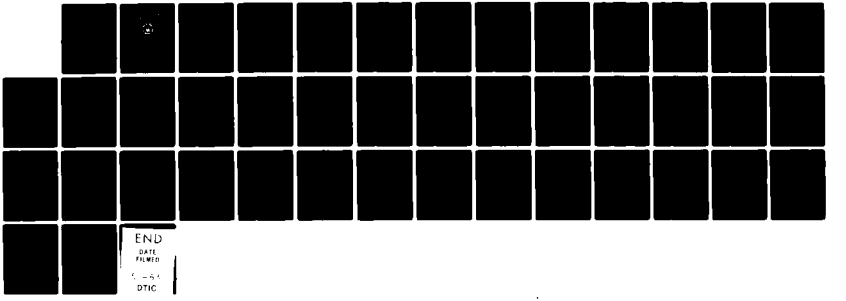
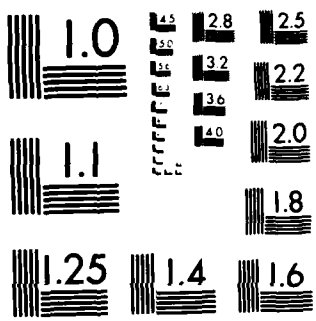
AD-A127 223 THE EFFECT OF FEEDBACK TO USERS OF VOICE RECOGNITION EQUIPMENT(U) NAVAL POSTGRADUATE SCHOOL MONTEREY CA G K POOCK ET AL. FEB 83 NPS55-83-003 MIPR-TB-024 1/1

UNCLASSIFIED F/G 17/2 NL



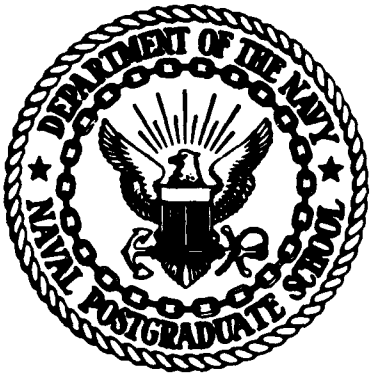
END
DATE FILMED
DTIC



NPS55-83-003

NAVAL POSTGRADUATE SCHOOL

Monterey, California



THE EFFECT OF FEEDBACK TO USERS OF VOICE RECOGNITION EQUIPMENT

by

G. K. Poock
B. Jay Martin
E. F. Roland

February 1983

Prepared for:
9th Infantry Division
Fort Lewis, WA 98433

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral J. J. Ekelund
Superintendent

D. A. Schrady
Provost

Reproduction of all or part of this report is authorized.

This report was prepared by:

G. K. Poock, Professor
Department of Operations Research

B. Jay Martin
Perceptronics

E. F. Roland
Rolands and Associates

Reviewed by:

K. T. Marshall, Chairman
Department of Operations Research

Released by:

William M. Tolles
Dean of Research

E.F. Roland worked on this project under a contract to NPS entitled "Research and development study of the feasibility of using computer voice entry", NPS Contract No. N--228-82-C-6418.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NPS55-83-003	2. GOVT ACCESSION NO. AD-A127223	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) THE EFFECT OF FEEDBACK TO USERS OF VOICE RECOGNITION EQUIPMENT		5. TYPE OF REPORT & PERIOD COVERED Technical 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) G. K. Poock B. Jay Martin E. F. Roland		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 MIPR TB-024
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, CA 93940		12. REPORT DATE February 1983 13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) 9th Infantry Division Fort Lewis, WA 98433		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 Speech Recognition
Voice Input/Output

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

This paper describes an experiment designed to study the effect of providing a user of VOICE recognition equipment with feedback concerning how well the voice recognizer is interpreting the user's voice commands.

The results indicated that users, who were preconditioned with no feedback, obtained better recognition performance when they were provided with more feedback. Users who were preconditioned with a lot of feedback degraded in performance when feedback information was taken away from them.

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

A

FOREWORD

The primary reason for doing this research was to examine the need for feedback of recognition results to operators in situations where they might move around and not always be in front of a computer terminal. Specifically, if operators were using voice entry into the Army's Artillery Control Console in the TACFIRE van, would their voice recognition accuracy degrade if they moved around in the van and didn't always have immediate feedback visually in front of them on the display console.

Generically, however, the results are applicable to any situation in which an operator may be somewhat mobile and not always receive direct visual feedback.

EXECUTIVE SUMMARY

The purpose of the present study was to determine the effects, if any, of feedback on the performance of a currently available voice recognition device (VRD). It is conceivable and likely that voice recognition equipment will be used in a variety of command, control, and communication (C^3) interfaces in the future. Different applications limit the type and amount of feedback that can be provided. For example, telephone input precludes the provision of visual feedback, sonar monitoring may prevent the use of auditory feedback, and remote input may eliminate feedback altogether.

The findings suggest that feedback has a limited effect on performance; subjects not accustomed to feedback reduced errors by 5% when feedback was introduced, while subjects accustomed to a lot of feedback encountered about 5% more errors when feedback was reduced. Across different types and levels of feedback, however, no major differences were found.

It was concluded that feedback reminds the user how to keep his voice inputs consistent with the speech patterns he created when training the device to recognize his voice. Voice recognition devices currently exist that tolerate greater inconsistency than the model used in this study. More sophisticated devices do not require extensively consistent voice inputs to reduce the number of errors as do less sophisticated VRD's and thus diminish the consideration of feedback in the human-machine interface. Still, errors are undersirable regardless of their frequency or consequences, and the results suggest that consistent feedback should be provided within practical limitations, to hold errors to a minimum.

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.

To maintain optimum performance in this increasingly diversified technology, it is imperative that human factors be carefully considered and accommodated. The amount and type of feedback supplied to the user is potentially an important variable in the human-machine interface. Feedback is commonly defined as knowledge of results. After making a voice input, there are three possible results: (1) a recognition, in which the correct utterance in memory is matched with the input; (2) a non-recognition, in which no acceptable match is found; and (3) a misrecognition, in which the computer matches the input with the wrong utterance in memory. Most VRD's are equipped to deliver auditory and visual feedback; nonrecognitions are accompanied by a beep, and in some VRD's, a message such as "NO MATCH, " "REPEAT," or "I DON'T UNDERSTAND" may be presented on a screen or verbally by a speech synthesizer. Misrecognitions are not normally identified as errors by the VRD, since the criterion for choosing a match is based only on spectrographic analysis (the sound characteristics of the utterance). However, in some applications, it is conceivable, and likely, that the VRD would submit the spectrographic match to programming capable of determining if the match is a member of currently acceptable inputs (Calcaterra, 1982). For example, in an interactive program, the computer may be awaiting a voice input of either "CALL MENU" or "EXIT PROGRAM." If the spectrographic match for the input "CALL MENU" was misrecognized as "CONTINUE," the computer could in

this case,supply feedback indicating that a misrecognition had occurred, since it knows that "CONTINUE" is not one of the 2 acceptable inputs at the current junction. As a result, misrecognitions could be accompanied by the same type of feedback as nonrecognitions. Finally, correct recognitions are usually presented on a screen and could also be verbalized via a speech synthesizer.

Unfortunately, in some applications, users may not have the luxury of multidimensional feedback. For example, speech input by telephone or radio eliminates use of the visual modality. In situations requiring a user to monitor auditory signals, such as sonar, or in situations where extraneous auditory signals are unacceptable, the auditory modality is unacceptable for feedback.

In any case, informed decisions will soon need to be made concerning the type and amount of feedback to supply, as well as what to expect (in terms of performance) as a result of situational limitations on feedback.

1.2 Problem

Feedback is generally associated with improvement in performance, i.e., a "learning curve." It is questionable, however, to what extent making speech inputs to a VRD constitutes a learning situation for the user. Rather, it is the goal of the VRD to "learn" to recognize the user's speech. Perhaps the most basic question about feedback is, does it have any effect on future performance? If the answer is "no," then the issue is academic, but if the answer is "yes," a series of new questions arise: Does feedback improve or hinder performance, and if so, by how much? Is there a particular optimum level of feedback? Does the sensory modality to which feedback is directed differentially affect performance? Do the type and amount of feedback affect nonrecognitions and misrecognitions in the same way?

The purpose of the current research was to determine the answers to these questions.

1.3 Objective

The specific objective of the present research was to assess the effects, if any, of various levels of feedback on recognition accuracy.

2. METHOD

2.1 Subjects

Forty-eight subjects (26 male, 22 female) were recruited from Monterey Peninsula College and the Navy Postgraduate School in Monterey, California. Eleven were military personnel and thirty-seven were civilians. The subjects' ages ranged from 18 to 75.

2.2 Apparatus

An Interstate Electronics Corporation VRT 101 voice recognition device was used in this study. It is important to note that the Threshold T600 model VRD was considered for use in this study. However, in a recent study, the T600 produced a total error rate of only 1% (Schwalm and Martin, 1982). Since the current study intended to examine the change in errors across feedback conditions, encountering a floor effect with the T600 seemed probable. Thus, the Interstate VRT 101 was used in the hope that this problem could be avoided. The Interstate allows manipulation of four parameters: reject threshold, delta level, speech input level, and number of training passes. Reject threshold is used to compare the degree of precision in the match between the input utterance pattern and the reference pattern. The value can be set from 0 to 100. A higher value results in better rejection of invalid words at the expense of a greater frequency of rejection of valid words. Interstate suggests a setting of 82 to 94 (Interstate Electronics Corporation, 1981). A slightly more liberal value of 80 was used in the present study since invalid words would not be included in the measurements. The delta level is used to reject words when the difference between the classified word and the second place word scores are less than this threshold. This level is usually in a range of 2 to 10 (Interstate Electronics Corporation, 1981). The delta level was set to 3 in the present study, based on information supplied by

Interstate and previous experiments at the Naval Postgraduate School (Poock et al, 1982). The speech input level has four settings, for loud, average, soft speakers, plus an experimental setting. The setting for average speakers was used except for 2 subjects who required the soft setting for acceptance of their inputs. Interstate suggests 5 to 7 training passes (Interstate Electronics Corporation, 1981). Six were used in the present study. The Interstate VRT 100 is capable of storing up to 100 utterances, and 100 utterances were used in the present investigation. These utterances appear in Appendix A.

A Shure model SM10 "boom" microphone (mounted on a headset) was u' ' as the input device. A solid-state resonator, attached to a telegr key, provided an auditory signal for feedback.

2.3 Experimental Design

A 2x8x5 mixed design was employed in this experiment. After training, subjects first tested the VRD under one or the other of 2 feedback conditions. These initial feedback conditions provided baseline error rates for each subject and will be referred to as preconditions. Thus, precondition was a two-level between group variable. In the first precondition, subjects received No Feedback concerning either recognitions, misrecognitions, or nonrecognitions. In the second precondition, subjects received Total Feedback. In the total feedback precondition, the following auditory and visual information was available:

> Visual Feedback -- the CRT would present the correctly recognized or the misrecognized word, and a "NO MATCH" indication was presented for nonrecognitions.
> Auditory Feedback -- the experimenter verbalized the information presented on the CRT and, in the case of nonrecognitions, a beep was sounded.

After obtaining baseline error rates in their respective preconditions, each subject entered one of eight test conditions. While the preconditions represented the extremes of feedback (all or none), the test conditions represented the extremes plus six intermediate levels of feedback. Thus, test condition was an eight-level between groups variable, occuring under each precondition. The eight test conditions were as follows:

(1) No feedback -- same as No Feedback precondition.
(2) Nonrecognition Beep -- a beep sounded for nonrecognitions only.
(3) Nonrecognition and Misrecognition Beep -- the same beep sounded for both nonrecognitions and misrecognitions.
(4) Different Nonrecognition and Misrecognition Beeps -- a low beep sounded for nonrecognitions, and a high beep sounded for misrecognitions.
(5) Nonrecognition Beep and Verbal -- a beep sounded for nonrecognitions, and the experimenter verbalized recognitions and misrecognitions (i.e., what appeared on the CRT).
(6) Visual Feedback -- all correct recognitions and misrecognitions were presented on the CRT, and a beep was sounded for nonrecognitions.
(7) Total Feedback -- same as Total Feedback precondition

The above feedback scheme is summarized in Table 2-1. Each subject performed 5 trials under a test condition, making trials the within variable with 5 levels. A summary of the experimental design appears in Figure 2-1.

TABLE 2-1

FEEDBACK SCHEME

Test Condition	Nonrecognitions			Misrecognitions			Correct Recognitions	
	Beep	Verbal	Visual	Beep	Verbal	Visual	Verbal	Visual
None*								
Nonrecognition Beep	♪							
Nonrecognition Misrecognition Beep	♪			♪				
Different Nonrecognition Misrecognition Beeps	♪			♪				
Nonrecognition Beep Verbal Feedback	♪				"Word"		"Word"	
Visual Feedback			No Match			Word		Word
Mixed Feedback	♪					Word		Word
Total Feedback*	♪	"No Match"	No Match		"Word"	Word	"Word"	Word

*Also a precondition

		TRIALS				
PRECONDITION	TEST CONDITION	1	2	3	4	5
NO FEEDBACK	No Feedback	S1, 2, 3				
	Nonrecognition Beep	S4, 5, 6				
	Nonrecognition & Misrecognition Beep	S7, 8, 9				
	Different Nonrecogntion & Misrecognition Beeps	S10, 11, 12				
	Nonrecognition Beep & Verbal Feedback	S13, 14, 15				
	Visual Feedback	S16, 17, 18				
	Mixed Feedback	S19, 20, 21				
	Total Feedback	S22, 23, 24				
TOTAL FEEDBACK	No Feedback	S25, 26, 27				
	Nonrecognition Beep	S28, 29, 30				
	Nonrecognition & Misrecognition Beep	S31, 32, 33				
	Different Nonrecognition & Misrecognition Beeps	S34, 35, 36				
	Nonrecognition Beep & Verbal Feedback	S37, S38, S39				
	Visual Feedback	S40, 41, 42				
	Mixed Feedback	S43, 44, 45				
	Total Feedback	S46, 47, 48				

FIGURE 2-1.
SUMMARY OF EXPERIMENTAL DESIGN

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 or her particular speech patterns for all the utterances he or she will be using. For the VRT 100, this training procedure consisted of entering 6 passes of the entire vocabulary (6x100 or 600 utterances for each subject) into the voice recognizer. Each time a particular utterance is entered, it is compared to the average pattern of the previous entries for that utterance. If not similar enough to the average of the previous patterns, the utterance is rejected and must be repeated. If three successive rejections occur, the average pattern (for that particular utterance) is erased, and reformation of an average pattern based on 6 entries starts anew. In other words, the speech pattern for a particular utterance is the average of 6 entries, interrupted by no more than 2 successive rejections. The VRD saves these patterns in its memory automatically for comparison with utterances in testing. Ideally, these subsequent utterances are matched with those in memory and the result is a correct response. In cases where the VRD cannot make this match, a nonrecognition (or rejection) occurs. Occasionally, however, the VRD "thinks" it has matched an utterance with one in memory, but the match is incorrect. This constitutes a misrecognition. Thus, two types of errors are possible: nonrecognitions (or rejections) and misrecognitions (misinterpretations) of an utterance. The training procedure took approximately 45 minutes for each subject.

2.4.2 Precondition Testing. Within 3 days after training, subjects began pretesting by making 4 passes (2 passes a day for 2 days) through the vocabulary list. The order of the vocabulary words was reversed for every other pass through the list to reduce order effects. Half the subjects received No Feedback and half received Total Feedback.

2.4.3 Final Testing. Within 3 days after precondition testing, subjects began final testing. Subjects in each precondition were divided into 8 groups of 3 subjects each, and randomly assigned to each of the 8 test conditions. Subjects made 5 (testing) passes through the vocabulary list at 1 pass a day for 5 days. The order of the vocabulary words was again reversed for every other pass through the list to reduce order effects.

2.5 Independent and Dependent Variables

The independent variables were precondition: No Feedback and Total Feedback; test condition: No Feedback, Nonrecognition Beep, Nonrecognition and Misrecognition Beeps, Different Nonrecognition and Misrecognition Beeps, Nonrecognition Beep and Verbal Feedback, Visual Feedback, Mixed Feedback; and trials.

The dependent variables were nonrecognitions (or rejections), misrecognitions, and total errors, which was a linear combination of nonrecognitions and misrecognitions.

Baseline error rates were computed for each subject by averaging their errors over the 4 precondition trials. Change in errors, or error differences, were then computed for each subject in each of the 5 test condition trials by subtracting the baseline error rate from the raw errors in each trisl. Thus, positive numbers indicate an increase in errors and negative numbers indicate a decrease in errors.

3. RESULTS

3.1 Overview

This section describes the results of the present study. All analyses of variance procedures were performed using the arc sin transformation of relative difference scores to stabilize the variance of the error terms (Neter and Wasserman, 1974). The mean change in error rates that appear in tables and figures, however, are untransformed.

As defined earlier, nonrecognitions and misrecognitions by the voice recognition system may have distinctly different applications in an applied setting. To take an extreme example, in a weapons deployment activity, it would be far more desirable for the system to respond to an input error by nonrecognition, where no action is taken, 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 (nonrecognitions + misrecognitions).

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

3.2 Total Errors

Table 3-1 presents the analysis of variance summary table for change in total errors (nonrecognitions + misrecognitions). A significant main effect of precondition ($F = 18.544$, $p < .001$) is evident. There were no significant main effects for test condition or for trials, and there

TABLE 3-1.

ANALYSIS OF VARIANCE SUMMARY TABLE
FOR CHANGE IN TOTAL ERRORS

SOURCE	df	MS	F
Precondition (P)	1	2.432	18.544*
Test Condition (C)	7	.184	1.402
P x C	7	.141	1.071
Error	32	.131	
Trials (T)	4	.022	1.045
T x P	4	.043	2.072
T x C	28	.023	1.098
T x C x P	28	.023	1.121
Error	128	.021	

*P < .001

were no significant interactions. Mean changes in total errors (in percent) are shown in Table 3-2, and the main effect of precondition is portrayed graphically in Figure 3-1.

Figure 3-2 portrays graphically the relationship of total errors for preconditions by condition. The figure shows a reduction in errors for the No Feedback precondition group and an increase in errors for the Total Feedback precondition group under the test condition. The crossing lines in Figure 3-2 indicate the No Feedback precondition group produced fewer errors than did the Total Feedback precondition group after transfer to the test condition.

3.3 Nonrecognitions (Rejections)

An analysis of variance was performed on the change in nonrecognitions alone to determine the effects, if any, of preconditions, trials, and test conditions. Table 3-3 presents the analysis of variance summary table for change in nonrecognitions.

A significant main effect of precondition ($F = 23.663$, $p < .001$) was found. As in the case of total errors, there were no significant main effects of test condition or trials, and there were no significant interactions. Mean change in nonrecognitions (in percent) are shown in Table 3-4, and the main effect of precondition is portrayed graphically in Figure 3-3.

Figure 3-4 portrays graphically the relationship of nonrecognitions for preconditions by condition. The figure shows a reduction in nonrecognitions for the No Feedback precondition group under the test condition and an increase in nonrecognitions for the Total Feedback precondition group under the test condition. As in the case of total errors, the No Feedback precondition group produced fewer nonrecognitions that the Total Feedback precondition group after transfer to the test condition.

TABLE 3-2

MEAN CHANGE IN TOTAL ERRORS (IN PERCENT)
FROM PRECONDITION TO TEST CONDITION

	Precondition		
Test Condition	No Feedback	Total Feedback	$\bar{x}\Delta$ Test Condition
No Feedback	3.[illegible]3	2.40	3.19
Nonrecognition Beep	-9.18	1.25	-3.97
Nonrecognition and Misrecognition Beep	-6.33	5.05	- .64
Different Nonrecogniton & Misrecognition Beeps	-1.70	10.78	4.54
Nonrecognition Beep & Verbal Feedback	-6.22	3.65	-1.28
Visual Feedback	-12.78	3.30	-4.74
Mixed Feedback	-3.42	7.03	1.81
Total Feedback	-2.65	3.08	.22
$\bar{x}\Delta$ Precondition	-4.79	4.57	Grand $\bar{x}\Delta$ - .11

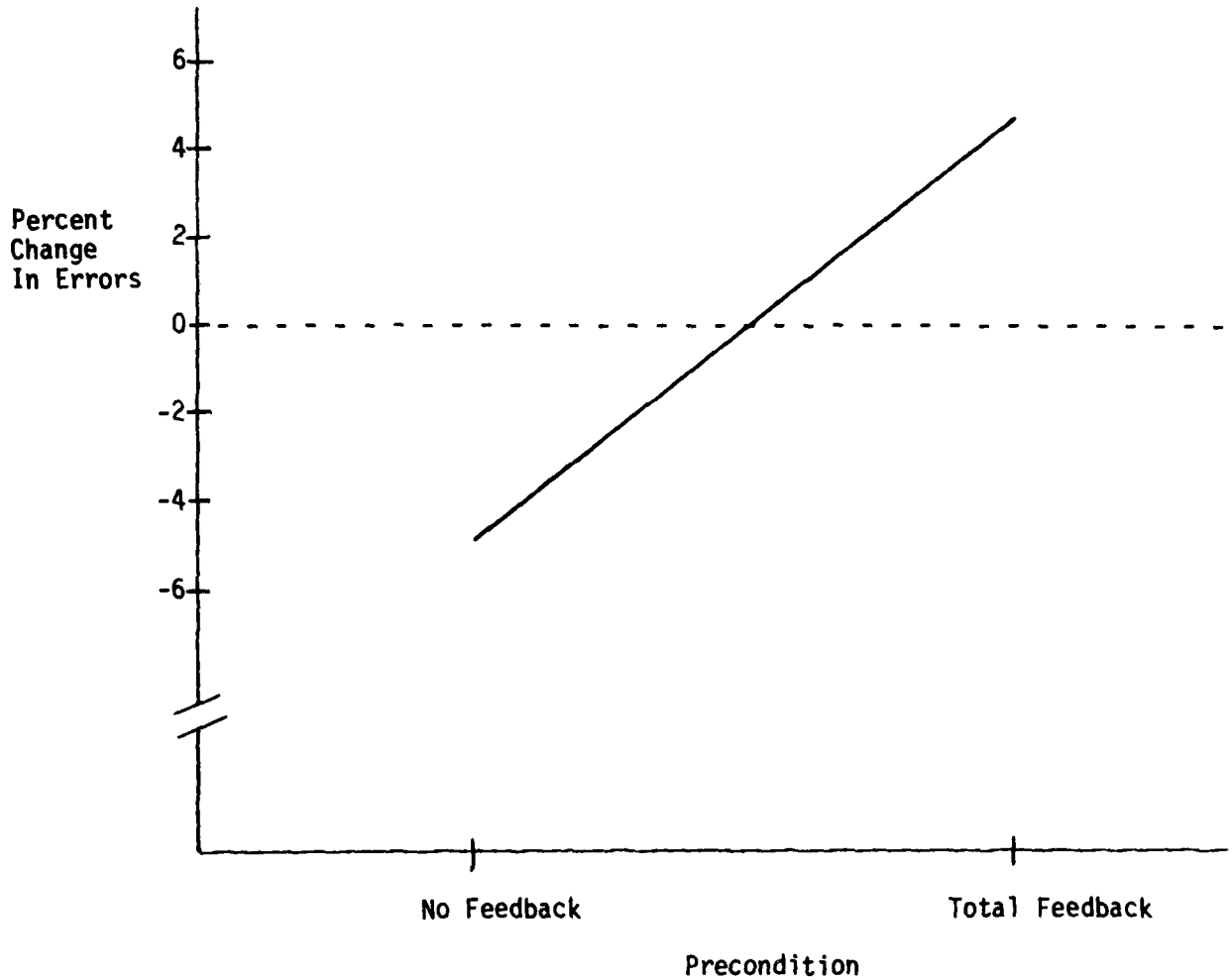


FIGURE 3-1.
CHANGE IN TOTAL ERRORS FROM PRECONDITION TO MEAN TEST CONDITION BY PRECONDITION

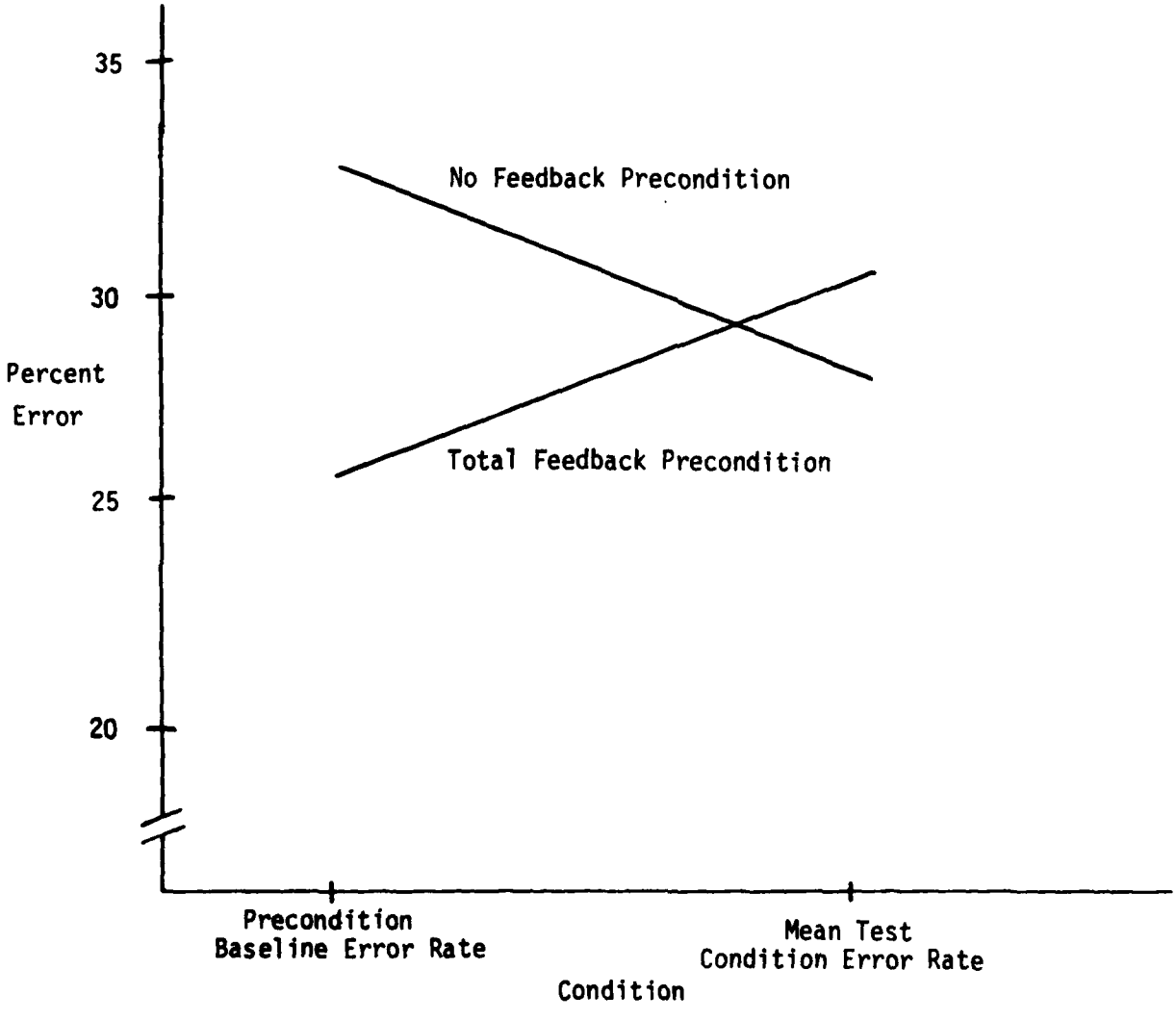


FIGURE 3-2.
TOTAL ERRORS FOR PRECONDITIONS BY CONDITION

TABLE 3-3

ANALYSIS OF VARIANCE SUMMARY TABLE
FOR CHANGE IN NONRECOGNITIONS

Source	df	MS	F
Precondition (P)	1	1.539	23.663*
Test Condition	7	.111	1.701
PxC	7	.045	.692
Error	32	.065	
Trials (T)	4	.019	1.381
TxP	4	.012	.866
TxC	28	.011	.844
TxCxP	28	.016	1.146
Error	128	.014	

*P<.001

TABLE 3-4

MEAN CHANGE IN NONRECOGNITIONS (IN PERCENT)
FROM PRECONDITION TO TEST CONDITION

	Precondition		
Test Condition	No Feedback	Total Feedback	$\bar{x}\Delta$ Test Condition
No Feedback	3.08	3.27	3.17
Nonrecognition Beep	-7.73	.82	-3.46
Nonrecognition and Misrecognition Beep	-4.40	4.67	.13
Different Nonrecogniton & Misrecognition Beeps	- .75	7.22	3.23
Nonrecognition Beep & Verbal Feedback	-5.27	2.70	-1.28
Visual Feedback	-9.30	2.43	-3.43
Mixed Feedback	-1.90	5.38	1.74
Total Feedback	-1.62	3.87	1.13
$\bar{x}\Delta$ Precondition	-3.49	3.79	Grand $\bar{x}\Delta$ - .15

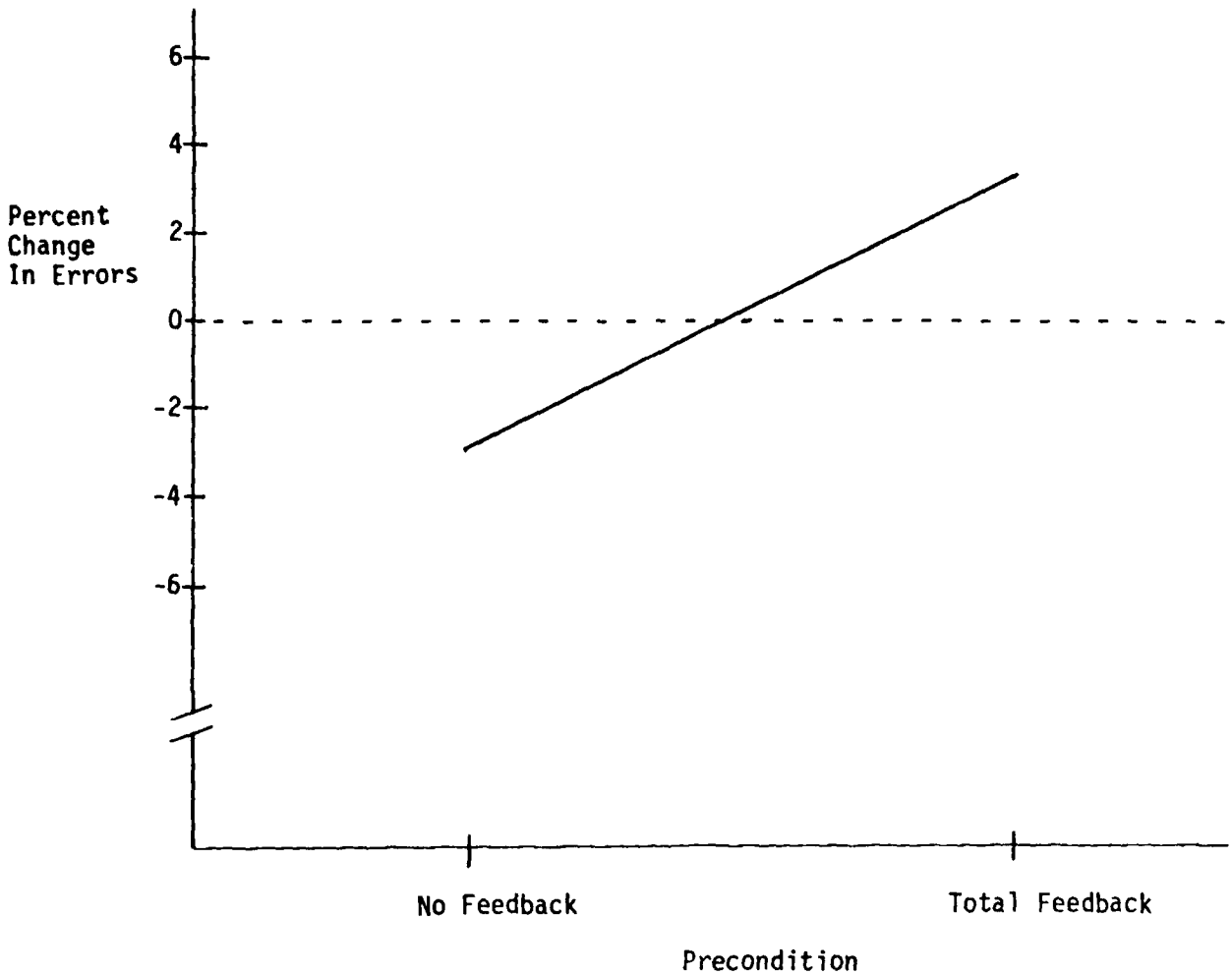


FIGURE 3-3.
CHANGE IN NONRECOGNITIONS FROM PRECONDITION TO MEAN TEST CONDITION BY PRECONDITION

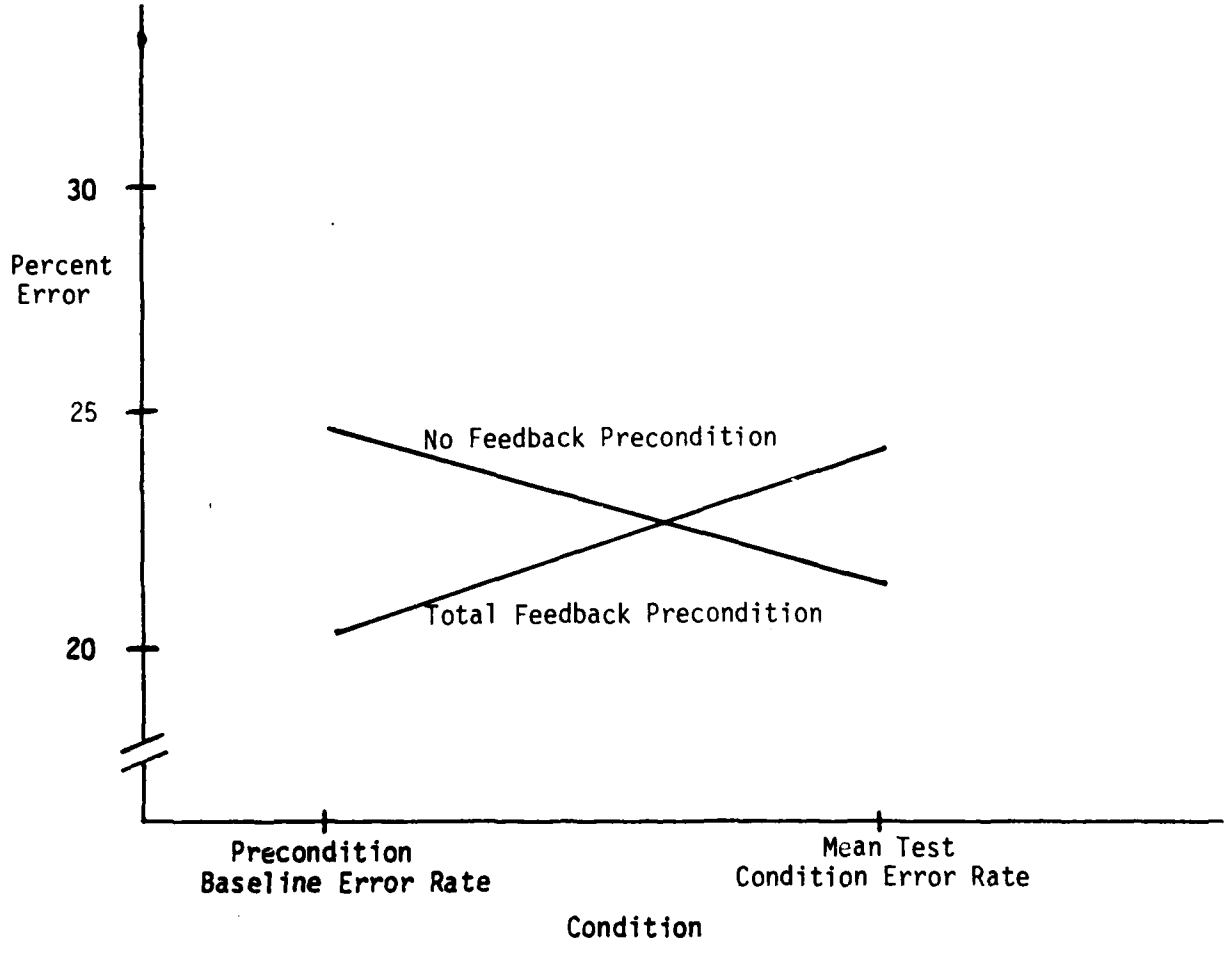


FIGURE 3.4
NONRECOGNITIONS FOR PRECONDITIONS BY CONDITION

3.4 Misrecognitions

As was done for nonrecognitions, an analysis of variance was performed on the misrecognitions alone, to determine the effects, if any, of preconditions, trials, and test conditions. Table 3-5 presents the analyses of variance summary table for change in misrecognitions.

A significant main effect of precondition ($F = 8.92$, $p < .01$) was found. As in the cases of total errors and nonrecognitions, there were no significant main effects of trials or test conditions. There was, however, an interaction of trials with precondition ($F = 7.732$, $p < .05$). Mean change in misrecognitions (in percent) are shown in Table 3-6 and the main effect of precondition is portrayed graphically in Figure 3-5.

Figure 3-6 portrays the relationship of misrecognitions for preconditions by condition. The figure shows a reduction in misrecognitions for the No Feedback precondition group under the test condition, and an increase in misrecognitions for the Total Feedback precondition group under the test condition. Unlike nonrecognitions and total errors, the misrecognitions of the Total Feedback precondition group remained lower than the No Feedback precondition group, even after transfer to the test condition.

Figure 3-7 portrays graphically the interaction of trials with preconditions for misrecognitions. It is apparent that from trial one to trial two, the No Feedback precondition group produced fewer misrecognitions (by about 1.5%) while the Total Feedback precondition group produced more misrecognitions (by about 1%).

TABLE 3-5

ANALYSIS OF VARIANCE SUMMARY TABLE
FOR CHANGE IN MISRECOGNITIONS

Source	df	MS	F
Precondition (P)	1	.125	8.012*
Test Condition	7	.010	.636
PxC	7	.017	1.091
Error	32	.016	
Trials (T)	4	.002	.659
TxP	4	.008	2.732**
TxC	28	.003	1.092
TxCxP	28	.003	.875
Error	128	.003	

*P<.01
**p<.05

TABLE 3-6

MEAN CHANGE IN MISRECOGNITIONS (IN PERCENT)
FROM PRECONDITION TO TEST CONDITION

	Precondition		
Test Condition	No Feedback	Total Feedback	$\bar{x}\Delta$ Test Condition
No Feedback	.90	- .87	.02
Nonrecognition Beep	-1.45	.43	- .51
Nonrecognition and Misrecognition Beep	-1.93	.38	- .77
Different Nonrecogniton & Misrecognition Beeps	- .95	3.57	1.31
Nonrecognition Beep & Verbal Feedback	- .95	.95	0
Visual Feedback	-3.48	.87	-1.31
Mixed Feedback	-1.52	1.65	.07
Total Feedback	-1.03	- .78	- .91
$\bar{x}\Delta$ Precondition	-1.30	.77	Grand $\bar{x}\Delta$ - .26

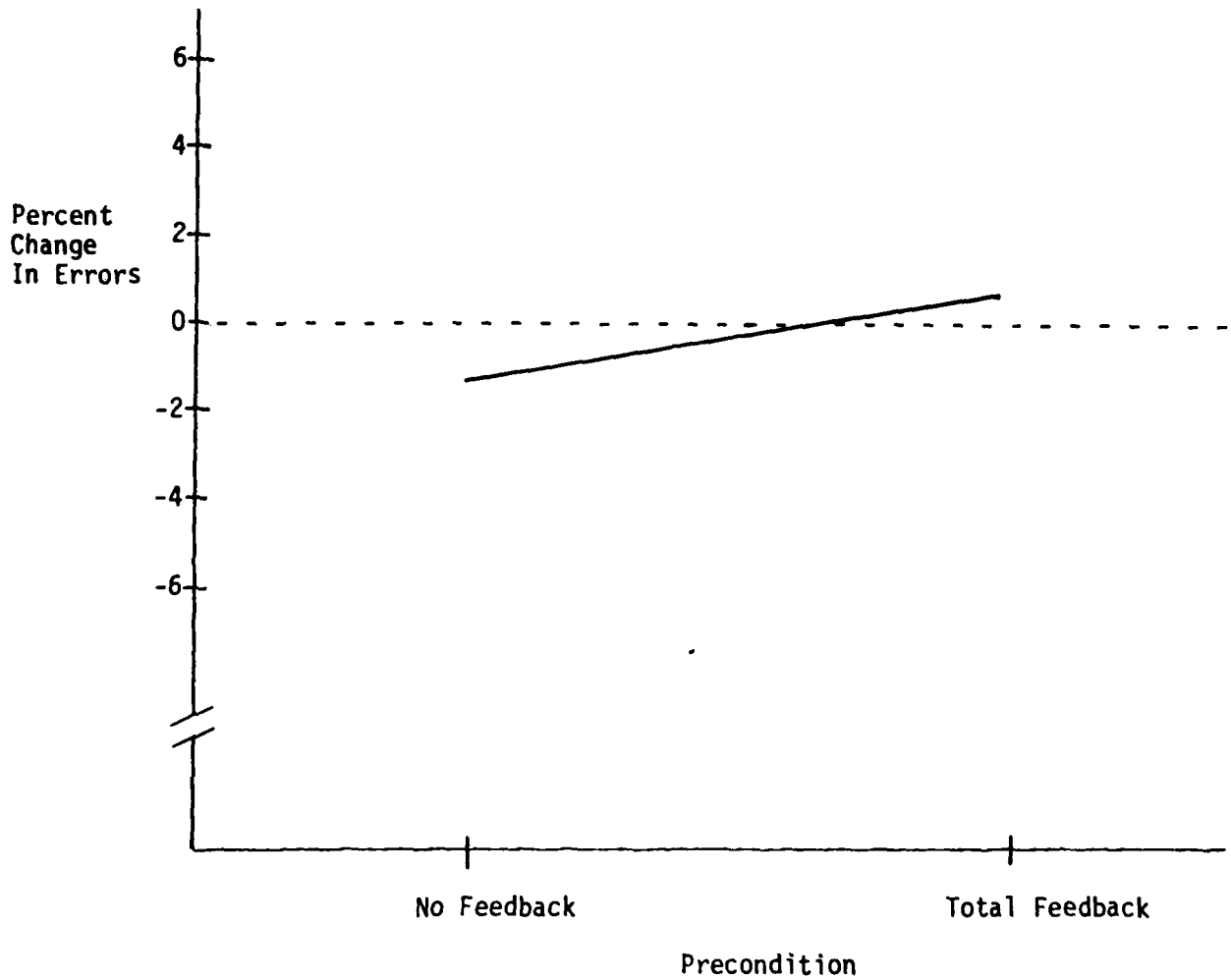


FIGURE 3-5
CHANGE IN MISRECOGNITIONS FROM PRECONDITION TO MEAN TEST CONDITION BY PRECONDITION

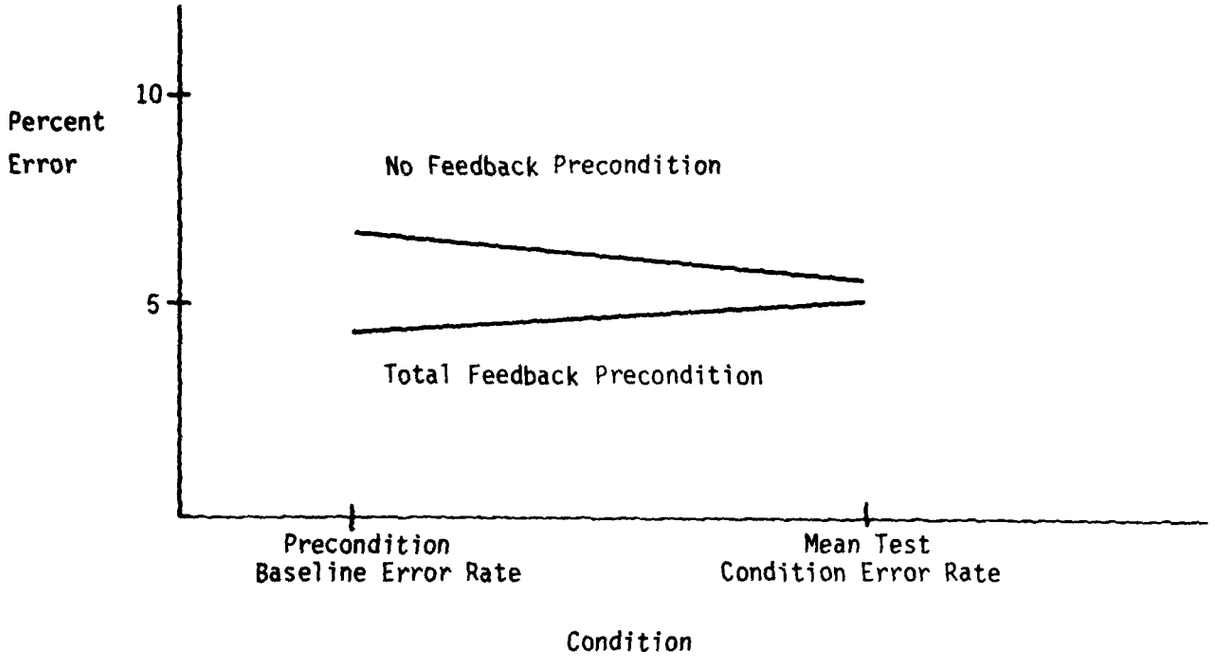


FIGURE 3-6.
MISRECOGNITIONS FOR PRECONDITIONS BY CONDITION

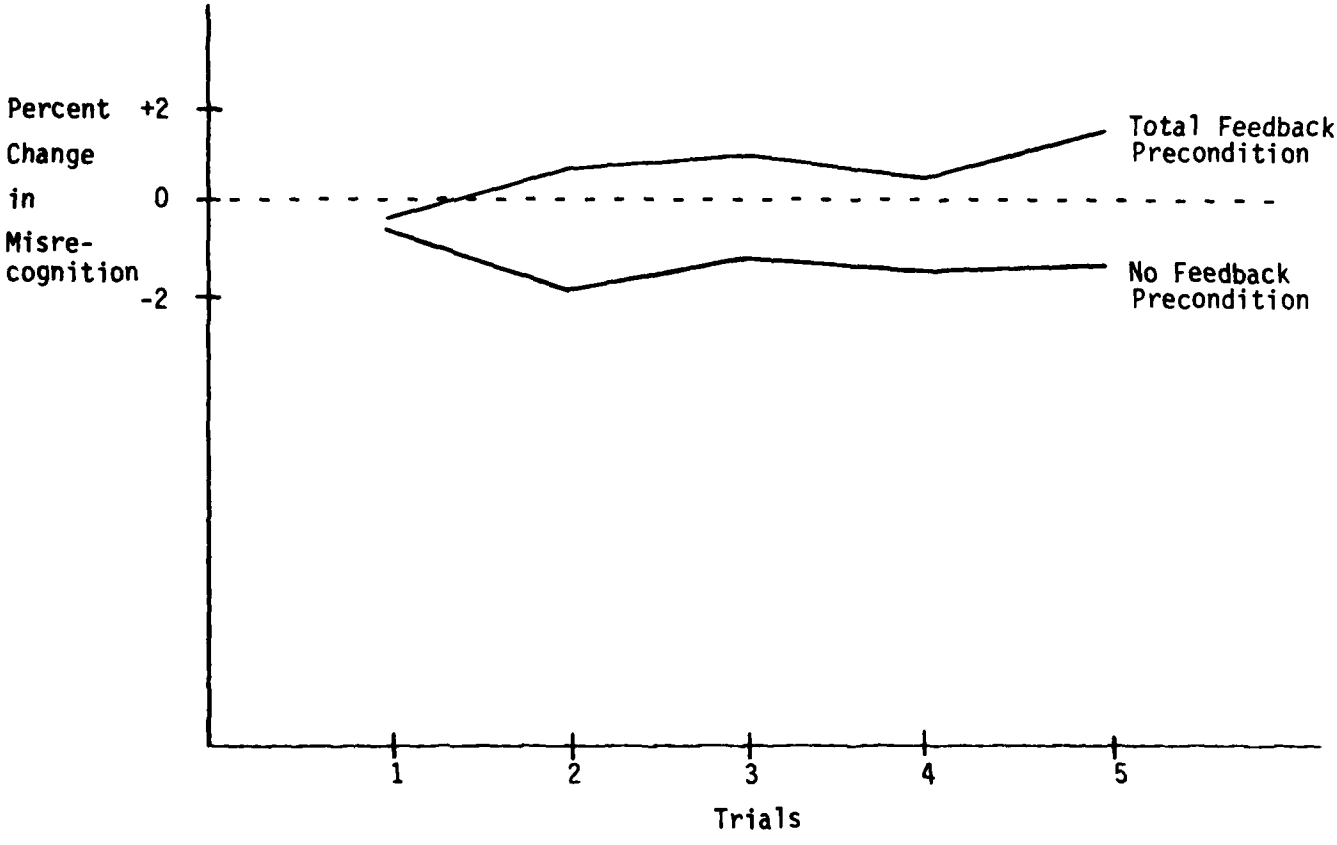


FIGURE 3-7.
INTERACTION OF TRIALS WITH PRECONDITIONS
FOR MISRECOGNITIONS

4. DISCUSSION

Having presented the results of the present study, some implications of those results are now discussed.

4.1 Effect of Precondition

There was a significant difference in the change and direction of change in errors, between subjects preconditioned with No Feedback and subjects preconditioned with Total Feedback. Further, the differences were consistent across nonrecognitions, misrecognitions, and total errors. While subjects from both groups received identical treatments in the test condition, this treatment represented an increase in feedback for the No Feedback subjects and a decrease in feedback for the Total Feedback subjects. Increasing feedback resulted in a reduction of nonrecognitions, misrecognitions, and total errors, for subjects preconditioned with No Feedback while decreasing feedback resulted in an increase in nonrecognitions, misrecognitions, and total errors for subjects preconditioned with Total feedback. Even though misrecognitions increased for subjects preconditioned with Total Feedback, while they decreased for subjects preconditioned with No Feedback, the latter still produced more misrecognitions in the test condition (as indicated by the converging lines in Figure 3-6).

However, nonrecognitions and total errors produced by subjects preconditioned with Total Feedback actually exceeded the reduced number of nonrecognitions and total errors produced by subjects preconditioned with No Feedback (as indicated by the crossing lines in Figures 3-4 and 3-2, respectively).

These results suggest some important considerations for future applications of voice input. First of all, feedback (or lack of feedback) is a contributing factor to error rate. With the equipment used in the present study, total errors increased significantly (about 5%) when feedback was decreased, and when feedback was increased, total errors decreased significantly

(about 5%). As a result, the amount and type of feedback to which a user becomes accustomed, perhaps in training, should not exceed or differ from that which will be used in the actual working situation. Supplemental feedback during training may reduce errors in training, but would be associated with cost (increased errors) rather than benefit (sustained reduction in errors) after transition to the actual work setting.

Recent research at the Naval Postgraduate School has been investigating remote voice input with the user in a room, building, or outside area, away from the VRD and feedback signals. Effective transmission looks promising insofar as hardware capabilities are concerned and the development of this capability will undoubtedly lead to increased remote voice input. However, users accustomed to making voice inputs at the immediate location of the VRD, which usually provides auditory and visual feedback, may face an increase in errors when using a remote system lacking feedback capabilities. Alternatively, the remote system should be equipped with feedback capabilities, or training should be structured so that feedback (if any) is consistent with that available on the remote system.

4.2 Effect of Test Condition

There were no significant differences between any of the 8 test conditions, nor was test condition involved in any significant interactions. As expected, with only 3 subjects from each precondition under each of the 8 test conditions, large discrepancies in error rates would have had to occur to reach acceptable levels for statistical significance. Indeed, the difference between Visual Feedback and Different Nonrecognition and Misrecognition beeps was 9.28%. This seemingly substantial difference was easily negated by high error variance and low degrees of freedom. (Nonparametric tests were also conducted and essentially supported the results of the analyses of variance.) However, to assume (based on the absence of statistical

significance between the 8 test conditions), that feedback has no effect, would be a tenuous conclusion at best. As seen in the case of precondition effects, feedback can have a significant effect. The useful information to come out of the 8 test conditions is simply that there are unlikely to be extremely large differences in performance due to different types of feedback.

4.3 Effects of Trials

There was no significant main effect of trials, but there was a significant interaction of trials with precondition. It may be seen in Figure 3-7 that from trial one to trial two the subjects preconditioned with No Feedback produced fewer misrecognitions while the subjects preconditioned with Total Feedback produced more misrecognitions. It is possible that the No Feedback group learned to reduce misrecognitions from trial one to trial two due to the introduction of feedback beginning in trial one. During the same phase, the Total Feedback group may have shown an increase in misrecognitions due to the withdrawal of some of the feedback to which they were accustomed. However, the absence of a similar interaction for nonrecognitions and total errors suggests that this conclusion is somewhat speculative. In any event, the magnitude of the divergence is so small that the author is led to believe that this effect may be spurious, thus making meaninful interpretation difficult at this time.

5. CONCLUSIONS

The present research has shown that feedback does affect performance in voice recognition. Performance of subjects not accustomed to feedback improved by about 5% when presented with some type of feedback. Subjects accustomed to a lot of feedback produced approximately 5% more errors when feedback was reduced. Without feedback, the user is free to forget various parameters of each utterance as stored in the training file, such as intonation, accented words or syllables, speed of delivery, pitch and range. In this respect it is impressive that the VRD was capable of fairly reliable recognition across feedback conditions.

The VRD chosen for experimentation yielded an average of approximately 25% total errors in the total feedback precondition. Fortunately, the more problematic misrecognitions occured at a rate of only 5%. It should be re-emphasized that these error rates do not reflect the capabilities of all currently available "VRD's." The VRT 100 was employed in this experiment to attempt to avoid the "floor" effect noted previously. One can only speculate as to how feedback would have affected performance using a VRD such as the Threshold T600, but it is reasonable to assume that VRD's that make fewer errors, can recognize greater variations (changes in intonation, pitch, etc.) in each utterance, while VRD's that require less variation for accurate recognition rely more on feedback to direct the user's speech. Interestingly, in a recent study the T600 produced only 2.67% total errors with a 240 utterance vocabulary that included 98 of the 100 utterances used in the current study (Poock, 1981). Accordingly, the importance of feedback should be determined by the capabilities of the particular VRD, and the cost of errors.

Still, errors are undesirable no matter how infrequent or how minute the consequences. The current study has shown that a consistent form of feedback can reduce errors, and should be provided when possible. The results were less conclusive concerning different levels and types of feedback provided, but suggested no large differences in performance as a function of these variables.

6. REFERENCES

Calcaterra, F.S. Applications of artificial intelligence in voice recognition systems in micro-computers, Masters thesis at Naval Postgraduate School, Monterey, CA, March 1982.

Interstate Electronics Corporation, Voice Recognition Terminal Model VRT-101, Operation and Maintenance Manual TM P00700298, November 1981.

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

Poock, G.K., Schwalm, N.D., and Roland, E.F. Wearing Protective Masks: Effects on Voice Recognition System Performance. Proceedings of the Voice Data Entry Systems Applications Conference, September 1982.

Schwalm, N.D., Martin, B.J., Poock, G.K. and Roland, E.F. Trying for speaker independence in the use of speaker dependent voice recognition equipment. Naval Postgraduate School, Monterey, California, Report No. NPS55-82-032, December 1982.

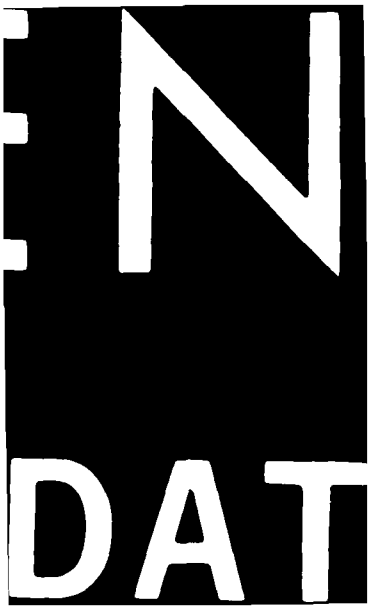
APPENDIX A

1. ONE
2. NINE
3. MOVE IT RIGHT
4. GARY POOCK
5. SPEECH RECOGNITION
6. LOAD G L D3
7. EUROPE
8. LOAD THE GANN
9. VIETNAM
10. KITTY HAWK
11. EFFICIENT TRANSMISSION
12. LEVEL TWO
13. BANGKOK
14. YANKEE
15. CONNECT TO CHARLIE
16. XRAY
17. DIEGO GARCIA
18. TOKYO
19. SAVE
20. LOAD THE SERVER
21. BLUE FORCE ONE
22. KILO
23. RADIOLOGY
24. BOMBAY
25. HONOLULU
26. ARKANSAS
27. BUSINESS MEETING
28. SEA OF JAPAN
29. PACIFIC DATA BASE
30. IRAN
31. RANGOON
32. WHISKEY
33. BRISBANE
34. YOKOHAMA
35. HOLLISTER
36. ADVISORY
37. INDIA
38. BANGLADESH
39. VICTOR
40. IBERIAN CARRIER
41. HOTEL
42. VLADIVOSTOK
43. TANGO
44. PLOT ALL SUBMARINES
45. NAPLES
46. UNITED AIRLINES
47. ACCAT TITLE
48. QUEBEC
49. STRAIGHT OF HORMUZ
50. ANTWERP

51. CONTINUOUS SPEECH
52. JAPAN
53. EIGHT
54. INTERACTIVE
55. GOLF
56. LIMA
57. DROP
58. OSCAR
59. ARABIAN TANKER
60. CHANGE DIRECTOR TO MARTIN
61. KRONOMETR
62. PORTLAND
63. IDENTIFICATION
64. PERCEPTRONICS
65. LOGIN POOCK
66. CARRIAGE RETURN
67. ASPRO
68. SCOPE
69. AFRICA
70. USER'S GUIDE
71. CALCUTTA
72. MAINE
73. SWEDEN
74. SUITABILITY
75. POPPA
76. SAIGON
77. CANTON
78. SYSTEM INTEGRATION
79. ZULU
80. AUTOMATIC RECOGNITION
81. JOHN KENNEDY
82. ADVANTAGES
83. WYOMING
84. CRITERIA
85. RED FOX
86. BALTIMORE
87. AIR ROUTES
88. CONTINUOUS
89. MOVE IT UP
90. KOREA
91. UNIFORM
92. INDONESIA
93. WEST GERMAN TORPEDO
94. DOWN IN DETAIL
95. KIEV
96. ACAPULCO
97. POOCK N P S PASSWORD
98. MIKE
99. TWO
100. CORRECTION

DISTRIBUTION LIST

	No. of Copies
COL Paul Cerjan 9th Infantry Division Fort Lewis, WA 98433	2
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	2
Professor Gary Poock Code 55Pk Naval Postgraduate School Monterey, CA 93940	150


N
DAT