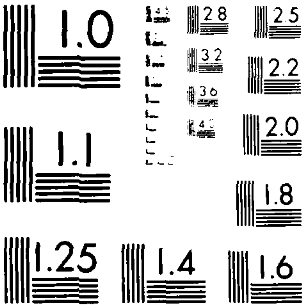
AD-A142 554 EFFECTS OF EMOTIONAL AND PERCEPTUAL-MOTOR STRESS ON A VOICE RECOGNITION S..(U) NAVAL POSTGRADUATE SCHOOL MONTEREY CA G K POOCK ET AL. FEB 84 NPS55-84-002 1/1

UNCLASSIFIED F/G 5/10 NL

END
DATE
FILMED
8 84
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

2

NPS55-84-002

NAVAL POSTGRADUATE SCHOOL

Monterey, California

EFFECTS OF EMOTIONAL AND PERCEPTUAL-MOTOR STRESS ON A VOICE RECOGNITION SYSTEM'S ACCURACY: AN APPLIED INVESTIGATION

by

Gary K. Poock
B. Jay Martin

February 1984

Approved for public release; distribution unlimited.

Prepared for:
Naval Electronics Systems Command
Code 613
Washington, DC 20363

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
NPS55-84-002

2. GOVT ACCESSION NO.
AD-A142554

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
EFFECTS OF EMOTIONAL AND PERCEPTUAL-MOTOR STRESS ON A VOICE RECOGNITION SYSTEM'S ACCURACY: AN APPLIED INVESTIGATION

5. TYPE OF REPORT & PERIOD COVERED
Technical

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Gary K. Poock
B. Jay Martin

8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Naval Postgraduate School
Monterey, CA 93943

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
62721N
N0003983WRDX083

11. CONTROLLING OFFICE NAME AND ADDRESS
Naval Electronics Systems Command
Code 613
Washington, DC 20363

12. REPORT DATE
February 1984

13. NUMBER OF PAGES
69

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distribution unlimited.

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

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
VTAG, Voice Recognition, Speech Recognition, Stress

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

This was an applied investigation examining the ability of a speech recognition system to recognize speakers' inputs when the speakers were under different stress levels. Subjects were asked to speak to a voice recognition system under three conditions:
1) Normal office environment
2) Emotional Stress and
3) Perceptual-Motor Stress.

Results indicate a definite relationship between voice recognition system performance and the type of low stress reference patterns used to achieve recognition.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

EXECUTIVE SUMMARY

The purpose of this research was to determine the effects of operator emotional stress and operator perceptual-motor stress on the recognition accuracy of a currently available voice recognition (VR) system.

The findings suggest if the operator was under no stress while training the VR system to recognize his voice, significantly more errors will result when he subsequently uses the VR system while he is experiencing emotional stress or perceptual-motor stress than when he uses the system under no stress. However, the increase in errors due to either type of stress can be reduced or eliminated when the operator trains the VR system under the corresponding stress condition.

In the present research, low levels of emotional stress and perceptual-motor stress were investigated, and although significant, the increase in errors due to mixing training and subsequent use conditions averaged about 2%.

It was concluded that current VR systems are negatively affected by using the system under a psychological environment different from the one under which it was trained. While the effects may be of small practical significance with low stress levels, the question was raised as to the potential for more practically significant increases in errors under high psychological stress environments.

1. INTRODUCTION

1.1 Background

In recent years, voice technology has developed to the extent that basic systems have now been used successfully in several industrial and military applications. Voice recognition devices that have been installed in "real world" situations have reduced input errors, cut task time, increased user friendliness, and proven cost effective in general (Nye, 1982; Poock, 1982). This successful climate, along with continued reductions in the cost of voice recognition systems, has made voice input an attractive alternative to motor input in a wide variety of settings.

Research and development are already in progress for the application of voice recognition in areas such as "walk up" electronic bank tellers, aids for the handicapped, and fighter jets. With each potential application, new questions and problems inevitably arise, usually with regard to system reliability. Different environmental conditions and task requirements introduce variables that may affect the human, the machine, or both. Noise, vibration, feedback techniques, training strategies, speech pattern access, response time, vocabulary size, and characteristics of particular populations of users are examples of such variables. So far, the state-of-the-art in voice recognition equipment has fared well in handling the kinds of problems that these variables can create.

While the effects of many environmental factors have been investigated, little information has been generated concerning psychological atmosphere, and the effects it may have on voice recognition accuracy. Within the domain of psychological atmosphere, one variable that may warrant special attention, especially in many military applications of voice recognition, is that of psychological stress, and in particular, emotional stress.

1.2 Problem

Although little work has been done to investigate the effects of emotional stress on VR, related studies indicate a definite need for further research. Armstrong and Poock (1981a) investigated the effects of mental loading on VR. They discovered a significant increase in recognition errors when subjects performed a concurrent mental task, compared to when no such task was performed. Armstrong (1980) found a similar increase in errors when subjects performed a concurrent motor task as compared to when they did not. Armstrong and Poock (1981b) found a significant increase in errors over time, similar to a vigilance decrement. The independent variables in these studies constitute specific types of stress. It is assumed that the increase in errors occurred because the users were speaking under conditions different from those under which they trained the VR system; conditions that altered their speech characteristics enough to increase errors.

Figure 1-1 presents a structure of some of the causes of stress. Clearly, items in one branch may induce stress in another branch, and the items are not exhaustive. The Armstrong (1980) and the Armstrong & Poock (1981a, 1981b) studies examined those branches of stress labeled "motor workload" under "Physical" stress, and "fatigue" and "processing demands" under "Unemotional Psychological" stress. The current research is intended to continue this line of investigation into the branch labeled "Emotional" stress.

Emotional stress may be viewed as a psychological variable described by an intensity continuum, similar to a continuous variable like pain. Just as the intensity of an identical pain stimulus (e.g., 5 volts to the forearm) may be perceived differently by two individuals, an identical emotional stressor (e.g., failing a driver's test) may be more severe for one person

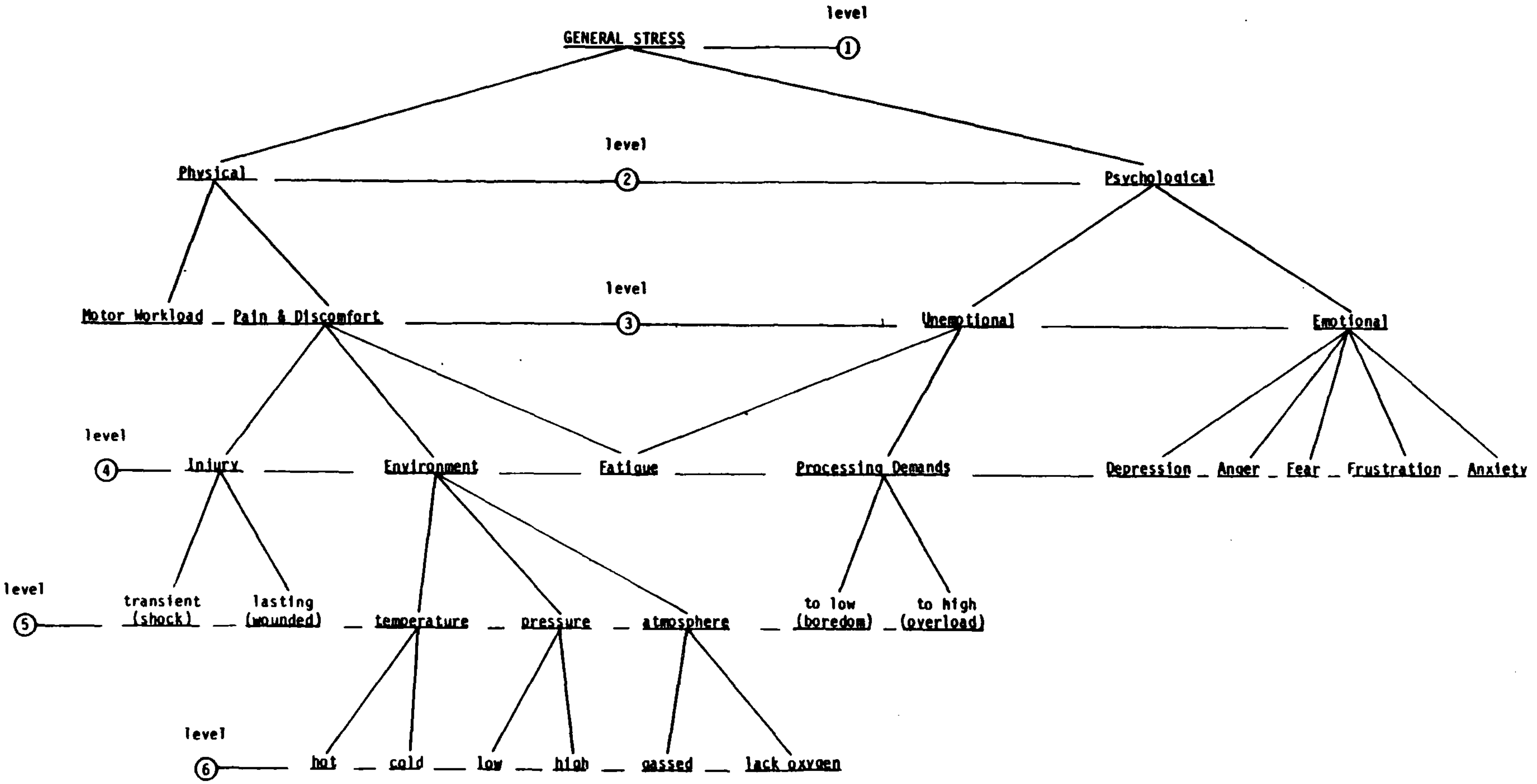


FIGURE 1-1
SOURCES OF STRESS

than another. However, even across individuals, some emotional stressors are more severe than others. For example, death of a spouse would clearly be a more intense emotional stressor than failing a driver's test (Holmes & Rahe, 1967). Further, there are different types of emotional stress (e.g., fear, frustration, anxiety, etc.) just as there are different types of pain (e.g., sharp, aching, burning, etc.).

Due to the prevalence of ethical and safety considerations involved with human research volunteers, the current experiment was aimed at investigating only a low intensity, short term state of emotional stress in the subjects.

A safe method of inducing a low intensity, short term emotional stress was explored by Glass & Singer (1972). Glass & Singer found that "exposure to unpredictable noise, in contrast to predictable noise, was followed by impaired task performance and lowered tolerance for post-noise frustrations" (p. 459). Furthermore, Glass & Singer found that "stress after effects" increase when the subject believes he is experiencing more noise than another subject under otherwise identical conditions. Glass & Singer indicated that exposing subjects to loud, intermittent, random noise, especially in the context described above, produced feelings of anxiety, frustration, and anger. Several other investigators have also used noise to produce stress in humans and other animals (see Selye, 1976). A method of inducing emotional stress similar to that used by Glass & Singer was implemented in the present study, for which a detailed description appears in the Procedure section.

In addition to an emotional stress condition (produced in part by noise), a perceptual-motor stress condition very similar to Armstrong (1980) was included in the experiment.

If emotional and perceptual-motor stress conditions do result in increased recognition errors, as was the case with the concurrent mental and motor tasks of Armstrong and Poock, the question arises as to whether or not there is a way to avoid such errors. Can the user be trained to speak consistently with his training, even under stress, or can the training be structured to accommodate inputs when the user is under stress? The fact that voice is now being used to measure stress (e.g., in lie detection) indicates that one has little control over the stressful dimensions of one's voice (Brenner, Shipp, Doherty, Morrissey, 1983). Therefore, research should concentrate on modifying the training format to accommodate inputs made under stress, rather than training operators to speak in a manner consistent with their original training. Armstrong and Poock (1981b) suggested that "training the recognizer under conditions similar to those that will be experienced during operation... would parallel Drennen's (1980) and Elster's (1981) research into training and operating a recognition system under various ambient noise levels." Drennen found that the recognizer performed best when trained under the same noise level present during testing. Perhaps the recognizer would also perform best if trained under the same motor and emotional stress levels that occur during testing.

Finally, if recognition errors increase under perceptual-motor stress and emotional stress, is the increase in errors under the separate stress conditions due to a single general stress response, or are the type of stress and corresponding errors caused by perceptual-motor stress qualitatively different from those caused by emotional stress?

In the investigation of the issues and questions raised above, a direct index of stress would be desirable. Questionnaires are often used to elicit subjects' ratings of the amount of stress they experienced. While this method is fairly direct, it is still filtered by the subjects' ability to answer accurately and willingness to answer honestly. Therefore, some additional measure of stress was sought. Unfortunately, some of the most widely accepted and reliable methods could not be implemented due to various practical limitations. Pupilometry, for example, is a well accepted measure of psychological stress, but was incompatible with the visual perceptual-motor condition (Brenner et al, 1983). Kalsbeek (1971) reported several studies in which sinus arrythmia and/or heart rate varied significantly with dynamic and static physical workload, mental workload, perceptual-motor workload, and emotional stress. Sinus arythmia is the irregularity of one's heart rate. Bonsper (1970) found a decrease in sinus arrythmia with increased information processing levels. Krol and Opmeer (1970) found sinus arrhythmia and heart rate varied significantly with different levels of perceptual-motor workload in a flight simulator. In an experiment with parachute jumpers (Krol and Opmeer, 1969) both sinus arrhythmia and heart rate differentiated between levels of emotional stress. It was decided, then, to employ sinus arrhythmia and heart rate as measures of emotional and perceptual motor stress.

1.3 Objectives

The specific objectives of this research were the following:

(1) To repeat a concurrent perceptual-motor task/voice input condition simular to Armstrong's (1980) to determine the reliability of his results.

(2) To introduce an emotional stress condition concurrent with voice input and examine the effects, if any, of emotional stress on recognition accuracy.

(3) To determine if training the recognizer under perceptual-motor stress and emotional stress conditions similar to those present during testing results in fewer recognition errors than those errors that result from differential training and testing conditions.

(4) To investigate the relationship between recognition errors produced by emotional stress and perceptual-motor stress.

(5) To explore sinus arrythmia and heart rate as physiological measures of stress.

2. METHOD

2.1 Subjects

Eighteen volunteers were recruited from the Naval Postgraduate School and Fleet Numerical Oceanography Center in Monterey, California. There were ten male officers, three female officers, two enlisted males, one enlisted female and two civilian females. Military volunteers represented the Navy (10), Air Force (3), Army (2), and Marines (1). One subject had four hours of previous experience with a voice recognition device and another subject had two hours prior experience with a VRD. The remaining sixteen subjects had never used VR equipment before.

2.2 Apparatus

Figure 2-1 provides a schematic of the apparatus. Most phases of the experiment took place with the subject inside an Industrial Acoustics Company, Inc. Controlled Acoustical Environments chamber. Also in the sound chamber was a Lafayette Instrument Co. Model 2203E Photoelectric Pursuit device used to induce operator perceptual-motor stress. The Pursuit device presented an approximately 2 cm by 2 cm square light target that traveled counter-clockwise around the circumference of a 26.5 cm diameter circle at 40 rpm. A light sensitive wand attached to the pursuit device was used to pursue and track the target. A Demco-Gray Gralab Universal Timer was wired to the pursuit device but was outside the sound chamber, allowing the experimenter to turn the target on and off.

An IBM programmable bell (basic school bell variety) was located inside the sound chamber for activation in the emotional stress condition. The bell produced noise at 100 db A. Outside the sound chamber was a remote button attached to a Lafayette Instruments Company, Inc. Model 52020 Eight Bank Program Timer. This program timer was wired to the bell inside the sound

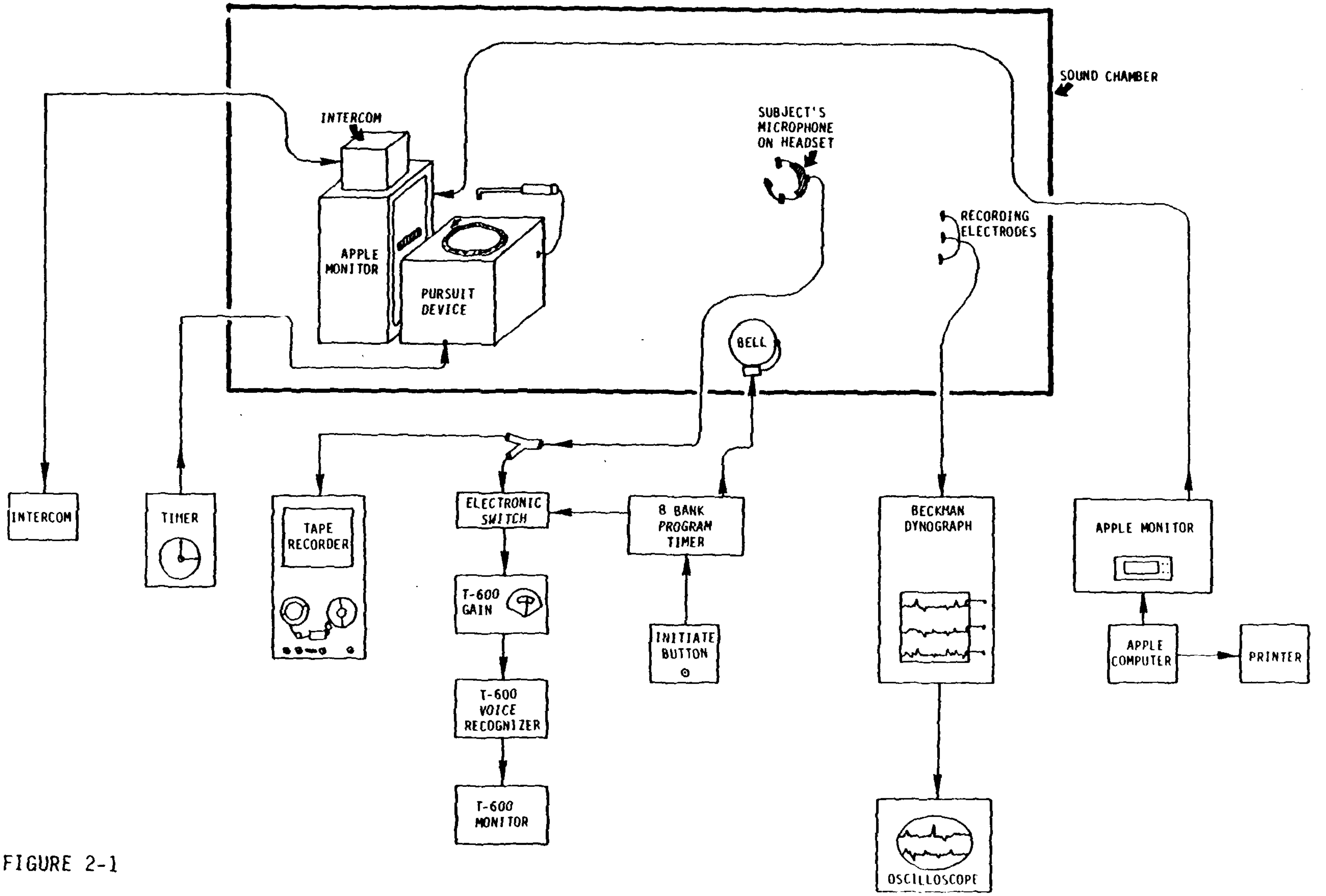


FIGURE 2-1

SCHEMATIC OF APPARATUS

chamber, and to an electronic switch (Sheridan Electronics Corp. model No. 4112-DAY-45-CN headset adapter MC-385-C) between the subjects' microphone and the voice recognizer. When the experimenter pushed the remote button to ring the bell, the program timer first opened the electronic switch preventing the voice recognizer from "hearing" anything, then rang the bell for .5 second, then paused an additional .1 second before closing the electronic switch. This system prevented the voice recognizer from erroneously accepting the sound of the bell as voice input during both training and testing.

A Threshold Technology model T600 voice recognition system was used in this study. The system was capable of storing 256 voice utterances of up to 2 seconds each. Thirty utterances were used in the present investigation. These utterances appear in Appendix A.

A Shure model SM10 "boom" microphone (mounted on the subject's headset) was used as the input device. This microphone is supplied as standard equipment with the T600. The microphone was wired to the T600 via the electronic switch described above, and to an Akai model 4000 DS MK II tape recorder so that both the T600 voice recognizer and the tape recorder received identical information (or "heard" the same thing).

Inside the sound chamber and directly behind the pursuit device was an Apple model CMI3L color monitor. The monitor faced the subject and the lower portion of its screen was obscured by the pursuit device. The prompts for the utterances appeared on the screen just above the back edge of the pursuit device. Therefore, in the perceptual-motor stress condition, the subject could briefly glance up to see the next prompt without losing track of the pursuit target.

The Apple monitor was wired to a video monitor outside the sound chamber in the experimenter's view. This monitor presented the same prompts as those presented to the subject, plus additional prompts to the experimenter in the lower portion of the screen.

An Apple II Plus computer was attached to the monitors. The Apple computer and original software generated the prompts to both monitors as well as some auditory prompts to the experimenter only. The computer was attached to a printer that provided hard copies of each prompt sequence.

A Beckman Type RM Dynograph Recorder, positioned outside the sound chamber, was used to record heartbeat and electrocardiogram rate. Both heartbeat and electrocardiogram rate were plotted simultaneously on stripcharts (and an attached Beckman Oscilloscope Type OE-10) via a Type 9806A A-C Coupler and a Type 9857 Cardiotachometer Coupler. Three Beckman recording electrodes were attached to the subjects with short term electrode disks and Beckman Electrode Electrolyte. Between uses, the electrodes were grouped together electrically at the post end and soaked in a 10% saline and distilled water solution at the electrode end to maintain the constancy of their electrical resistance.

One Fanon FI-3 intercom was located inside the sound chamber, and another outside to provide communications between the subject and the experimenter.

A Hewlett-Packard 9874A Digitizer attached to a Hewlett-Packard 9845A computer was used to reduce the stripchart information to numeric data.

2.3 Experimental Design

This experiment employed a 3x3x4 within subjects design. Three training conditions were crossed with the same three conditions under testing. The conditions were: No Stress, Perceptual-Motor Stress, and Emotional Stress. Each subject performed four trials under each test condition. A summary of the experimental design appears in Figure 2-2.

	TRAINING																																			
	No Stress												Perceptual-Motor Stress												Emotional Stress											
TESTING SUBJECT #	NO STRESS *				PERCEPTUAL MOTOR STRESS				EMOTIONAL STRESS				NO STRESS				PERCEPTUAL MOTOR STRESS *				EMOTIONAL STRESS				NO STRESS				PERCEPTUAL MOTOR STRESS				EMOTIONAL STRESS *			
TRIAL # →	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1																																				
2																																				
3																																				
4																																				
5																																				
6																																				
7																																				
8																																				
9																																				
10																																				
11																																				
12																																				
13																																				
14																																				
15																																				
16																																				
17																																				
18																																				

* colums in which "live" error data was recorded

FIGURE 2-2
SUMMARY OF EXPERIMENTAL DESIGN

2.4 Procedure

2.4.1 Counterbalancing and Scheduling. All 18 subjects experienced each of three training conditions and each of three test conditions. The training condition sequence was fully counter balanced with three subjects in each of six possible sequences. The test condition was also fully counter balanced with three subjects in each of six possible sequences. Training condition sequence was partially counter balanced with test condition sequence so that each training condition sequence was followed by three different test condition sequences:

Counterbalancing Technique	Training Condition Sequence	Test Condition Sequence
Same train/test sequence	N, Pm, E	N, Pm, E
Reversed train/test sequence	N, Pm, E	E, Pm, N
Middle Exchange train/test sequence	N, Pm, E	N, E, Pm

N = no stress   Pm = Perceptual-Motor Stress   E = Emotional Stress

Subjects were required to make six appointments over a two week period, with a limit of one appointment in a given day. The first three appointments were for training conditions. The first took about one hour, and the second and third appointments took about 40 minutes. The last three appointments were for test conditions and each took about 25 minutes.

2.4.2 Introduction. At the onset of each subject's first session, the subject was asked to read the INSTRUCTIONS AND INTRODUCTORY REMARKS (see Appendix B). The experimenter then demonstrated the procedure for attaching the three recording electrodes and their placement. One electrode was attached near the middle of the sternum and one on each side of the subject's waist just above the hips. For a few subjects, this triangulation did not yield measurable ECG, and one of the side electrodes was alternatively placed further up their side, nearer the underarm. The subject's electrodes were then attached to the Dynograph outside the sound chamber and the experimenter recalibrated the machine until heartbeat and heartrate were being measured and recorded accurately. During this time the subject was asked to read the VOICE RECOGNIZER VOCABULARY TRAINING information (see Appendix C). After the Dynograph was operating properly and the subject had finished reading, the experimenter reiterated the written instructions in detail, then elicited and answered questions from the subjects. The subject then practiced training an utterance on the T600.

2.4.3 Training

2.4.3.1 General Training Format. The term "training," as used in discussions of voice recognition studies, refers to the process by which the speaker makes known to the recognizer the characteristics of his particular speech patterns for all the utterances he will be using. For the T600, this training procedure consists of entering 10 passes of each utterance (10x30 or 300 utterances per training condition, per subject) into the voice recognizer. The recognizer automatically averages the ten passes of each utterance into a single template, enters these templates into its "memory," and matches any subsequent utterances (in testing) with the templates in memory. Ideally, these subsequent utterances are matched with the template for the same utterance in memory, resulting in correct response output on a CRT. In cases where a match is not possible a nonrecognition or rejection occurs, signified by a "beep" from the recognizer. In effect, the machine is saying "I don't understand that utterance--please say it again." Occasionally, however, the recognizer makes an incorrect match. In this case, an incorrect response is output on the CRT, constituting a "misrecognition." Thus, two types of errors are possible: nonrecognitions (or rejections) and misrecognitions (or misinterpretations) of an utterance.

Once the subjects understood the training format in general, they were re-connected to the Dynograph from inside the sound chamber and issued instructions pertaining to the particular training condition.

2.4.3.2 No Stress, Perceptual-Motor Stress, and Emotional Stress Training Conditions. Subjects were given the INSTRUCTIONS FOR NORMAL AND MOTOR CONDITIONS (see Appendix D) for the No Stress and Perceptual-Motor Stress Training, or the INSTRUCTIONS FOR FEEDBACK TRAINING CONDITION (see Appendix E) for Emotional Stress training; and asked to read them while the experimenter checked Dynograph and audio recording levels outside the sound chamber. In the Emotional Stress training condition, subjects were led to believe that the bell would ring once for each "bad" voice input they made to the recognizer. A "bad" input was described as an input that did not contribute to better recognition accuracy than could be expected from the template that had already been formed from the previous training inputs for that utterance. Subjects were told that the determination of a good or "bad" input was based on the T600's standard algorithms. Furthermore, subjects were informed that various feedback schedules were under investigation, therefore this feedback (the bell ringing) could occur immediately after the "bad" input, or up to three inputs later, making it impossible for them to directly determine which inputs were "bad." Finally, each subject was told that although this feedback schedule might seem complex, not to be concerned, because most subjects make only a few

21

"bad" inputs, and thus, the bell will only ring a few times. In actuality, no distinction was ever made concerning good or bad inputs, and the bell was always rung after 70 of the 300 training inputs for each subject. The location of the 70 rings was randomly generated for each subject.

The purpose of this charade was to induce emotional stress in the subjects. Telling the subjects that the bell rang as a result of their voice inputs implied that they were responsible for the bell, yet there was little they thought they could do (in actuality there was nothing they could do) to control the bell. Responsibility without control typically leads to frustration. To enhance the effect even further, the bell per se was quite loud and irritating, and rang unpredictably. These facits of inducing emotional stress parallel those mentioned by Glass & Singer (1972). Also, each subject heard 70 rings after being told that most other subjects make only a few "bad" inputs. The implication is apparent to each subject that other subjects are not being exposed to nearly as much noise, another ingredient that induces emotional stress according to Glass & Singer (1972). Finally, the simple impression of doing poorly, especially compared to most other subjects, was expected to enhance emotional stress.

To attribute any difference between training conditions to type of stress, it was important to hold the timing or rhythm of voice inputs constant across training conditions. Otherwise, a difference in the emotional stress training condition could be due to the interruptions in the training rhythm caused by the bell ringing, rather than emotional stress. Therefore, in the Perceptual-Motor Stress and No Stress training conditions, a "STAND BY" message was displayed for an equivalent duration and number of times as the bell rang in the Emotional Stress condition. These "STAND BY" messages were randomly generated in the same fashion as the bell ringings. Subjects were instructed not to make any voice inputs when the "STAND BY" message was on the screen, since they were told in these conditions timing was one of the variables under investigation.

In the Perceptual-Motor Stress training condition subjects were instructed to track the target as accurately as possible. Subjects were told the pursuit task should be given equal priority with making voice inputs, and that their time on target would be recorded.

Once the subjects were given the above information (for the appropriate condition), they were asked to sit quietly in the sound chamber for five minutes before the training session started.

During this time, outside the sound chamber, the experimenter initiated the Apple program that randomized the presentation order of the 30 utterances. When the five minute period was over the actual training began. In the Perceptual-Motor Stress condition, the subjects began tracking on the pursuit device at this point. The prompt for the first utterance appeared on the experimenter's monitor along with numeric prompts indicating when the bell or "STAND BY" message should be activated. The experimenter keyed the appropriate utterance into the T-600 to prepare the voice recognizer to receive training passes for that utterance. Then the utterance prompt appeared on the subject's monitor in the sound chamber. The subject would make voice inputs of the utterance displayed on the monitor until interrupted by either the bell ringing or the "STAND BY" message (depending on the training condition). When the bell stopped ringing or the utterance prompt reappeared on the monitor, the subject would continue entering training passes again until interrupted again, or until training of that utterance was complete. At no time was the bell ringing allowed to be interpreted (by the VR system) as part of the voice pattern training. When training for one utterance was completed the subject awaited the display of a new utterance prompt on the monitor, at which time the process was repeated until all 30 utterances had been trained.

At the termination of each training session, each subject had created a file of 30 utterance templates which were recorded (in digital form) on tape cartridges.

2.4.4 Testing

2.4.4.1 Live Testing

Each subject was scheduled to make four passes through the 30 utterances under each of the three test stress conditions. At the onset of each session the subject first attached his electrodes as described previously and the experimenter re-calibrated the Dynograph to insure an accurate measurement and recording. The T-600 cartridge containing the trained utterances for the current subject under the corresponding stress condition was loaded into the voice recognizer.

In the Emotional Stress test condition the subjects were told that the bell would ring immediately after any voice input that was not accurately recognized. The subjects were further informed that in this condition only, their recognition accuracy scores would be rank ordered with the other 17 subjects, and posted by their name on the outside of the sound chamber door. As an example, the experimenter presented a paper (which had been posted on the door throughout the entire experiment) that appeared to be the rank ordering of accuracy scores from a previous experiment (see Appendix F). The experimenter pointed out that most scores were above 90%, that the lowest was a 73%, and that in general, this range was representitive of the performances in the current experiment.

In actuality, the bell was activated after an average of one in every three (40 of 120) voice inputs, regardless of whether or not the input utterance was correctly recognized.

In this manner, it seemed obvious to each subject that they were producing far more recognition errors and experiencing far more noise than most other subjects. As in the Emotional Stress training condition, this contrived feedback, coupled with the aversive nature of the bell per se, was intended to induce low level emotional stress in the subjects concurrent with their voice inputs to the recognizer.

In the Perceptual-Motor Stress test condition the subjects performed the same pursuit task as they had done in the Perceptual-Motor Stress training condition.

In the No Stress test condition, the subjects simply input each utterance as it's prompt appeared on the monitor.

In the Perceptual-Motor and No Stress test conditions, "STAND BY" messages were not necessary to control timing of inputs since, as in the Emotional Stress test conditions, timing was controlled by the prompt-presentation rate of the Apple program. Utterance prompts were presented once every five seconds. Each presentation sequence of the 30 utterances was randomized by the Apple, as were the signals to the experimenter to activate the bell. To the subject, the beginning and end of the 4 trials was transparent, however, the Apple program insured that each trial contained exactly 10 randomly located bell signals.

During the test sessions the experimenter tape recorded all voice inputs (at 7-1/2 fps); at the same time, the experimenter recorded on paper the recognitions, nonrecognitions, and misrecognitions of the subjects live voice inputs to the T-600.

After each test session the subjects filled out a POST SESSION QUESTIONNARIE (see Appendix G).

2.4.4.2 Taped Testing

After the training conditions were completed each subject had produced three training files, each stored on a cartridge that could be loaded into the T-600 memory. The three training files were created under: 1) No Stress, 2) Perceptual-Motor Stress, and 3) Emotional Stress. During testing, only one of the training files could be accepted at a time. Therefore, to find out which training file produced the highest number of recognitions when tested, (for example, under the No Stress test conditions), required three individual tests:

No Stress Test condition to No Stress Training file

No Stress Test condition to Perceptual-Motor Stress Training file

No Stress Test condition to Emotional Stress Training file

Further, three more tests would be required to discover which training file produced the highest recognition rate for utterances made under Perceptual-Motor Stress test conditions, and 3 more tests for utterances made under Emotional Stress test conditions.

Without tape recording, each subject would have to undergo each of the three test conditions three times, for a total of nine test sessions.

However, by tape recording each subject under each of the three test conditions, the No Stress test condition tape could be played back to each of the three training files;

No Stress Training file

No Stress test conditioning tape — Perceptual-Motor Stress Training file

Emotional Stress Training file

and the same could be done with the tapes of the Perceptual-Motor Stress test conditions and the Emotional Stress test condition.

There are 2 distinct advantages to using tape recorded test conditions: 1) the subjects had to complete only three stress test conditions rather than nine; 2) any differences between the recognition rate obtained by inputting utterances from one test condition to the 3 different training files would have to be due to differences in the training files, since the recorded test utterances were always identical. Had a subject actually undergone the Emotional Stress condition (or any of the conditions for that matter) three times, once to each training file, it seems likely that his stress level would vary with the successive test occasions, introducing a confounding that was avoided by tape recording.

The first step was to insure that the T-600 responded the same way to tape recorded input as it did to live input. Although the investigator's pretests indicated that the T-600 did respond to taped voices the same as to live voices, more extensive testing was done with the actual audio tapes generated in the live test phase of the experiment. Each of the 54 test condition audio tapes (18 subjects x 3 test conditions each) was played directly into the T-600, under the same conditions that prevailed during live testing. For example, the audio tape of Subject 1 in the No Stress test condition was played to the T-600 with the No Stress training file for Subject 1 loaded into the T-600's memory. The T-600's responses (correct recognitions, nonrecognitions, and misrecognitions) were noted and compared to the responses noted during live testing. This procedure confirmed the investigator's pre-test results by indicating that the T-600 did in fact respond to taped voice input in a manner consistent with live voice inputs.

Once the reliability of the taped testing method was verified, each subject's voice tapes were played to each of the training files to obtain the balance of the error data.

2.5 Independent and Dependent Variables

The independent variables in this study were training condition: No Stress, Perceptual-Motor Stress, and Emotional Stress. The dependent variables were nonrecognitions, misrecognitions, total errors (which was a linear combination of nonrecognitions and misrecognitions), sinus arrhythmia heart rate and subjective stress.

3. RESULTS

3.1 Overview

For error data all analyses of variance procedures and post hoc range tests were performed using the arcsin transformation of raw data to stabilize the variance of the error terms (Neter and Wasserman, 1974). The mean error rates that appear in the tables and figures are untransformed. All a posteriori tests for significance between pairs of means were performed using the Scheffe procedures described in Bruning and Kintz (1977), and Hays (1963, p. 465). Subjects source of variance (not represented in ANOVA summary tables) account for 17 df.

As defined earlier, nonrecognitions and misrecognitions by the voice recognition system may have distinctly different implications in an applied setting. In a weapons deployment activity, for example, it would be far more desirable for the system to respond to an input error by nonrecognition (a "beep"), where the speaker is told to repeat or correct the input than for the system to misinterpret the input and to carry out some incorrect (and perhaps critical) command in error. Thus, it was considered essential to determine the effects of the independent variables on nonrecognitions and misrecognitions separately, as well as on total number of errors.

Section 3.2 presents the data on total number of errors. Section 3.3 presents the results of analyses done on nonrecognitions, while Section 3.4 presents the results of analyses done on misrecognitions.

The remaining sections will present stress data from the test phase. Section 3.5 presents the analyses done on sinus arrhythmia, section 3.6 presents the analyses done on heart rate, and section 3.7 presents the analyses done on the POST SESSION SURVEYS.

3.2 Total Errors

Table 3-1 presents the analysis of variance for total errors (nonrecognitions + misrecognitions). A significant main effect of trials was found (F=3.102, P<.05) and there was a significant interaction of training condition with test condition (F=8,238, P<.001). No other main effects or interactions reached statistical significance. Mean total errors (in percent) for training condition by test condition are shown in Table 3-2. The main effect of trials and the interaction of training condition with test condition are portrayed graphically in Figure 3-1 and 3-2, respectively.

With regard to the main effect of trials, a Scheffe test for significance between pairs of means detected no significant differences between any two trials. This result is not surprising considering the conservative nature of the Scheffe test and the borderline significance of trials in the analysis of variance (see Myers, 1972).

With regard to the interaction of training condition with test condition, Scheffe tests were performed to detect simple effects between test conditions within training conditions. The following effects were significant at the .05 level:

Under No Stress Training - No Stress Testing versus Perceptual-Motor Stress Testing (for No Stress Testing versus Emotional Stress Testing P<.06)
Under Perceptual-Motor Training - Perceptual-Motor Testing versus No Stress Testing
Perceptual-Motor Testing versus Emotional Stress Testing
Under Emotional Stress Training - Emotional Stress Testing versus Perceptual-Motor Stress Testing

TABLE 3-1

ANALYSIS OF VARIANCE SUMMARY TABLE FOR TOTAL ERRORS

SOURCE	df	MS	F
TRAINING CONDITION (A)	2	.01686	.043
ERROR	34	.39175	
TRIALS (T)	3	.19740	3.102*
ERROR	51	.06363	
AT	6	.01293	.767
ERROR	102	.01686	
TEST CONDITION (B)	2	.06412	.494
ERROR	34	.12985	
AB	4	.34918	8.238**
ERROR	68	.04238	
AT	6	.04356	1.109
ERROR	102	.03275	
ATB	12	.01972	.830
ERROR	204	.02375	

*P<.05
**P<.001

TABLE 3-2

MEAN TOTAL ERRORS (IN PERCENT) FOR TRAINING CONDITION BY TEST CONDITION

		TRAINING CONDITION			
		NO STRESS	PERCEPTUAL-MOTOR STRESS	EMOTIONAL STRESS	$\bar{X}$ TRAINING CONDITION
TEST CONDITION	NO STRESS	2.546	4.491	4.120	3.750
	PERCEPTUAL-MOTOR STRESS	4.630	2.778	5.000	3.982
	EMOTIONAL STRESS	4.074	4.676	3.750	4.290
	$\bar{X}$ TEST CONDITION	3.719	4.136	4.167	4.01 Grand $\bar{X}$

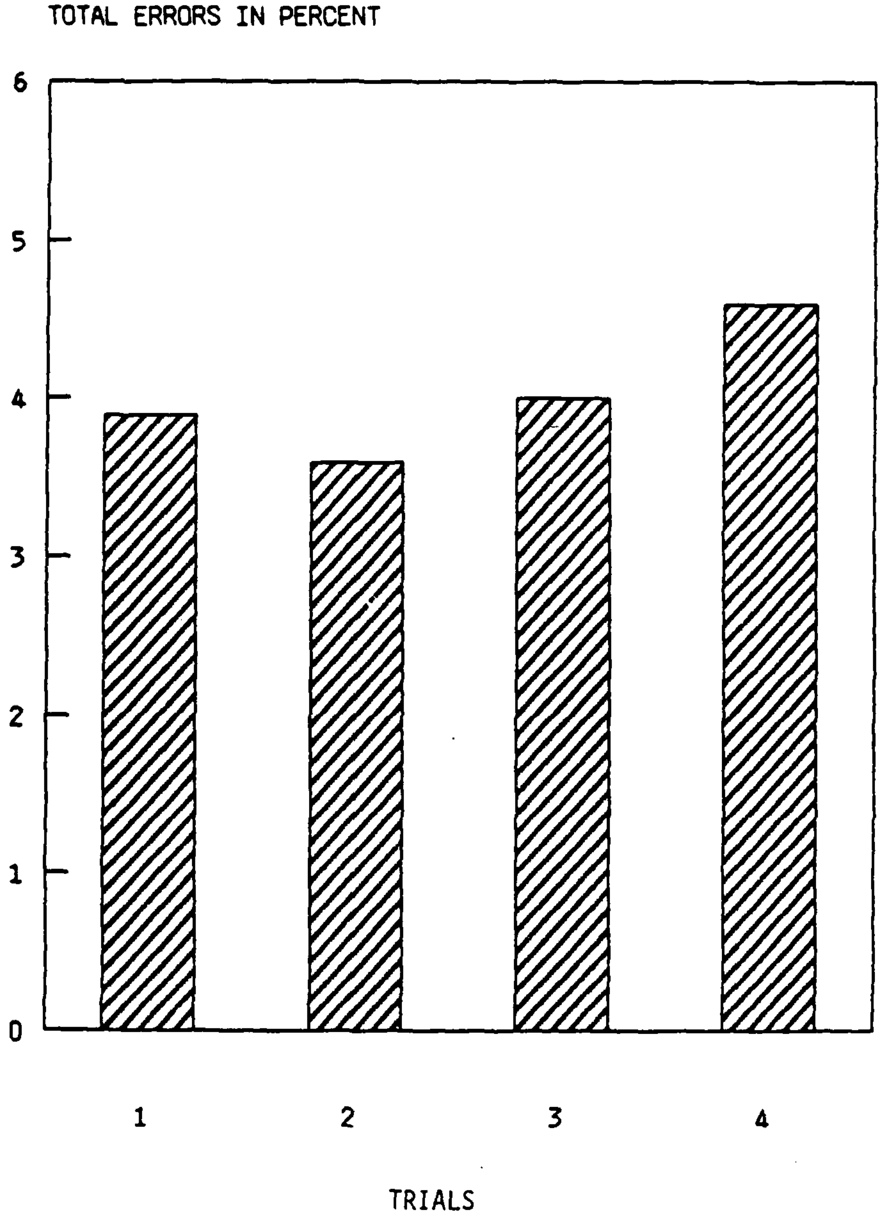
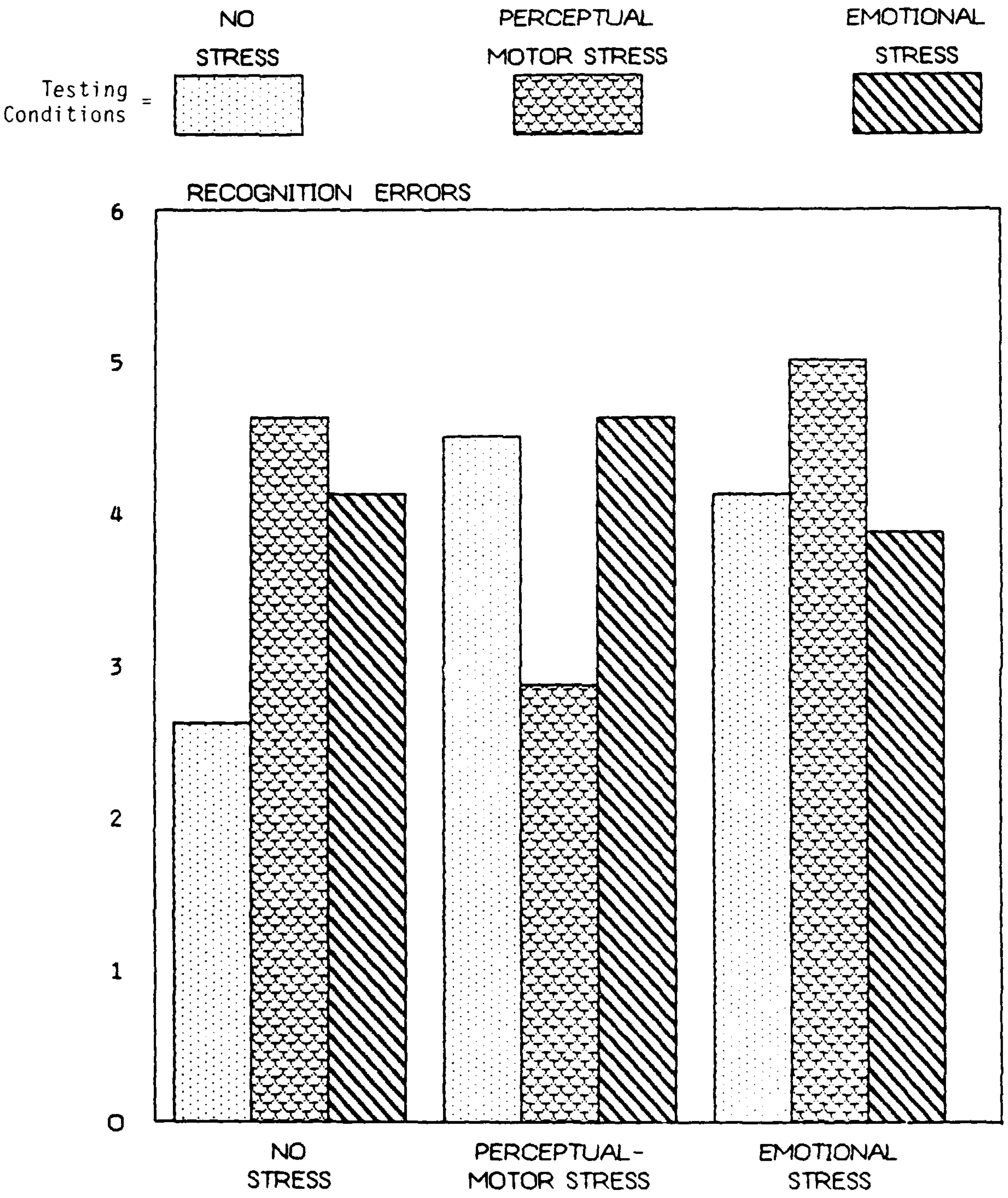


FIGURE 3-1
MEAN TOTAL ERRORS (IN PERCENT) BY TRIALS



TRAINING CONDITION
ERRORS ARE IN PERCENT

FIGURE 3-2.
MEAN TOTAL ERRORS (IN PERCENT) INTERACTION OF TRAINING CONDITION WITH TEST CONDITION

In general, the significant interaction and simple effects just described indicate that using the recognizer under the same stress condition as it was trained under will produce significantly fewer errors than errors produced using the recognizer under stress conditions different from those under which it was trained. Further, the greatest incompatability seems to exist between Perceptual-Motor Stress and both No Stress and Emotional Stress, while the least incompatability exists between No Stress and Emotional Stress.

3.3 Nonrecognitions

Table 3-3 presents the analysis of variance for nonrecognitions. A significant interaction of training condition with test condition was found (F=4.150, P<.005). No other interactions or main effects reached statistical significance. Mean nonrecognitions (in percent) for training condition by test condition are shown in Table 3-4, and the interaction is portrayed graphically in Figure 3-3.

Scheffe tests were performed to detect simple effects between test conditions within training conditions. The only significant difference between means occurred under the No Stress training condition between No Stress testing and Perceptual-Motor Stress testing. Still, the relationships between nonrecognition means closely resembled those of total errors. However, nonrecognitions accounted for only 25% of the total errors with misrecognitions contributing the balance of 75%. In previous experiments the reverse was true, nonrecognitions outweighed misrecognitions by at least 3 to 1. (Martin, 1983; Poock, Martin, and Roland, 1983; Poock et al, 1983; Poock, Schwalm, and Roland, 1981) Probable reasons for this reversal will be discussed in the next section.

TABLE 3-3

ANALYSIS OF VARIANCE SUMMARY TABLE FOR NONRECOGNITIONS

SOURCE	df	MS	F
TRAINING CONDITION (A)	2	.00950	.124
ERROR	34	.07647	
TRIALS (T)	3	.03565	.871
ERROR	51	.03510	
AT	6	.00704	.420
ERROR	102	.01675	
TEST CONDITION (B)	2	.01591	.213
ERROR	34	.07465	
AB	4	.11728	4.150*
ERROR	68	.02826	
BT	6	.00810	.323
ERROR	102	.02505	
BTA	12	.00926	.616
ERROR	204	.01504	

*P<.005

TABLE 3-4

MEAN NONRECOGNITIONS (IN PERCENT)
FOR TRAINING CONDITION BY TEST CONDITION

		TRAINING CONDITION			
		NO STRESS	PERCEPTUAL-MOTOR STRESS	EMOTIONAL STRESS	$\bar{X}$ TRAINING CONDITION
TEST CONDITION	NO STRESS	.417	1.250	1.157	.941
	PERCEPTUAL-MOTOR STRESS	1.343	.556	1.157	1.019
	EMOTIONAL STRESS	1.111	1.019	.509	.880
	$\bar{X}$ TEST CONDITION	.957	.941	.941	Grand $\bar{X}$ .947

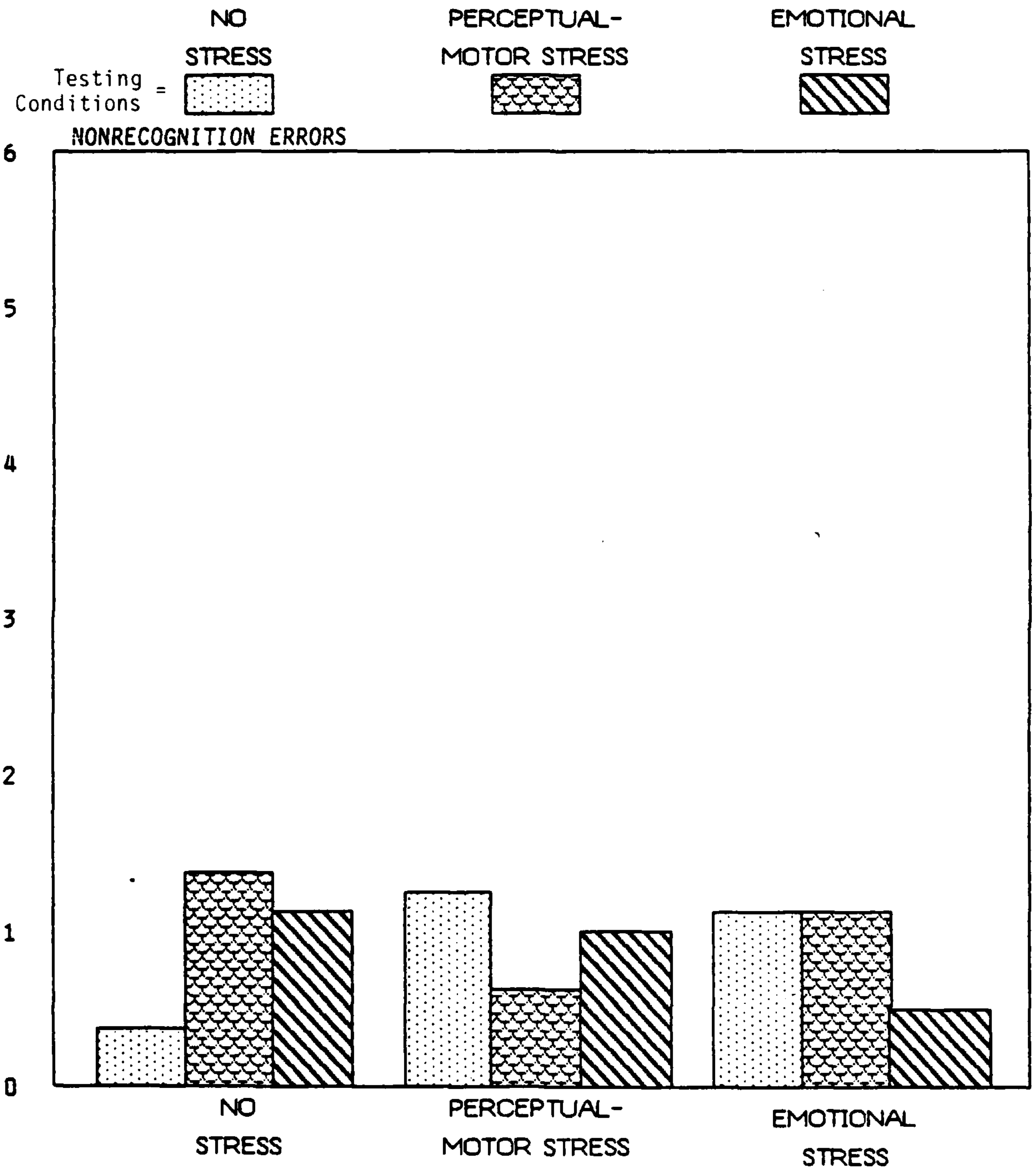


FIGURE 3-3
MEAN NONRECOGNITIONS (IN PERCENT) INTERACTION
FOR TRAINING CONDITION BY TEST CONDITION.

3.4 Misrecognitions

Table 3-5 presents the analysis of variance summary table for misrecognitions. A significant main effect of trials was found (F=2.895, P<.05) and there was a significant interaction of training condition with test condition (F=4.326, P<.005). No other main effects or interactions reached statistical significance. Mean misrecognitions (in percent) for training condition by test condition are shown in table 3-6. The main effect of trials and the interaction of training condition with test condition are portrayed graphically in Figure 3-4 and 3-5, respectively.

With regard to the main effect of trials, a Scheffe test for significance between pairs of means detected no significant differences between any two trials as with total errors, this result is not surprising since the main effect was of borderline significance in the analysis of variance and the per-comparison alpha employed by the Scheffe test is quite low.

Further Scheffe tests were performed with regard to the interaction, to detect simple effects between test conditions within training conditions. The only significant difference between means occurred under the Perceptual-Motor Stress Training condition; between Perceptual-Motor Stress testing and Emotional Stress Testing. However, the relationships between means are generally the same as those obtained for total errors, indicating that the best recognition accuracy was obtained when subjects tested the VRD under the same stress conditions as they trained it under.

Misrecognitions outnumbered nonrecognitions and accounted for 75% of the total errors, constituting a reversal of previous findings as discussed earlier. The utterances used in the present research were selected from a vocabulary of 250 utterances used by Poock (1981). The size of the vocabulary was restricted to 30 utterances in the current research to avoid lengthy test sessions per subject. However, in an attempt to avoid floor

TABLE 3-5

ANALYSIS OF VARIANCE SUMMARY TABLE FOR MISRECOGNITIONS

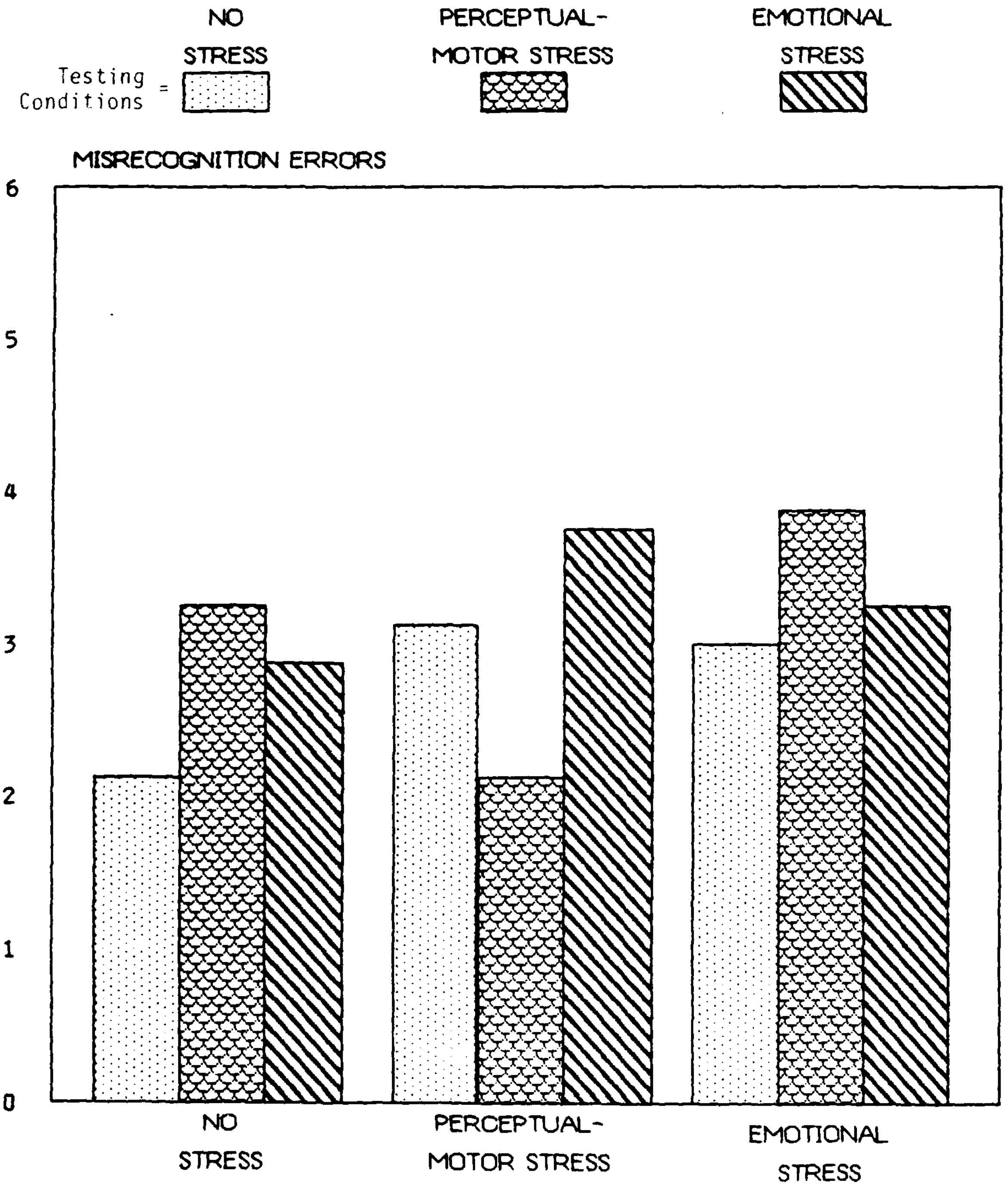
SOURCE	df	MS	F
TRAINING CONDITION (A)	2	.01772	.052
ERROR	34	.33765	
TRIALS (T)	3	.15692	2.895*
ERROR	51	.05420	
TA	6	.01356	.747
ERROR	102	.01815	
TEST CONDITION (B)	2	.10113	1.299
ERROR	34	.07782	
AB	4	.17312	4.326**
ERROR	68	.04002	
BT	6	.04884	1.462
ERROR	102	.03340	
ATB	12	.01039	.429
ERROR	204	.02421	

*P<.05
**P<.005

TABLE 3-6

MEAN MISRECOGNITIONS (IN PERCENT) FOR TRAINING CONDITION BY TEST CONDITION

		TRAINING CONDITION			
		NO STRESS	PERCEPTUAL-MOTOR STRESS	EMOTIONAL STRESS	$\bar{X}$ TRAINING CONDITION
TEST CONDITION	NO STRESS	2.13	3.24	2.96	2.79
	PERCEPTUAL-MOTOR STRESS	3.29	2.22	3.84	3.04
	EMOTIONAL STRESS	2.96	3.66	3.24	3.35
	$\bar{X}$ TEST CONDITION	2.78	3.12	3.29	Grand $\bar{X}$ 3.06



TRAINING CONDITION ERRORS ARE IN PERCENT

FIGURE 3-4.
MEAN MISRECOGNITIONS (IN PERCENT) FOR INTERACTION OF TRAINING CONDITION WITH TEST CONDITION

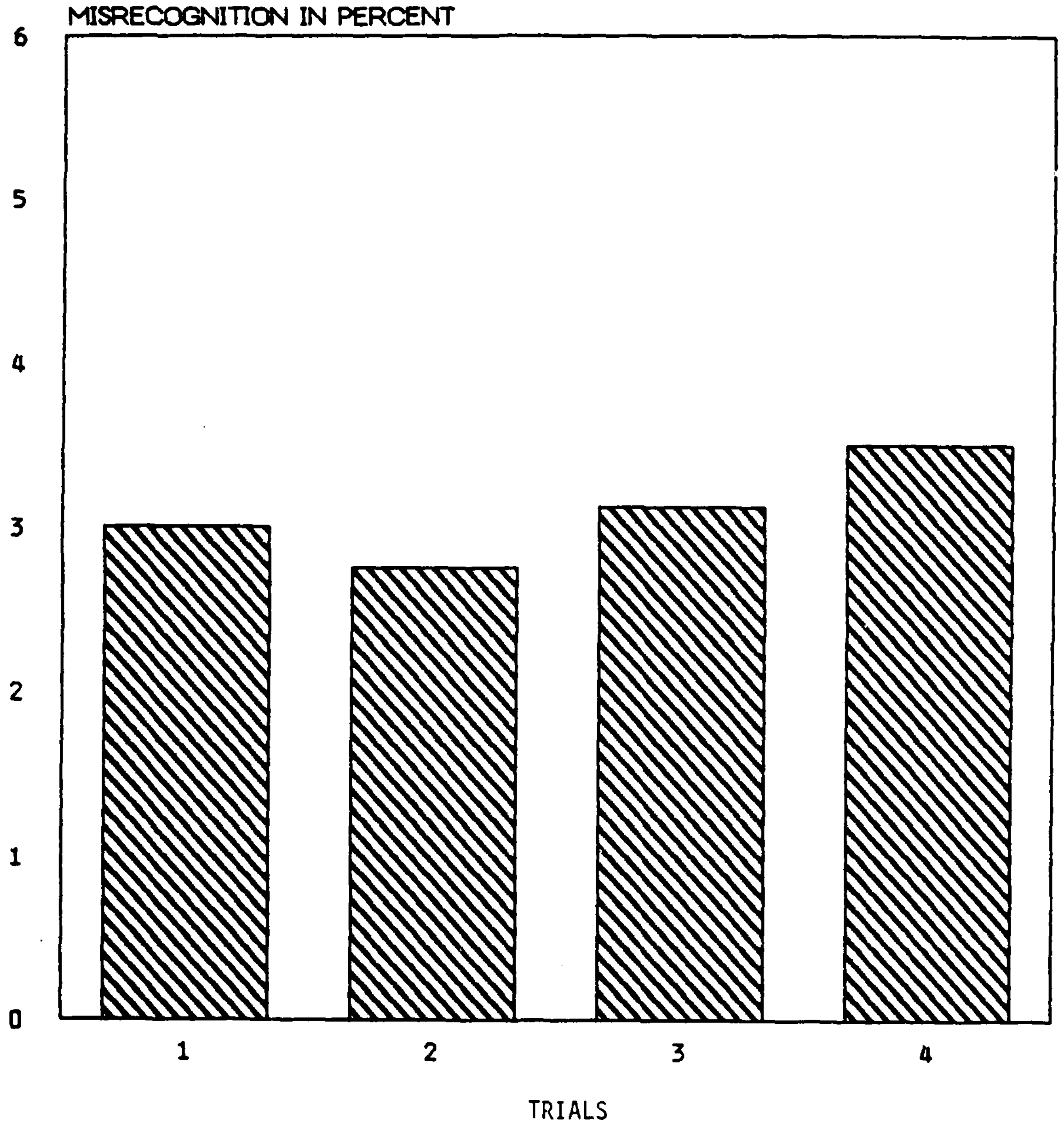


FIGURE 3-5
MEAN MISRECOGNITIONS (IN PERCENT) BY TRIALS

43

effects in the error data, a sub set of Poock's vocabulary was chosen that contained utterances with high error rates, primarily "confusions", which are misrecognitions. This is probably the main factor contributing to the abnormally high misrecognition rate in the present study. Another factor may be a difference between the training method used in the current research and the training method used in previous studies:

In a typical training session, after all utterances have been initially trained, the subject recites each utterance to the recognizer to see if all utterances are recognized at least two out of three times. Those utterances that do not meet this criterion are then retrained until at least two out of three passes are correctly recognized. However, this methodology was incompatible with the contrived feedback phases of the current study, and was therefore omitted completely to allow consistent training criteria across the stress training conditions. It is conceivable, but speculative, that training to a two out of three criterion would have filtered out a greater number of misrecognitions than nonrecognitions, resulting in a typical high nonrecognition to low misrecognition ratio.

3.5 Sinus Arrhythmia

Sinus arrhythmia is the irreegularity of the heart beat. It is normal for healthy people to have a certain degree of irregularity (or arrhythmia) in their heart beat, especially during relaxation. Typically, under stress, the heart beat attains better rhythm or regularity, representing a reduction in sinus arrthymia. Test condition means for sinus arrhythmia were observed in the expected direction, high (associated with low stress) in the No Stress test condition and low (associated with high stress) in the Perceptual-Motor and Emotional Stress conditions. However, this main effect did not reach statistical significance in the analysis of variance. The test condition means for sinus arrhythmia are presented numerically and graphically in Figure 3-6.

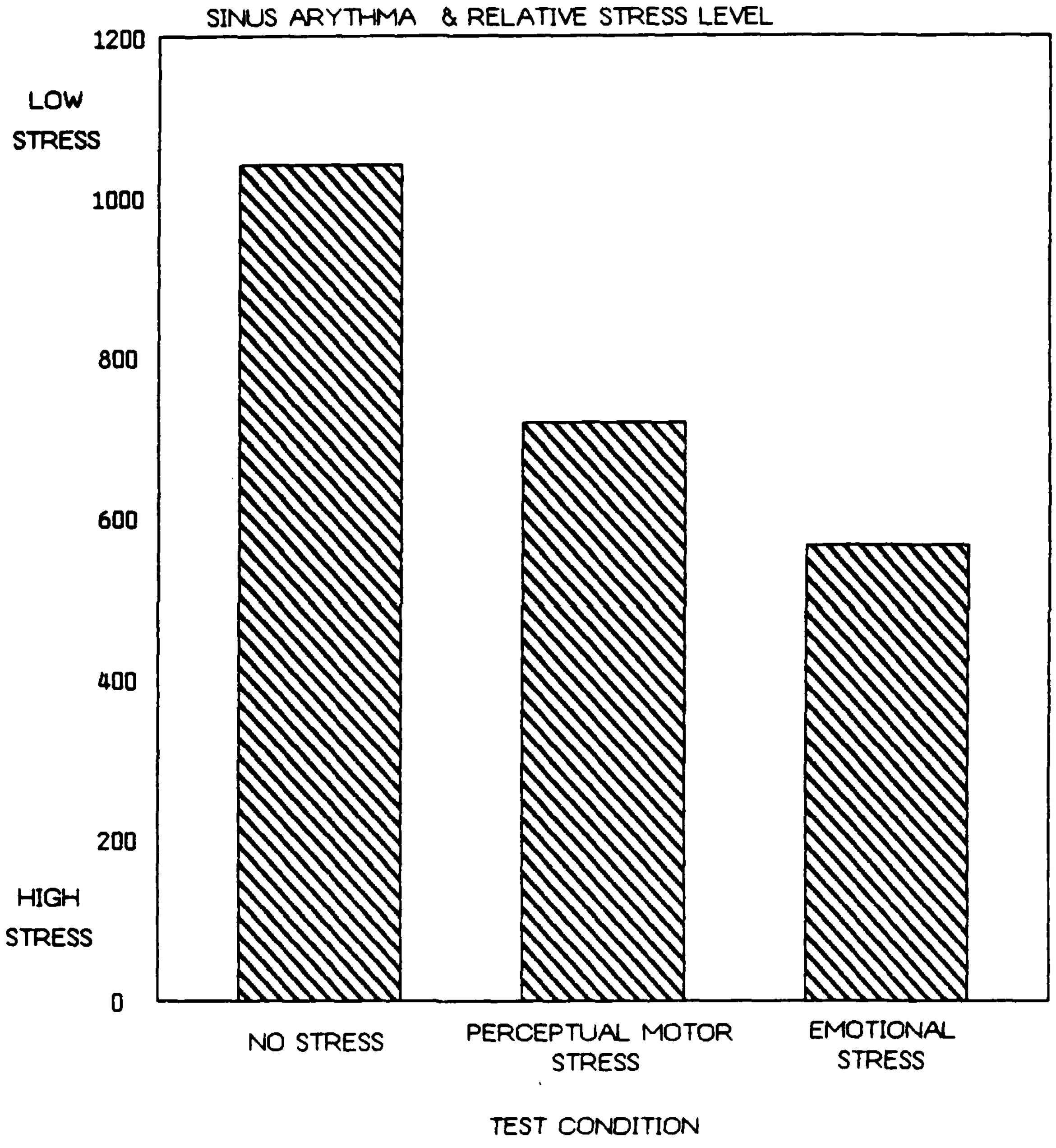


FIGURE 3-6.
SINUS ARRTHYMIA BY TEST CONDITION

3.6 Heartrate

An analysis of variance on heartrate in the test conditions yielded significant main effects for trials (F=5.159, P<.005) and test conditions (F=4.256, P<.025). The analysis of variance summary totals for heartrate is presented in Table 3-7. Mean heartrate for trials by conditions are presented in Table 3-8 and Figure 3-7.

A Scheffe test indicated that heartrate in trial four was significantly higher than in trial one and trial two. The increase in heartrate under the Perceptual-Motor Stress condition was the primary contributor to this trials effect. Interestingly, a similar increase of less magnitude occurred under the No Stress condition. The reason for this is unknown.

A Scheffe test on the test condition means showed that heartrate under the Perceptual-Motor Stress condition was significantly higher than heartrate under the Emotional Stress condition. This finding reinforces the distinction between qualitatively different types of stress, especially in light of the fact the Perceptual-Motor Stress elevated heartrate, (compared to No Stress) and Emotional Stress depressed heartrate (compared to No Stress).

3.7 Subjective Stress

Freidman Tests were conducted on ranks to each of the five survey questions/dimensions (and ties were treated as described by Bradley, 1976). These analyses showed that in four of the five dimensions, subjects ranked the three test conditions significantly differently (at the .01 level). Subjects responses to "Enjoyment" did not vary significantly over the 3 test conditions. Mean rankings for dimension by test condition appear in Figure 3-8.

TABLE3-7
ANALYSIS OF VARIANCE SUMMARY TABLE
FOR HEART RATE

SOURCE	df	MS	F
TRIALS (T)	3	81.042	5.159*
ERROR	51	15.710	
TEST CONDITIONS (C)	2	1206.532	4.256**
ERROR	34	283.470	
CT	6	16.773	1.203
ERROR	102	13.948	

*p < .005
**p < .025

TABLE 3-8

MEAN HEARTRATE FOR TEST CONDITON BY TRIALS

		TRIAL				
		1	2	3	4	$\bar{X}$ TEST CONDITION
TEST CONDITION	NO STRESS	78.17	78.67	78.97	81.31	79.28
	PERCEPTUAL-MOTOR STRESS	81.81	82.36	84.31	86.14	83.65
	EMOTIONAL STRESS	75.97	74.47	75.39	76.06	75.47
	$\bar{X}$ TRIALS	78.65	78.50	79.56	81.17	GRAND $\bar{X}$ 79.47

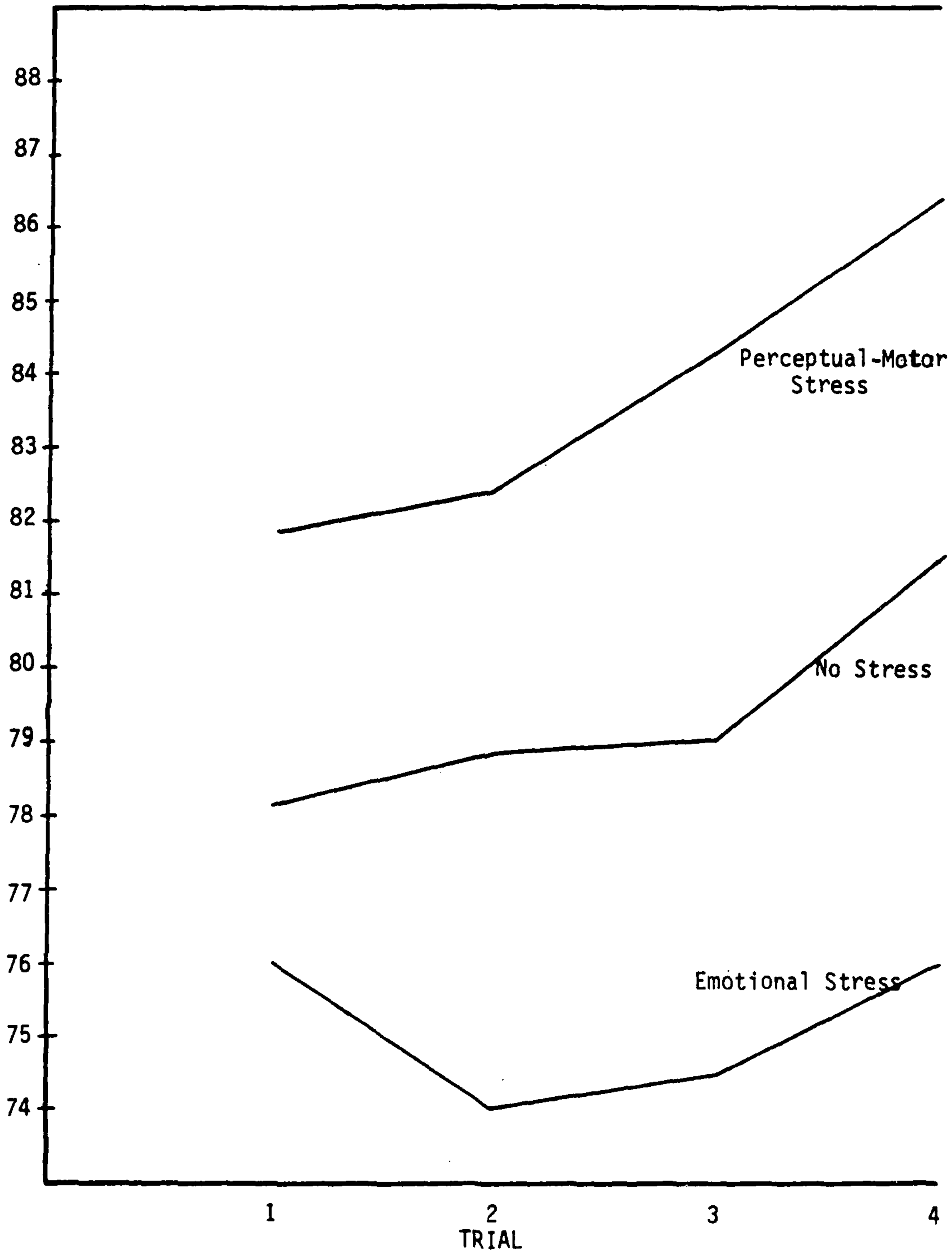


FIGURE 3-7.
MEAN HEART RATE FOR TRIALS BY TEST CONDITION

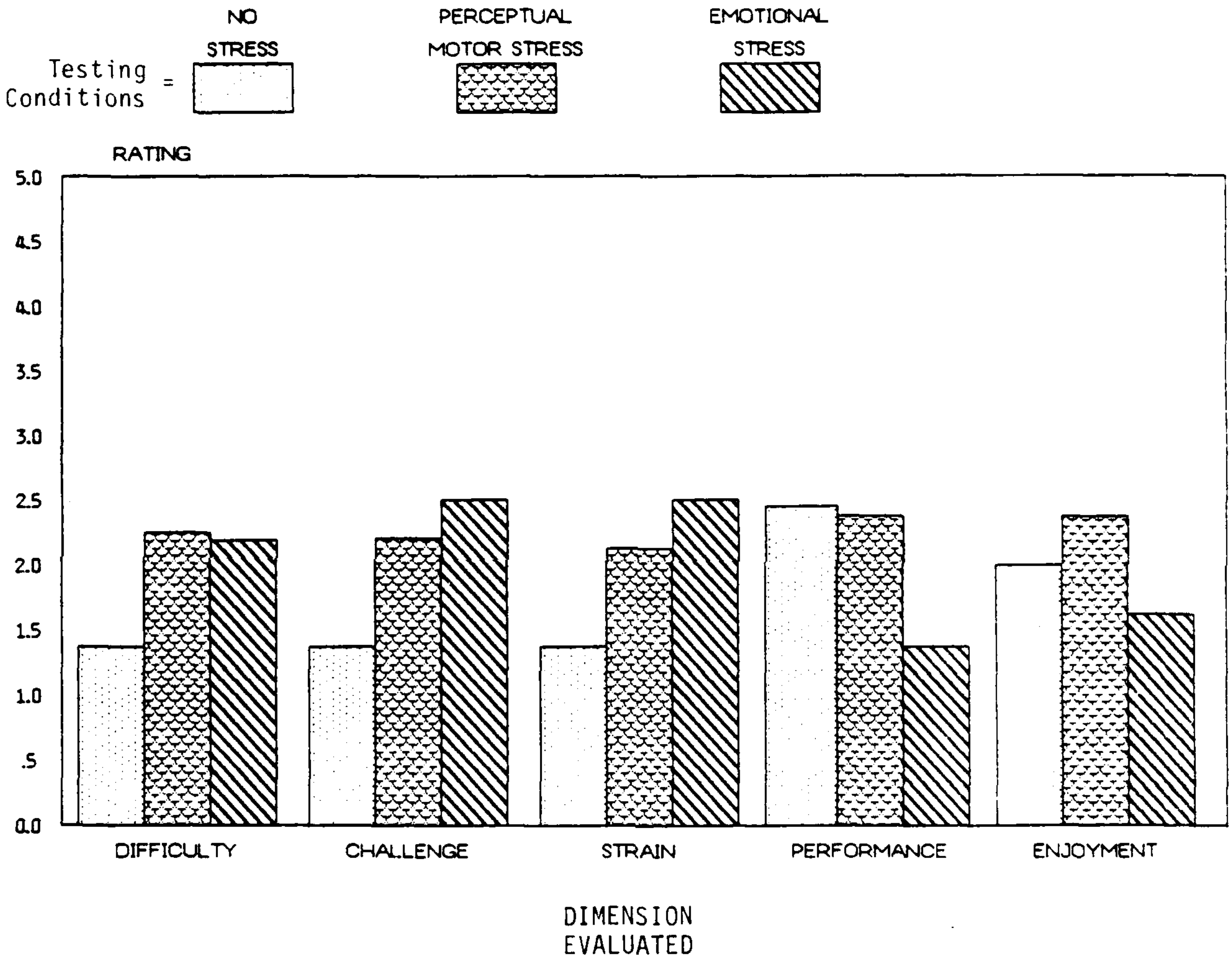


FIGURE 3-8
MEAN RATINGS FOR TEST CONDITION BY DIMENSION

Pearson correlations between difficulty, challenge, and strain, were high and positive ($r=.79$, $P<.001$) and will be collectively referred to henceforth as subjective stress. Mean responses all remained below the intensity midpoint on the subjective stress continuum, a result that corresponds well with the experimental intent of inducing only a low level of stress in our subjects. However, as indicated by the Freidman Tests, subjective stress was significantly lower in the No Stress Test condition than in the Perceptual-Motor Stress and Emotional Stress Conditions.

Subjective Stress had a lower negative correlation ($r=.27$, $P<.005$) with perceived performance, and subjects believed they performed significantly poorer under the Emotional Stress condition than under the No Stress and Perceptual-Motor Stress conditions, even though they received no feedback whatsoever under the later conditions!

4. DISCUSSION

This section will discuss the current findings with regard to the objectives put forth earlier in this paper.

4.1 Replication of Effects of Perceptual-Motor Stress Concurrent with Voice Input

Armstrong had subjects train a VRD under normal (no stress) conditions. He then had the subjects test the recognizer under the same normal conditions, and while performing a pursuit task (perceptual-motor stress condition). There were significantly more errors under the perceptual-motor stress condition than under the normal condition. The current research confirms Armstrong's findings. After training the VR system under No Stress, 2.5% errors resulted under No Stress testing, while 4.6% errors resulted under Perceptual-Motor Stress testing. This 2% increase is significant, and corresponds to the increase obtained by Armstrong for a similar vocabulary.

4.2 Emotional Stress

To study the effects of voice input under emotional stress required a safe and effective method of inducing low level emotional stress in our subjects. To meet this end, subjects were exposed to loud, aversive noise, and various misinformation regarding their "poor" performance. In surveys completed after each test condition, subjects indicated that while they experienced relatively low levels of subjective stress (strain, difficulty, and challenge) they experienced significantly greater stress under the Emotional Stress condition than under the No Stress condition. At the end of the experiment subjects were informed of the actual nature of the Emotional Stress condition and of those portions of the condition in which they had been intentionally misled. At this point it was common for subjects to offer informal, unsolicited, statements regarding the effectiveness of our charade in the Emotional Stress condition. Typically, subjects expressed feelings of considerable frustration and some anger with the relentless bell, including a few subjects who also said they had been suspicious as to whether the bell ringing was actually associated with input errors on a one to one basis. These subjective measures clearly support the effectiveness of our Emotional Stress condition.

Less clearly, but still supporting of the effectiveness of our Emotional Stress condition, were the physiological measures of stress. Sinus arrhythmia under the Emotional Stress condition was only 54% of sinus arrhythmia under the No Stress condition. Although the direction of this finding was consistent with an interpretation of greater stress in the Emotional Stress condition, the value was not statistically significant.

Heart rate under the Emotional Stress condition was somewhat subdued, but did not vary significantly from heart rate under the No Stress condition. The sinus arrhythmia and heart rate findings may reflect the low level nature of the Emotional Stress condition. Comparable findings were obtained by Brenner et al (1983) between two levels of psychological stress. In one level subjects were supposed to remember and repeat two-number strings, (virtually no stress) while in the second level they tried to remember and repeat seven-number strings, representing "increasing degrees of anxiety and stress associated with increased memory load" (p.4). Two physiological voice stress measures indicated non-significant (P>.05) but higher levels of stress under the seven-number strings condition, resulting in "a tendency towards identifying acoustic correlations of stress but with a sufficient variablility in the experimental data to prohibit establishing statistical reliability" (p.10).

Brenner et. al, then analyzed taped voices of pilots under no stress communications (routing flight info) and high stress communications (emergency information prior to unsuccessful landings). The same two voice measures were performed on these tapes as were performed in the memory task. In this case however, the differences between the no stress and high stress conditions were significant (P<.05 or better). Brenner's et al observations are brought forth here to support to contention that the effects of emotional stress lie on an intensity continuum, and that the results of our Emotional Stress condition are a reflection of sampling from the low end of that continuum.

The error data reinforce this standpoint. Emotional Stress testing of No Stress training files resulted in more errors (4.1%) than No Stress testing of No Stress training files (2.5%). The difference, however, was of borderline significance (P<.06).

4.3 Same Versus Differential Training/Testing

Having determined that Perceptual-Motor and Emotional Stress testing of No Stress training files (Differential) results in more errors than No Stress testing of No Stress training files (same), we turn to a new question: Can the increase in errors associated with Perceptual-Motor and Emotional Stress testing be counteracted by including Perceptual-Motor or Emotional Stress in the training file? In general, the answer is yes. Perceptual-Motor Stress testing of Perceptual-Motor Stress training files resulted in about the same number of errors (2.8%) as did No Stress testing of No Stress training files (2.%), and compared to 4.6% errors for Perceptual-Motor Stress testing of No Stress training files.

Emotional Stress testing of Emotional Stress training files only reduced errors to 3.75% compared to 4.1% for Emotional Stress testing of No Stress training files. While errors were always lower under same training/testing conditions than differential training/testing conditions, it appears that the effect emotional stress has on the voice is not as easily counteracted as the effect of perceptual-motor stress. This issue will be discussed in more detail in the next section.

4.4 Relationship Between Errors Produced Under Perceptual-Motor Stress and Emotional Stress

A question posed earlier asked if the errors produced by perceptual-motor stress and emotional stress were a result of some underlying general stress response in the voice, or two fairly distinct stress responses in the voice. If the effect of perceptual-motor stress in the voice was the same as the effect of emotional stress, then differential training/testing between the two should result in an equal number of errors as same training/testing within either. However, such was not the case. In testing Perceptual-Motor Stress training files, Emotional Stress testing resulted in significantly more errors than Perceptual-Motor testing. Similarly, in testing Emotional Stress training files, Perceptual-Motor Stress testing produced significantly more errors than Emotional Stress testing. We also obtained a significant difference in heart rate for subjects during Perceptual-Motor Stress versus Emotional Stress testing. Collectively, these results lend clear support to the idea that perceptual-motor stress and emotional stress have qualitatively different effects on the voice. (For a physiological viewpoint, see Brenner et al, 1983.)

4.5 Sinus Arrhythmia and Heart Rate

While sinus arrhythmia and heart rate offered some expected trends and significant differences, these measures did not seem to be sensitive enough to reflect changes induced by the Emotional Stress condition. Conversely, our manipulations were not strong enough to affect, for example, the sinus arrhythmia index. Krol and Opmeer (1969) obtained significant differences in sinus arrhythmia between levels of emotional stress. However, they were probably sampling from the high end of the emotional stress intensity continuum eluded to previously, in that their measurements were made on first time parachute jumpers, 2 minutes before a jump. With this in mind we would not discard sinus arrhythmia as an objective measure of emotional stress, but suggest reserving it for high to low emotional stress comparisons, and levels of information processing comparisons. Similar conclusions were drawn for heartrate, which is probably most useful in measuring motor stress.

5. CONCLUSION

Previous research has shown that various factors in the voice recognition system environment affect recognition accuracy, especially when those factors are inconsistent between training and subsequent use of the system. Drennen (1980) and Elster (1980) found an increase in errors due to using the VR system under different noise levels than those present during training. Other investigators found similar effects due to psychological factors such as information processing load (Armstrong and Poock, 1981a), perceptual-motor load (Armstrong, 1980), and task duration (Armstrong and Poock, 1981b). The present research has shown further evidence of the importance of the psychological environment in VR systems training and use. Three stress conditions were examined; No Stress, Perceptual-Motor Stress, and Emotional Stress. Recognition errors typically increased when the system was used in a stress condition other than the stress condition in which training occurred. However, if training and use occurred under the same stress condition, errors returned to a nominal level, regardless of the condition. It appears then, that human factors, specifically those in the psychological environment, such as frustration, anger, attention allocation and fatigue may parallel the effects of environmental factors like noise (as it affects the microphone), with regard to training and subsequent use of VR systems.

These results suggest that VR system training should be carefully constructed to include as many human factors (at the appropriate levels) as are foreseeable in actual VR system use.

In some situations, certain factors are likely to change levels during VR systems use. For example, aircraft controllers may experience several levels of emotional stress in a single shift. Training the system under no emotional stress will result in poorer performance under emotional stress. Training the system under emotional stress will result in poorer

57

performance when the operator is not under emotional stress. The interpretation of the current research then, would obviously prescribe including voice samples from as many emotional stress levels as possible in the training file, to achieve optimum performance. This procedure is not without cost, however.

Attempts to include a high resolution of samples, for each of several pertinent factors (noise, frustration, mental fatigue, boredom, etc.) could quickly use up available computer memory, in addition to being tedious, time consuming, and difficult to quantify. Clearly, these considerations must be weighed against the type and criticality of errors.

In the worst-case example of the present study (Emotional Stress Training/Perceptual-Motor Stress Testing) recognition accuracy was still 95%, compared to an average improvement to 97% recognition accuracy when training/testing were under the same condition. In this light the VRD performed quite well under our training and testing cross-manipulations. Our main concern is with the fact that changing stress levels between training and testing resulted in statistically significant increases in errors, with low intensity stress levels. The potential for more practically significant increases in errors under high stress is not yet known, and is suggested as a topic for future research.

6. REFERENCES

Armstrong, J.W. The Effects of Concurrent Motor Tasking on Performance of a Voice Recognition System. Naval Postgraduate School Master's Thesis, September 1980.

Armstrong, J.W. and Poock, G.K. Effect of Operatoı Mental Loading on Voice Recognition System Performance. Naval Postgraduate School Technical Report NPS55-81-016, August 1981(a).

Armstrong, J.W. and Poock, G.K. Effect of Task Duration on Voice Recognition System Performance. Naval Postgraduate School Technical Report NPS-55-81-017, September 1981(b).

Bonsper, D.E. The Effect of Increased Information Processig on Sinus Arrhythmia and Heartbeat. Naval Postgraduate School Master's Thesis, September 1970.

Bradley, J.V. Probability; Decision; Statistics. Prentice Hall, Englewood Cliffs, N.J., 1976.

Brenner, M., Shipp, T., Doherty, E.T., Morrissey, P. Voice Measures of Psychological Stress-Laboratory and Field Data. Invited Paper-Iowa Conference on Physiology and Biophysics of Voice, May 1983.

Bruning, J.L. and Kintz, B.L. Computational Handbook of Statistics (2nd ed.), Glenview, Illinois: Scott, Foresman and Co., 1977.

Cohen, S.I. Central Nervous System Functioning in Altered Sensory Environments, Psychological Stress. Appley and Trumbull, eds. Meredith Publishing Company, New York, 1967.

Drennen, T.G. Voice Technology in Attack/Fighter Aircraft, Voice-Interactive Systems: Applications & Payoffs. May 1980.

Elster, R.S. The Effects of Certain Background Noises on the Performance of a Voice Recognition System. Naval Postgraduate School Technical Report NPS54-80-010, September 1980.

Glass, D.C. and Singer, J.E. Behavioral Aftereffects of Unpredictable and Uncontrollable Quosive Events. American Scientist 60:457-465, 1972.

Hays, W.L. Statistics for Psychologists. Hott, Rinehart and Winston, Inc., 1966.

Holmes, T.H. and Rahe, R.H. The Social Readjustments Scale. Journal of Psychosomatic Research, 11:213-218, 1967.

Kalsbeek, J.W.M. Sinus Arrhythmia and the Dual Task Method in Measuring Mental Load, in Measurement of Men at Work. Eds: Singleton, W.T., Fox, J.G., Whitfield, D., Taylor & Francis, Ltd., 1971.

Krol, J.P. and Opmeer, C.H.J.M. Psychological Parameters During Different Phases of Flight in a Flight Simulator. Report TNO-Laboratory of Ergonomic Psychology, 1970.

Krol, J.P. and Opmeer, C.H.J.M. Sinus Arrhythmia, Heart Rate and Respiratory Rate with Parachute Jumpers. Report TNO-Laboratory of Ergonomic Psychology, 1969.

Martin, B.J. Current Capabilities in Voice Recognition Technology. Experiment IV: Use of Voice Recognition Equipment by Naive Versus Practiced Speakers. (PFTR-1121-83-4) Man-Machine Systems Design Laboratory. Naval Postgraduate School, Monterey, CA., April 1983.

Myers, J.L. Fundamentals of Experimental Design (2nd ed.). Allyn and Bacon, Inc., 1972.

Neter, J. and Wasserman, W. Applied Linear Statistical Models, Homewood, Illinois: Richard D. Irwin, Inc., 1974.

Nye, J.M. Human Factors Analysis of Speech Recognition Systems, Speech Technology, Vol. 1, 2, April 1982.

Poock, G.K. Voice Recognition Information Sheet. Naval Postgraduate School.

Poock, G.K., Martin, B.J. and Roland, E.F. The Effect of Feedback to Users of Voice Recognition Equipment. Naval Postgraduate School Technical Report NPS55-83-003, February 1983.

Poock, G.K., Schwalm, N.D., and Roland, E.G. Toward Speaker Independence Voice Recognition. Paper presented at the 26th Annual Human Factors Society Meeting, Seattle, WA., October 1981.

Poock, G.K., Schwalm, N.D., Martin, B.J. and Roland, E.F. Trying for Speaker Independence in the Use of Dependent Voice Recognition Equipment. Naval Postgraduate School Technical Report NPS55-82-032, December 1982.

Schwalm, N.D., Poock, G.K., and Roland, E.F. Using Voice Recognition Technology: A Psychological Perspective. Submitted for publication to Speech Technology, February 1983.

Selye, H. Stress in Health and Disease. Butterworth (Publishers) Inc.: Boston, MA., 1976.

APPENDIX A

TRACK ENEMY
VIETNAM
KILO
UNIFORM
BUSINESS MEETING
AVAILABLE
EIGHT
PROCEED
SYSTEM INTEGRATION
POPPA
EFFICIENT TRANSMISSION
ALTITUDE
COURSE
ENEMY DETECTION
NINE
COMMAND
COMMAND AND CONTROL
INTERACTIVE
RELOCATE
LIMA
MOVE IT RIGHT
CONTINUOUS SPEECH
ADVISORY

```
----------------------
HOTEL
----------------------
BINGO
----------------------
CONTINUOUS
----------------------
SPEECH RECOGNITION
----------------------
INDIA
----------------------
KOREA
----------------------
OSCAR
----------------------
```

APPENDIX B

APPENDIX B

INSTRUCTIONS AND INTRODUCTORY REMARKS

First a reminder about what to expect in the experiment:

(1) Your voice will be recorded during some phases of the experiment.

(2) Three recording electrodes will be attached to your torso during nearly all phases of the experiment, and your heart beat and rate will be recorded at these times.

(3) During some phases of the experiment you will be exposed to a loud bell (about 100 db.).

(4) You will be informed that your name and scores for some phases of the experiment will be rank ordered and posted.

If you object to any of these aspects of the experiment (or any other aspects not mentioned here) please notify the experimenter immediately.

This experiment involves analysis of a combined human operator/voice recognition equipment system under various conditions. The actual experiment will be carried out in a sound-proof booth and subject-experimenter communication during the actual experiment will be via the booth intercom system.

Please carry out the experiment exactly as directed and do not discuss your performance with anyone other than the experimenter as inappropriate subject prior knowledge could invalidate the results.

APPENDIX C

APPENDIX C

VOICE RECOGNIZER VOCABULARY TRAINING

The 30 word vocabulary being used with the voice recognizer in this experiment is attached to these instructions. You will be required to repeat each word of this vocabulary ten times to train the recognizer to recognize your particular vocalizations of each word. To facilitate recognition by the voice recognizer, you should include in the ten repetitions as many as possible of the different ways you might say the word in normal speech; for example, use different intonations and emphasis, and small variations in volume.

Please observe the following guidelines while inputting voice data to the recognizer both during training and later during the actual experiment.

(1) Speak each word crisply and quickly but do not overpronounce; for example, words ending in "t" - delete final "t" if more natural.

(2) Also, do not leave a period of silence within an utterance or the recognizer will mistake it for two separate utterances.

(3) Microphone location is very important and should be kept constant throughout the experiment, i.e., adjust it if it gets out of place. The experimenter will initially demonstrate correct microphone placement.

(4) Whenever a word is on the screen, you should avoid coughing, clearing your throat, or asking questions, since these sounds would be taken as training passes of the word on the screen.

APPENDIX D

APPENDIX D

INSTRUCTIONS FOR NORMAL AND MOTOR CONDITIONS

In these conditions you will not get any feedback concerning your performance, and the parameters that determine performance are different from in the feedback condition, so good performers in the feedback condition are sometimes poor performers in the motor and normal conditions and vice-versa. In the motor condition we want to see how a physical task affects voice recognition accuracy. In the motor and normal conditions, we want to examine the affect of timing on training. Therefore, a STAND BY signal will occasionally appear on your screen in the place of the current word. When this happens you should stop making training inputs until the training word re-appears. Otherwise, just continue making inputs until the word disappears or the experimenter tells you to stop.

APPENDIX E

APPENDIX E

INSTRUCTIONS FOR FEEDBACK TRAINING CONDITION

In the feedback training condition you will get feedback concerning the quality of your verbal training inputs to the voice recognizer. Your feedback will be either the silence or ringing of a bell after each pass. Silence means everything is OK, so continue with the next training pass. Ringing means that one of the last four passes was no good (the recognizer has determined that it will not contribute to better recognition accuracy). When the bell rings, you should wait until it stops ringing, then pause a second before continuing with the next pass.

We are using this type of feedback based on information from past experiments:

(1) People who get feedback can monitor and improve their inputs, and therefore get higher recognition accuracy than people who do not get feedback.

(2) People who get delayed feedback (generalized feedback) do better than people who get immediate (specific) feedback.

You will get delayed feedback, and the bell is fairly loud, but most subjects will get "rung" relatively few times.

APPENDIX F

RESULTS OF EXPERIMENT 1-83 VRD

The following are voice acceptability/accuracy scores from the feedback phase of the experiment, in rank order.

	NAME	% ACCURACY
1	Jorgensen, Ron	98
2	Morgens, David	97
3	Chapman, Allan	95
4	deLaTorre, Mike	95
5	Reddert, Tom	92
6	Price, Scott	91
7	Cooke, Kathy	90
8	Maxwell, Roger	86
9	Schvaneveldt, Ken	81
10	Hibbert, Vincent	80
11	Reese, Scott	77
12	Erickson, Mike	73

Thank you for your participation.

APPENDIX G

POST SESSION QUESTIONNAIRE

NAME ______________________ SUBJECT # _____ TRAINING TEST

NORMAL MOTOR FEEDBACK

PLEASE ANSWER THE FOLLOWING QUESTIONS TRUTHFULLY AND AS ACCURATELY AS POSSIBLE.

IF FOR SOME QUESTIONS YOU FEEL YOU NEED MORE INFORMATION TO BASE YOUR ANSWER ON, THEN YOU MAY JUST GUESS.
CIRCLE A NUMBER FOR EACH ITEM.

1) HOW DIFFICULT DID YOU PERCEIVE THE SESSION TO BE?

0-------1-------2-------3-------4-------5

NOT DIFFICULT AT ALL EXTREMELY DIFFICULT

2) HOW MUCH DID YOU ENJOY THE SESSION?

0-------1-------2-------3-------4-------5

DID NOT ENJOY IT AT ALL ENJOYED IT VERY MUCH

3) HOW CHALLENGING WAS THE SESSION?

0-------1-------2-------3-------4-------5

NOT CHALLENGING AT ALL EXTREMELY CHALLENGING

4) HOW MUCH STRAIN DID YOU EXPERIENCE DURING THE SESSION?

0-------1-------2-------3-------4-------5

NO STRAIN AT ALL VERY MUCH STRAIN

5) HOW WOULD YOU RATE YOUR PERFORMANCE (ABILITY TO MAKE VOICE INPUTS ACCEPTABLE TO THE VOICE RECOGNIZER) IN THE SESSION?

0-------1-------2-------3-------4-------5

VERY POOR PERFORMANCE EXCELLENT PERFORMANCE

DISTRIBUTION LIST

	NO. OF COPIES
Library, Code 0142 Naval Postgraduate School Monterey, CA 93943	2
Director of Research Administration Code 012A Naval Postgraduate School Monterey, CA 93943	1
Library, Code 55 Naval Postgraduate School Monterey, CA 93943	1
Defense Technical Information Center Cameron Station Alexandria, VA 22314	2
Professor G. K. Poock Code 55Pk Naval Postgraduate School Monterey, CA 93943	150