AD-A040 994 THRESHOLD TECHNOLOGY INC DELRAN N J F/G 17/2
VOICE INPUT CODE IDENTIFIER (VICI).(U)
JUN 77 P B SCOTT F30602-75-C-0111
UNCLASSIFIED RADC-TR-77-190 NL

1 OF 1
AD
A040994

END
DATE
FILMED
7-77

RADC-TR-77-190
Final Technical Report
June 1977

# VOICE INPUT CODE IDENTIFIER (VICI)

Threshold Technology, Inc.







ROME AIR DEVELOPMENT CENTER
Air Force Systems Command
Griffiss Air Force Base, New York 13441

This report has been reviewed by the RADC Information Office (OI) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public including foreign nations.

This report has been reviewed and is approved for publication.

APPROVED: *Richard S. Vonusa*
RICHARD S. VONUSA
Project Engineer

APPROVED: *H Davis*
HOWARD DAVIS
Technical Director
Intelligence & Reconnaissance Division

FOR THE COMMANDER: *John P. Huss*
JOHN P. HUSS
Acting Chief, Plans Office

Do not return this copy. Retain or destroy.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

# REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

| | | |
|---|---|---|
| 1. REPORT NUMBER: RADC-TR-77-190 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle): VOICE INPUT CODE IDENTIFIER (VICI) | | 5. TYPE OF REPORT & PERIOD COVERED: Final Technical Report |
| | | 6. PERFORMING ORG. REPORT NUMBER: N/A |
| 7. AUTHOR(s): Phillips B. Scott | | 8. CONTRACT OR GRANT NUMBER(s): F30602-75-C-0111 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS: Threshold Technology, Inc. Delran NJ 08075 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 62702F 40270509 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS: Rome Air Development Center (IRAP) Griffiss AFB NY 13441 | | 12. REPORT DATE: June 1977 |
| | | 13. NUMBER OF PAGES: 59 |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Same | | 15. SECURITY CLASS. (of this report): UNCLASSIFIED |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE: N/A |

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)

Same

18. SUPPLEMENTARY NOTES

RADC Project Engineer: Richard Vonusa (IRAP)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Speech Recognition
Pattern Recognition
Acoustic Phonetics

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

This report describes the development, operation and performance characteristics of an improved Advanced Development Model of a Voice Input Code Identifier (VICI). The VICI is an isolated word recognition system capable of recognizing the English digits and four control words, CANCEL, ERASE, VERIFY and TERMINATE. The improved VICI system is based upon equipment previously developed under Contract F30602-74-C-0171. This original VICI system was intended for use by male talkers only and required wide bandwidth speech data input. The

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

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

390 998

advanced development model is capable of operation with input speech which has been limited to telephone bandwidth. It will accept digits and control words as spoken by either male or female talkers. No training of the system by a speaker is necessary. By the use of an alpha-numeric output display, a speaker using the system can verify that each digit spoken into the system was correctly recognized. Errors can be corrected through the use of the control words. In addition, the system can detect and correct many single digit errors which occur within six digit code groups. This error correction algorithm uses check digits which are included in the code groups to detect errors. The system then corrects the errors when possible or requests a reentry of the data by the talker.

To confirm system performance, final tests were made by the use of tape recorded digits and control words and by a group of talkers directly inputting data. Individual digit accuracy in the tests conducted by the use of tape recordings was 96.85 percent for 182 talkers. Tests of control words spoken by 37 talkers resulted in an accuracy of 95.74 percent. In both types of tests all vocabulary words were viable. In final tests at RADC eight talkers directly inputting 300 digits each had a combined accuracy of 95.4 percent. All tape recorded data were passed over actual telephone loops which included two centrals and a connecting trunk as well as lines to and from centrals. The system was tested with a total of over 56,000 words spoken by 193 male and female talkers.

A speaker dependent software package also was developed which provides for the recognition of up to 200 words. This software is a structured vocabulary program which allows recognition of up to 30 words in any node of the structure. Up to 30 nodes can be included in the sentence structure. This program operates in the VICI system.

TABLE OF CONTENTS

## LIST OF ILLUSTRATIONS



## LIST OF TABLES

EVALUATION

This report represents a significant achievement in the area of automatic speech processing. It has proved that it is possible to attain high word recognition scores in real time using a limited vocabulary with words spoken in a discrete manner and independent of both male and female speakers regardless of geographic accent. In addition the system has the capability to successfully recognize the vocabulary over telephone bandwidth speech.

The Voice Input Code Identifier (VICI) system has the capability of recognizing the English digits and several command words independent of speaker. The purpose of this automatic word recognizer is to develop a front-end for the Base and Installation Security System's (BISS) automatic speaker verification system. In this manner the BISS requirement for a completely voice-oriented technique for a requester to claim his identity will be fulfilled. The VICI subsystem would eliminate the need for picture badges, keypunching code numbers, and other fallible mechanical methods of entering an identification number. The speaker would simply utter his code numbers (sequence of four digits and one or two check digits) and if correctly entered into the system automatic speaker verification (ASV) would then be performed by having the speaker utter a group of key phrases which would be compared to his reference file. Based on this comparison the speaker is either verified or rejected as an impostor to the ASV system.

Richard S. Vonusa
RICHARD S. VONUSA
Project Engineer

## Section I

## BACKGROUND AND INTRODUCTION

A very accurate spoken digit recognition system previously was developed for the Air Force by Threshold Technology Inc. (TTI) under contract F30602-74-C-0171.[1] This system was capable of recognizing English digits spoken in isolation by male talkers with an average accuracy of nearly 98 percent. This was achieved without any type of adaptation to any talker's voice. The Voice Input Code Identifier (VICI), as the digit recognition system has been called, was developed to provide a front-end for the Base and Installation Security System's (BISS) automatic speaker verification system. This was accomplished in order to provide a voice oriented system to allow a person requesting base entry to claim his identity and be verified. A complete voice entry system could obviate the need for picture badges, keypunching of code number and other fallible mechanical identify verification. The original VICI system development was based upon the TTI VIP-100 isolated word system. This system normally requires training (adaptations) for each talker.

The original VICI system while achieving high accuracy was limited to male talkers and required high quality input speech with a bandwidth of 200 to 8 kHz. For an operational application in an actual BISS system, a VICI system must recognize digits spoken by female personnel as well as males. Also, it may be necessary to locate the major portion of the VICI system at a considerable distance from the input microphone with the connection between the input and recognition system provided by local telephone lines. Because these important constraints tend to reduce recognition accuracy the use of error correction could be helpful. During the effort reported herein, the VICI system was modified to accommodate female as well as male talkers, to allow good accuracy with speech input limited to telephone bandwidth and to provide a measure of error correction by the use of check digits. Extensive modifications have been made in both software and hardware to achieve these improvements. The software operates in less than 8K of core memory in the Nova 1200 computer which is included in the system.

Also developed during this program was a software system with the capability of recognizing up to 200 words as spoken by a particular operator. This system was designed to recognize up to a maximum of 30 words in any node in a syntactic structure. Up to 30 nodes are possible with any arrangement of the 200 vocabulary words. The 200 word software system is speaker dependent but will run on the VICI system provided that the operator trains the system for his or her voice for the vocabulary. The system modification to allow speaker-independent VICI operation did not preclude operation as a speaker dependent system. It has been necessary for TTI to supply 8K of additional core memory to implement this software on the VICI system.

Section II of this report describes the original wideband VICI system followed by the modifications in both hardware and software to allow operation over telephone bandwidth lines, to accommodate female as well as male talkers and to provide error correction by the use of check digits. A description of final system tests both live and from tape is included in Section III.

Conclusions and recommendations are listed in Section IV.

## Section II

## TECHNICAL DISCUSSION

### A. Introduction

In order to expand and improve upon the capabilities of the VICI digit recognition system, previously developed for the Air Force by TTI, a series of modifications based on expanded studies have been performed. Principal modifications performed upon the VICI system include redesign of feature recognition networks to accommodate both male and female speakers as well as provide for operation with telephone bandwidth speech, development of new master reference arrays suitable for recognition of both male and female speech, development of an error correction algorithm which operates by the use of check digits added to the basic four digit code, and development of a structured 200 word speaker-dependent program which will operate on the VICI system without hardware modifications. In order to test the effectiveness of the modifications made to the system, tape recordings were made of nearly 200 talkers both male and female speaking digits in code groups. Extensive testing was conducted by the use of large number of speakers from this group of recordings for each test series throughout the development program. A special set of transmitting and receiving modules was developed which allowed testing over actual telephone line connections. Both terminals provided complete electrical isolation from the telephone line by the use of input and output transformers so that there would be no disturbance of normal telephone service by connection of the terminals. All testing throughout the major portion of the program was conducted by the use of speech data passed either directly over telephone lines by the use of these terminal modules or by tape recorded speech previously passed over the telephone lines by the use of these modules. In the following paragraphs, a review of the previous program leading to the development of the original VICI system is presented first. Considered in succession are problems encountered in universal talker recognition over telephone line bandwidths connections, recognition of both male and female talkers on the system previously designed for male talkers only, and the development of an algorithm which accomplished error correction by the use of check digits.

### B. Previous VICI Experiments

The automatic speech recognition system developed for the Air Force by TTI under the first VICI program (contract F30602-74-C-0171) was based on the VIP-100 speech recognition system manufactured by TTI for commercial applications. The VIP-100 was originally designed to recognize a vocabulary essentially unrestricted in content but restricted in size by the storage limitations of the core memory of the associated minicomputer. The VIP-100 is a speaker dependent isolated word recognition system which utilizes a training (adaptation) routine for individual speakers and words. VIP-100 system consists of a speech preprocessor and feature extraction section in which all processing is done in hardware and a classifier in which further processing is performed in software by a Data General Nova 1200 minicomputer. A block diagram of the VIP-100 system is shown in Figure 1. The minicomputer also time normalizes word durations and provides core storage for the reference

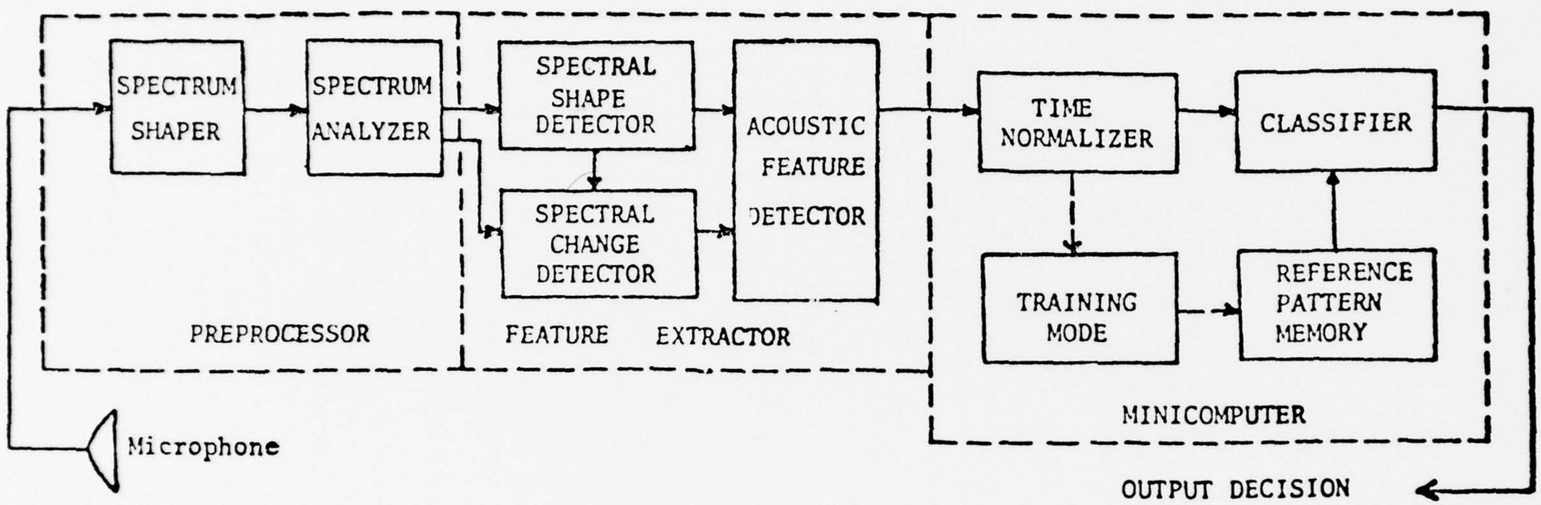



Fig. 1 Block diagram of VIP-100 speech recognition system.

arrays necessary for recognition of each word in the vocabulary. A detailed description of the operation of the VIP-100 is included in the final report for the above referenced contract.[1] In order to achieve the goal of the first VICI program of recognition of digits and control words by an unlimited set of male talkers (high quality speech), it was necessary to perform extensive modifications in the hardware and the software associated with the VIP-100 system.

A principal series of investigations during the development of the original VICI system was concerned with achieving a universal reference array set which would be applicable to the recognition of single digits and for control words as spoken by any male talker with little or no training by each user. In the normal speaker-dependent operation of the VIP-100 system, reference arrays of acoustic features are established for each word in the vocabulary set by the person using the system at any particular time. These reference arrays are generated during an adaptation or training phase in which the speaker preparing to use the system pronounces several (usually 10) repetitions of each word in order to insure maximum recognition accuracy. During the recognition of an unknown input word a feature array is established for the input word. A correlation process is then used to decide which reference array most closely resembles the array caused by the input word. The highest correlation score, if above a fixed threshold, indicates which word was spoken.

In order to develop a system which will accept identification codes as spoken by any male talker it was necessary to greatly modify the procedure used to establish reference data. It has been observed that a single training word for each vocabulary word is often adequate for good accuracy with the VIP-100 if training words are spoken in close time proximity to test data input. Therefore, a possible mode of operation of the VIP-100 system in the VICI application would be to require a complete single-word training phase prior to the inputting of the VICI four-digit identification code by a person desiring to have his identity verified by the VICI system. Such a procedure is undesirable from an operational standpoint, however, because of the time required for inputting the required number of samples. Therefore, it became necessary to develop a reference array set which would allow accurate recognition from an unlimited speaker set without elaborate training. A number of different approaches were explored before the final technique was devised for an optimum array set. One of these involved the use of several alternate reference arrays for each vocabulary word. These arrays were chosen such that each array represented a wide variety of expected pronunciations for each word. In many of the commercial applications in which the VIP-100 has been used, it has been noted that a particular talker has often been able to achieve highly accurate recognition for a number of words, especially digits, when using another talker's stored reference arrays. This phonomenon occurs most frequently when the two speakers are natives of the same geographical area so that their pronunciations are similar.

Another technique explored involved merging of reference data from each of several talkers for each vocabulary word. This merging process is similar to the normal training routine used with individual talkers when forming a single reference array by pronouncing 10 repetitions of a vocabulary word.

An extensive number of experiments were conducted by the use of reference arrays derived from each of 20 male talkers for each of the 10 digits and the four control words in the required VICI vocabulary. These experiments explored the use of alternate reference arrays from multiple speakers, merging arrays from specific sets of speakers, use of merged arrays for each of several alternate arrays and finally the merging of all data from 20 speakers to form one average reference array. The latter technique resulted in the best recognition accuracy achieved in all of the experiments which were conducted. Several additional experiments were conducted to determine quantatively, if any additional improvement in recognition accuracy might be afforded by the use of a limited number of single training samples to augment the universal reference array which was achieved by merging arrays from 20 talkers. The most extensive of these experiments involved 50 speakers each inputting 50 groups of four digits without any training data and then with the single repetition train on the digits one, three and nine, which had proved most troublesome in recognition accuracy. Recognition accuracy was marginally improved by the use of three single training digits.

To augment these studies which led to a universal reference array, a number of major maodifications in both hardware and software of the conventional VIP-100 system were necessary. For example, a major modification of the VIP-100 recognition software was made to allow additional correlations to take place after the initial recognition decision. These additional correlations known as a "second-look" involve only the initial portion of the feature array of an input word and selected reference arrays. Hardware modifications involved development of alternate feature recognition networks for phoneme and phoneme-like sounds and a major rearrangement of the set of 32 acoustic features used to construct reference arrays. The VIP-100 originally included in its set of 32 recognition features, 17 spectral maxima which covered almost the entire wide-band spectrum capability of the system from 200 to 8000 Hz. It was found that the spectral maxima above 2 kHz vary greatly from talker to talker. Therefore, although maxima above this frequency region serve as effective recognition features in a speaker dependent system, they had a negative effect in a speaker independent design for VICI. The seven maxima features in the region above 2 kHz were eliminated and replaced with additional phoneme and spectral features.

The resulting VICI system had the ability to recognize with very high accuracy (single word accuracy just under 98%) digits and control words spoken in isolation by a total of 85 male talkers ranging in age from 16 to 65 years. The hardware and software developed during the initial VICI program served as the basis of the further experiments conducted during the improvement program reported here.

## C. System Optimization for Operation Over a Wide Range of Conditions

Modifications to the VICI system to allow operation under less than ideal conditions were concerned principally with operation over telephone line bandwidths and response to either male or female talkers. Of these two problem areas the more difficult was operation over telephone line bandwidths. Figure 2 illustrates a typical frequency response characteristic measured at Threshold Technology Inc. (TTI) over an area telephone connection

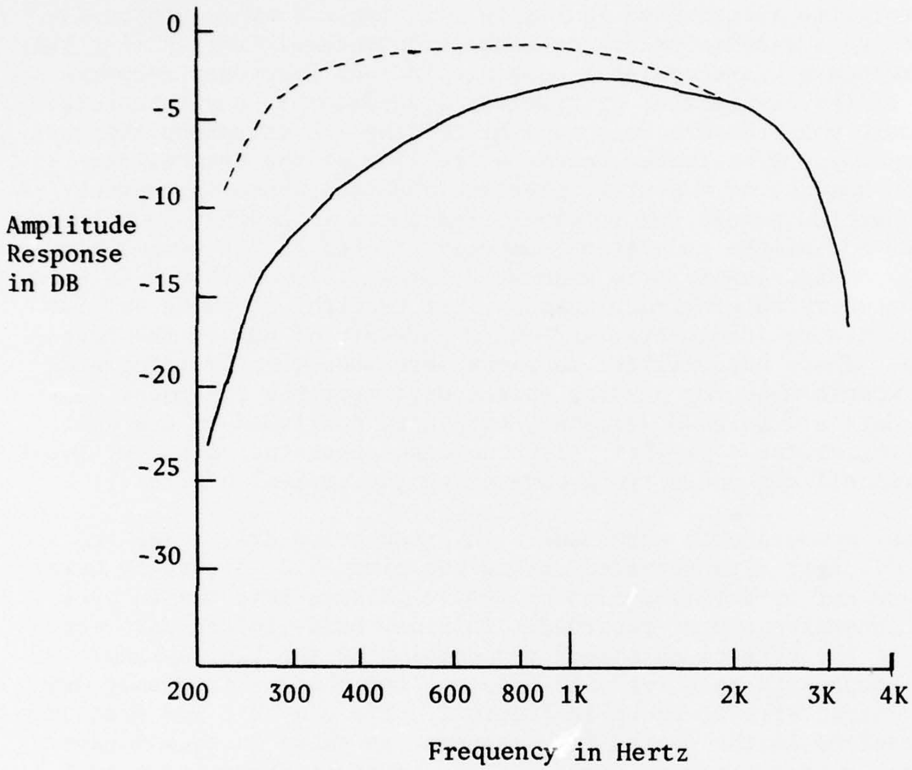



Fig. 2. Typical frequency response characteristic of telephone loop used for band-limiting VICI data. Solid curve is response before compensation. Broken curve shows response with low frequency boost added by receiver module.

which was subsequently used to band-limit all data used for both experimental and for final test purposes. The telephone connection included two exchanges, 461 and 829 in the area code 609 (Southern New Jersey). During the time period in which this effort was conducted, TTI maintained several lines to each of these two exchanges. It was therefore, possible to place a call to a TTI number in one exchange from a number in the other exchange. The frequency response characteristic illustrated in Figure 2 is typical of a number of measurements made on a similar telephone connection several times during this program. This response characteristic does not include frequency response characteristics of the transmitter or receiver transducer in any telephone instruments. These measurements were made by driving the telephone instrument through a matching transformer connected in lieu of the transmitter transducer in the handset of a Western Electric 500D telephone instrument. A voltmeter was connected across the receiver transducer in another 500D instrument at the other end of the connection, usually located in the same laboratory area at TTI. Measurements were made at a level of lower than 9 dB below 1 mw. It was necessary to construct notch filter sections at 60 Hz and 180 Hz in order to eliminate fundamental and third harmonic of hum on the telephone connection. These notch filter sections were subsequently integrated into a receiver module used for passing speech data over the telephone connection. These data are in good agreement with data published by the Bell Telephone Laboratories for a similar telephone connection including two local loops, two central offices and a trunk connecting the central offices.[2]

After initial measurements were made with bread-board driver and receiver elements, a set of digit data recorded during the first VICI effort by male talkers plus a new set of data recorded by female talkers were passed over this telephone connection and re-recorded. This new band-limited data were used in a study of the effects on speech recognition by the VICI system. It was possible to compensate to a reasonable degree for the low frequency droop in the response characteristic shown in Figure 2. The use of a low frequency boost circuit resulted in the dotted line response as shown in this figure. The high frequency cutoff however, could not be compensated for because of the steepness of the curve. The principal recognition effects due to the very limited high frequency response were noted in the digits 6 and 7 as well as the control words ERASE and CANCEL. The major portion of the energy in the fricatives in these four words was effectively eliminated by the severe high frequency restriction on the telephone connection as would be expected. During examinations of the spectral energy of these fricatives it was difficult in many cases to discern even a trend of increasing energy with frequency (positive slopes) because of the effects of the close talking noise cancelling microphone with which the data were recorded. Blasting effects were often noted which generated spurious low frequency energy during fricatives. It was obvious that an alternative method of fricative detection as compared with energy ratios previously utilized in the VICI development would be necessary. Ideally a voiced-unvoiced decision which could be based on the periodicity or lack of periodicity of the speech waveform to indicate the presence of fricatives could be used. The development of a satisfactory voiced-unvoiced detector based on periodicity, however, was beyond the scope of this effort. Therefore, it was determined that the next best approach would be to include a small amount of preprocessing at the transmitting end of the telephone connection. Because of the intended application of the VICI system as an input terminal to the BISS system, operational control over the type of terminal to be placed at an input station would be possible. Therefore, it is not unreasonable to propose a terminal at an input station which can accomplish some preprocessing of the speech signal before it is transmitted over telephone lines.

A simple zero-crossing network detector for high frequency energy was constructed which served to gate a 2.8 kHz oscillator the output of which could be impressed upon telephone lines to signal the presence of a fricative at the transmitting end of the connection. Subsequently, a complete transmitter module with self-contained power supply was constructed to be used in further experiments and ultimately to be supplied at the end of the program with the VICI system for further use by the Air Force. Figure 3 illustrates a block diagram of the finished transmitter module. The microphone preamplifier provides proper compensation for a Telex 1200 or similar noise cancelling microphone. The zero-crossing detector is adjusted to provide an output to control the gated oscillator whenever fricatives with substantial energy above 3 kHz are spoken. The 2.8 kHz tone from the oscillator then is mixed with the input speech for transmission along over the telephone lines to the receiver module which is then connected to the VICI system. The receiver module built for the project is shown in block form in Figure 4. It includes the preamplifier, two notch filters, and low frequency boost previously mentioned as well as a differential output enabling direct connection to the line input to the VIP preprocessor. These two units were used at ends of the telephone connection for all subsequent dubbing of tape recording data done to band-limit the data and for live tests at TTI and final tests at RADC. Tape recording dubs were made by connecting the tape recorder at the receiving end across one side of the differential output which was connected to the VIP-100 preprocessor.

As outlined in paragraph B of this section, an extensive study during the first VICI contract of the possible means of generating a master reference array for male talkers resulted in the conclusion that the best array was generated in a relatively simple manner. That is, a reference array which is generated by the merging of arrays derived from a fairly large set of talkers. Data from 20 male talkers were used for final arrays. Therefore, during this program early experiments conducted to determine the recognition accuracy achievable for female talkers involved the use of master reference arrays derived from a number of female talkers, in this case, 10. This number of female talkers was used initially because 10 female talkers were readily available to record data which could be used for a generation of individual reference arrays and because studies during the preceding program had disclosed 10 as being the minimum number of talkers who could be used to effectively to derive merged arrays. Comparisons of the reference array matrices generated by female talkers through the merging process, and those generated by male talkers disclosed significant differences between the comparable arrays. These differences were especially obvious in the set of 10 spectral maxima as might be expected. First and second vowel formants which are reflected by the maxima features were generally higher in vowels for the female talkers. Also, the phoneme-like feature recognition networks which were optimized for males speaking digits and control words in high quality speech showed significantly different responses for many female phonemes.

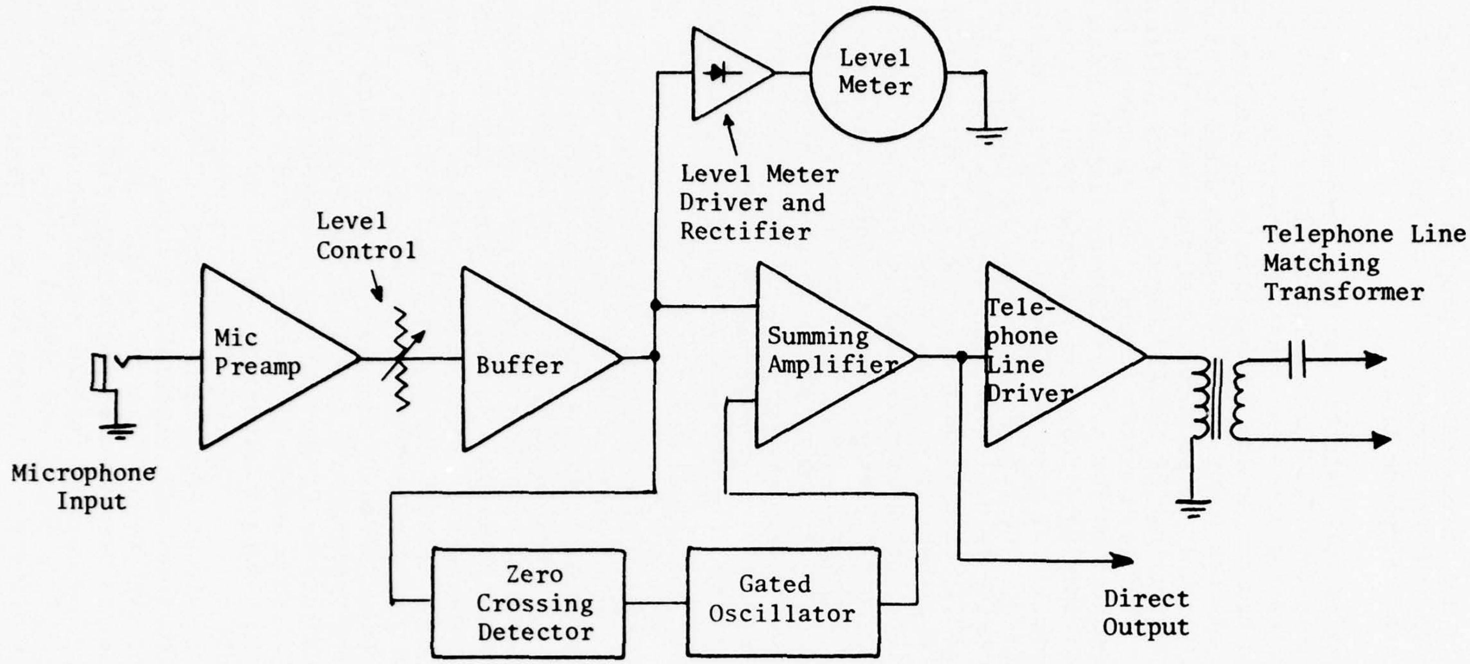



Fig. 3 Block diagram of telephone line transmitter module. Power supply is self-contained.

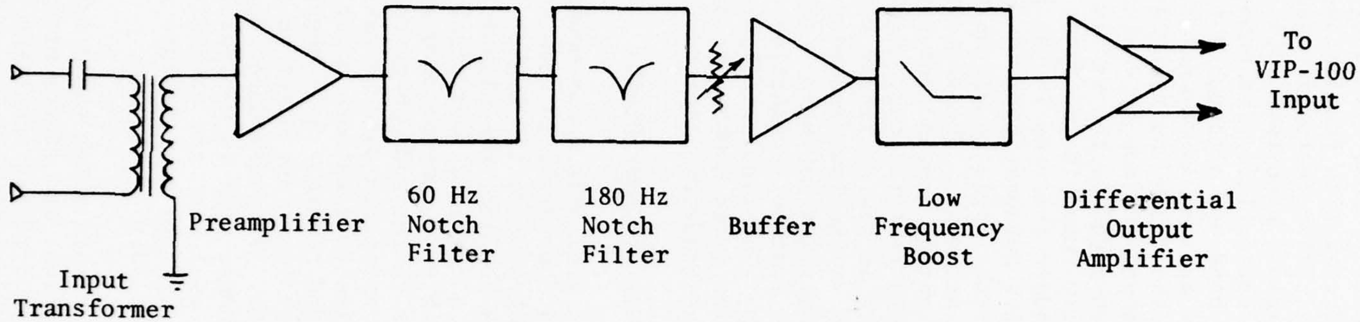



Fig. 4 Block diagram of telephone line receiver module.
Power is provided by VIP-100.

The initial tests were conducted with high quality speech. However, subsequent investigations and especially those involving modification of phoneme recognition networks were conducted with bandlimited speech. Table I illustrates the two classes of recognition features, spectral features and phoneme-like features for both the original wide-band VICI and the modified feature set for narrow-band male and female VICI. The number of spectral features has been decreased for narrow-band use although the number of maxima has been increased by one channel. The maximum in channel 1 has been deleted for narrow-band operation because the telephone bandwidth restriction at low frequencies renders the first channel maximum as virtually useless. Two maxima in channels 11 and 12 have been added to the set in order to accommodate the higher second formants generally found in the front vowels of the female talkers. In addition to the added features in the phoneme-like category, all phoneme-like feature recognition networks have been modified to some extent. Appendix A of this report presents logic equations for the modified and newly developed feature recognition networks. The most obvious characteristic of these networks is the lack of energy inputs from any channel above 16. Because of the telephone bandwidth restrictions, no high frequency energy was available so the channels above 16 are of little use except to indicate end effects. The gradual roll off at the low frequency portion of the spectrum due to telephone line characteristics mandated a rearrangement of spectral feature inputs to phoneme-like feature recognition networks in the low frequency portion of the spectrum. The final configuration represents features which gave best test results for large numbers of male and female talkers whose speech had been telephone bandwidth limited.

## D. Error Correction by the Use of Check Digits

As previously stated, the VICI system was modified to consistently recognize with high accuracy a four digit ID code spoken by any speaker under less than ideal conditions. Obviously, the higher the recognition accuracy achieved for each digit in the ID code, the higher the resultant accuracy will be obtained for entering a complete four digit code. For a given digit accuracy, however, it is possible to improve code recognition accuracy by adding additional check digits to the code which can be used to correct errors. The proliferation of automatic data processing in business and financial transactions has given rise to a variety of self-checking number systems for identifying transactions and people. Self-checking number systems are based on the introduction of at least one digit to a basic number code to generate a self-checking number. The introduced digit is calculated such that when the self-checking number code is manipulated by a specified mathematical procedure the result will meet the established criteria for the system. Self-checking number systems vary because manipulations and/or criteria used in different systems vary.

There are three types of errors which these systems are designed to reduce: transcription, transposition, and random errors. Transcription (or substitution) errors involve incorrectly inputting or transmitting one number of a code; for example, the substitution of the number 3 for the number 8 in the four digit ID code 8574 to produce the erroneous code 3574. Transposition errors occur when two digits are interchanged, that is, the 5 and the 7 in the aforementioned code to result in the erroneous code 8754. Random errors

TABLE I RECOGNITION FEATURES FOR WIDE-BAND AND NARROW-BAND VICI

| Spectral Features | | Phoneme-Like Features | |
|---|---|---|---|
| Wide Band | Narrow Band | Wide Band | Narrow Band |
| MAX 1 | MAX 2 | Long Pause | Long Pause |
| MAX 2 | MAX 3 | Slope Gap | Slope Gap |
| MAX 3 | MAX 4 | Energy Gap | Energy Gap |
| MAX 4 | MAX 5 | UVNLC | UVNLC |
| MAX 5 | MAX 6 | Zero Crossing | Zero Crossing |
| MAX 6 | MAX 7 | i | i |
| MAX 7 | MAX 8 | I | I |
| MAX 8 | MAX 9 | $\varepsilon_2$ | $\varepsilon_2 + \varepsilon_3$ |
| MAX 9 | MAX 10 | w | w |
| MAX 10 | MAX 11 | ɜ | ɜ |
| PSB 2 | MAX 12 | $n_1 + n_3$ | n |
| PSB 3 | NSB 2 | ʌ$_1$ + ʌ$_3$ | ʌ |
| NSB 5 | PSB 3 | $\varepsilon_1$ | ʌ$_4$ |
| NSB 6 | NSB 8 | r | ɔ |
| NSB 7 | | s | ɑ$_4$ |
| NSB 8 | | | $U_2$ |
| NSB 10 | | | u |
| | | | Stop-Fricative |

are any multiple numbers of errors of either or both of the first two types. Commonly used check digit systems apply a weight or multiplier to each columnar position of the entry. The sum of the products of the weights and corresponding digits is divided by a modulus and the remainder must meet the previously specified criteria.

In many commercial and financial applications, check digits are added to a basic identification or transaction number to serve as an error flag, that is, to serve as an indication that an error has occurred in the transcription of a set of numbers but not in any way to modify or attempt to correct the error. Because the goal of the VICI program has been to achieve efficient entry of a four digit ID code with a high accuracy, a self-checking number system in this application should have the power to correct at least some of the errors attributed to the speech recognition system.

The goal of error correction imposes additional limitations on the self-checking number system to be used for VICI. In order to implement a reasonable system with a small number of check digits, assumptions must be made as to the type of errors which may occur during the speech recognition processing. Therefore, all errors have been assumed to be of the transcription type as described above. Transposition errors will not be considered because the latter type would not be expected to be made by the speech recognition equipment since each digit is individually considered. Transposition errors can be expected to be made only by the person speaking the digits and will not be further considered. Furthermore, no attempt will be made to correct more than one error in a four digit group. Any attempt to correct more than one error per group by the use of check digits would be unweildy and probably unnecessary. Test results achieved during the first VICI development program indicated that only in a very small number of cases did more than one recognition error per four digit code group occur. In the live tests held at RADC, with 21 speakers each speaking 75 four digit groups, during final tests for the first VICI development program there were a total of 134 incorrect four digit codes caused by a total of 142 incorrect digits.

Another important assumption which has been made is that the error correction digit or digits will always be recognized correctly. A number of additional steps have been taken to insure the accuracy of the error correction digits as will be explained.

The check-digit correction algorithm which has been established for VICI consists of two principal parts; first, detection that an error has occurred in the four digit ID code portion of the total inputted number, and second, an attempt to make a correction by substitution for digits which are likely to be an error. The routine which has been developed for the initial detection of errors in the ID code is of the prime number weight type with modulo 13. This type of error checking system was ranked second in a group of seven self-checking number systems in a recent paper by Herr.[5] Actually, his ranking was based on a modulo 11 rather than modulo 13 system. However, the larger the modulus, the fewer errors will go undetected. Because the modulus is greater than 10, it has been necessary to use two check digits rather than one. There is an added advantage in addition to the higher modulus in the use of two check digits. It is now possible to use combinations of just four digits which will be least likely to be confused with each other in the two check digit positions. The four digits which have been found in previous VICI investigations to be least likely to be confused with each other are 0, 4, 6 and 8.

The format for the augmented identification code is simple, the first four digits in the six digit code are the four identification digits with the fifth and sixth digits representing a coded version of the numbers 0 through 12, the check digit. The recognition algorithm for the fifth and sixth digits limits the possible choices to the four digits listed above for further increased accuracy. The actual operation of the check digit scheme is as follows. The weights for the first four code digits are one, three, five and seven respectively. For a particular code group the check digit is calculated by multiplying each ID code digit by the associated weight, suming these products and then subtracting the sum of the products from the next largest multiple of 13. For example, the check digit for the code 6859 would be calculated as follows:

$$\begin{array}{r} 6 \times 1 = 6 \\ 8 \times 3 = 24 \\ 5 \times 5 = 25 \\ 9 \times 7 = \underline{63} \\ 118 \end{array} \qquad \begin{array}{r} 13 \times 10 = 130 \\ -\underline{118} \\ 12 \end{array}$$

so the check digit is 12.

The check digits calculated by the above scheme are encoded with the four allowed digits shown in Table II.

TABLE II CHECK DIGIT ENCODING PAIRS

| Check Digit | Encoding Pair |
|---|---|
| 0 | 00 |
| 1 | 04 |
| 2 | 06 |
| 3 | 08 |
| 4 | 40 |
| 5 | 46 |
| 6 | 48 |
| 7 | 60 |
| 8 | 64 |
| 9 | 68 |
| 10 | 80 |
| 11 | 84 |
| 12 | 86 |

Table III illustrates 50 groups of six digits used in all system tests including the final tests as is described in Section 3 of this report. These

TABLE III LIST OF 50 SIX DIGIT GROUPS USED FOR TESTING

| | | |
|---|---|---|
| 5 2 5 1 6 8 | 6 3 1 4 4 0 | 0 3 3 8 8 4 |
| 7 5 9 0 8 4 | 3 4 9 5 6 8 | 4 7 7 0 4 6 |
| 1 0 1 7 8 0 | 5 6 5 9 4 8 | 6 8 0 3 0 4 |
| 6 2 6 2 6 8 | 1 1 3 4 4 6 | 3 0 6 8 0 6 |
| 2 0 2 7 4 0 | 4 6 0 7 6 0 | 9 1 5 4 0 0 |
| 7 2 7 0 4 0 | 8 9 2 6 4 0 | 7 8 2 3 0 8 |
| 3 6 6 8 8 0 | 9 6 4 2 4 0 | 2 4 8 3 0 8 |
| 0 4 4 1 0 0 | 0 7 6 9 0 8 | 8 8 7 9 0 0 |
| 8 4 3 2 0 8 | 2 2 8 3 6 8 | 9 3 9 8 8 4 |
| 4 1 8 5 6 8 | 7 3 7 0 0 4 | 6 9 2 4 8 4 |
| 9 9 0 7 4 8 | 0 0 5 7 4 0 | |
| 5 8 3 1 0 4 | 1 4 0 6 8 0 | |
| 1 7 1 4 8 0 | 8 1 9 5 0 0 | |
| 0 9 8 6 6 4 | 9 7 4 6 8 6 | |
| 2 3 2 9 6 0 | 3 5 7 8 6 4 | |
| 6 7 0 9 0 4 | 2 1 2 5 0 6 | |
| 8 5 4 1 0 6 | 5 5 1 4 8 6 | |
| 3 8 6 2 6 0 | 1 6 1 5 4 8 | |
| 7 9 5 0 4 8 | 4 5 3 1 8 4 | |
| 4 2 9 6 6 0 | 5 0 8 3 8 6 | |

digit groups are the same as used during the initial VICI development program with the addition of the two check digits to each four digit ID code.

The error correction mode is one of two modes available in the operation of the modified VICI system. The other mode is normal recognition with no error correction. When the error correction mode operation is selected, an orderly digit inputting routine must be closely followed in order to properly input the four identification digits of this six digit code. The routine requires that each digit of the six digit code be inputted within 10 seconds of the preceding digit and that any errors made by the machine or the speaker which are obvious to the speaker be cancelled by restarting the inputting of the code group by the use of the control word "Cancel". For purposes of this investigation, the error checking mode of operation is initiated by a command from the Teletype console. In an operational installation this initialization would be automatic. By the use of the control word CANCEL, it is possible at any time to reinitialize the cycle. The following paragraphs describe the complete recognition and error checking cycle with the aid of the flowcharts in Figures 5 and 6.

The error correction algorithm is initialized either by Teletype keyboard input command or by recognition of the control word CANCEL and the Self-Scan display associated with the VICI system displays the message "Input Code". The ID code storage registers in the program are initialized and the word input and recognition section of the program is enabled. At this point a 10 second counter is started. Failure to input a digit within 10 seconds will result in reinitialization of the program with the Self-Scan display showing the message "Repeat Code". After the first word is inputted the word is tested to determine whether the word CANCEL was recognized which will reinitialize the system as previously mentioned. If CANCEL is not recognized, the first digit is stored in a buffer and once again a timer is started to insure input of the next digit within 10 seconds. This process is iterated until four digits have been inputted. After the four digits which comprise the ID code section of the complete code group have been inputted, the system expects two more digits each of which must be either zero, four, six or eight. CANCEL is also acceptable. After an initial recognition pass with each of the fifth and sixth inputted digits, a test is made to determine whether the digit recognized is one of these four or some other digit. If a prohibited digit is recognized in either instance the correlation table associated with the inputted word is modified, that is, the correction score for the prohibited word initially recognized is set to zero. The normal correlation maximum selection routine used for recognition then selects the highest correlation score of those remaining and a recheck is made to determine whether an allowed or prohibited word has been recognized. If again a prohibited word has the highest correlation score remaining, the process is iterated in an attempt to establish a correct recognition of an allowed digit. After three iterations through this process for either the fifth or sixth digits (the first or second check digits) the algorithm will be reinitialized with the display showing the message "REPEAT CODE". It is then necessary to input the complete six digit code.

If one of the four allowed digits is recognized in each of the two check digit positions the two recognized check digits are then converted to the

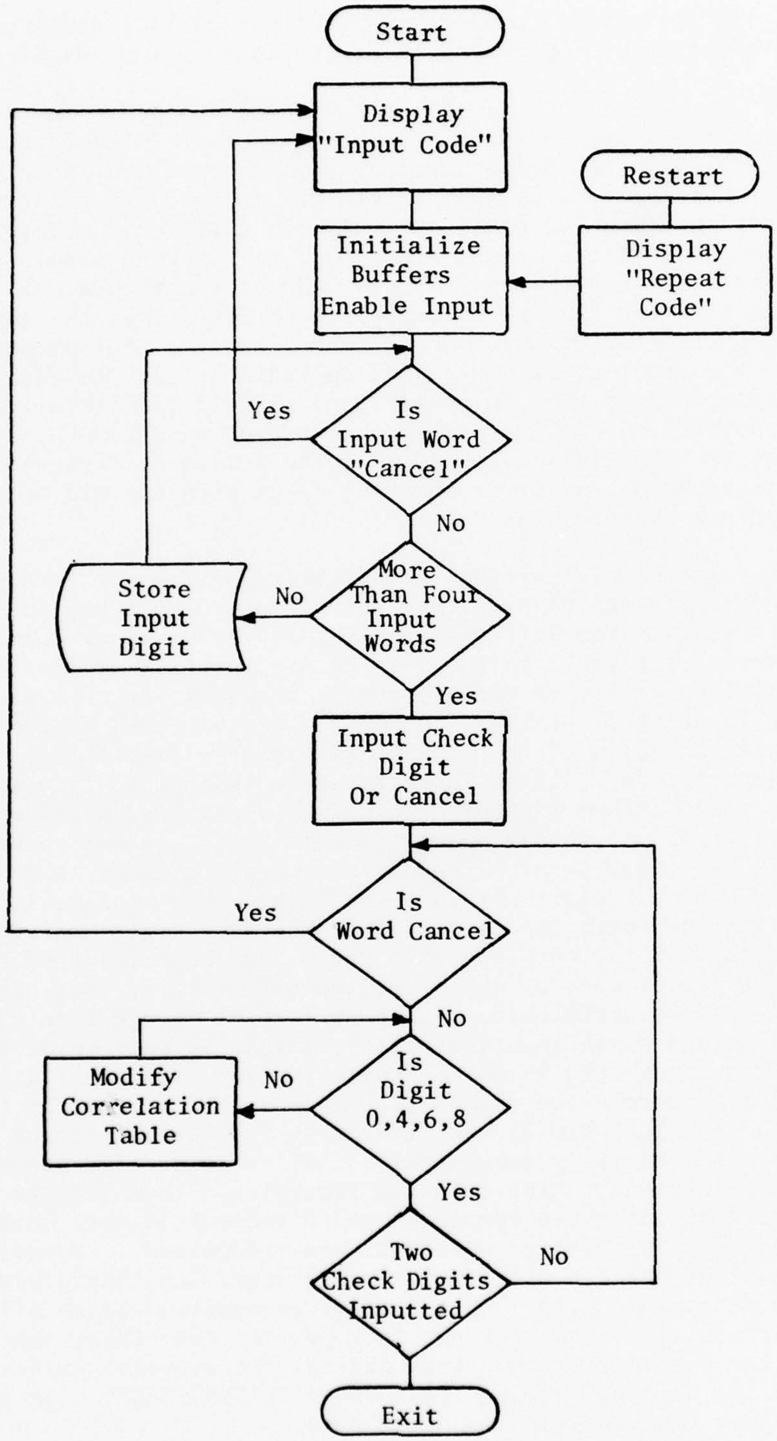



Fig. 5 Flow chart of error detection routine of check-digit error correction algorithm. Exit from this routine to start of error correction routine.

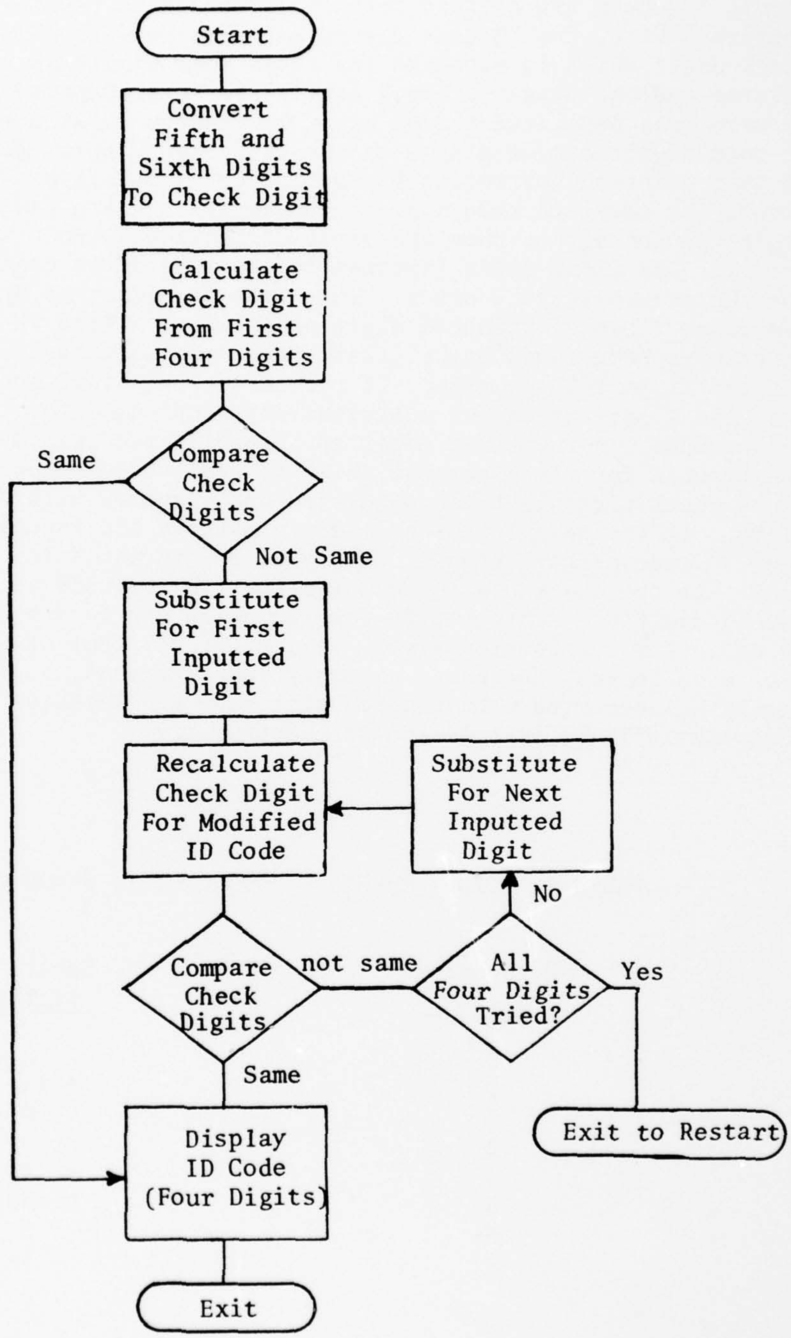



Fig. 6 Flowchart of error correction routine of check-digit error correction algorithm. Entrance to this routine is from the exit of the error detection routine shown in Fig. 5.

modulo 13 check digit (zero through 12) which is used by the error checking routine. Next, the ID code digits are retrieved from buffer storage and the check digit which is expected for those four digits is calculated. The calculated and the inputted check digits are then compared to determine whether an error has been detected or not. If the two check digits agree, the four ID code digits are displayed on the Self-Scan display for the user to verify. At this point no correction by the talker is possible. If a two digit error occurs the complete code must be reentered after a CANCEL. If the check digits do not agree, then the error correction subroutine is initiated (Figure 6). The first digit inputted in the identifier code of four is tested for the possibility of error. This test is performed by comparison to a reference table of probable digit confusions. Table IV illustrates the digits which have been found to be likely to be confused based during this and during the first VICI program. If the recognized digit appears in this table, then the substitution (or substitutions) indicated for this digit is made. For example, if the first digit is initially recognized as 8, a 3 would be substituted for the 8 because this has been found to be a likely confusion. A new check digit is then calculated and compared with the inputted check digit. If the calculated check digit matches the inputted check digit the four ID code as corrected is displayed for verification on the Self-Scan display. If the check digits do not match, the process is iterated for the second digit and subsequently for the third and fourth digits in an attempt to correct a recognition error. If, after all four of the ID code digits have been tested, and using probably subsituttions, and no match of check digits has occurred, the display will show the message "REPEAT CODE" and the algorithm will be initialized for another try.

TABLE IV DIGIT SUBSTITUTIONS USED IN ERROR CORRECTION

| Recognized Digit | Substitute Digit(s) |
|---|---|
| 0 | 2 |
| 1 | 4 and 5 |
| 2 | 0 |
| 3 | 8 |
| 4 | 1 |
| 5 | 1 and 9 |
| 8 | 3 |
| 9 | 5 |

This approach occasionally can lead to the generation of erroneous corrected codes. For example, the codes 5831 and 5335 have the same check digit, that is, 1. If the former code was inputted with the final digit of the four recognized as a 5 rather than a 1, then the error subroutine would sense an error and go into operation. The first substitution would be a one for the initial 5. However, the resultant check digit would be 3 so the substitution for the second digit of the code would be tested. In this case, a 3 would be substituted for 8 and the proper check digit would be calculated and no further substitutions would be made. Therefore, the display would show the code 5335, a code with not one but two errors. The person desiring verification would then be required to reenter the complete code after viewing the faulty code on display. During the development of the error correction algorithm a study was made of the number of possible occurrences of this phenomenon. As a basis for this study it was assumed that the digits 0, 2, 6, 7 and 8 would be either correctly identified or would be misrecognized on a random basis. During the first VICI development contract, occasional 0-2, 0-7 and 2-8 confusions were noted. Such confusions were so infrequent relative to the confusions for which the error correction algorithm was designed that these possible confusion pairs were ignored. Of the 10,000 codes possible with a group of four digits, 625 include only the five "good" digits, 0, 2, 6, 7 and 8. For these codes, the check digit algorithm would be unnecessary. However, 2500 groups will contain three good digits and one "bad" (error prone) digit. The "bad" digits were defined as 1, 3, 4, 5 and 9. In previous VICI tests confusion between 1 and 4, 1 and 5, 5 and 9 were found in both directions in addition to the misrecognition of 3 as an 8 (but rarely the reverse). The remaining 6875 code groups contain two or more "bad" digits. The erroneous corrections mentioned above can occur in approximately 1100 of these remaining groups or 11% of all groups. Therefore, error correction by the use of check digits should correct all single digit errors in 8225 of 9325 codes from digit groups which contain one or more "bad" digits, an effectiveness of 88.27%.

The number of groups including non-correctable digits (1100) was calculated by the use of a computer program written in the BASIC language. This program examined each entry in a previously constructed table of 10,000 four-digit codes with check digits and detected each group for which an erroneous correction could be made by the check digit error correction algorithm. It should be noted that all calculations have been based upon the premise that only one recognition error would occur in a four-digit group, that the two check digits would always be correctly recognized, and that 0, 2, 6, 7 and 8 would be correctly recognized. Error correction could occur in many of these 1100 groups because many include one or more correctable as well as non-correctable digits. For instance, in the example shown above with the identification code 5831, if the 5 were recognized as a 1 the error correction algorithm would function properly.

This study was made before the bulk of the system tests were conducted and, therefore, some of the basic assumptions on which the study was based are no longer valid. System tests disclosed a significant number of 0-2 confusions (in both directions). Also, the digit 8 was recognized incorrectly as a 3 with about the same frequency as the reverse which was not the case with the wide-band VICI tests. An analysis has been made of the effectiveness of error correction when applied to the list of 50 code groups used for all tests for this VICI program. It should be noted that most of the system accuracy tests conducted with data generated by nearly 200 talkers did not include error correction. System performance figures are based on single digit accuracy rather than code group accuracy resulting from the action of the error correction. Therefore, analysis of the effective error correction upon the 50 code groups can be used as a guide to project overall system accuracy when error correction is applied. For this analysis, the substitutions shown in Table IV which were included in the final error correction program were assumed in all cases. Correct recognition of the last two digits in the six digit code group which represented the error checking digit was assumed. It was also assumed that the digits 6 and 7 occuring in the ID code portion of the group would be correctly recognized. The results of this study are shown in Table V. The left column in each section of the table shows the complete six digit code group including the identification code and the check digit code. Eighteen of the 50 code groups appear with one of the first four digits underlined. Two groups show two of the first four digits underlined. The remaining 30 groups show no digits underlined. An underlined digit in a group indicates a digit which could not be corrected by the error correction routine and would result in the erroneous code being displayed as shown in the center column of each section of the figure. For example, the first group in the left hand section of the table is 525168. Misrecognition of the second five in the first group of four digits would result in the erroneous code 5011. This phenomenon would occur because the iterative substitution process begins with the first digit and ends with the fourth of the ID code. Anytime that a correctable error occurs in the first digit position with no other errors, then the error correction algorithm will function properly. However, errors occurring in the second, third, or fourth position of the identifier section of the group can result in an erroneous code. In this example, it was assumed that the system recognized 5251 incorrectly as 5211. Substitution of a one or a nine in the first position would not result in a correct check digit of nine (encoded by the last two digits in the group, 68). However, the substitution of a zero for a two in the second digit position would result in a correct check digit and, therefore, the display of the erroneous code 5011.

For the above example, it is possible for the check digit algorithm to make a maximum of 7 substitutions in an attempt to correct the code group. This number of substitutions is possible because three of the four digits in a group can be substituted for by more than one number, that is, each digit 5 can be substituted for a 1 and a 9 and the digit 1 can be substituted for by a 4 and a 5. The only allowable substitution for the 2 is the digit 0. The maximum number of substitutions possible with the four identification digits would be eight. This phenomenon could only occur with the group composed of 1's, 5's or a combination thereof in which the fourth digit was an error and in which the algorithm did not generate an erroneous code by substitutions in the first three positions. The number of substitutions possible for each of the 50 test code groups is shown in the right column of each section of the table. There is a total of 200 substitutions possible in these 50 codes. Twenty-two of these substitutions involve the underlined digits which would result in erroneous codes being generated. Of the 50 codes,

TABLE V NON-CORRECTABLE DIGITS IN TEST CODE GROUPS

| Code Group | Resulting Erroneous Code(s) | No. Subs | Code Group | Resulting Erroneous Code(s) | No. Subs. |
|---|---|---|---|---|---|
| 5 2 5 1 6 8 | 5011 | 7 | 8 9 2 6 4 0 | | 3 |
| 7 5 9 0 8 4 | 7992 | 4 | 9 6 4 2 4 0 | | 3 |
| 1 0 1 7 8 0 | 1257 | 5 | 0 7 6 9 0 8 | 2765 | 2 |
| 6 2 6 2 6 8 | | 2 | 2 2 8 3 6 8 | | 4 |
| 2 0 2 7 4 0 | | 3 | 7 3 7 0 0 4 | | 2 |
| 7 2 7 0 4 0 | | 2 | 0 0 5 7 4 0 | 0297 | 4 |
| 3 6 6 8 8 0 | | 2 | 1 4 0 6 8 0 | 4426 | 4 |
| 0 4 4 1 0 0 | 2411, 0415 | 5 | 8 1 9 5 0 0 | | 6 |
| 8 4 3 2 0 8 | | 4 | 9 7 4 6 8 6 | | 2 |
| 4 1 8 5 6 8 | 4535 | 6 | 3 5 7 8 6 4 | | 4 |
| 9 9 0 7 4 8 | | 3 | 2 1 2 5 0 6 | 0129 | 6 |
| 5 8 3 1 0 4 | 5335 | 6 | 5 5 1 4 8 6 | | 7 |
| 1 7 1 4 8 0 | | 5 | 1 6 1 5 4 8 | 1641 | 6 |
| 0 9 8 6 6 4 | | 3 | 4 5 3 1 8 4 | 4181 | 6 |
| 2 3 2 9 6 0 | 0829, 2825 | 4 | 5 0 8 3 8 6 | 9088 | 5 |
| 6 7 0 9 0 4 | | 2 | 0 3 3 8 8 4 | | 4 |
| 8 5 4 1 0 6 | 8515 | 6 | 4 7 7 0 4 6 | | 4 |
| 3 8 6 2 6 0 | | 3 | 6 8 0 3 0 4 | | 3 |
| 7 9 5 0 4 8 | | 4 | 3 0 6 8 0 6 | | 3 |
| 4 2 9 6 6 0 | 4056 | 3 | 9 1 5 4 0 0 | | 6 |
| 6 3 1 4 4 0 | | 4 | 7 8 2 3 0 8 | | 3 |
| 3 4 9 5 6 8 | | 5 | 2 4 8 3 0 8 | 2188 | 4 |
| 5 6 5 9 4 8 | | 5 | 8 8 7 9 0 0 | | 3 |
| 1 1 3 4 4 6 | 5434 | 6 | 9 3 9 8 8 4 | 5398 | 4 |
| 4 6 0 7 6 0 | | 2 | 6 9 4 2 8 4 | 6540 | 3 |

20 or 40% could contain one or more digits which could not be corrected. This figure is considerably higher than that derived from the study mentioned above. However, on a single digit basis, 22 out of 200 single digits could not be corrected (11%).

## Section III

## FINAL SYSTEM TESTS

### A. Background of Test Data

The VICI system, modified for operation under adverse conditions, was tested with a total of more than 56,000 words, both digits and control words spoken by 193 different talkers. This group included 139 male talkers and 54 female talkers. The ages of the talkers ranged from 16 to 65 years. Most tests were from tape recordings although a limited number of live input tests were conducted at RADC upon equipment delivery. All tape recorded data were telephone band-limited by dubbing over actual telephone lines.

Data were initially tape-recorded by use of a Telex 1200 noise canceling microphone connected to either a reel-to-reel recorder or a cassette recorder. Data were then passed over an area telephone loop by the use of the two terminals constructed for this purpose. Figure 7 illustrates the complete loop configuration. These terminals are described in detail in Section II of this report. The telephone loop included two exchanges, 461 and 829, in area 609 (Southern New Jersey. Dubbing of data tapes for use in testing and for establishing the original reference data base was conducted over a period of approximately a year. Therefore, the tests results do not reflect any short term conditions of the telephone circuits. Frequency response tests made at several times during this period show reasonably similar telephone line characteristics. Background noise did vary from dubbing session to dubbing session, but did not have a discernible effect upon recognition results. In order to provide additional data for acceptance tests at RADC, 12 new talkers were recorded. Each talker spoke the 50 groups of six digits. These data were then passed through the base telephone system by the use of the transmit and receive terminals previously described. No frequency response measurements were made on the loop used for this bandwidth limiting dub.

The live tests of eight talkers held at RADC did not include telephone bandwidth limiting. The transmit terminal was used with the direct connection to the VIP-100 preprocessor. An internal low pass filter provided in the VIP-100 was used to bandlimit the high frequency portion of the speech spectrum. For the test, the element from a hand-held noise microphone (Shure model 488B) mounted in a telephone handset was used.

### B. Test Results

Final test results involving several conditions are shown in Table VI. In this table, the data is divided into six groupings on the basis of the sex of the talkers, telephone line conditions, and the type of the data input, that is, four digit groups, six digit groups or control words. The bulk of the data used for testing was recorded at TTI through the two local exchanges with the connecting trunk line. It included a total of 170 talkers both male and female most of whom spoke a total of 300 digits, that is, the 50 groups of six digits each. Included in the group of data from 170 talkers were four

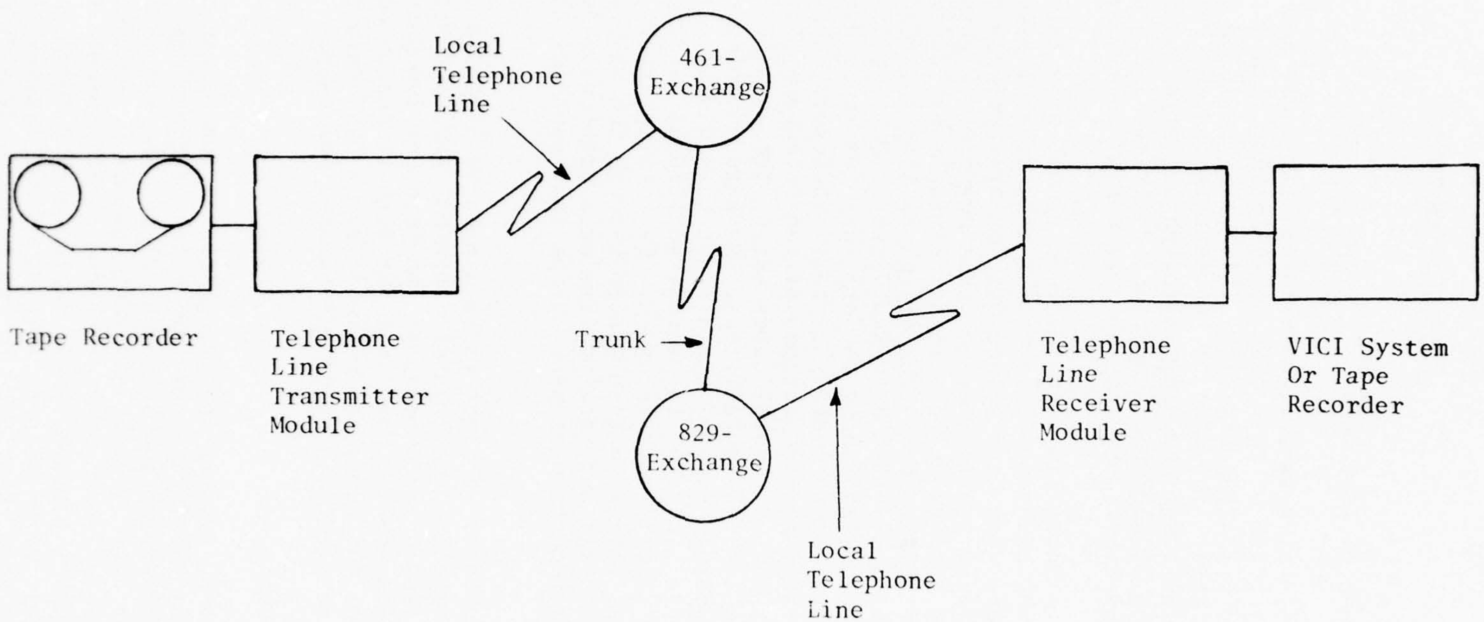



Fig. 7 Complete system block diagram for dubbing data or testing over telephone lines at TTI.

TABLE VI FINAL TEST RESULTS

| Data Set | Number of Talkers | Sex of Talkers | Type of Data | Total Valid Words | Total Errors | Percent Correct | Remarks |
|---|---|---|---|---|---|---|---|
| 1 | 101 | M | 6 digit group | 30172 | 895 | 97 | Recorded and tested at TTI |
| 2 | 21 | M | 4 digit groups | 4189 | 115 | 97.25 | Same as Data 1 |
| 3 | 48 | F | 6 digit groups | 14348 | 458 | 96.8 | Same as Data 1 |
| 4 | 12 | M&F | 6 digit groups | 3588 | 178 | 95.12 | Recorded and tested at RADC |
| 5 | 8 | M&F | 6 digit groups | 2400 | 110 | 95.4 | Live band-limited at RADC |
| 6 | 37 | M&F | Control Words | 1480 | 63 | 95.74 | Same as Data 1 |

digit groups recorded during the previous VICI program by 21 male talkers. The measure of recognition accuracy derived from tests involving the six digit groups is biased somewhat when calculated on a per digit basis. Two of the six digits in these groups (the check digits) were combinations of only four different numbers, that is, zero, four, six and eight. Therefore, any tendency of the recognition system toward higher, or lower accuracy, for these four digits as compared with the other six would bias the results obtained with the six digit groups. Testing with a set of four digit groups provided a measure of reference because the digits in shorter groups were the same as the first four of the six in the longer groups. Single digit accuracies for the total group of 101 males who recorded six digit groups were nearly the same as 21 talkers speaking four digit groups as is shown in the table. Twenty-six male and 11 female talkers were tested on control words only. The control word recordings were taken from a group recorded during the previous VICI contract. Bandlimiting was the same as for the first three groups of digits recorded and dubbed over phone lines at TTI.

The two tests conducted at acceptance of the equipment at RADC included in most cases the same talkers but were conducted under somewhat different circumstances. Therefore, they have been grouped in two different categories. The RADC data which was first tape recorded was also passed over the local telephone loop. However, telephone bandlimiting was not used during live tests, but bandlimiting at the high frequency portion of the spectrum was accomplished with a low-pass filter. As mentioned above, the live tests were also conducted with a different microphone. This hand-held microphone was subsequently found to have variable high frequency characteristics depending upon its position relative to the talker's lips. The effect was rendered erratic the ability of the transmit module to detect high frequency energy caused by fricatives in the digits 6 and 7 and the control word CANCEL. Therefore, the recognition results for four talkers who participated in both tape recorded and live tests varied considerably because of the microphone and transmission line differences. The average accuracy, however, was similar in both cases and was somewhat lower than the results achieved at TTI.

Error matrices for the various conditions shown in Table VI are illustrated in Figures 8 through 12. Figure 13 is the error matrix for 26 male talkers talkers speaking a random mixture of digits (not in groups) and control words. These data were used to derive the results of Data Set 6 in Table VI. A composite error matrix for all final tests involving both four and six digit groups is shown in Fig. 14.

| Recognized \ Spoken | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | | 3 | 163 | 2 | 8 | | 8 | 5 | 12 | 2 |
| 1 | | | | 4 | 40 | 46 | 7 | 1 | | 1 |
| 2 | 40 | | | 3 | 2 | | 12 | 5 | 1 | 31 |
| 3 | | 1 | 16 | | | | | | 41 | 28 |
| 4 | 1 | 23 | 1 | 3 | | 18 | 6 | | | 1 |
| 5 | 1 | 9 | 1 | | 3 | | | 4 | | 52 |
| 6 | | | | | | | | | | |
| 7 | 2 | | 4 | 1 | 5 | 3 | 6 | | | |
| 8 | | | 1 | 15 | | | 3 | | | |
| 9 | | 6 | | | | 42 | 1 | 1 | | |
| C | | | | | 1 | | 11 | 1 | 4 | 1 |
| E | 1 | | | 1 | 1 | | 5 | | 4 | |
| V | 12 | 2 | | | 16 | 2 | | 2 | | 3 |
| T | | | | 6 | | 1 | | 2 | | 8 |
| REJ/NR | 1 | | 17 | 10 | 41 | 19 | 3 | 12 | 2 | |

Fig. 8 Error matrix for 101 male talkers speaking six digit groups, tested at TTI from tape recordings.

| RECOGNIZED \ SPOKEN | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | | | 23 | | 1 | | | 3 | | |
| 1 | | | | 1 | | 18 | | | | |
| 2 | 2 | | | | | | | | | 1 |
| 3 | | | 1 | | | | | | 1 | |
| 4 | | 1 | | | | 15 | | | | 10 |
| 5 | | 3 | | | | | | 1 | | 9 |
| 6 | | | | | | | | | | |
| 7 | | | | | | | | | | |
| 8 | | | 1 | | | | | | | |
| 9 | | 3 | | | | 1 | | | | |
| C | | | 1 | | | | | | | |
| E | | | | 1 | | | 1 | | 1 | |
| V | 3 | | | | 3 | | | | | |
| T | | | | | | | | | | 5 |
| REJ/NR | | | | | 1 | | 1 | 1 | | |

Fig. 9 Error matrix for 21 male talkers speaking four digit groups, tested at TTI from tape recordings.

SPOKEN

| RECOGNIZED | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | | | 89 | | 1 | | | 2 | 2 | |
| 1 | | | | | 27 | | 1 | 6 | 4 | |
| 2 | 6 | | | 5 | 1 | | | 3 | | 1 |
| 3 | | 1 | 5 | | | | | | | 3 |
| 4 | | 5 | 6 | 2 | | | 2 | 1 | 3 | |
| 5 | | 81 | 2 | 5 | 17 | | | 3 | | 25 |
| 6 | | | | 2 | | | | | 2 | |
| 7 | 1 | | | | | | | | | |
| 8 | | | 1 | 4 | | | 1 | | | 1 |
| 9 | | 9 | | | 1 | 4 | | 1 | | |
| C | | | 1 | 2 | 1 | | 2 | | 3 | |
| E | | | | 3 | | | | | 5 | |
| V | | | | | 17 | | | 1 | | 1 |
| T | | 3 | | 15 | | | | 13 | 2 | 7 |
| REJ/NR | | 3 | 2 | 5 | 4 | | 6 | | 18 | |

Fig. 10 Error matrix for 48 female talkers speaking six digit groups, tested at TTI from tape recordings.

SPOKEN

| RECOGNIZED | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | | | 45 | | | | | 1 | | |
| 1 | | | | | 31 | 3 | | | | 2 |
| 2 | 5 | | | | | | 1 | | | 1 |
| 3 | | | | | | | | 1 | 2 | 1 |
| 4 | | | | | | 2 | | | | |
| 5 | | 6 | | | 15 | | | 1 | | |
| 6 | | | | | | | | | | |
| 7 | | | | | | 1 | | | | |
| 8 | | | | 15 | | | 1 | | | 1 |
| 9 | | 3 | | | | 9 | | | | |
| C | | | | | | | | | | |
| E | | | | 1 | | | | | | |
| V | | | | | | | | | | |
| T | | | | 1 | | 1 | | | | 2 |
| REJ/NR | | | 4 | 1 | | 1 | 1 | 6 | 8 | |

Fig. 11 Error matrix for 12 talkers speaking six digit groups, both recorded and tested at RADC

SPOKEN

| RECOGNIZED | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | | | 7 | | | | | 1 | | |
| 1 | | | | | 9 | 13 | | | | 2 |
| 2 | 3 | | | | | | | | | 3 |
| 3 | | | 1 | | 1 | | | | | 4 |
| 4 | | | | | | | | | | |
| 5 | | 3 | 2 | | 15 | | | | | 25 |
| 6 | | | | | | | | | | |
| 7 | 1 | | | | 1 | | | | | 3 |
| 8 | 1 | | | 4 | | | | | | 1 |
| 9 | | | | | | | | | | |
| C | | | | | | | | | | |
| E | | | | | | | | | | |
| V | 2 | | | | | | | | | |
| T | | | | | | | | | | 3 |
| REJ/NR | | | | | 1 | | | | 3 | |

Fig. 12 Error matrix for 8 talkers speaking six digit groups, talking directly to VICI at RADC.

| RECOGNIZED \ SPOKEN | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | C | E | V | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | | | 10 | | | | | | 1 | | 1 | | 1 | 1 |
| 1 | | | | | 1 | | | | | | | | | |
| 2 | | | | | | | | | | 2 | 1 | | | |
| 3 | | | | | | | | | | | | 5 | | 10 |
| 4 | | 3 | | | | 8 | | | | | | | 10 | |
| 5 | | 2 | | | | | | | | 3 | | | 4 | 1 |
| 6 | | | | | | | | | | | | | | |
| 7 | | | | | | | | | | | 2 | | | 2 |
| 8 | | | | 1 | | | | | | | | 2 | | |
| 9 | | | | | | 4 | | | | | | | 3 | 1 |
| C | | | | | | | | | | | | | 2 | |
| E | | | | | | | | | 1 | | | | | |
| V | | | | | | | | | | | | | | 1 |
| T | | | | 3 | | | | | 1 | | | | 2 | |
| REJ/NR | | | | | | | | | 2 | | | | | |

Fig. 13 Error matrix for 26 male talkers speaking digits and control words in random order, tested at TTI from tape recordings.

| RECOGNIZED \ SPOKEN | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | | 3 | 327 | 2 | 10 | | 8 | 12 | 14 | 2 |
| 1 | | | | 5 | 110 | 80 | 8 | 7 | 4 | 5 |
| 2 | 56 | | | 8 | 3 | | 13 | 8 | 1 | 37 |
| 3 | | 2 | 23 | | 1 | | 6 | 1 | 44 | 36 |
| 4 | 1 | 29 | 7 | 5 | | 35 | 8 | 1 | 3 | 11 |
| 5 | 1 | 102 | 5 | 5 | 50 | | | 9 | | 111 |
| 6 | | | | 2 | | | | | 2 | |
| 7 | 4 | | 4 | 1 | 6 | 4 | 6 | | | 3 |
| 8 | 1 | | 3 | 38 | | | 5 | | | 3 |
| 9 | | 21 | | | 1 | 56 | 1 | 2 | | |
| C | | | 2 | 2 | 2 | | 13 | 1 | 7 | 1 |
| E | 1 | | | 6 | 1 | | 6 | | 10 | |
| V | 17 | 2 | | | 36 | 2 | | 3 | | 4 |
| T | | 3 | | 22 | | 2 | | 15 | 2 | 25 |
| REJ/NR | 1 | 3 | 23 | 16 | 47 | 20 | 11 | 19 | 31 | |

Fig. 14 Composite error matrix for all final digit tests.

## Section IV

## CONCLUSIONS AND RECOMMENDATIONS

### A. Conclusions

The VICI system developed during a previous program as a front-end for a BISS automatic speaker verification system has been modified during this program for operation under less than ideal conditions. The investigations conducted during this program which subsequently led to the modifications of the system have been centered upon accurate recognition of both male and female speech as well as accommodation for the effects of telephone line bandwidth restrictions. An important factor in the enhancement of system accuracy has been the development of an error correction algorithm which operates by the use of check digits. Error correction provides an increase in ID code accuracy for a given digit accuracy . Another highlight of this effort has been development of the software to allow speaker dependent operation of the VICI system with the vocabulary of up to 200 words.

The modified system will now operate using telephone line bandwidths with good recognition accuracy for both male and female talkers without adaptation. It has been necessary to provide a small measure of preprocessing at the transmission end of the telephone connection in order to insure good accuracy over a telephone line bandwidth. A transmitting module for connection at the sending end of a telephone connection was designed and constructed to allow this preprocessing and to provide for compensation for a noise cancelling microphone as a speech input device to this system at a remote terminal. A receiver module to allow interfacing at the receiving end of a telephone line connection to the VIP-100, upon which the system is based, has also been constructed. This module provides compensation for the low frequency losses of the telephone line. These two terminal modules allow the system to be used with any area telephone connection without interfering with normal telephone operations.

The error correction algorithm has been developed to increase the speed and accuracy of the inputting of four digit talker identification code groups. This algorithm necessitates the addition of two digits to the four digit ID group to form a new code group of six digits. This algorithm has been designed to detect all single digit errors occurring in the four digit ID portion of the code group and to correct between 80 and 90% of these single digit errors occurring in the ID group. This automatic error correction is independent of an error correction which can be made by a speaker viewing on an alphanumeric display the recognition decision immediately after a digit is spoken. The latter method of error correction is still possible with the VICI system as an alternative to automatic error correction.

The modified VICI system is now capable of single digit accuracy under less an ideal condition within approximately 2% of the very high accuracy achieved by the VICI system originally developed by TTI for high quality speech and male talkers only.

B. Recommendations

The ultimate goal for the VICI system developed under two programs by TTI has been as a front end for the BISS speaker verification system. This system as currently configured will fulfill these requirements. A field operational system could be the result of a modest additional design effort expended on the development of a suitable remote terminal which could be located at an input station for the BISS system. Such an input terminal would incorporate the present telephone transmitting module together with an automatic level control which would obviate the need for any gain adjustment at the transmitting terminal. This input station would also provide a means for controlling the user display via signals suitable for telephone-line transmission from the processor end of the telephone connection. In the present configuration of the advanced development model VICI the display used by a talker to verify the accuracy of recognition is hardwired to the VIP-100 processor. For field operation of the system it would be necessary to control the display by the use of tones which could be impressed upon the telephone connection at the processor end of the loop. The design of such a complete remote output station would be relatively simple and involves no state-of-the-art technique advances.

The capabilities of this system to operate in a wide range of BISS environments could be enhanced by an additional development program which would eliminate the use of preprocessing at the remote terminal end of the telephone connection and would lead to the use of a telephone handset as an input transducer rather than a noise cancelling microphone. Such a development program would explore the design of an accurate voiced-unvoiced detector for the VICI processor which could be used to detect the presence of friction in the several words in which fricatives appear in the VICI vocabulary. If an ordinary telephone were to be used as the complete remote input station, an alternative to the display now included in the system would be necessary to indicate recognition decisions to the talker at the remote input station. A relatively simple voice-response unit operating over the telephone connection from the central processor back to the input station would suffice to fulfill this requirement at modest cost. The speed with which an input code could be entered and thus the verification process expedited could be improved slightly if the system were modified for operation in a continuous speech mode. However, in order to maintain the versatility of the VICI system as currently implemented such modifications for continuous speech must include concatenations of all digits. The present approach with four digit codes spoken in isolation allow up to 10,000 codes to be used and recognized by the system. Any restriction of digit concatenations to accommodate a continuous speech recognition algorithm would reduce the usefulness of the system.

## References

1. P.B. Scott, "Voice Input Code Identifier", Rome Air Development Center, Griffiss AFB, New York; RADC-TR-75-188 dated July 1975 (AD B006851L).

2. Members of the Technical Staff, Bell Telephone Laboratories, Transmission Systems for Communications. Bell Telephone Laboratories, 1971.

3. J.R. Herr, "Self-Checking Number Systems", Computer Design, June 1974.

## Appendix A

## NARROW BANDWIDTH FEATURE RECOGNITION LOGIC

The modified and new feature recognition networks developed during this program to respond to talkers speaking over telephone line bandwidths can be described by the use of logic equations. The logic equations for all of the new feature networks are presented in Table A-1. These logic equations can be translated into equivalent logic diagrams. The notational rules for these logic equations are as follows:

1. An expression of the form $(\int_{T1} XQ1 - \int_{T2} YQ2)$ indicates that the excitatory quantity $Q_1$ and the inhibitory (subtractive) quantity $Q_2$ are integrated with time constants $T_1$ and $T_2$ and employ gain factors X and Y, respectively.

2. The analytical expression for the binary AND function will be of the form $C = A \cdot B$, where C represents the digital output of the AND gate for the two inputs A and B which can be in analog or digital form.

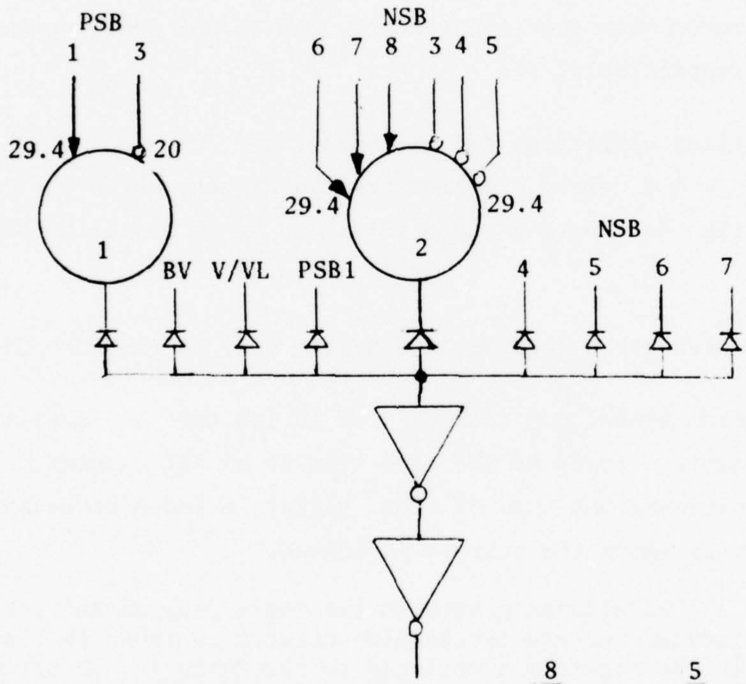
3. The expression for a logical OR function will be the form $C = A + B$.

4. The summation symbol $\sum_{m}^{n} Q$ will be used to indicate a plurality of (analog) input signals of the same type to an ATL element. In each case Q represents the type of input signal, m and n represent the interval over which the feature is summed.

An example of the relationship between the logic diagram and the logic equation for a particular feature recognition network is shown in Fig. A-1. The network shown in the figure was designed to recognize /U/ in narrow-band speech. This network includes as inputs both binary and analog representation of negative slopes. Design considerations for this network are outlined in the following paragraph.

This vowel is characterized typically by a first formant in the region of filter channels 2 to 3, with a second formant in the vacinity of filter channel 6. The effect of telephone bandwidth on this vowel is not severe. The expected positive slope in channel 1 is enhanced by the rolloff of the telephone line at low frequencies. Energy above the second formant can be expected to decrease more rapidly than above the first formant because the third formant is located at the upper end of the telephone pass band in the vacinity of filter channel 12.

This requirement is implemented by an ATL element with negative slopes in channels 6 to 8 as excitatory inputs and negative slopes in channels 3 to 5 as inhibitory inputs. The numbers shown next to the ATL inputs in Fig. A-1 indicate the input resistors (in thousands of ohms). The unity gain resistor values for the ATL elements is 34K ohms. Lower values of resistance will result in gain factors of greater than unity as can be seen in Fig. A-1.




$$U = (1.15\ PSB1 - 1.7\ PSB3) \bullet BV \bullet V/VL \bullet PSB1 \bullet 1.15\left(\sum_{6}^{8} NSB - \sum_{3}^{5} NSB\right) \bullet NSB\ (4 \cdot 5 \cdot 6 \cdot 7)$$

Figure A-1 Logic diagram and equivalent logic equation for narrow-band /U/ recognition network.

The logic equations shown in Table A-1 illustrate all of the new recognition networks as well as the modifications made in many of the networks for operation under less than optimum conditions. For every modified network, two equations are shown; the equation for the original wide-band male talker condition; and the equation for the network as modified for telephone speech with both male and female talkers. It is very difficult to categorize the changes in the networks as being either for telephone speech or for female talkers. The extensive testing and experimentation process which led to the development of these new and modified networks was conducted by the use of telephone band-limited data from both male and female talkers. Therefore, the resulting networks combine compensation for both conditions together. The following networks were modified: V/VL, /i/, /I/, /$\varepsilon_1$/, /ɜ/, /w/, UVNLC and /ʌ/. These networks were not changed: BV, Energy Gap, and Slope Gap. The new networks include /$\varepsilon_2 + \varepsilon_3$/, /u/, /U/, /a/ and /$\wedge_4$/.

TABLE A-1 PHONEME-LIKE FEATURE RECOGNITION LOGIC EQUATIONS (SHEET 1 of 7)

$$V/VL\ (new) = \left[ \left(\int_5 \sum_1^4 1.7E - \int_5 \sum_{13}^{16} 1.7E\right) + \left(\int_5 \sum_1^3 1.7E - \int_5 \sum_5^7 1.7E\right) \right] \bullet 1.27 \left[ \left(\int_5 \sum_1^4 1.7E - \int_5 \sum_{14}^{16} 1.7E\right) + \right.$$

$$\left(\int_5 \sum_7^{11} 1.7E - \int_5 \sum_{13}^{16} 1.7E\right) + \left(\int_5 \sum_{10}^{13} 2.75E - \int_5 \sum_{14}^{16} 2.75E\right) \bullet 1.27 \left[ \left(\int_5 \sum_{12}^{15} 1.7\ NSB - \int_5 \sum_3^4 1.7\ PSB\right) + \right.$$

$$\left. \left(\int_5 \sum_6^9 1.7E - \int_5 \sum_{10}^{13} 1.7\right) + \left(\int_5 \sum_1^4 1.7E - \int_5 \sum_8^{11} 1.7E\right) - \int_9 \left(\int_5 \sum_{13}^{16} 1.4E - \int_5 \sum_9^{12} 1.4E\right) \right]$$

$$V/VL\ (old) = \left[ \left(\int_5 \sum_1^3 1.7E - \int_5 \sum_4^5 1.16E\right) + \left(\int_5 \sum_1^3 1.7E - \int_5 \sum_{10}^{12} 1.7E\right) \right] \bullet 1.27 \left[ \left(\int_5 \sum_1^4 1.7E - \int_5 \sum_{15}^{19} 1.7E\right) + \right.$$

$$\left. \left(\int_5 \sum_7^{11} 1.7E - \int_5 \sum_{16}^{19} 1.7E\right) - \int_9 \left(\int_5 \sum_{15}^{19} 1.4E - \int_5 \sum_{10}^{14} 1.4E\right) \right] \bullet 1.27 \left[ \left(\int_5 \sum_{11}^{14} 1.7E - \int_5 \sum_{15}^{18} 1.7E\right) + \right.$$

$$\left. \left(\int_5 \sum_6^9 1.7E - \int_5 \sum_{10}^{13} 1.7\right) + \left(\int_5 \sum_1^4 1.7E - \int_5 \sum_8^{11} 1.7E\right) - \int_9 \left(\int_5 \sum_{15}^{19} 1.4E - \int_5 \sum_{10}^{14} 1.4E\right) \right]$$

$$BV = \left[ \left(\int_5 \sum_6^8 2.75E - \int_5 \sum_9^{11} 2.75E\right) + \left(\int_5 \sum_7^9 2.75 - \int_5 \sum_{10}^{12} 2.75E\right) + NSB8 \right]$$

TABLE A-1 PHONEME-LIKE FEATURE RECOGNITION LOGIC EQUATIONS (Sheet 2 of 7)

$$\text{i (new)} = \Big[ \big(\int_5 \sum_9^{14} 1.7E - \int_5 1.7 \sum_2^3 E\big) \bullet NSB2 \bullet NSB3 \cdot \overline{(MAX5+6+7)} \bullet V/VL \Big] \bullet \Big[ PSB\ (7 + 10) \cdot$$

$$PSB8 \bullet PSB9 \bullet \big(\sum_3^4 NSB - \sum_6^7 NSB\big) \bullet \big(\sum_8^{10} PSB - NSB4 - NSB7\big) \bullet \big(\int_5 \sum_{11}^{13} 2.7E - \int_5 \sum_{15}^{16} 2.75E\big) +$$

$$NSB4 \bullet \big(\int_5 \sum_9^{11} 2.75E - \int \sum_6^8 2.75E\big) \bullet \Big\{ 1.16 \big(\sum_9^{10} PSB - \sum_7^8 NSB - \sum_7^8 PSB\big) +$$

$$1.16 \big(\sum_7^9 PSB - \sum_{10}^{12} PSB - \sum_7^8 NSB\big) \Big\} \bullet \int_5 \sum_3^5 MAX \Big]$$

$$\text{i (old)} = \Big[ \big(\int_5 \sum_9^{14} 1.7E - \int_5 1.7 \sum_2^3 E\big) \cdot NSB2 \cdot NSB3 \cdot \overline{(\underline{MAX5+6+7})} \cdot V/VL \Big] \bullet \Big[ \underline{BRST} \Big] \bullet \Big[ (PSB7 + 10) \cdot$$

$$PSB8 \cdot PSB9 \cdot \big(\sum_3^4 NSB - \sum_6^7 NSB\big) \cdot \big(\sum_8^{10} PSB - NSB4 - NSB7\big) \cdot \big(\int_5 \sum_{11}^{14} 2.75E - \int_5 \sum_{17}^{19} 2.75E\big) +$$

$$NSB4 \cdot \big(\int_5 \sum_7^{10} 1.75E - \sum_{15}^{17} 1.75E\big) \cdot 1.16 \big(\sum_9^{10} PSB - \sum_7^8 NSB - \sum_7^8 PSB\big) +$$

$$1.16 \big(\sum_7^9 PSB - \sum_{10}^{12} PSB - \sum_7^8 NSB\big) \Big]$$

TABLE A-1 PHONEME-LIKE FEATURE RECOGNITION LOGIC EQUATIONS (SHEET 3 of 7)

$$I\ (\text{old}) = V/VL \cdot NSB3 \cdot NSB4 \cdot PSB7 \cdot PSB8 \cdot (1.15\ PSB1 - 1.7\ PSB2) \cdot \left(\int_5 \sum_{10}^{13} 2.75E - \int_5 \sum_{16}^{19} 2.75E\right) \cdot \left(\int_5 \sum_{9}^{11} 2.75E - \int_5 \sum_{6}^{8} 2.75\ E\right)$$

$$\varepsilon_1(\text{old}) = V/VL \cdot PSB\ (7+8) \cdot NSB4 \cdot (1.16\ PSB1 - 1.7\ PSB3) \cdot \left(\int_5 \sum_{9}^{11} 5.63E - \int_5 \sum_{5}^{7} 5.63E\right)$$

$$I\ (\text{new}) = V/VL \cdot NSB3 \cdot NSB4 \cdot PSB7 \cdot PSB8 \cdot \left(\int_5 \sum_{9}^{11} 2.75E - \int_5 \sum_{6}^{8} 2.75E\right) \cdot \overline{NSB9}$$

$$\varepsilon_1(\text{new}) = V/VL \cdot PSB\ (7+8) \cdot NSB4 \cdot (1.16\ PSB1 - 1.7\ PSB3) \cdot \left(\int_5 \sum_{9}^{11} 5.63E - \int_5 \sum_{5}^{7} 5.6E\right) \cdot \overline{PSB10}$$

$$\varepsilon_2 + \varepsilon_3 = V/VL \left[MAX3 \cdot (MAX9 + MAX10) \cdot (MAX2 + MAX3) + PSB1 \cdot PSB2 \cdot PSB9 \cdot PSB10\right]$$

TABLE A-1 PHONEME-LIKE FEATURE RECOGNITION LOGIC EQUATIONS (SHEET 4 of 7)

$$3\ (\text{old}) = V/VL \cdot BV \cdot NSB9 \cdot NSB10 \cdot NSB11 \cdot (NSB8 + NSB12) \cdot \left(1.4 \sum_{8}^{9} NSB - 1.4 \sum_{6}^{7} NSB\right)$$

$$3\ (\text{new}) = V/VL \bullet PSB1 \bullet PSB2 \bullet \overline{PSB10} \bullet PSB\ (6 + 8) \bullet PSB7 \bullet (1.12\ PS1 - 1.65\ PS\ 2) \bullet \overline{NSB2} \bullet$$

$$\left(.86\ PSB8 + \sum_{6}^{7} 1.12\ PSB - \sum_{4}^{5} 1.65\ PSB\right)$$

$$w\ (\text{old}) = V/VL \bullet BV \bullet \left[\left(\int_5 \sum_{1}^{3} 1.15E - \int_5 \sum_{7}^{10} 1.15E\right) \bullet \left(\int_5 \sum_{4}^{6} 1.15\ NSB - \int_5 \sum_{8}^{11} 1.15\ NSB\right) + \right.$$

$$\left. \left(\int_5 \sum_{1}^{3} 1.15E - \int_5 \sum_{5}^{9} 1.15E\right) \bullet \left(\int_5 \sum_{3}^{5} 1.15\ NSB - \int_5 \sum_{8}^{11} 1.15\ NSB\right)\right]$$

$$w\ (\text{new}) = V/VL \bullet BV \bullet \left[\left(\int_5 \sum_{1}^{3} 1.15E - \int_5 \sum_{7}^{10} 1.15E\right) \bullet \left(\int_5 \sum_{4}^{6} 1.15\ NSB - \int_5 \sum_{8}^{10} 1.15\ NSB\right) + \right.$$

$$\left. \left(\int_5 \sum_{1}^{3} 1.15E - \int_5 \sum_{5}^{9} 1.15E\right) \bullet \left(\int_5 \sum_{3}^{5} 1.15\ NSB - \int_5 \sum_{8}^{10} 1.15\ NSB\right)\right]$$

TABLE A-1 PHONEME LIKE FEATURE RECOGNITION LOGIC EQUATIONS (SHEET 5 of 7)

$$UVNLC(old) = \overline{V/VL}\left[\left(\int_{14}\sum_{15}^{19} 1.4E - \int_{14}\sum_{10}^{14} 1.4E\right) + \left(\int_{14}\sum_{15}^{19} 1.4E - \int_{14}\sum_{5}^{9} 1.4E\right)\right] \cdot \left[\underline{BRST}\right] \cdot$$

$$\left[\underline{\left(\int_{14}\sum_{15}^{19} 1.4E - \int_{14}\sum_{1}^{4} 1.7E\right)}\right]$$

$$UVNLC(new) = \overline{V/VL}\left[\left(\int_{14}\sum_{13}^{16} 1.7E - \int_{14}\sum_{9}^{12} 1.7E\right) + \left(\int_{14}\sum_{13}^{16} 1.7E - \int_{14}\sum_{5}^{8} 1.7E\right)\right] \cdot \left[\left(\int_{14}\sum_{13}^{16} 1.4E - \int_{14}\sum_{1}^{4} 1.7E\right)\right]$$

$$\wedge(new) = V/VL \cdot BV \cdot \left(1.4\sum_{6}^{7} PSB - 1.4\sum_{8}^{9} PSB\right) \cdot \left(1.4\sum_{8}^{9} NSB - 1.4\sum_{6}^{7} NSB\right) \cdot PSB2 \cdot PSB3$$

$$\wedge(old) = V/VL \cdot BV \cdot \left[\left(1.4\sum_{6}^{7} PSB - 1.4\sum_{8}^{9} PSB\right) + \left(1.4\sum_{8}^{9} NSB - 1.4\sum_{6}^{7} NSB\right)\right] \cdot$$

$$\left[PSB\ (2+3) + PSB\ (3+4) \cdot PSB2 \cdot \left(\int_{5}\sum_{10}^{13} 2.75E - \sum_{16}^{19} 2.75E\right)\right]$$

TABLE A-1 PHONEME-LIKE FEATURE RECOGNITION LOGIC EQUATIONS (SHEET 6 OF 7)

u $= V/VL \cdot BV \cdot \overline{PS\ (9+10)} \cdot NSB4 \cdot (2.3\ NSB5 - 2.3\ MAX\ 6) \cdot \left(\int_5 \sum_1^4 E - \int_5 \sum_5^8 E\right) \cdot$

$$1.15 \left[ \left(\int_5 \sum_2^6 2.75E - \int_5 \sum_7^{11} 2.75E\right) - \left(\int_5 \sum_8^{10} 2.75E\right) - \left(\int_5 \sum_9^{11} 2.75E - \int_5 \sum_6^{8} 2.75E\right) \right]$$

Energy Gap ($EGap_1 + EGap_2$) $= \left[ \int_{2.2} 3.4 \int_{55} \left(\int_5 \sum_5^9 .25E - 5.63\ UVNLC\right) - \int_{2.2} 3.4 \int_5 \sum_5^9 .25E) \right] +$

$$\left[ \left(\int_{2.2} 3.4 \int_{35} \left(\int_5 \sum_{10}^{14} .25E - 5.63\ UVNLC\right) - \int_{2.2} 3.4 \int_5 \sum_{10}^{14} .25E\right) \right]$$

Slope Gap $= \int_{2.2} 2.3 \int_{35} (x) - \int_{2.2} 3.4\ (x)$

where $(x) = \left[ \sum_7^{10} 0.25\ NSB - \int_{19} (0.5\ NSB1 + 0.4\ NSB2 + 0.25\ NSB3 + 0.4\ NSB4 + 0.5\ PSB5) \right]$

TABLE A-1 PHONEME-LIKE FEATURE RECOGNITION LOGIC EQUATIONS (SHEET 7 OF 7)

$$U = V/VL \bullet BV\ PSB1\ NSB\ (4 \cdot 5 \cdot 6 \cdot 7) \bullet (1.15\ PSB1 - 1.7\ PSB3) \bullet 1.15 \left( \sum_{6}^{8} NSB - \sum_{3}^{5} NSB \right)$$

$$\text{ɔ} = V/VL \bullet BV \bullet PSB2 \bullet NSB6 \bullet NSB7 \bullet (1.15\ PSB1 - 1.7\ PSB3) \bullet 1.15 \left( \sum_{6}^{7} NSB - \sum_{4}^{5} NSB \right)$$

$$\text{ɑ} = V/VL \bullet PSB2 \bullet PSB3 \bullet NSB8 \bullet NSB9 \bullet \quad PSB4 + NSB\ (6 \cdot 7)$$

$$\wedge_4 = BV \cdot 1.8\ (NSB4 - NSB2 - NSB3 - \int_5 \sum_{5}^{6} NS)$$

## Appendix B

## DESCRIPTION AND OPERATING INSTRUCTIONS - STRUCTURED VOCABULARY WORD RECOGNITION PROGRAM

This speech recognition program is designed to operate with the VICI system or any other TTI VIP-100 which includes a Nova 800, Nova 1200 or Nova 2 with 16K of core memory. This program has the capability of recognizing up to 200 separate words in a syntactic structure. This structure allows recognition of a maximum of 30 words in any one node of the structure. Up to 30 nodes can be included in the sentence structure. The words to be included in each mode as well as the node sequences can be changed at will by the use of Teletype input. The program is speaker dependent; it therefore, requires the input of training data by a speaker.

Once a particular node structure has been established, two types of operation are possible, Sequential and Optional. In the sequential operation a talker must follow the predetermined sequence of nodes when inputting speech data. Any number of words in a particular mode may be inputted. Each node is terminated and the next node in the sequence is made active by the command word GO. The optional node type of operation allows the operator to choose by voice command, any of up to 30 nodes for use at a particular time. The node chosen is made active by inputting the two digits of the node number. Another node may be subsequently made active by exiting from the current node by the use of the GO command and then speaking the two digit node number. Figure B-1 is a set of flowcharts of the program.

The following paragraphs describe operation of the structured program.

A. General Operating Procedures for Program

1. Vocabulary

The total vocabulary capability is 200 words. The first 13 words are fixed; the remainder of the vocabulary may be selected by the user. Up to and including 30 nodes may be included in the structure with up to and including 30 words in each node. A particular vocabulary word may be included in any number of nodes. Every vocabulary word has a number from 0 thru 199 which must be used for constructing nodes, for training and for constructing display messages. Any vocabulary word (except the first 13) may be represented on the output display by up to 16 letters (or digits or combination thereof). Vocabulary words 0 to 9 are reserved for digits 0 to 9, word 10 is the command GO word, 11 is the command ERASE and word 12 is the command CANCEL. All other words are chosen by the operator. The command GO is always active. It serves to terminate the active node and to make active the next node in sequential or allow input of a node number in optional operation. The command ERASE is always active and erases the last spoken word. In sequential operation, the node number and name cannot be erased. The command CANCEL is always active but has slightly different functions in optional and in sequential operation. In optional operation CANCEL deletes all spoken inputs in the active node as well as the node itself. In sequen-

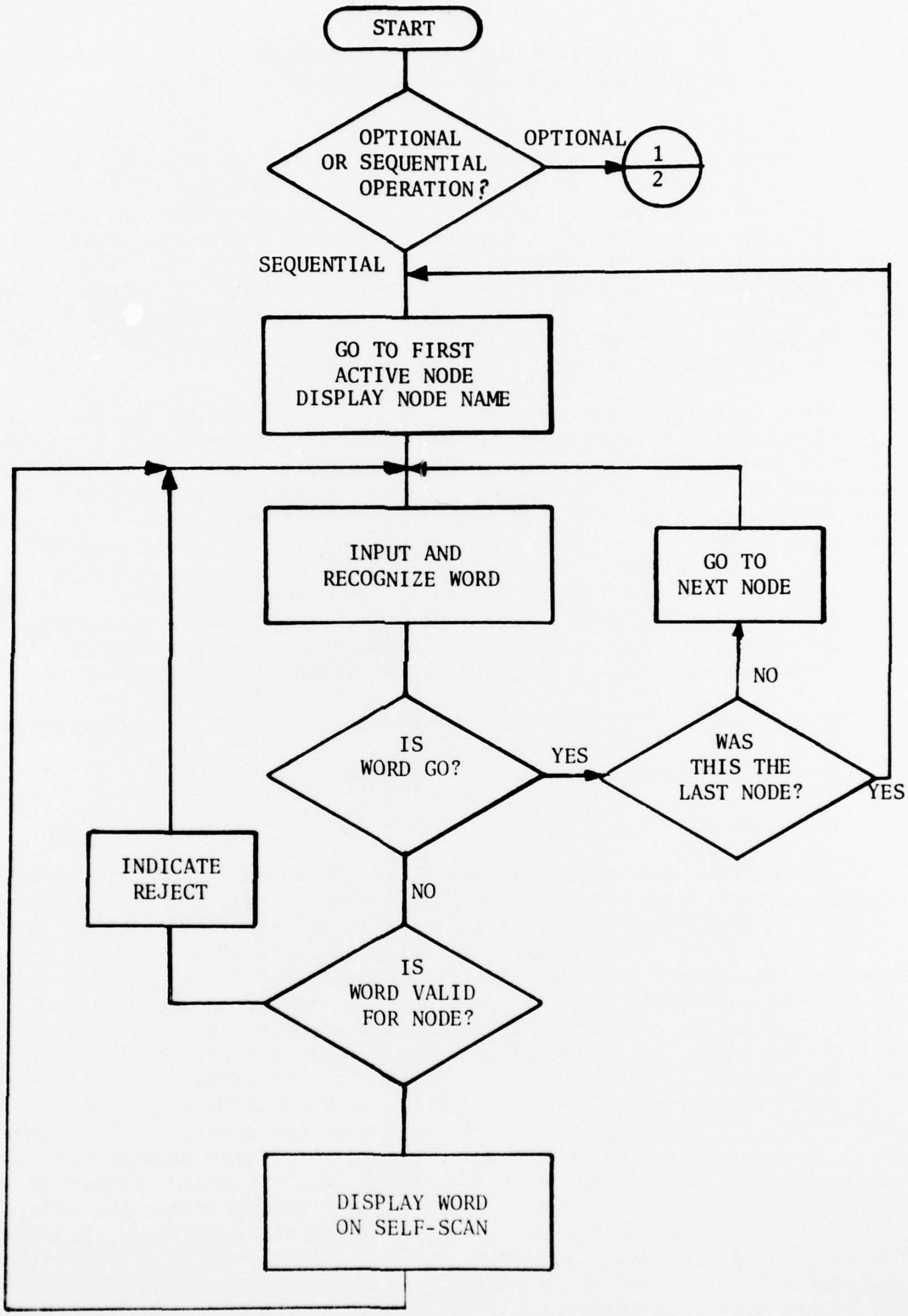



Fig. B-1 Flowchart of 200 word program (First Sheet)

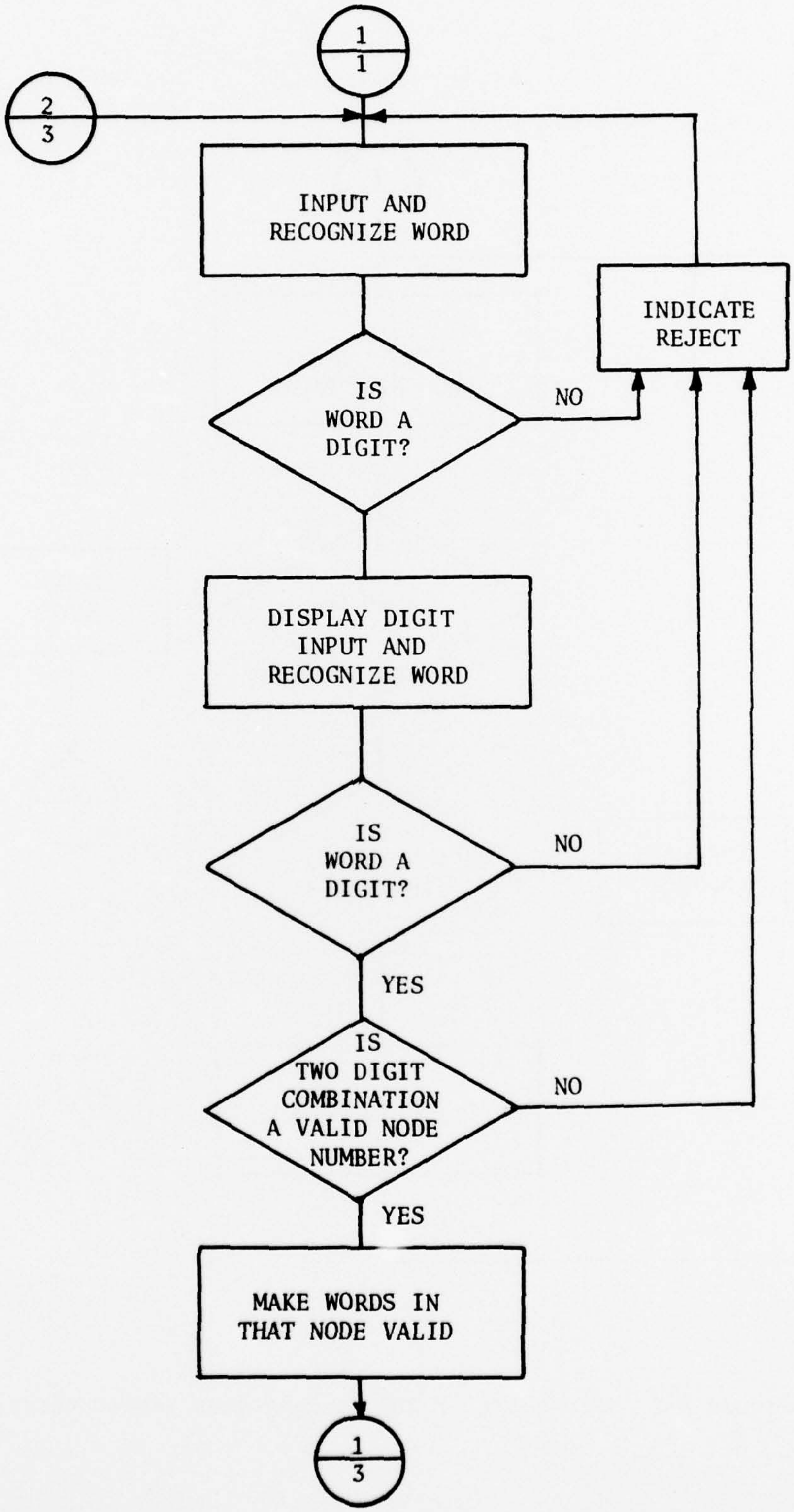



Figure B-1 Flowchart of 200 word program (second sheet)

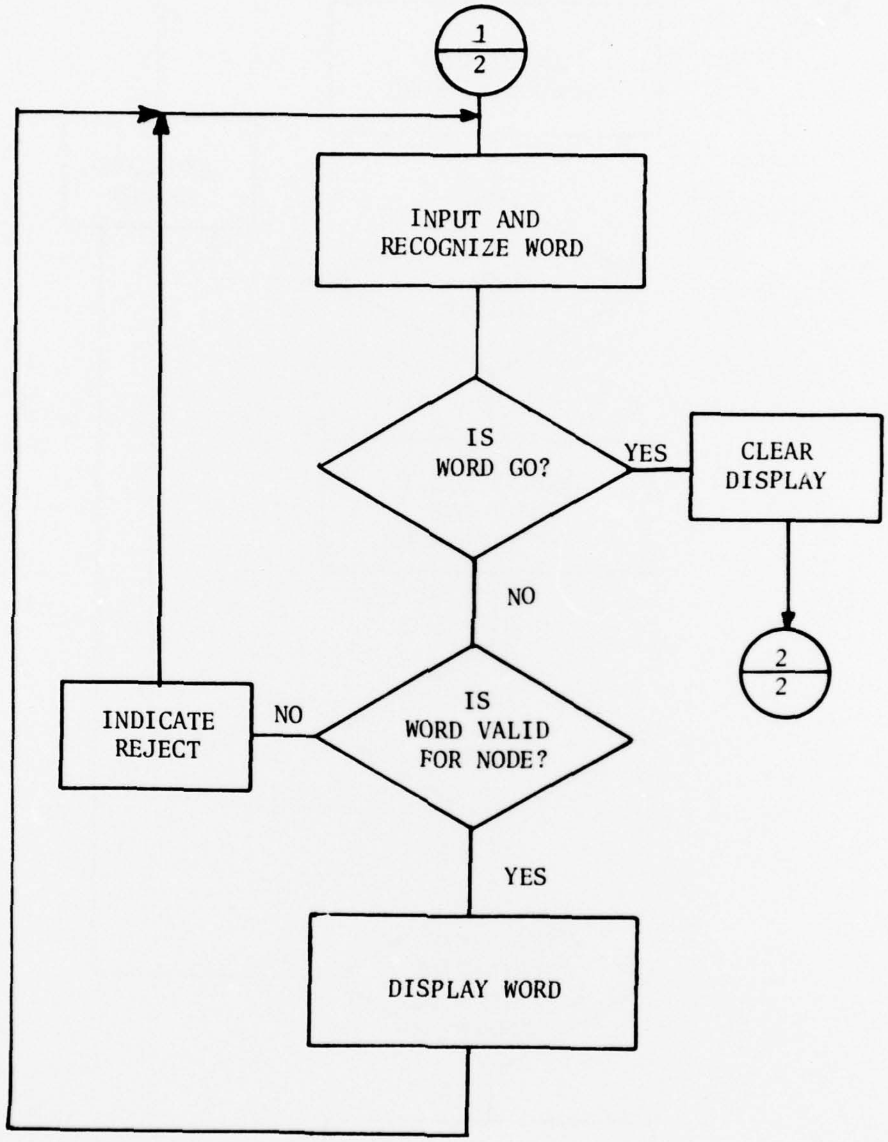



Figure B-1 Flowchart of 200 word program (sheet three) tiȧl operation, CANCEL deletes all spoken inputs in a node but cannot delete the node.

B. Node Structure

As mentioned above, up to 30 nodes may be included in a sentence structure. Each node must be identified with a two digit number from 00 to 99. Any node may be also given a name for display purposes. A node name may be up to 14 characters in length. A node name will appear on the output display along with the node number when the node is made active in either the sequential or optional nodes of operation. In sequential operation, the order of nodes is from lower to higher node numbers. When the program starts in the sequential mode, the lowest numbered node is always active first. After all nodes have been completed, the program reverts again to the lowest numbered node.

C. Starting the Program

The starting address of the program is 40 (octal). This address is automatically available on the Turnkey Console of the VICI Nova 1200 computer. Pressing the Reset and the Start switch causes the following message on the TTY:

TYPE 1 FOR INSTRUCTIONS

Typing a 1 will result in the following Teletype message:

STRUCTURED RECOGNITION PROGRAM FOR 2ØØ WORDS

TYPE:

T - TRAIN
I - INPUT TRAINING DATA
O - OUTPUT TRAINING DATA
A - INPUT NODE DATA
B - OUTPUT NODE DATA
G - GO TO RECOGNITION PHASE
S - START THE SYSTEM
M - MODIFY DISPLAY MESSAGES FOR WORDS IN VOCABULARY
H - PRINT ACTIVE NODES TOGETHER WITH WORDS IN EACH ACTIVE NODE
V - PRINT VOCABULARY
Q - EDIT NODE STRUCTURE
C - MODIFY NODE DISPLAY MESSAGES

The above commands and the operations which they each activate are explained in the following paragraphs.

1. Command T - for training the system

The training routine is usually the first to be executed. It has two primary functions:

(1) To specify the number of words in the vocabulary;

(2) To adapt the recognition system for the voice characteristics of the particular user.

The training mode is accessed by typing the letter "T" (remember that all keyboard entries must be terminated with a carriage return). The Teletype will reply "NO. OF REPS?". The user should reply with a number from 1 to 10 to indicate the number of repetitions of each vocabulary word which will be spoken to train the system. The most reliable recognition results are obtained when 10 training samples are used. Once the number of training samples has been specified, the Teletype will print "A OR I" to ask whether the operator desires to train all words (A) or an individual word or words (I). In initially operating the system, train all words by responding with the letter "A". The Teletype will respond by printing "VOCABULARY SIZE?". The operator should respond with a number between 1 and 200 to indicate the first word to be trained. At this time the operator for which the system is to be trained should properly position the microphone and set the volume to the appropriate level. Speech entered into the system prior to pressing the return key will be ignored by the system. The system is now ready to accept the previously specified number of training repetitions for each of the vocabulary words. Consecutive samples of a given vocabulary word are entered in sequence. That is, all samples of the first vocabulary word should be entered first. The display will then indicate the second word to be trained and continue displaying that word until all training samples have been entered. The process will be continued until the entire vocabulary is trained. After training one or more words individually (I) training is terminated by replying with the word number 200.

The system will after training respond in the same manner as it will to the S command. The Teletype will type one of the following two messages (depending on past operation):

(1) SYSTEM IS SET FOR SEQUENTIAL NODES, IF OK TYPE 1; or
(2) SYSTEM IS SET FOR OPTIONAL NODES, IF OK TYPE 1

If a node structure has been already selected (or inputted from tape), as outlined below, then the desired operation may be selected by typing a 1 (or another character) as is appropriate, followed by a carriage return. If nodes have not yet been selected, then they must be selected in the following manner. First type a "P" while holding the CONTROL (CTRL) button. Then type Q for edit and proceed as outlined below.

2. Command Q - for editing nodes.

This allows a simple editing function for the nodes and the words associated with each node.

The message:

A,D,T?

is typed and the user types one of the above to select a function.

A - Add a node
D - Delete a node
T - Terminate the edit

Note - to change the vocabulary of a node, first delete the node then add it.

(a) The ADD function - A - typing an A causes the TTY to respond: "NODE" - the valid response is any number 0-99. If a NODE exists with that number, the program repeats its request. Every NODE must have a number because in the optional node operation nodes are activated by saying the two digit number.

"Name of this NODE" - input is up to 14 alphanumeric characters and is the message which will appear on the display when this NODE is activated in either sequential or optional operation. No more than 14 characters will be accepted. If no name is required, type a carriage return.

"Word Set Same As NODE #?" - allows this NODE to have the same conditions as an existing NODE. Typing a valid NODE # completes the ADD function, and the TTY responds with "A, D, T?" typing a non-existent causes the program to advance.

"1 FOR DIGITS" - Type "1" to attach the single digit vocabulary 0-9 as a set of conditions for this category. Typing anything other than 1 skips the digit vocabulary. In both cases, the program advances.

"WORD #" - Typing any number which corresponds to a vocabulary word attaches that word as a condition to the category. To terminate this process the user types 200 and the program requests:

A,D,T?

Message which may appear during the ADD function:

"NODE LIST FILLED" - there is no room for the node. The program can accommodate up to 30 nodes.

NOTE - It is extremely important that only a well ordered exit be made from the ADD function. If an error is made during ADD, complete the function and do not type "Control P".

(b) The DELETE function - D -

Typing a D causes the TTY to respond with: "NODE #" - the number of the category to be deleted, an illegal number causes the request to be repeated. It is also possible to make a well ordered exit from the Edit function at this point by typing "Control P".

Messages which may appear during DELETE:

ALL NODES DELETED

(c) The TERMINATE function - T - typing a "T" causes the editing of nodes to be terminated and the TTY will type the message:

TYPE 1 FOR INSTRUCTIONS"
after which another command can be issued.

3. Command C - for modifying node names without changing node structures. The TTY responds to a "C" by outputting the message:

"NODE #" - the number of the node whose message is to be changed. Typing a non-existent category number causes this function to terminate. Typing a valid number causes a line feed and the program waits for input of up to 14 characters followed by a carriage return.

4. Command B - for outputting node data to paper tape.

The node structure entered by the use of the command Q may be saved for future use, on paper tape. After turning on the paper tape punch, type the command B. The node structure data tape will be punched with leader at both ends. Turn off the punch before further operation. Display messages for vocabulary data is included on node structure data tapes.

5. Command A - for inputting a node structure from paper tape.

A node structure and reference which have previously been outputted on tape as outlined above may be inputted when needed. Place the node structure data tape in the tape reader of the TTY and turn on the reader. Then type the command, A and a carriage return. The tape will be read.

6. Commands S and G - for inputting speech data.

The system can be put into one of two speech recognition states by the use of the command S - to start the system. As previously mentioned (under training) the system will respond with a message telling the operator that the system is set for either sequential or for option operation. The operator can either accept the type of operation indicated by typing 1 and a CR or can go to the other choice by typing any other character and CR. The command G will go directly to recognition without the option of changing the manner of operation.

7. Command O - for outputting to paper tape of training data.

The reference data compiled during training may be saved on punched paper tape for future use. The resulting tape will retrain the system for the particular operator and vocabulary when it is read into the system with the appropriate command. The reference data tape is produced with the output "O" command. Type the "O" command followed by carriage return and turn the Teletype punch on. The computer will punch the paper tape. The reference training data will be punched out complete with leader at both ends of the tape. The Teletype will print "TYPE 1 FOR INSTRUCTIONS" when the tape is completed. Turn the punch off before further operation. The system will still be trained for the operator when the output routine is completed since execution of this routine does not modify the training data.

8. Command I - for inputting training data from tape.

The system may be trained from a previously produced reference data paper tape by use of the "I" command. The reference data tape should be placed in the tape reader first; the reader control should then be set to the start position. The "I" command should then be entered on the keyboard followed by a carriage return. The paper tape will be read. The training data from the tape will replace the current training data (including vocabulary size) for the selected speaker. CAUTION, do not press any Teletype keys while the tape is being read.

9. Command M - to modify display messages for vocabulary words (regardless of node structure).

The display characters and the corresponding Teletype keys are shown in Fig.B-1. The characters enter the display at the right and overflow from the left. Thus, if less than 16 characters (including spaces) are entered into the display, the previously displayed message will not be erased but merely shifted to the left a corresponding number of character spaces. This mode of operation allows the results of several consecutive word recognitions to be displayed; thus, an entire string of individually entered digits or other commands may be displayed simultaneously.

Two message modify entry techniques are available if the operator does not wish to retain any of the previously displayed message. He may enter a full 16 characters during the message modify instruction by using spaces to fill character positions not needed for the actual message. As an alternative, he can enter control A (hold down control key while pressing the A key) as the first character of the message and then enter the message he wants displayed. The control A character will cause the display to be cleared of all previously displayed characters before the new message is displayed. The operator must enter the correct number of leading spaces with either technique if the message is to be centered in the display.

The message modify routine is called by typing M on the Teletype. The Teletype will respond with "WORD NO.?". The operator should reply with the number of the first word for which the display message is to be modified (remember that the first word is word number Ø). The Teletype will respond with a carriage return and line feed. The operator should then enter the new message to be displayed for the particular vocabulary word. A total of 16 or less characters including spaces and control characters should be entered; the message should be terminated with a carriage return. The Teletype will then ask for the word number of the next word for which the display message is to be modified. The rub-out feature is not operational during the character entry procedure; if pressed, it will appear as a "?" in the displayed message.

Three special control characters are available during the character entry procedure. They are called:

| Display Character | Keyboard Entry | Display Character | Keyboard Entry | Display Character | Keyboard Entry |
|---|---|---|---|---|---|
| 0 | Shift P | U | U | + | + |
| A | A | V | V | , | , |
| B | B | W | W | - | - |
| C | C | X | X | . | . |
| D | D | Y | Y | / | / |
| E | E | Z | Z | Ø | Ø |
| F | F | | Shift K | 1 | 1 |
| G | G | | Shift L | 2 | 2 |
| H | H | | Shift M | 3 | 3 |
| I | I | | Shift N | 4 | 4 |
| J | J | | Shift O | 5 | 5 |
| K | K | Space | Space | 6 | 6 |
| L | L | ! | ! | 7 | 7 |
| M | M | " | " | 8 | 8 |
| N | N | # | # | 9 | 9 |
| O | O | $ | $ | : | : |
| P | P | % | % | ; | ; |
| Q | Q | & | & | | Shift , |
| R | R | ' | ' | = | = |
| S | S | ( | ( | | Shift . |
| T | T | ) | ) | ? | ? |
| | | * | * | | |

Fig. B-2 Display Character Set with Corresponding Teletype™ Entry Symbol

Control A--clears the display of current message
Control B--blanks the display for approximately 250 milliseconds
Control C--backspaces the current message one position

Control may be returned to the selection routine when all desired message modifications have been completed. This is accomplished by answering the "WORD NO.?" request with 200.

10. Command H - for printing node vocabulary.

Typing the command H will result in a Teletype printout of each node vocabulary list. The nodes will appear in numerical order. Each node list will be headed by the node number (to the left) and the node name (to the right). Below the node number will be numbers of the words in the node vocabulary. Below the node name will be the display messages for the node vocabulary words.

# METRIC SYSTEM

## BASE UNITS:

| Quantity | Unit | SI Symbol | Formula |
|---|---|---|---|
| length | metre | m | ... |
| mass | kilogram | kg | ... |
| time | second | s | ... |
| electric current | ampere | A | ... |
| thermodynamic temperature | kelvin | K | ... |
| amount of substance | mole | mol | ... |
| luminous intensity | candela | cd | ... |

## SUPPLEMENTARY UNITS:

| Quantity | Unit | SI Symbol | Formula |
|---|---|---|---|
| plane angle | radian | rad | ... |
| solid angle | steradian | sr | ... |

## DERIVED UNITS:

| Quantity | Unit | SI Symbol | Formula |
|---|---|---|---|
| Acceleration | metre per second squared | ... | m/s |
| activity (of a radioactive source) | disintegration per second | ... | (disintegration)/s |
| angular acceleration | radian per second squared | ... | rad/s |
| angular velocity | radian per second | ... | rad/s |
| area | square metre | ... | m |
| density | kilogram per cubic metre | ... | kg/m |
| electric capacitance | farad | F | A·s/V |
| electrical conductance | siemens | S | A/V |
| electric field strength | volt per metre | ... | V/m |
| electric inductance | henry | H | V·s/A |
| electric potential difference | volt | V | W/A |
| electric resistance | ohm | | V/A |
| electromotive force | volt | V | W/A |
| energy | joule | J | N·m |
| entropy | joule per kelvin | ... | J/K |
| force | newton | N | kg·m/s |
| frequency | hertz | Hz | (cycle)/s |
| illuminance | lux | lx | lm/m |
| luminance | candela per square metre | ... | cd/m |
| luminous flux | lumen | lm | cd·sr |
| magnetic field strength | ampere per metre | ... | A/m |
| magnetic flux | weber | Wb | V·s |
| magnetic flux density | tesla | T | Wb/m |
| magnetomotive force | ampere | A | ... |
| power | watt | W | J/s |
| pressure | pascal | Pa | N/m |
| quantity of electricity | coulomb | C | A·s |
| quantity of heat | joule | J | N·m |
| radiant intensity | watt per steradian | ... | W/sr |
| specific heat | joule per kilogram-kelvin | ... | J/kg·K |
| stress | pascal | Pa | N/m |
| thermal conductivity | watt per metre-kelvin | ... | W/m·K |
| velocity | metre per second | ... | m/s |
| viscosity, dynamic | pascal-second | ... | Pa·s |
| viscosity, kinematic | square metre per second | ... | m/s |
| voltage | volt | V | W/A |
| volume | cubic metre | ... | m |
| wavenumber | reciprocal metre | ... | (wave)/m |
| work | joule | J | N·m |

## SI PREFIXES:

| Multiplication Factors | Prefix | SI Symbol |
|---|---|---|
| 1 000 000 000 000 = $10^{12}$ | tera | T |
| 1 000 000 000 = $10^{9}$ | giga | G |
| 1 000 000 = $10^{6}$ | mega | M |
| 1 000 = $10^{3}$ | kilo | k |
| 100 = $10^{2}$ | hecto* | h |
| 10 = $10^{1}$ | deka* | da |
| 0.1 = $10^{-1}$ | deci* | d |
| 0.01 = $10^{-2}$ | centi* | c |
| 0.001 = $10^{-3}$ | milli | m |
| 0.000 001 = $10^{-6}$ | micro | μ |
| 0.000 000 001 = $10^{-9}$ | nano | n |
| 0.000 000 000 001 = $10^{-12}$ | pico | p |
| 0.000 000 000 000 001 = $10^{-15}$ | femto | f |
| 0.000 000 000 000 000 001 = $10^{-18}$ | atto | a |

* To be avoided where possible.

# *MISSION*
## *of*
## *Rome Air Development Center*

*RADC plans and conducts research, exploratory and advanced development programs in command, control, and communications ($C^3$) activities, and in the $C^3$ areas of information sciences and intelligence. The principal technical mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, ionospheric propagation, solid state sciences, microwave physics and electronic reliability, maintainability and compatibility.*