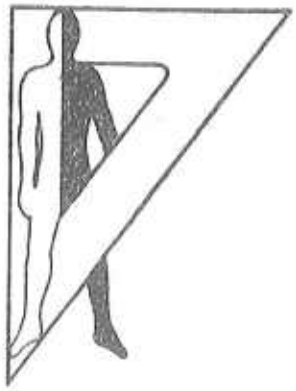

Technical Memorandum 7-83

THE EFFECTS ON COMPUTER RECOGNITION OF SPEECH WHEN SPEAKING THROUGH PROTECTIVE MASKS

Frank J. Malkin

May 1983
AMCMS Code 612716.H700011

U. S. ARMY HUMAN ENGINEERING LABORATORY

Aberdeen Proving Ground, Maryland

Destroy this report when no longer needed.
Do not return it to the originator.

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Use of trade names in this report does not constitute an official endorsement or approval of the use of such commercial products.

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Memorandum 7-83	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* THE EFFECTS ON COMPUTER RECOGNITION OF SPEECH WHEN SPEAKING THROUGH PROTECTIVE MASKS		5. TYPE OF REPORT & PERIOD COVERED Final 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Frank J. Malkin		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS US ARMY HUMAN ENGINEERING LABORATORY ABERDEEN PROVING GROUND, MD 21005		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS AMCMS Code 612716.H700011
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE May 1983 13. NUMBER OF PAGES 16
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)*

Speech Recognition
Protective Mask
Computer Recognition of Speech
Advanced Speech Technology
Microphone
Aviation
Human Factors

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

The purpose of this experiment was to investigate the effects on computer recognition of speech when speaking through aviator protective masks as compared to the standard boom-mounted microphone.

Twelve Army aviators were tested with a 50-word vocabulary under three conditions: no mask (boom-mounted microphone), M24 aviator protective mask which has an internally mounted microphone, and the XM33 developmental aviator protective mask which has an externally mounted microphone.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

ABSTRACT (cont'd)

Although there was no significant difference in performance between the boom-mounted microphone and either of the aviator protective masks, there was a significant difference between masks. The M24 mask experienced a significantly lower error rate than did the XM33.

The performance of the speech recognition system is discussed in terms of the three test conditions and the experience level of the speakers.

The conclusions were (1) of the two masks designs, the M24 mask with the internally mounted microphone provided better results with the speech recognition system and (2) although there was no significant difference in performance between the boom-mounted microphone and the protective masks, there was an indication of inconsistent performance obtained with the boom-mounted microphone which was not clearly understood and warrants further investigation.

AMCMS Code 612716.H700011 Technical Memorandum 7-83

THE EFFECTS ON COMPUTER RECOGNITION OF SPEECH WHEN SPEAKING THROUGH PROTECTIVE MASKS

Frank J. Malkin

May 1983

APPROVED: _____
JOHN D. WEISZ
Director
U.S. Army Human Engineering Laboratory

U.S. ARMY HUMAN ENGINEERING LABORATORY
Aberdeen Proving Ground, Maryland 21005

ACKNOWLEDGEMENT

The author is grateful to Mr. Harry J. Reed for performing the computer programming in support of this study and to Dr. Jon J. Fallesen for his assistance with the Biomedical Computer Program and technical advice.

CONTENTS

THE EFFECTS ON COMPUTER RECOGNITION OF SPEECH WHEN SPEAKING THROUGH PROTECTIVE MASKS

INTRODUCTION

Computer technology is being introduced into US Army helicopter cockpits purportedly to decrease crew workload. Currently, the primary means of interacting with a computer is through a keyboard or function keys of some type. Helicopter crew members flying below tree-top level in order to avoid enemy detection do not have "eyes and hands free" for keyboard operations.

Computer recognition of speech is being considered as an alternative means of computer operation because it permits the operator to interact with onboard computers while leaving the eyes and hands free to perform primary flight and navigation tasks. The pilot "talks" to the computer and it "understands."

However, there are issues related to airborne applications which need to be investigated (Lea, 1980).

One of these issues is the use of protective masks. In view of the enemy chemical warfare threat, it is likely that helicopter crew members will be required to wear protective masks in flight. The microphone in the current M24 aviator mask is located directly in front of the diaphragm inside the mask. In the XM33 developmental aviator mask, the microphone is located behind the diaphragm outside the mask. What is the effect on computer recognition of speech when the crew member speaks through one of these masks as opposed to the standard Army aviation boom-mounted microphone?

The Naval Postgraduate School, Monterey, California, has recently conducted studies using a stenographer mask and tank crew mask (Poock et. al., 1982). The conclusion of these studies was that although the use of masks contributed to an increase in the percent of errors made, this increase in errors may be mitigated to an extent when the individuals speaking into the speech recognition system have experience with masks or microphones.

OBJECTIVE

The objective of this study was to determine the effect of three mask conditions (no mask, M24 mask, XM33 mask) on speech recognition accuracy when the speakers are experienced aviators.

METHOD

Subjects

Twelve male Army aviators assigned to Aberdeen Proving Ground, Maryland, participated in this study. The mean age was 36 with a range from 28 to 47. The mean experience level in flight hours was 3,627 with a range from 1,100 to 6,000. None of the aviators had previous experience using speech recognition equipment.

Apparatus

The following apparatus was used:

a. Interstate Electronics VRT 103, Voice Recognition Terminal (Figure 1).

This is a speaker-dependent voice recognition system which requires that, prior to use, the system be provided with a sample of how each user pronounces the utterances in a predetermined vocabulary. This is referred to as training the system. Each sample is stored in memory as a reference for later comparisons. When in use, the system recognizes words by comparing current utterances with the samples stored in memory and selecting the closest match.

It is referred to as an isolated word recognition system because the longest utterance it can sample is 1.25 seconds in duration, and a pause is required between each vocabulary item.

Levels for Gain, Reject Threshold, and Delta Value can be selected (Interstate Electronics Corp., 1982).

The VRT-103 has four different input gain levels. The first level has the least amount of gain (for louder voices) and the fourth level has the most amount of gain. For this experiment, the gain was set at level two--the setting for average speakers. Using this setting, the system was not sensitive to natural breathing sounds inside the masks.

The Reject Threshold Level and Delta Value are used to select the closeness of fit that is required between the current utterance and the stored sample in order for the utterance to be classified and accepted as a vocabulary word. The Reject Threshold Level and Delta Value were set at zero. This allowed all inputs into the microphone to be classified as vocabulary words. There were no rejections of utterances.

b. Headset with M87 Microphone (Figure 2).

The M87 microphone is the standard Army aviator boom-mounted microphone used with headsets and flight helmets.

Figure 1. Aviator seated at the Interstate Electronics VRT 103.

Figure 2. Headset.

c. M24 Aviator Protective Mask with M101 Microphone (Figure 3).

This is the standard mask currently used by Army aviators. The microphone is mounted inside the mask directly in front of the mouth.

d. XM33 Developmental Aviator Protective Mask with M101 Microphone (Figure 4).

The microphone is mounted behind the diaphragm outside the mask. The voice is emitted through a flapper valve to the microphone.

The characteristics of the M87 microphone used with the headset and the M101 microphone used with the masks are generally the same. However, the M101 is designed for mask installation.[1]

Procedures

Each subject was given a briefing concerning the purpose of the experiment and the procedures to be followed. Computer recognition of speech was explained, and the advantages and disadvantages of potential airborne applications were discussed.

The subject practiced using the speech recognizer by training and testing a sample six-word vocabulary. Words used in the practice vocabulary were not used in the actual test vocabulary. This practice session was conducted using the headset with the boom-mounted microphone.

Because the Interstate VRT 103 is speaker-dependent, each subject was required to train the recognizer to his voice with the selected vocabulary prior to testing. The vocabulary consisted of 50 words of the type that might be used in Army helicopter applications (see Appendix A). The 50-word vocabulary list received five training passes by each subject in each mask condition. This is the number of training passes recommended by the manufacturer.

Testing of the speech recognition system consisted of four trials per mask condition; that is, the 50-word vocabulary list was repeated four times by each subject under each mask condition. The subject was prompted by scrolling the vocabulary words on the video screen of the voice recognition terminal. After each utterance, the display scrolled to the next word. No feedback was provided to the subject as to whether or not the word spoken into the microphone was recognized by the system. The microprocessor within the voice recognition terminal was programmed to detect and maintain a record of words recognized and words misrecognized for each subject on each trial under each mask condition. This information was later made available for print-out.

---

[1] Mayer, M. Telephone Communication, November 1982.

Figure 3. M24 aviator protective mask.

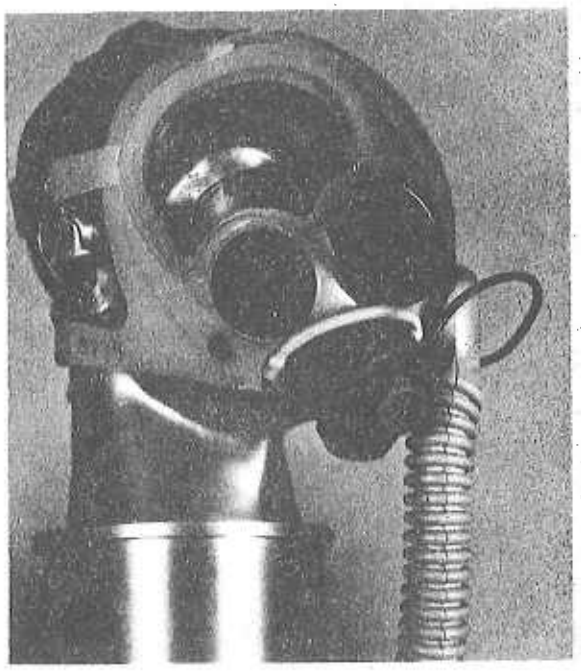

Figure 4. XM33 aviator protective mask.

Each subject completed training and testing with the speech recognizer under one mask condition before proceeding to the next. In order to control for ordering effects, mask conditions were counterbalanced. Between training and testing, the subject removed the headset or mask for a short "breather" (5 minutes or less). To preclude any changes in voice patterns, the subject remained seated and was not permitted to smoke, drink, or eat.

When using the boom-mounted microphone, the subject was instructed to position the microphone as it normally would be during actual flight. All subjects positioned the microphone directly in front of the mouth barely brushing the lips. The subject was also instructed to place the microphone in the same position for both training and testing.

Rest periods of 10 to 15 minute duration were provided between mask conditions during which time subjects were permitted to smoke and drink.

The total time required to train and test each subject was approximately 2-1/2 hours.

Design

A 3 x 4 x 12 (mask condition x trials x subjects) factorial design with repeated measures was used. All subjects were tested with the voice recognition system under all mask conditions. The presentation order of the mask conditions was counterbalanced.

The independent variable was mask condition, and the dependent variable was the percent of utterances misrecognized (error rate).

RESULTS

The data analyses were performed on the percent of utterances misrecognized. The percentage was calculated by the ratio of the number of misrecognitions to the number of utterances.

The Biomedical Computer Program (BMDP) (Jennrich & Sampson, 1979) was used to perform the analysis of variance (ANOVA).

A compound symmetry test indicated that no adjustment to the degrees of freedom was required (Horton, 1979, pp. 155-158) and suggested that there was no need for transformation of the proportional data.

The calculated means and standard deviations are shown in Table 1.

TABLE 1

Mean Precent Misrecognitions

	Mask Condition		
	No mask	M24	XM33
Mean	12	8.17	15.17
Standard Deviation	7.8	6.7	6.2

The results of the ANOVA are shown in Table 2. A significant main effect is indicated for mask condition; $F(2, 22) = 6.79$, $p < .005$. There is no significant effect for trials, nor is there a significant interaction.

TABLE 2

Analysis of Variance Summary Table

Source	DF	MS	F
Subjects (S)	11	.0397	
Mask (M)	2	.0590	6.79*
S x M	22	.0087	
Trials (T)	3	.0028	.89
S x T	33	.0032	
M x T	6	.0014	.57
S x M x T	66	.0024	

*$p<.005$

In view of the main effect for mask condition, a Scheffe' test (Keppel, 1973, pp. 208-211, p. 430) was performed for comparisons among the pairs of means. The results indicated that there is no significant difference in misrecognition error rates between the boom-mounted microphone and either of the protective masks. However, there is a significant difference between the two protective masks; CR=3.086, $p.<.01$.

Figure 5 shows the mean error rate across the four trials for each subject and mask condition.

DISCUSSION

Mask Performance

The performance of the boom-mounted microphone with the speech recognizer did not differ significantly from that of the protective masks. Because different masks and microphones were used, direct comparisons between the findings of this study and that of the Naval Postgraduate School studies are not appropriate. However, the results tend to support the conclusion of the Naval Postgraduate School that the increase in errors in the masked condition may be mollified by experience with speaking into masks or microphones.

Curiously, the performance of the boom-mounted microphone was rather inconsistent. Looking at the percent misrecognitions for subjects and mask conditions (see Figure 5), it appears that both the M24 and XM33 masks provided more consistent performance with the speech recognition system than did the no-mask condition. Except for one or two outliers, the "scores" for the two masks are clustered in a relatively tight grouping when compared with the no-mask condition. The reason for this is not clear.

It is interesting, though, that there is somewhat of a clustering of the seven subjects under the no-mask condition who had misrecognition rates of 10 percent or less.

The question remains as to why speech recognition performance, in the context of this study, is less consistent when speaking through the boom-mounted microphone. It can be conjectured that one or a combination of the following points contributed to this inconsistency.

a. The masks attenuate noise. Although the experiment was conducted in a relatively quiet laboratory setting and the boom-mounted microphone is noise-cancelling, it may be possible that obtuse noises or reverberations off the walls and floor may have affected performance of the microphone with the speech recognition system.

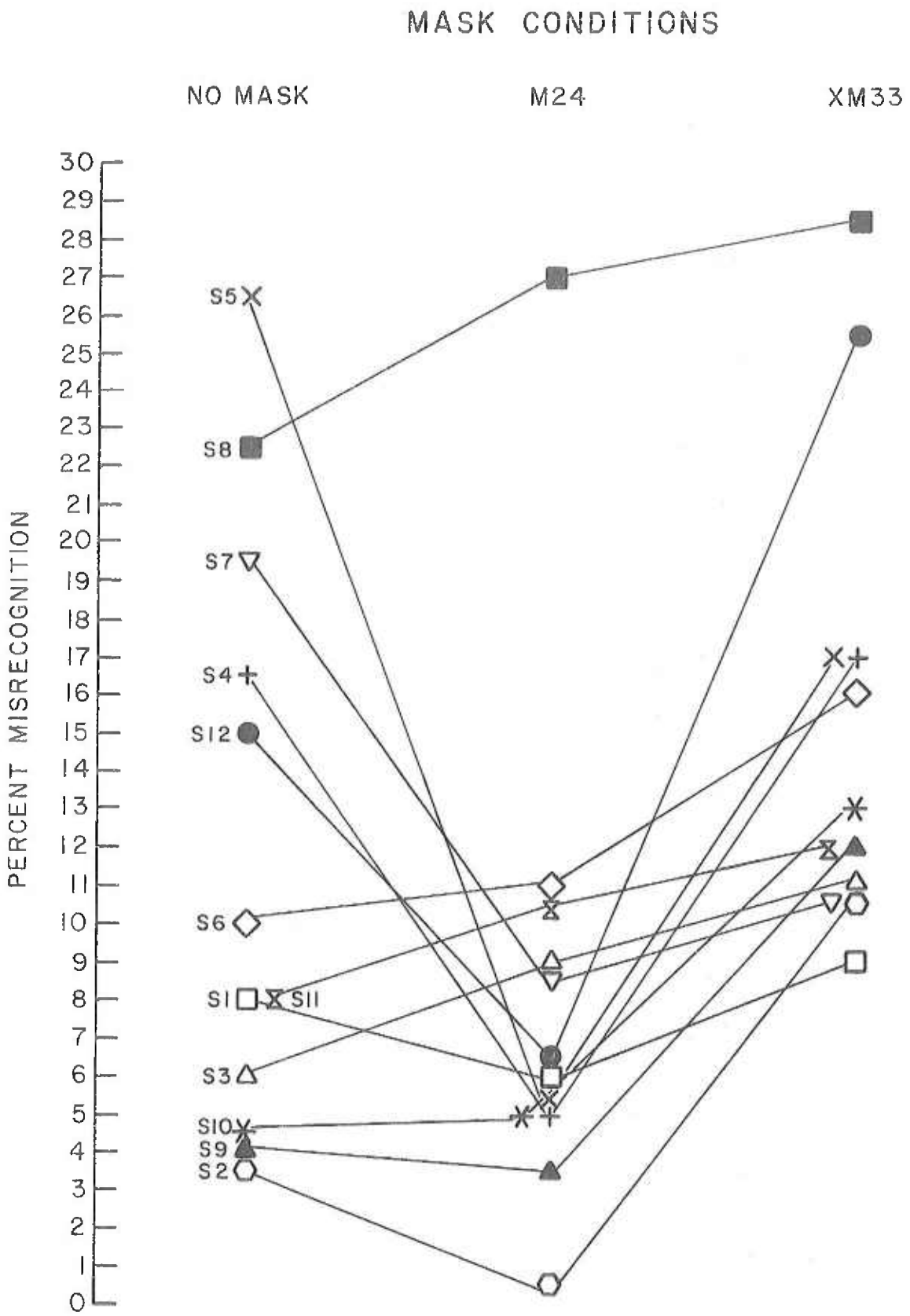

Figure 5. Misrecognitions (mean percent across trials) for each subject and mask condition.

b. In order to obtain good recognition accuracy rates, the microphone should be in the same relative position for both training and testing. Although this was controlled under all three conditions, it may possibly have been acomplished in a more precise manner with the two masks as a function of fitting the mask to the face.

c. There may be some differences between the M87 and M101 microphones that were not anticipated.

Recognition Accuracy Rates

The mean recognition accuracy rates (as opposed to the misrecognition error rates) obtained during this experiment, in round figures, were 92% with the M24 mask, 88% with the boom-mounted microphone, and 85% with the XM33 mask.

In view of the Interstate Electronics VRT-103 capability of achieving recognition accuracy rates of 99% or better, the recognition accuracy rates reported above may be questioned. However, in addition to the speech recognizer, the recognition accuracy rate is dependent upon several other variables, such as the speaker, the microphone, and the vocabulary.

In this case, although experienced with speaking into microphones, the speakers were using speech recognition equipment for the first time. The microphones were designed for use in military aircraft and are not necessarily ideal for use with speech recognition systems. Prior to the experiment, the vocabulary had received only a cursory examination to determine if any of the words would present recognition difficulty to the system.

However, Subject 2 obtained a 99.5% recognition accuracy rate when speaking through the M24 aviator protective mask. The author also obtained a 99% recognition accuracy rate with the vocabulary when using a commercially available Shure Brothers SM10 microphone.

In the Naval Postgraduate School stenographer mask study, a technique was used in which, immediately after training, subjects made two test passes on the vocabulary to identify any problems in the training of any particular utterance. If the system responded correctly on those two passes, the utterance was considered adequately trained. If errors occurred, a third pass was made. When less than two of three passes of any utterance was correct, the utterance was retrained.

If the retraining technique just described had been used in the present study, it is possible that there would have been a decrease in the percent of utterances misrecognized resulting in higher recognition accuracy rates for all mask conditions.

CONCLUSIONS AND RECOMMENDATIONS

Of the two mask designs, the M24 with the microphone installed inside the mask provided better results with the speech recognition system. The 92% recognition accuracy rate obtained with the M24 mask is encouraging. However, research is required to determine the degree of accuracy considered acceptable for airborne applications.

Although there was no significant difference in performance between the boom-mounted microphone and either of the protective masks, a comprehensive evaluation is needed in order to more fully understand the relatively inconsistent performance experienced with the M87 boom-mounted microphone.

The Army has an improved microphone under development which initially will be used in the AH-64 attack helicopter. An improved protective mask designed specifically for the AH-64 is also under development. When these new items become available, they should be tested with speech recognition equipment.

REFERENCES

Horton, R. L. The General linear model. New York: McGraw-Hill, Inc., 1979.

Interstate Electronics Corp. Voice recognition terminal Model VRT-101 operation and maintenance manual. Anaheim: Author, 1982.

Jennrich, R., & Sampson, P. Analysis of variance and covariance, including repeated measures. In W. J. Dixon and M. R. Brown (Eds.), Biomedical Computer Programs P Series. Berkeley: University of California Press, 1979.

Keppel, G. Design and analysis: A researchers handbook. Englewood Cliffs: Prentice-Hall, Inc., 1973.

Lea, W. A. Critical issues in airborne applications of speech recognition (Final Task Report NADC, DTIC No. AD A084703, Contract Number N62269-78-M-3770). Warminister, PA: Naval Air Development Center, May 1980.

Poock, G. K., Schwalm, N. D., & Roland, E. F. Wearing protective masks: Effects on voice recognition system performance. In L. Lerman (chair), Voice Data Entry Systems Applications Conference, 1982.

APPENDIX A

VOCABULARY

COURSE	CHARLIE
THREE	SIX
DISPLAY	HEADING
DELTA	CODE
ENGINE	RANGE
EIGHT	AREA
GOLF	FOUR
SELECT	BRAVO
ECHO	HYDRAULIC
SYSTEM	ATTACK
FARP	WAYPOINT
HOTEL	ENTER
STATUS	BATTLE
POSITION	FOXTROT
BASE	ALPHA
LEADER	STEER
FIVE	DESIGNATE
LASER	ASSEMBLY
TWO	SEVEN
TRANSMISSION	ENGAGE
TARGET	HOLD
MISSILE	NINER
TRACKER	ZERO
FUEL	SCOUT
ONE	DELETE