTT. These ranking criteria, and the rational for their use were discussed in the previous chapter. For simplicity, the NO-RAID-1 and NO-RAID-2 events were combined to provide a baseline reference point.

TABLE 2. TEMPORAL MEASURES FROM COMPONENT RAIDS

	NORAID	MR3	MR11	MR12	MRMAX	MF4E	MF7
n	305	217	696	400	495	325	326
MTT	2.92	2.85	2.79	2.72	2.62	2.36	2.72
σ^2	7.37	4.90	4.43	4.67	4.43	3.69	3.96
σ	2.71	2.21	2.10	2.16	2.10	1.92	1.99
ST (MIN)	14.8	10.3	32.4	18.1	21.6	12.8	14.8
TT (MIN)	75.0	41.2	93.1	60.0	71.3	38.0	43.6
ST/TT	.20	.25	.35	.30	.31	.34	.34
SPR	.25	.34	.53	.43	.44	.52	.52

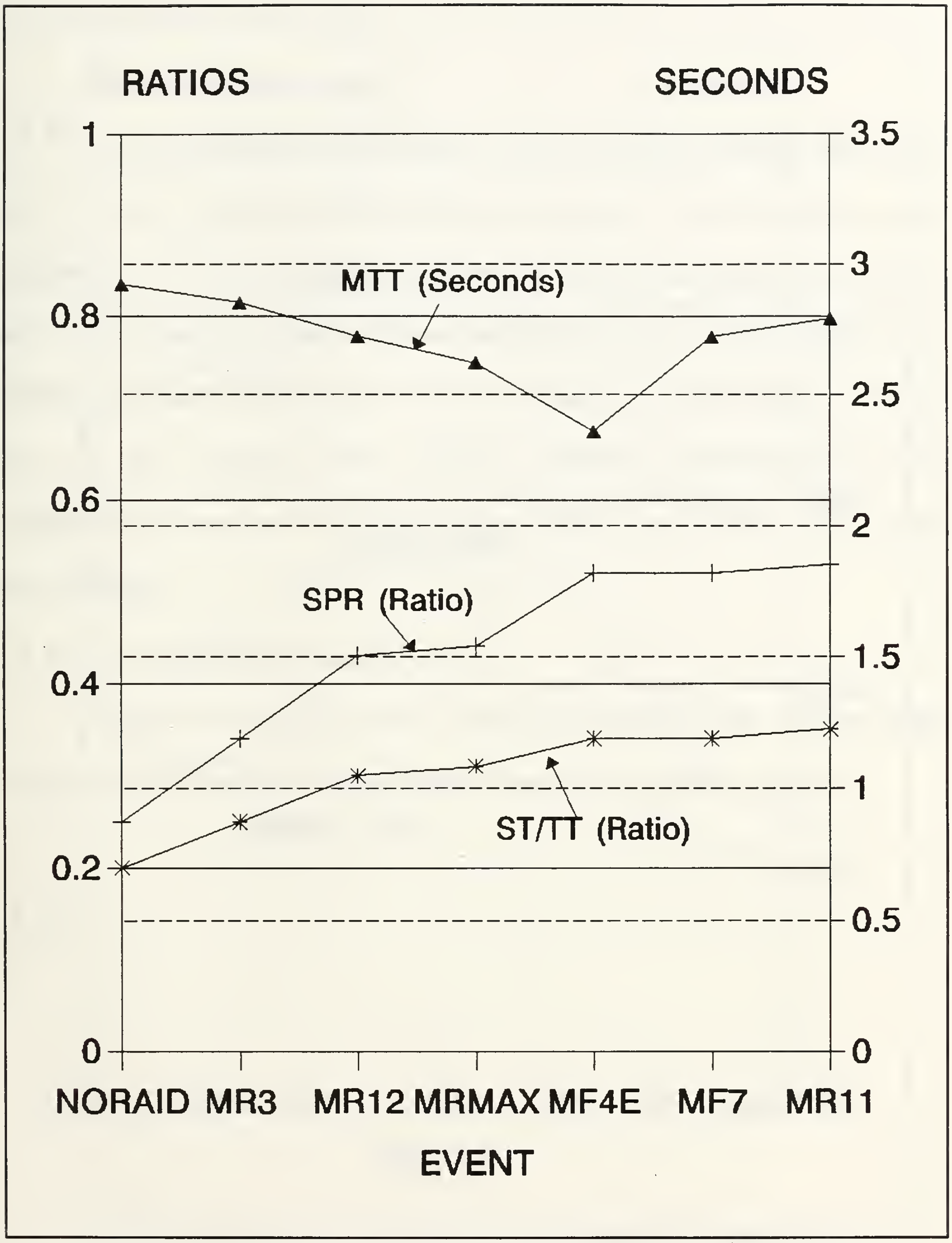


Figure 4. Rankings Based On SPR And ST/TT Ratio

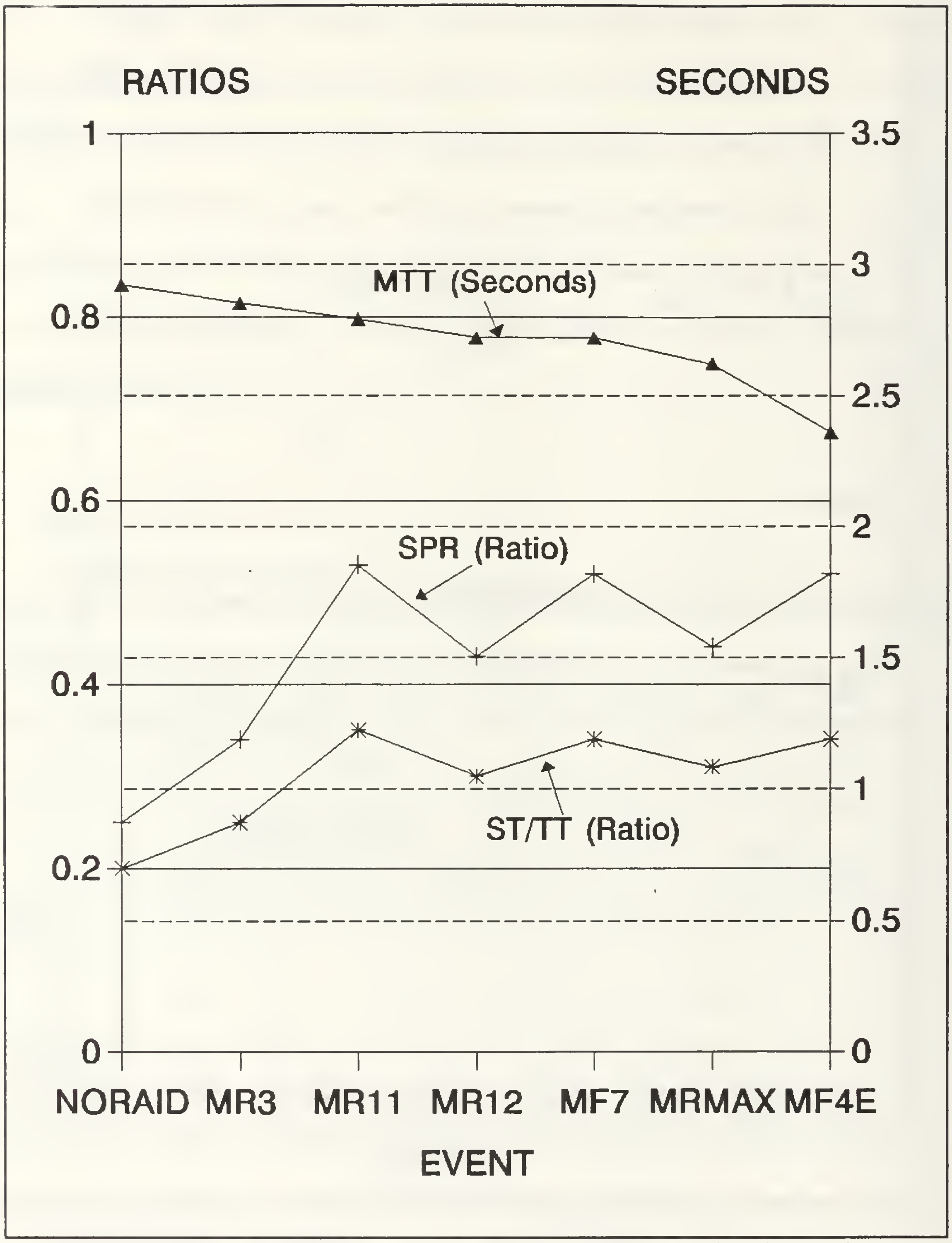


Figure 5. Rankings Based On Mean Transmission Time

B. STATISTICAL ANALYSIS

1. Composite Raid Analysis

The null hypothesis that the lengths of verbal utterances were the same during each of the three Composite Raid categories was rejected by the chi-square test for homogeneity ($X^2 = 27.7$; df = 12; $p < 0.01$). The three Kolmogorov-Smirnov two sample tests performed on each different pairing of the three Composite Raids also produced significant differences at $\alpha = 0.01$. The duration of verbal utterances collected during the three Composite Raid groupings (NO-RAID, MANNED-RAID, and MISSILE-RAID) came from statistically different distributions. Figure 6 depicts these three distributions.

2. Component Raid Analysis

TABLE 3 shows the results of the subject matter experts' ranking of the Component Raids according to their subjective impression of relative workload.

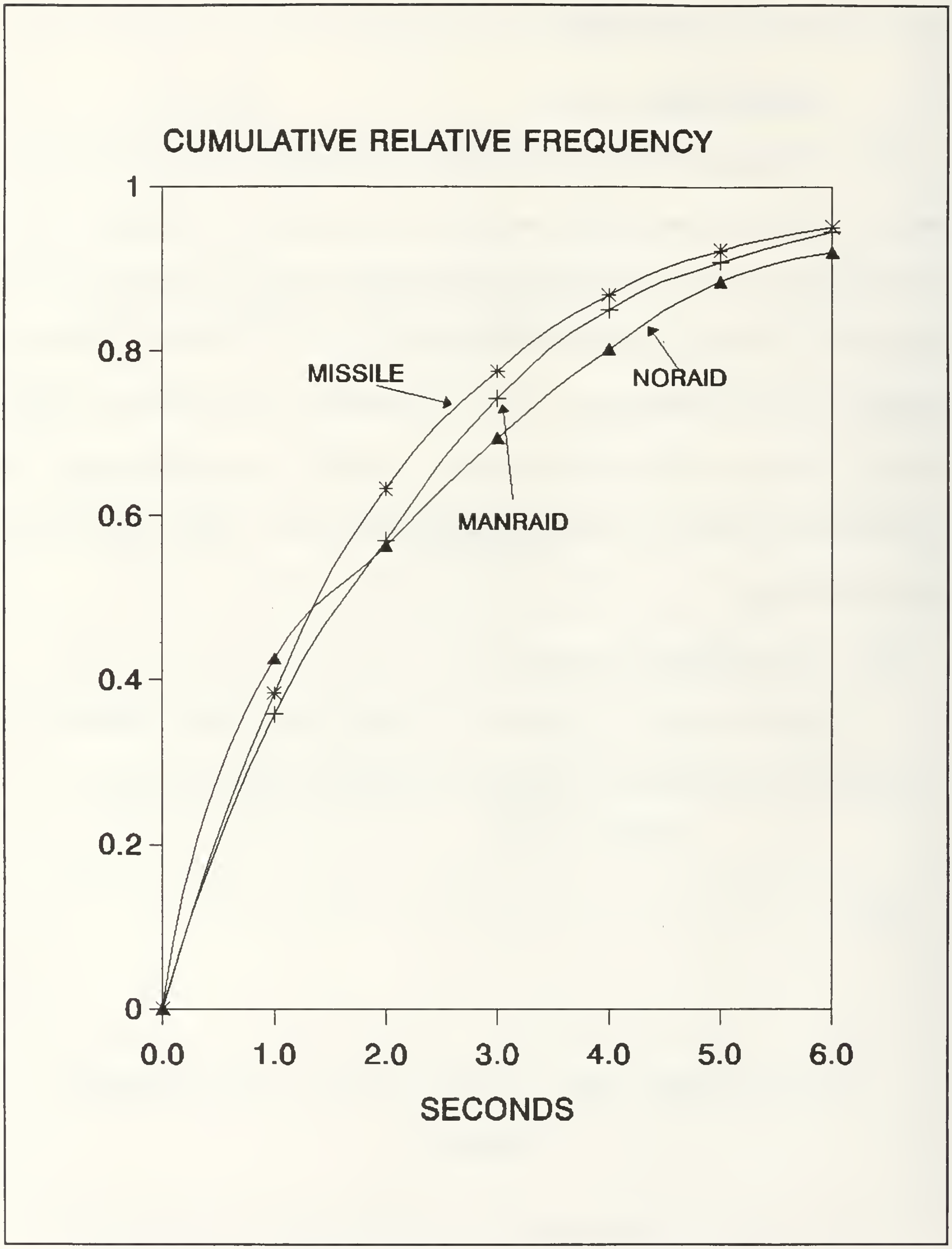


Figure 6. Composite Raid Cumulative Distributions

TABLE 3. SUBJECTIVE RANKINGS OF COMPONENT RAIDS

Ss	NORAID	MR3	MR11	MR12	MRMAX	MF4E	MF7
1	54	98	119	110	87	122	31
2	28	33	69	142	53	130	40
3	26	98	141	116	137	153	149
4	26	143	157	164	174	100	149
5	10	45	150	90	137	52	114
6	26	52	72	78	75	34	31
7	32	75	145	110	49	63	122
8	39	83	133	128	120	100	125
9	12	98	123	131	145	31	75
10	11	73	110	115	100	114	120
MEAN	26	75	122	118	108	91	100
RANK	7	6	1	2	3	5	4

IV. DISCUSSION

The data revealed that the measures derived from communications between CIC team members during simulated Composite Raids showed systematic quantitative differences as a function of varying workload levels. The data also revealed, however, that when these Composite Raids were decomposed into their Component Raids, the relative ranking of workload reflected by each Component Raid varied as a function of the temporal measure or subjective scale values chosen as the criterion for the workload ranking. Different ranking criteria produced different workload rankings for the Component Raids. This chapter will discuss three themes: (a) the general finding from the Composite Raid data, (b) the inconsistencies in the Component Raid ranking data, and (c) a comparison of findings from other related studies.

A. FINDINGS

1. Composite Raid

Figure 3 shows the temporal measures plotted against the three Composite Raid scenarios. The data show that communication patterns among CIC team members were significantly altered as a function of increasing workload. Based on the assumption that there is a monotonic increasing level of workload associated with the NO-RAID, MANNED-RAID, and MISSILE-RAID scenarios, the temporal measures followed the same monotonic relationship; that is, as workload increased, two of the measures, SPR and ST/TT ratio, also increased, and one, MTT, decreased. This finding substantiates that there were, in fact, quantitative differences in communication patterns and that the temporal measures systematically varied as a function of workload imposed on the CIC team.

Simply stated, Figure 3 shows that as workload increased, the *frequency* of transmissions also increased, but the *duration* of transmissions decreased. Moreover, as TABLE 1 revealed, as Mean Transmission Time decreased with increasing workload, the variability of transmission length also decreased. The variability of SPR and ST/TT decreased with increasing workload, simply because Mean Transmission Time was used to derive these two measures.

2. Component Raids

Figures 4 and 5 show the rank order of the seven Component Raids ranked as a function of two different criteria. In Figure 4, the Component Raids were rank ordered according to increasing SPR and increasing ST/TT ratio. In Figure 5, the Component Raids were rank ordered according to decreasing Mean Transmission Time. As noted in the previous section, the two sets of rankings of relative workload made on the basis of these two candidate measures of workload are not the same.

a. Workload Rankings

The three Composite Raid scenarios were logically ordered according to increasing workload; that is, NO-RAID imposed the lowest level of workload and MISSILE-RAID imposed the highest level. As Figure 3 shows, the temporal measures track this workload ordering. In a very rigorous sense, one would expect the same relative order to hold among the Component Raids when they are decomposed from the Composite Raids. However, as reported above, that order was not invariably retained. When decomposed, there were minor transpositions in the rank order of workload associated with the seven Component Raids depending upon the criterion used for the ordering.

Inspecting the three criteria (MTT, SPR and ST/TT ratio, and subjective rankings) used to rank workload of the Component Raids, reveals that MR-3 and NO-RAID consistently ranked sixth and seventh, respectively; the lowest workload levels on the scale. Of the five remaining Component Raids, MR-11 ranked at the top when ranked by subject matter experts and also when SPR and ST/TT ratio were used as the criteria. MF-7, MF-4E, and MR-MAX ranked high on workload when temporal measures were used as the criterion, but not as high when they were ranked by subject matter experts. This transposition of relative position in workload ranking provided by the subject matter experts could be attributed to a combination of the scenario descriptions provided in the scaling survey and the subjective interpretation of them by the respondents.

The minor inconsistencies and transpositions in the relative workload rankings of the Component Raids has a straightforward explanation. The raids' unique characteristics and conditions, which clearly contributed variability, were suppressed when the data were collapsed into the three Composite Raid scenarios. These conditions and characteristics included unique environmental conditions, unplanned or imposed equipment failures, uncertain tactical picture, and varying levels of workload within a Component Raid. Given the constraints imposed on experimental rigor by the operational situation associated with an OPEVAL, however, the central thesis still holds: there are, in fact, systematic quantitative changes in communications patterns among CIC team members as a function of increased workload.

B. COMPARISON TO NRaD FINDINGS

NRaD was the lead test agent for the stress analysis portion of DDG 51's OPEVAL. Their three methodologies for measuring stress were considerably different than the one explored in this thesis. As previously discussed, NRaD used subjective workload assessments from CIC watchstanders, subjective assessments of performance pressure by experts observing video and audio tape recordings of the CIC team, and objective measures of workload using console use patterns reconstructed from onboard computers.

A comparison of the results from the NRaD measurement approach and the present approach serves three purposes.

- If the two independent approaches produce similar conclusions, then the validity of the general finding that stress affected operator performance in DDG 51's OPEVAL is increased.
- The original objective of both studies was to demonstrate that stress was present and measurable in the OPEVAL. If the two methods meet that objective, then both can be considered reliable starting points for future use in TADMUS field experiments.

- If the analyses produce dissimilar conclusions, then either one or both methods could be considered insensitive to changes in workload induced by the OPEVAL.

Any outcome would renew efforts to resolve the difficult methodological task of unobtrusively measuring workload in an operational setting.

According to their *subjective analyses* of the simulated raids, NRaD concluded that there were ". . . no overt indications of excessive individual or team workload or performance pressure stress." This conclusion was caveated by reporting that ". . . it was clear that periods of medium workload intensity and short periods of high intensity occurred in the CIC." (NRaD 1992, p. 18)

The NRaD report did not specifically name the Component Raids which exhibited medium or high intensity workload. However, the Component Raids that NRaD did report three or more times as exhibiting noteworthy error rates, response times, and objective workload were MR-11, MF-4E, MF-7, and MR-MAX. Figure 3 of the present study identifies the same four events as having the greatest amount of workload compared to the baseline NORAID events. The difference between the NRaD approach and the present approach is that while both methods produced similar conclusions, the measures used to substantiate these conclusions were different. NRaD determined relative workload principally by subjective means, while the present study determined the same relative workload by an analysis of human communication patterns.

The two studies together underscore the fact that DDG 51's CIC team experienced periods of medium to high intensity workload and that these periods occurred in at least four of the six raids. Moreover, these events were predicted to produce the highest levels of workload during the OPEVAL and were designed consistent with the policy of stressing human operators as well as the machine. Considering data was collected on only eight scenarios, two of which were considered baseline measures (NO-RAID), the NRaD and the present study could provide a potentially productive point of departure for further research into the measurement of workload.

C. COMPARISON TO PREVIOUS TEMPORAL ANALYSIS

One of the few studies that used temporal aspects of verbal communications patterns to assess the impact of stress upon those patterns was conducted by Hicks (1979). The investigation examined both laboratory induced stress (electrical shock administered randomly while subjects read a passage) and situational stress (undergraduate students delivering speeches to an audience). Besides acoustical measures, which will not be considered here, Hicks' analysis derived two of the three temporal measures used in the present study; that is, SPR and the ST/TT Ratio. The third measure used by Hicks was Speech Rate. (Hicks, 1979, pp. xviii-xix)

Speech Rate is not equivalent to the present study's Mean Transmission Time. Hicks defined Speech Rate as the number of syllables produced per second. The present study defined Mean Transmission Time as the average duration of discrete verbal transmissions over the entire combat simulation.

Hicks' findings showed that speech produced under stressful conditions exhibited quantitatively different temporal patterns than speech produced under non-stressful conditions. The situational stress experiment revealed that SPR and ST/TT increased and Speech Rate decreased. The increased SPR and ST/TT measures were significant ($p < 0.05$). The decrease in Speech Rate was not statistically significant at $\alpha = 0.05$. Hicks concluded that stress tends to decrease Speech Rate and the number of speech bursts and pauses, which results in longer continuous speech periods. Simply stated, Hicks found that subjects in his situational experiments *communicated slower* and their *verbal utterances were longer.* (Hicks, 1979, pp. xix-xx)

Hicks' findings seem contrary to the present study's findings, but there are three plausible explanations for the apparent contradiction. First, Hicks' experiments did not analyze communication patterns elicited from a team. He analyzed communication patterns from individual speakers. Second, Hicks neither imposed multiple tasks on his subjects which required them to allocate their attentional resources across those tasks, nor did he tax their short term memory capacities. He simply had his subjects perform one task. Third, Hicks' subjects were not trained to use a highly disciplined, highly codified tactical language. His subjects were free to use any style and any rhetoric in their speech to their peers.

Despite the differences between Hicks' findings and the present study's findings, two central findings stand out. First, communication patterns from both individuals and teams tend to show quantitative changes as a function of stress. Second, these changes seem to be temporal in nature; that is, the frequency and duration of verbal transmissions with which humans communicate are affected by workload and its associated level of stress.

V. CONCLUSIONS AND RECOMMENDATIONS

A. CONCLUSIONS

The present study's findings stem from quantitative analyses of the 2,700 verbal transmissions made by members of DDG 51's CIC team while they were exposed to different levels of workload during their ship's OPEVAL. The Composite Raid data produced the clearest findings: as workload increased, the frequency of transmissions increased while the duration of transmissions decreased. There were more transmission per unit time, but the transmissions were shorter.

When the Composite Raids were decomposed into their Component Raids, and those Component Raids were ranked according to increasing or decreasing magnitudes of the temporal measures, the rank order of the raids tracked reasonably well with three other independent workload rankings of the same raids. The three different rankings were made by (a) a sample of Surface Warfare qualified officers, (b) operational experts at OPTEVFOR, and (c) behavioral researchers at NRaD. There was, therefore, convergent validity; that is, different rankings based on different criteria, including the temporal measures, produced like rank orderings of workload associated with the Component Raids.

Finally, when the present study's findings were compared to an open literature study of temporal measures in voice communication and stress, the comparison produced seemingly contradictory results. While the present study showed that verbal transmissions were more frequent, but shorter, the open literature study showed just the opposite: transmission bursts were less frequent, but longer in duration. The apparent contradiction probably derives from very dissimilar experimental conditions; that is, each study's subjects performed very different tasks that imposed significantly different cognitive demands. Despite the apparent contradictory findings, however, both studies did, in fact, show that communication patterns are affected by stress and that these changes are quantifiable.

Workload and stress effect changes in human communication patterns. That finding, which in the present study is based on naturalistic observations collected by unobtrusive, noninvasive means; that is, recording human speech, provides a basis for further research into first, the isolation of these patterns; and second, demonstrating that they are reliable and valid indices of workload and stress.

B. RECOMMENDATIONS

Congress directed that research into stress in team coordination be conducted to prevent tragedies similar to the 1988 VINCENNES incident. DDG 51's OPEVAL was the first OPEVAL that used findings from the TADMUS research project. There are four important lessons learned.

1. Experimental Realism

Laboratory experiments must be more realistic. They must closely mirror the environment which they purport to model. Laboratory experiments, while methodologically rigorous and tightly controlled, typically do not produce findings that are easily generalizable to the operational environment. These experiments will continue to produce critically needed information, but they must not be considered ends in themselves. Sailors operating complex equipment on a daily basis at sea could provide invaluable information to behavioral researchers. The DDG 51 OPEVAL should mark the beginning of a regular series of operational opportunities to verify the methods and results produced by laboratory experiments.

2. Front-End Planning

Because defense budget is shrinking, workload data must be extracted from the precious few opportunities available to gather it. An OPEVAL is a reasonable time to gather workload data provided adequate planning and operationally acceptable performance measures are considered early in its test plan development. Unobtrusively collecting performance data from which reliable estimates of workload could be later derived should be considered at the very beginning of the test plan development and not be included as an after thought. OPTEVFOR must be complimented for their efforts to incorporate this "first of its kind" data collection evolution into such a detailed test plan on short notice.

3. Human Factors in System Design

The Surface Warfare community needs to follow Naval Aviation's outlook on the importance of human factors in system design. With the advent of more complex combat systems in the surface community, the community should increasingly attend to the broad range of human factor requirements necessary to accommodate the increasing complexity and the demands it imposes in the human operator and maintainer.

4. Application

As applied to analysis of internal communications, at least two areas are recommended for further review.

- Although still in an experimental stage, voice stress analysis could provide an insight to and/or verification of methods analyzed in this study.
- Communication data from inport team trainers, fleet exercises, and actual combat events; for example, the tapes from USS VINCENNES, should be analyzed to produce reference points on a line representing stress effects on CIC team communications. This reference line could be used in future team trainer design as a gauge for evaluating the presence and amount of stress.

C. SUMMARY

It is important to note that an analysis of communications from a CIC team exposed to different levels of workload has never been conducted in an operational test environment and the findings must be considered tentative. This study, however, probably would have been further delayed had not the VINCENNES incident occurred and Congressional pressure been applied.

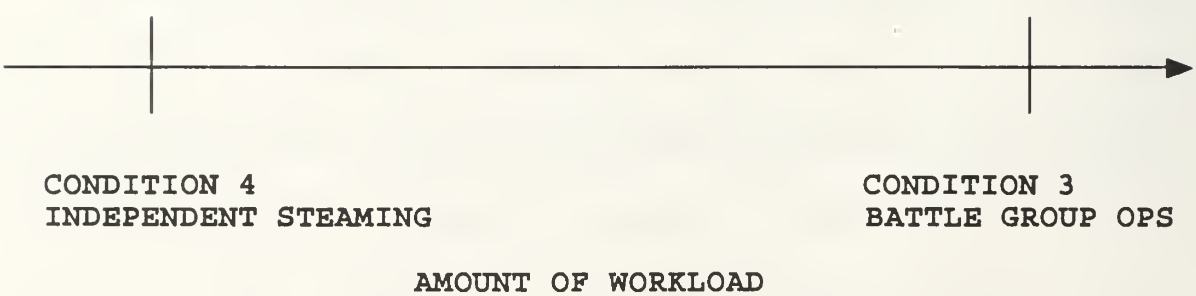
The motivation notwithstanding, as advances in technology increase naval combat system complexity, the chances of catastrophic error also increases dramatically. In the past, the air arm of the U.S. Navy has lead the way in human factors related research because of the potentially catastrophic consequences of mistakes in the cockpit of advanced jet aircraft. With the advent of AEGIS, New Threat Upgrade, and the extended ranges and lethality of surface-to-air, surface-to-surface, and surface-to-subsurface weapons, it is paramount that the Surface Warfare Community attend to the human interfaces to these devastating weapon systems, and the human information processing requirements that support them. The Navy is at a crossroads with respect to downsizing and decreasing budgets, but if this area of study is neglected, events such as what occurred in the Persian Gulf will become more commonplace and more tragic as technology outpaces the ability of man to control it.

APPENDIX A

MAGNITUDE ESTIMATION OF EXERCISE SCENARIOS

Please rate the following seven scenarios according to the amount of workload you would expect to experience as part of a CIC team in an AAW ship. The two points on the line are provided as reference points for your convenience.

Mark your selection with the appropriate abbreviation from the list of scenarios on the following page. Place your selection **ABOVE** the line and draw an arrow to the point on the line where you would like it to appear. You do not have to rate all the scenarios. If you are unfamiliar with a scenario, feel free to ignore it and move on.



SCENARIOS

MF4E. MISSILE EXERCISE (Firing event). No other ships in company. Approximate launch time known with threat sector of 90 degrees. Heavy jamming and chaff present. Eight targets presented and 16 Standard Missiles (SM-2) available to counter. Targets are air and surfaced launched drones and anti-ship missiles at varying altitudes and speeds. CPA times are within thirty seconds.

NORAID. UNDERWAY WATCH. Steaming in company of FFG during multi-threat exercise. HF data link established. No known contacts of interest. Helo ops scheduled within 30 minutes.

MR11. MULTI-THREAT EXERCISE. Steaming in company of FFG. Weather deteriorating rapidly. Events in exercise include: heavy airborne jamming and chaff corridors, six attack aircraft simulating attacks, simulated loss of weapon control system, possible submarine contact in area, and simulated TASM strike in progress on constructive Kirov.

MR3. MULTI-THREAT EXERCISE. Steaming in company of FFG. Data link established and FFG reporting its helo has gained contact on a submarine (outside enemy attack range). Four aircraft attacking with no jamming support.

MRMAX. AAW EXERCISE. Steaming independently. Heavy jamming and chaff present. Threat consists of a 50-60 manned aircraft stream raid at varying directions, altitudes and speeds. Ship is using decoys and high speed maneuvering.

MF7. MISSILE EXERCISE (Firing event). No other ships in company. Approximate launched time known. Heavy jamming and chaff present. Targets consist of two high speed, high altitude air launched drones and one unmanned aircraft. Targets' CPA time very close to simultaneous.

MR12. MULTI-THREAT EXERCISE. Steaming in company of FFG and controlling P3C. Data link established. Hostile submarine in area and unlocated. Weather deteriorating rapidly. Multiple, but spaced, three aircraft raids with medium airborne jamming and chaff. Simulated TASM strike in progress on constructive Kirov.

LIST OF REFERENCES

Allard, C.K., *Command, Control, and the Common Defense*, Yale University Press, 1990.

Barry, John and Charles, Roger, "Sea of Lies", *Newsweek,* July, 13 1992.

Binkin, M., *Military Technology and Defense Manpower*, The Brookings Institute, 1986.

Congressional Record, H.A.S.C. No. [illegible]-94, House Armed Services Committee Hearing on the Office of Operational Test and Evaluation, Government Printing Office, 14 September 1988.

Department of the Navy, Sea Systems Command, *CG 47 Class Services: Student's Guide Ship Familiarization*, Litton, Ingalls Shipbuilding, 1987.

Hicks, J.W. JR., *An Acoustical/Temporal Analysis of Emotional Stress In Speech*, Ph.D. Dissertation, University of Florida, 1979.

Hill, M., "The Vincennes: Seven Minutes In July," *San Diego Magazine*, February 1989.

Hockey, R., *Stress and Fatigue in Human Performance*, John Wiley & Sons, Ltd., 1983.

Kantowitz, B.H. and Sorkin R.D., *Human Factors: Understanding People-System Relationships*, John Wiley & Sons, Inc., 1983.

Keegan, J., *The Price of Admiralty: the Evolution of Naval Warfare*, Penguin Books, 1988.

Koopmans, L.H., *Introduction to Contemporary Statistical Methods*, Duxbury Press, 1987.

McGrath, J.J., "Performance Under Stress," Lecture for Prospective Commanding Officers and Prospective Executive Officers, April 1989.

Naval Postgraduate School Report 55-84-002, *Effects of Emotional and Perceptual-Motor Stress on a Voice Recognition System's Accuracy: An Applied Investigation*, by G.K. Poock and B.J. Martin, February 1984.

Navy Command, Control and Ocean Surveillance Center, Research, Development, Test and Evaluation Division, *Tactical Performance Under Stress During USS ARLEIGH BURKE (DDG 51) Operational Test*, May 1992.

Office of Chief of Naval Operations Memorandum CNO 1500: Serial 11B1/14-89 to Deputy Chief of Naval Operations (Manpower, Personnel, and Training), Subject: Sitrep On Stress Training Efforts, 27 April 1989.

Office of Naval Technology, *Tactical Decision-Making Under Stress (TADMUS)*, July 1991.

President's Blue Ribbon Commission on Defense Management, *A Report to the President on Defense Acquisition: A Formula for Action*, Government Printing Office, Washington D.C., 1986.

Petho, F.C., "Introduction to OA 3401: Human Factors in System Design," class notes, April 1992.

Pfeiffer, J., "The Secret of Life at the Limits: Cogs Become Big Wheels," *Smithsonian*, July 1989.

Polmar, N., *The Ships and Aircraft of the U.S. Fleet*, Naval Institute Press, 1987.

Squires, S., "Stress, Poor Communication Cited In Vincennes' Downing of Jet," *Washington Post*, October 7, 1988.

Telephone conversation between Commander J. Kren, 74, Commander Operational Test and Evaluation Force and the author, 7 July 1992.

Trost, C.A., "The Navy Is Living Up to Its Moral Imperative," *Washington Post*, August 20, 1988.

Wickens, C.D., *Engineering Psychology and Human Performance*, Harper-Collins Publishers, Inc., 1992.

Zatkin, J.L., "Scaling by Magnitude Estimation", class notes, September 1983.

INITIAL DISTRIBUTION LIST

1.	Defense Technical Information Center Cameron Station Alexandria, Virginia 22304-6145	2
2.	Library, Code 52 Naval Postgraduate School Monterey, California 93943-5002	2
3.	Commander, Operational Test and Evaluation Force Code 76 Norfolk, Virginia 23511-6388	1
4.	Commander Frank C. Petho, MSC, USN Naval Postgraduate School (OR/PE) Monterey, California 93943-5000	1
5.	Lieutenant Thomas M. Conlon, USN Department Head Class 126 Surface Warfare Officers School Command Newport, Rhode Island 02841-5012	2
6.	Commanding Officer Naval Command, Control and Ocean Surveillance Center RDT&E Division, Code 44 San Diego, California 92152-5000	1
7.	Department of the Navy Chief of Naval Operations (N86) Washington D.C. 20350	1
8.	Department of the Navy Chief of Naval Operations (N1) Washington D.C. 20350	1
9.	Captain George W. Conner, USNR Naval Postgraduate School (OR/CO) Monterey, California 93943-5000	2

Thesis
C7033 Conlon
c.1 DDG 51 operational evaluation : measures of workload from Combat Information Center communication patterns.