

AD-A130 155

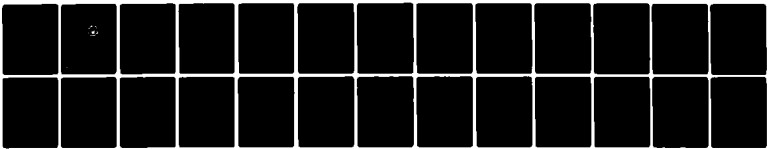
VOICE RECOGNITION PERFORMANCE WITH NAIVE VERSUS  
PRACTICED SPEAKERS(U) NAVAL POSTGRADUATE SCHOOL  
MONTEREY CA G K POOCK ET AL. JUN 83 NPSS55-83-016

1/1

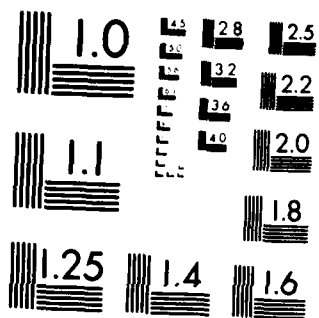
UNCLASSIFIED

F/G 17/2

NL



END  
DATE  
FILMED  
7 83  
DTIC

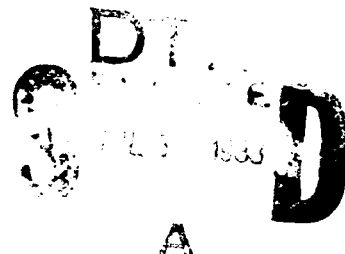


MICROCOPY RESOLUTION TEST CHART  
NATIONAL BUREAU OF STANDARDS 1963-A

NPS55-83-016

# NAVAL POSTGRADUATE SCHOOL

Monterey, California



ADA130155

DTIC FILE COPY

VOICE RECOGNITION PERFORMANCE WITH  
NAIVE VERSUS PRACTICED SPEAKERS

by

Gary K. Poock  
B. Jay Martin

June 1983

Approved for public release; distribution unlimited

Prepared for:  
Naval Electronics Systems Command  
Code 613  
Washington, D.C. 20363

83 07 6 016

NAVAL POSTGRADUATE SCHOOL  
Monterey, California

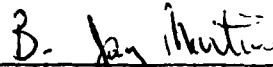
Rear Admiral J. J. Ekelund  
Superintendent

D. A. Schrady  
Provost

This research was supported and funded by the Naval Electronic Systems  
Command.




Gary K. Poock, Professor  
Department of Operations Research




B. Jay Martin  
Perceptronic

Reviewed by:



Alan R. Washburn, Chairman  
Department of Operations Research

Released by:



William M. Tolles  
Dean of Research

\* A portion of this work was performed for the Naval Postgraduate School  
under Contract No. N00014-82-C-0252.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NPS55-83-016	2. GOVT ACCESSION NO. A130155	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) VOICE RECOGNITION PERFORMANCE WITH NAIVE VERSUS PRACTICED SPEAKERS	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 93940	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS N0003983WRDX083	
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, CA 93940	12. REPORT DATE June 1983	
	13. NUMBER OF PAGES 27	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Naval Electronics Systems Command, Code 613 Washington, D.C. 20360	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, Automatic Word Recognition, Practice vs. Nonpracticed, Independence		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This study examined the accuracy of a current voice recognition device when used in a speaker independent mode by naive and practiced speakers. Neither group of speakers ever tested the voice recognizer with their own speech patterns in memory. It was interesting to discover that the naive speakers obtained a recognition accuracy at least as good as that of practiced speakers, with both groups attaining about 96% accuracy with other voice patterns in memory than their own.		

DD FORM 1 JAN 73 1473

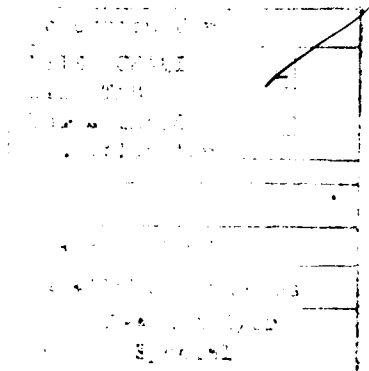
EDITION OF 1 NOV 65 IS OBSOLETE  
S/N 0102- LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TABLE OF CONTENTS

	<u>Page</u>
EXECUTIVE SUMMARY	ii
1. INTRODUCTION	1-1
1.1 Background	1-1
1.2 Problem	1-1
1.3 Objective	1-2
2. METHOD	2-1
2.1 Subjects	2-1
2.2 Apparatus	2-1
2.3 Experimental Design	2-1
2.4 Procedure	2-2
2.4.1 Training	2-2
2.4.2 Testing	2-2
2.5 Independent and Dependent Variables	2-4
3. RESULTS	3-1
3.1 Overview	3-1
3.2 Total Errors	3-1
3.3 Nonrecognitions	3-3
3.4 Misrecognitions	3-9
4. DISCUSSION	4-1
4.1 Total Errors	4-1
4.2 Nonrecognitions	4-1
4.3 Misrecognitions	4-1
5. CONCLUSION	5-1
6. REFERENCES	6-1
APPENDIX A	A-1



A



## EXECUTIVE SUMMARY

The purpose of the current study was to determine the accuracy of a current voice recognition device (VRD) when used by naive speakers versus practiced speakers, in a speaker independent mode (one in which the VRD device relies on the speech patterns of individuals other than the current speaker). It is conceivable that in future applications of VR technology, it may be costly or impractical to provide practice and training to all users.

The findings suggest that first time users of VR equipment, will obtain 96.85% recognition accuracy, a level at least as high as that obtained by users who have received training or practiced speaking to the VRD. Neither nonrecognitions (e.g., errors where the system rejects the input and responds, in effect, with "I don't understand you, say it again") or misrecognitions (e.g., errors where the system accepts the input but mistakes it for a different input) differed significantly for naive speakers versus practiced speakers. Furthermore, the misrecognition rate for naive speakers was only 1.11%.

It was concluded that training and practice may not always be necessary in order to obtain optimum performance in the human-VRD system. Without the need for practice, which implies modifying the human's behavior, the human-machine interaction is more natural, the "friendliness" of the VRD is enhanced, and the cost of the VR system use is reduced.

## 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. With constant improvements being made in the capabilities of voice recognition systems, their use in a wider variety of settings is already being contemplated.

As the variety of settings widens, the requirements for the VRD become more diversified. One situation may require a VRD to recognize the speech of only one user who has thoroughly "trained" the system. Another situation might require the VRD to recognize the speech of several users, and, in some instances, to recognize the speech of a user for whom the VRD has no speech patterns recorded, in effect, a speaker independent situation. In the latter cases it would be desirable for the VRD to be capable of recognizing the speech of as many users as possible, without an increase in errors due to the variance of speech patterns from user to user.

For purposes of this paper, we will refer to speaker independence as meaning where we use a speaker dependent recognizer but when a user talks to the recognizer, that user's voice patterns are never in memory. In any case, decisions must be made concerning the variety of stored speech patterns necessary for recognition of a user's speech in particular settings.

### 1.2 Problem

In recent experiments, Schwalm and Martin (1982) found that a currently available VRD performed with 95% recognition accuracy under speaker independent conditions. Their results were based on data from subjects who

had undergone a training session in which they practiced speaking to the VRD. This, in turn, could have optimized the VRD's recognition accuracy. While 95% recognition accuracy is impressive regardless of the possible effects of practice, the contribution that practice makes to recognition accuracy deserves investigation. Future applications of VR technology may involve users who have never trained a VRD or practiced speaking to one. In some applications the VRD may be required to interact with a user population large enough to make training by all users impractical.

The purpose of the present research was to determine the effects, if any, of training/practice on recognition accuracy.

### 1.3 Objective

The specific objective of the present research was to assess empirically the accuracy with which currently available VRDs could interpret utterances made by: (1) speakers who had received practice by training the VRD, and (2) speakers who had never trained or used a VRD.

## 2. METHOD

### 2.1 Subjects

Thirty volunteers (all males) were recruited from the Naval Postgraduate School in Monterey, California. Twenty-seven were students and three were staff. None had ever used voice recognition equipment before.

### 2.2 Apparatus

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

A Shure model SM10 "boom" microphone (mounted on a headset) was used as the input device. This microphone is supplied as standard equipment with the T600.

The Threshold system was linked to an IBM computer via a modem, allowing the experimenter to manipulate which set of speech patterns the Threshold would access when attempting to recognize the 50 utterances.

### 2.3 Experimental Design

A 2x3x6 mixed design was employed in this experiment. Experience was a two-level between group variable. One group received practice by training the VRD (henceforth, "practiced" group) and the other group did not (henceforth, "naive" group). Each subject performed six trials, making trials the within group variable with six levels. Subjects in each experience level were divided into three groups, each of which accessed a different set of voice patterns in the VRD, making pattern set the second

between variable with three levels. A pattern set is a group of reference patterns, called templates, that the VRD refers to in determining what utterance has been made. These templates are created in the training phase, as described below. Each pattern set consisted of four templates for each of the fifty utterances in the vocabulary (4 voices (templates) x 50 utterances = 200 templates per pattern set). In other words, a pattern set contained the trained templates from four random speakers on the same identical utterances listed in Appendix A. The use of three different pattern sets, each based on four different voices, provided internal replication of the experience by trials design, and allowed greater generalization of the results. A summary of the experimental design appears in Figure 2-1.

## 2.4 Procedure

2.4.1 Training. 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 (10x50 or 500 utterances 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 of the same vocabulary (in testing) with their templates in memory. Ideally, these subsequent utterances are matched with their templates 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.

				TRIALS					
				T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>	T <sub>5</sub>	T <sub>6</sub>
E X P E R I E N C E	PRACTICED	Group 1 Pattern Set 1	S <sub>1</sub>	→					
			2						
			3						
			4						
			5						
		Group 2 Pattern Set 2	6						
			7						
			8						
			9						
	10								
	NAIVE	Group 4 Pattern Set 1	11						
			12						
			13						
		Group 5 Pattern Set 2	14						
			15						
16									
E	Group 6 Pattern Set 3	17							
		18							
		19							
	NAIVE	Group 4 Pattern Set 1	20						
			21						
			22						
		Group 5 Pattern Set 2	23						
			24						
			25						
Group 6 Pattern Set 3	26								
	27								
	28								
		29							
		S <sub>30</sub>	→						

FIGURE 2-1.  
SUMMARY OF EXPERIMENTAL DESIGN

2.4.2 Testing. Each subject was scheduled to make two passes through the entire vocabulary list on each of three successive days. Subjects in the practiced group made 2 additional passes through the vocabulary list each day, providing further practice not received by the naive group. For the practiced group, these sessions were administered on Wednesday, Thursday, and Friday of the same week in which training took place. Testing sessions for the naive group were scheduled on Wednesday, Thursday, and Friday of a different week. Thus, a total of six testing trials were run for each subject. Both practiced and naive speakers were able to complete the experiment within one week. Subjects in the practiced group and the naive group never tested against a pattern set containing their own speech patterns, thus, both experience groups tested in the speaker independent mode.

2.4.3 Summary. Fifteen subjects who had never used VR equipment before (naive subjects) tested a VRD along with 15 subjects who had trained and practiced using VR equipment (practiced subjects). Subjects in both groups tested the device in the speaker independent mode, and both practiced and naive speakers accessed identical pattern sets. Recognition accuracy was recorded for 300 critical utterances by each subject. While critical utterances were the only inputs naive speakers ever made to the VRD, each practiced speaker had made 1,100 additional inputs to the VRD as a result of training and practice sessions.

## 2.5 Independent and Dependent Variables

The independent variables in this study were pattern set, trials, and experience: practiced or naive. The dependent variables were nonrecognitions (or rejections), misrecognitions, and total errors, which was a linear combination of nonrecognitions and misrecognitions.

### 3. RESULTS

#### 3.1 Overview

This section describes the results of the present study. All repeated measures analyses of variance procedures 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 Scheffé procedures described in Bruning and Kintz (1977).

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.

#### 3.2 Total Errors

Table 3-1 presents the analysis of variance for total errors (nonrecognitions + misrecognitions). There were no significant effects of experience, pattern set, or trials, nor were there any significant

TABLE 3-1  
ANALYSIS OF VARIANCE SUMMARY TABLE  
FOR TOTAL ERRORS

Source	df	MS	F
Experience (E)	1	.02053	.053
Pattern Set (P)	2	.08908	.231
Ex P	3	.13846	
Error	24	.38519	
Trials (T)	5	.03760	1.743
T x E	5	.03193	1.480
T x P	10	.02778	1.288
T x P x E	10	.04021	1.865
Error	120	.02157	

interactions. Mean total errors for experience by trials are shown in Table 3-2.

### 3.3 Nonrecognitions

An analysis of variance was performed on the nonrecognitions alone to determine the effects, if any, of experience, trials, and pattern sets. Table 3-3 presents the analysis of variance summary table for nonrecognitions.

A significant main effect of trials ( $F=2.36$ ,  $p<.05$ ) was found, as was a significant three-way interaction of trials by pattern set by experience ( $F=2.219$ ,  $P<.05$ ). No other main effects or interactions were statistically significant. Mean nonrecognitions for experience by trials are shown in Table 3-4. The main effect of trials, and the three-way interaction of trials by pattern set by experience are portrayed graphically in Figures 3-1 and 3-2, respectively.

With regard to the main effect of trials, although the analysis of variance indicated a significant trials effect, review of Figure 3-1 reveals no apparent systematic change over trials. A Scheffé test for significance between pairs of means detected no significant differences between any two trials. Evidently, the analysis of variance is sensitive to the spurious nature of errors across trials. However, the difference between even the highest and lowest error rates over trials is not large enough to reach statistical significance in the post hoc Scheffé test. For further discussion on post hoc range tests, and lack of significance in post hoc tests where significance was reached in an analysis of variance, see J.L. Myers, 1972.

TABLE 3-2  
 MEAN TOTAL ERRORS (IN PERCENT)  
 FOR EXPERIENCE BY TRIALS

		TRIALS						$\bar{x}$ Trials
		1	2	3	4	5	6	
EXPERIENCE	PRACTICED	5.20	3.60	5.60	5.33	4.27	5.20	4.87
	NAIVE	4.00	3.60	2.67	2.80	2.80	3.07	3.15
	$\bar{x}$ EXPERIENCE	4.60	3.60	4.14	4.07	3.53	4.1	Grand $\bar{x}$ 4.01

TABLE 3-3  
ANALYSIS OF VARIANCE SUMMARY  
TABLE FOR NONRECOGNITIONS

Source	df	MS	F
Experience (E)	1	.05712	.158
Pattern Set (P)	2	.02264	.063
Ex P	2	.05488	.152
Error	24	.36168	
Trials (T)	5	.04666	2.356*
T x E	5	.03194	1.613
T x P	10	.03147	1.589
T x P x E	10	.04395	2.219*
Error	120	.01980	

\*P < .05

TABLE 3-4  
 MEAN NONRECOGNITIONS (IN PERCENT)  
 FOR EXPERIENCE BY TRIALS

		TRIALS						$\bar{x}$ Trials
		1	2	3	4	5	6	
EXPERIENCE	PRACTICED	3.60	2.27	3.73	4.13	3.47	4.13	3.56
	NAIVE	3.47	2.13	1.60	1.47	1.60	2.00	2.04
	$\bar{x}$ EXPERIENCE	3.53	2.20	2.67	2.80	2.53	3.07	Grand $\bar{x}$ 2.80

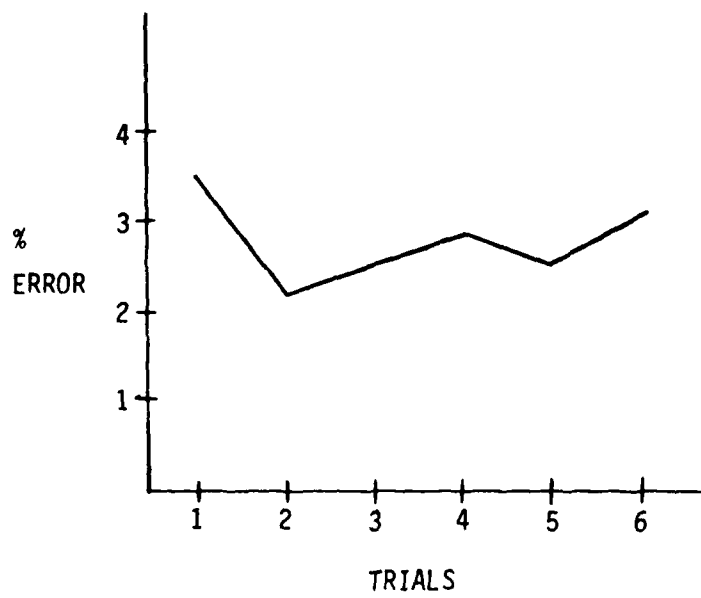


FIGURE 3-1  
NONRECOGNITIONS BY TRIALS

— Practiced  
 - - - Naive

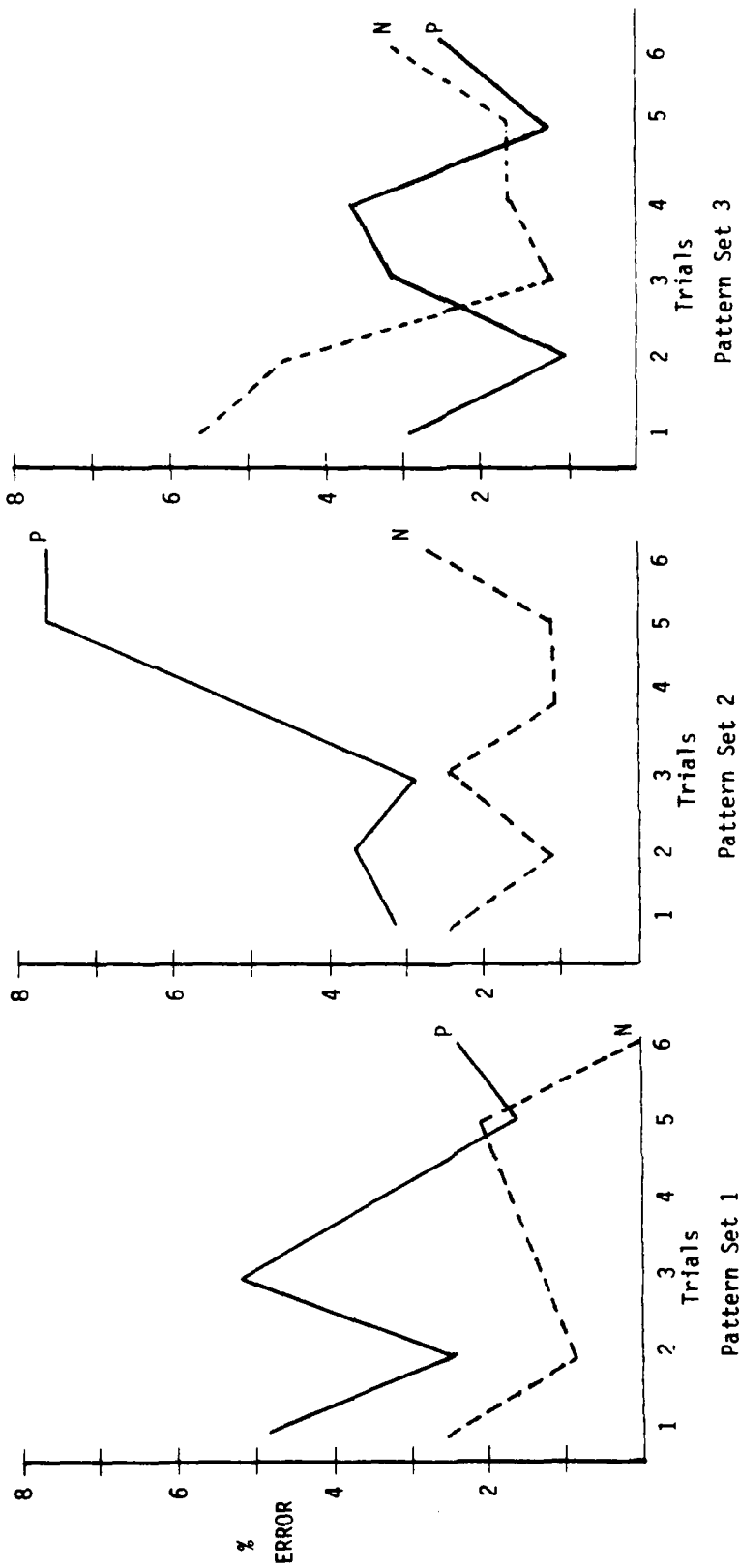


FIGURE 3-2  
 NONRECOGNITIONS FOR EXPERIENCE  
 BY TRIALS BY PATTERN SET

The experience by trials by pattern set interaction also reached significance in the analysis of variance. Again, there were no interpretable or systematic effects, and the authors attach no practical significance to either the trials or the experience by trials by pattern set interaction.

#### 3.4 Misrecognitions

As for nonrecognitions, an analysis of variance was performed on the misrecognitions alone to determine the effects, if any, of experience, pattern sets, and trials. Table 3-5 presents the analysis of variance summary table for misrecognitions.

A significant main effect of pattern sets ( $F=6.02$ ,  $p<.01$ ) is evident. The main effects of experience and trials were not significant, nor were any of the interactions. Mean misrecognitions for experience by pattern set are shown in Table 3-6, and the effect of pattern sets is portrayed graphically in Figure 3-3.

With regard to the main effect of pattern sets, a Scheffé test for significance between pairs of means was performed to determine where such differences lie. Again, as was the case for nonrecognition trials, the main effect of misrecognitions by pattern sets, reported in the analysis of variance, could not be detected in the Scheffé test. (Review Figure 3-3 for further clarification.) Misrecognitions do vary somewhat as a function of pattern set. However, the greatest number of errors (pattern set 1) was 2.23%, leaving little range for variability with a floor of zero. With the stringent per comparison alpha level imposed by the Scheffé test, the difference in range between pattern set one and pattern set three (where the least errors occurred) did not reach significance. All statistical results considered, the effect of pattern sets may be attributed to greater dissimilarity between the voices of subjects and contributors of pattern set one, than between voices of subjects and contributors of pattern sets 2 and 3.

TABLE 3-5  
ANALYSIS OF VARIANCE SUMMARY  
TABLE FOR MISRECOGNITIONS

Source	df	MS	F
Experience (E)	1	.00000	0
Pattern Set (P)	2	.39584	6.02*
Ex P	2	.08367	1.272
Error	24	.06575	
Trials (T)	5	.01504	.728
T x E	5	.03154	1.525
T x P	10	.02492	1.205
T x P x E	10	.01496	.724
Error	120	.02067	

\*P < .01

TABLE 3-6  
 MEAN MISRECOGNITIONS (IN PERCENT)  
 FOR EXPERIENCE BY PATTERN SET

		PATTERN SET			$\bar{x}$ Pattern Sets
		1	2	3	
E X P E R I E N C E	PRACTICED	2.93	.53	.47	1.31
	NAIVE	1.53	1.13	.67	1.11
	$\bar{x}$ EXPERIENCE	2.23	.83	.57	Grand $\bar{x}$ 1.21

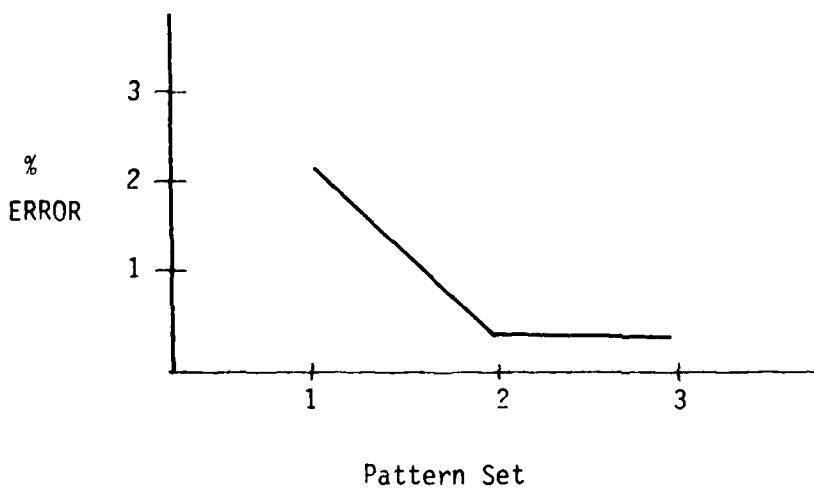


FIGURE 3-3  
MISRECOGNITIONS BY PATTERN SETS

## 4. DISCUSSION

The following section discusses some implications of the aforementioned results.

### 4.1 Total Errors

There were no significant differences in the number of total errors produced by practiced speakers versus naive speakers. In positive terms, naive speakers obtained recognition accuracy of 96.85%, with the VRD relying on the speech patterns of four independent speakers. This performance represents a slight (1.72%) but statistically non-significant improvement over practiced speakers, and lends further support to previous findings of greater than 95% recognition accuracy in the speaker independent mode in general (Schwalm & Martin, 1982).

### 4.2 Nonrecognitions

Nonrecognitions accounted for 70% of the total errors. As was the case with total errors, there were slightly fewer (1.52%) nonrecognitions produced by naive speakers, however, this difference was non-significant.

### 4.3 Misrecognitions

As was the case with total errors and nonrecognitions, naive speakers produced slightly fewer misrecognitions (.2%) than practiced speakers, again the difference was non-significant. Misrecognitions accounted for only 30% of the total errors, a fortunate finding since misrecognitions are the more problematic of the two types of errors, as explained earlier.

The question arises as to why, even though not statistically significant, naive speakers seem to make less errors than practiced speakers.

An explanation for the apparently better performance of naive subjects as opposed to practiced subjects may be linked to the effects of stress on voice recognition performance. In a previous study (Schwalm, 1983), it was found that speakers' attitudes about their performance in the initial stages of using voice recognition technology appeared to contribute to their subsequent performance. It is entirely possible that subjects who had used voice recognition equipment before felt that they should be able to use that equipment with a high level of proficiency (even though there may be no real objective reasons to expect this). If subjects really felt that this should be the case, they may have entered the experiment with some self-imposed expectations of achieving a high level of performance during the experiment. It is therefore possible that when the subjects made their first few errors, they became frustrated (or stressed, in the general sense) and that the quality of their subsequent inputs was degraded (see Schwalm, 1983). Thus, poorer performance for the practiced group might be expected.

It is important to note that the above explanation based on self-imposed (psychological) stress is speculative at this point. The authors feel that the entire area of psychological (as well as other sources of) stress, as it applies to performance with voice recognition technology, deserves considerable research attention in the future. If individuals will be required to use voice recognition equipment in a growing number of applications, and if (as it appears at this time) stress changes the quality of voice input, there is significant value in determining just how stress affects the users of voice recognition equipment and their performance.

## 5. CONCLUSION

The present research has shown that a person who has never trained or practiced speaking to a VRD can obtain 96.85% recognition accuracy with the VRD relying on the speech patterns of four independent speakers. This degree of accuracy does not differ significantly from speakers who did train the VRD and practiced speaking the vocabulary. In the speaker independent mode, training is not associated with any significant cost or benefit in recognition accuracy. In other words, training and practice may not be necessary, a situation favorable to the potential applications of VR technology.

Some human-machine systems involve very high "friendliness" demands. In some applications, the need for all users to train or practice speaking to the VRD represents an acceptable cost. However, in other applications (with large or unspecified populations) the need for all users to train and practice speaking to the VRD could be so impractical that it would eliminate voice as a method of input. The current findings suggest that voice is a viable method of input, not requiring training and practice for successful operation.

The reader is reminded of some pertinent qualifications to these findings. All subjects were male, native English speakers from the Naval Postgraduate School, ranging from about 25 to 35 years of age. The three pattern sets that the subjects tested against were created by subjects who met these same criteria. Under a conservative interpretation, the 96% average recognition rate might decrease in a real world situation involving a more diversified user population. However, if the pattern sets were constructed selectively, rather than by random assignment, the 96% recognition rate might logically be expected to increase. Future research at the Naval Postgraduate School will investigate spectrographic speech characteristics

in an effort to qualify and optimize the speech patterns stored in the VRD's memory. All things considered, the authors are confident that the current findings reflect the capability of state of the art VRDs to interact successfully with untrained, unpracticed users such as those who participated in the present investigation.

## 6. REFERENCES

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

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.

Schwalm, N.D., Pooch, G.K., and Roland, E.F. Using voice recognition technology: A psychological perspective. Submitted for publication to Speech Technology, February, 1983.

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

APPENDIX A

WORD #	UTTERANCE	WORD #	UTTERANCE
1	ONE	26	SIERRA
2	YANKEE	27	APPLICATION
3	GARY POOCK	28	HUMAN FACTORS
4	CARRIAGE RETURN	29	CENTRAL EXPRESSWAY
5	IRAN	30	FILE TRANSFER PROTOCOL
6	SWEDEN	31	NINE
7	LOGIN POOCK	32	INDIA
8	ACCAT TITLE	33	LIMA
9	LOAD GLD3	34	POPPA
10	POOCK NPS PASSWORD	35	UNIFORM
11	THREE	36	KOREA
12	LOGOUT	37	INTERACTIVE
13	RED SPHERE	38	CONTINUOUS
14	SEVEN	39	CONTINUOUS SPEECH
15	MOVE IT DOWN	40	SYSTEM INTEGRATION
16	SPIROGRAPH	41	MIKE
17	CLOSE OUT CHARLIE	42	TANGO
18	UNITED STATES	43	WHISKEY
19	NORTH ATLANTIC MAP	44	ZULU
20	MEDITERRANEAN MAP	45	BANGLADESH
21	SIX	46	HOLLISTER
22	BRAVO	47	CORPORATION
23	DELTA	48	ADVANTAGES
24	FOXTROT	49	RADIOLOGY
25	ROMEO	50	AUTOMATIC RECOGNITION

DISTRIBUTION LIST

	No. of Copies
Library, Code 0142 Naval Postgraduate School Monterey, CA 93940	4
Dean of Research Code 012A Naval Postgraduate School Monterey, CA 93940	1
Library, Code 55 Naval Postgraduate School Monterey, CA 93940	1
Professor Gary Poock Code 55Pk Naval Postgraduate School Monterey, CA 93940	150

