VOXBOX™ Speech Recognition Device

Catalog Number 26-1181

Radio Shack

TRS-80™ MICRO COMPUTER SYSTEM

HARDWARE

CUSTOM MANUFACTURED IN U.S.A. FOR RADIO SHACK A DIVISION OF TANDY CORPORATION

Congratulations for selecting this Radio Shack Computer Product!

The TRS-80 VOXBOX speech recognition unit allows you to enjoy this new technology on your Level II TRS-80 computer. Now you may use words and phrases to control and instruct your computer — and to enter data. Focus on the work or play at hand — such as on the Video Display — without the distraction of having to type on the keyboard.

- Simple to use — connects directly to TRS-80 or to Expansion Interface TRS-80 Bus Card Edge.
- Includes dynamic push-to-talk mic with coiled cord for communication with the VOXBOX.
- Includes machine-language "driver" program and three application programs.
- Using the examples and instructions in this manual, you can write your own programs for custom applications.
- Use the training mode to teach the VOXBOX up to 32 words or phrases, each word/phrase up to 1.2 seconds long.
- Re-train it at any time to recognize a different 32-word vocabulary . . . . in any language!

Note: Speech recognition is a new technology. In fact, your Radio Shack VOXBOX is one of the first such devices to be both available and affordable to general consumers. For this reason, Radio Shack recommends that the unit be used primarily for entertainment and experimentation. Proceed advisedly before committing the unit to any serious application.

Required Equipment

TRS-80 with 16K RAM
Cassette Recorder
Expansion Interface optional

Contents

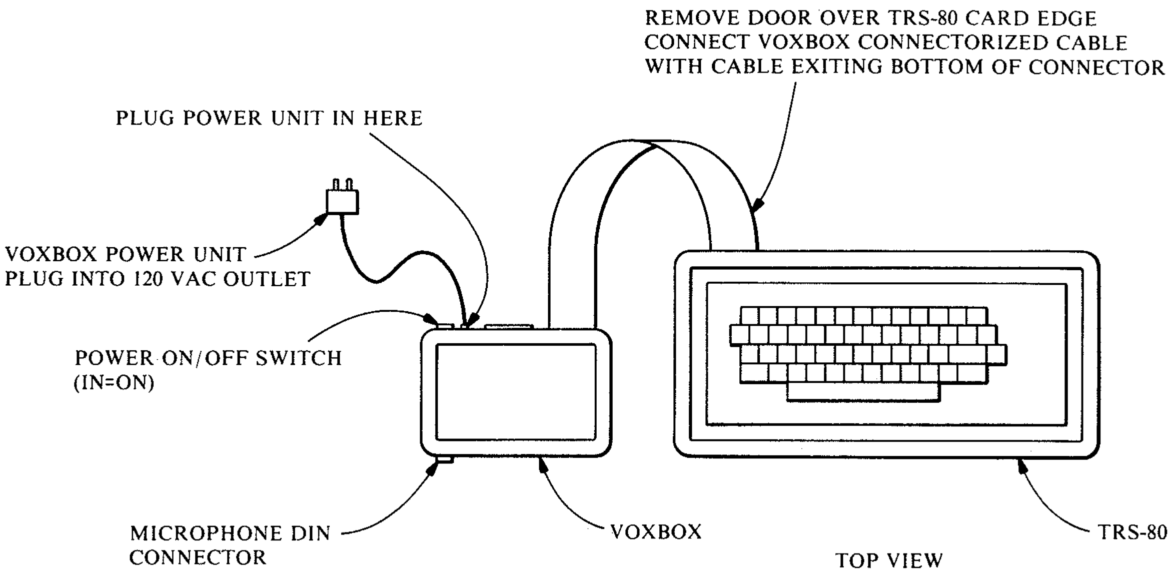
Appendices

1. Setting Up

Connecting the VOXBOX to the TRS-80

(See Figure 1.)

1. Turn off the TRS-80; leave the VOXBOX unplugged.
2. Remove the curved hood on the left rear of the TRS-80, exposing the TRS-80 card edge. Connect the VOXBOX cable to this card-edge, with the ribbon cable exiting from the **bottom** of the connector.
3. Plug the microphone into the DIN jack on the front of the VOXBOX. Take care to make the correction properly.
4. Connect the power-supply mini-plug into the back of the VOXBOX.
5. Plug the VOXBOX power supply into a source of 120 VAC.
6. Follow the power-up instructions below.



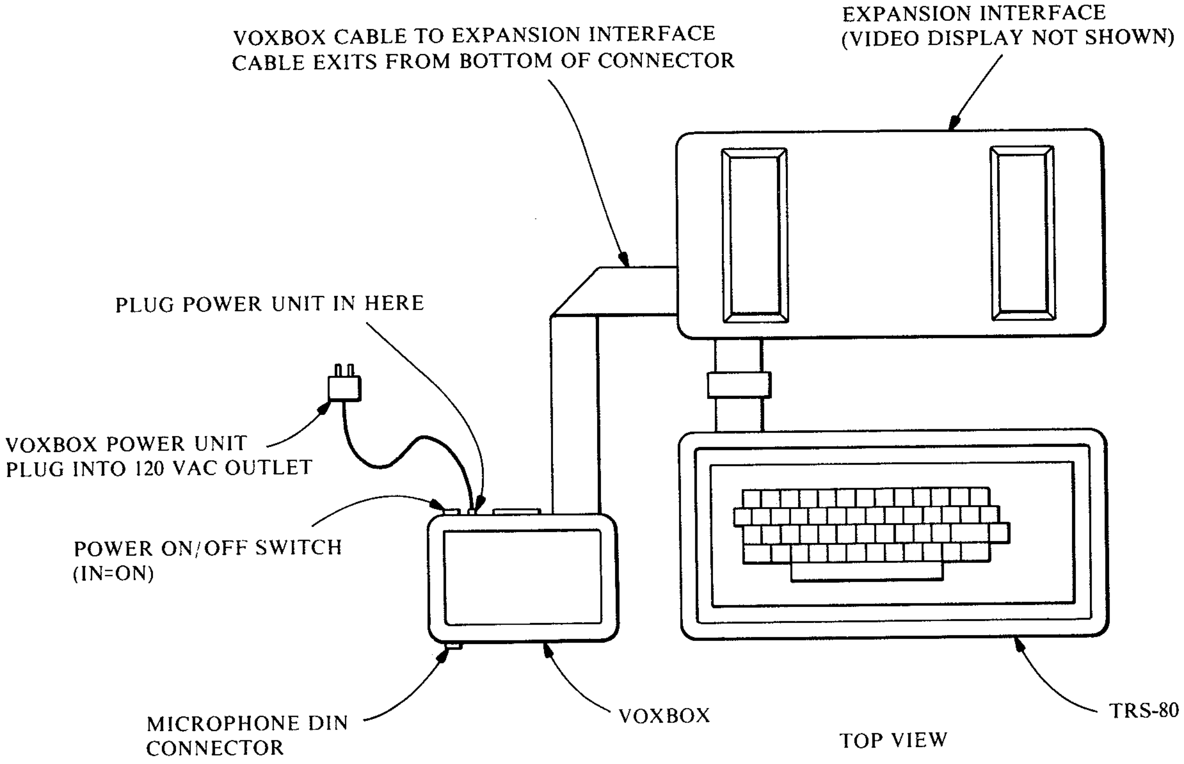
CONNECTING VOXBOX TO YOUR TRS-80

FIGURE 1

Connecting the VOXBOX to the Expansion Interface

(See Figure 2.)

1. Turn off the TRS-80; leave the VOXBOX unplugged.
2. Connect the other end of the VOXBOX cable to the TRS-80 Bus Card-Edge (left side, front, on the Expansion Interface), with the ribbon cable exiting from the **bottom** of the connector.
3. Plug the microphone into the DIN jack on the front of the VOXBOX. Take care to mate the connection properly.
4. Connect the power-supply mini-plug into the back of the VOXBOX.
5. Plug the VOXBOX power supply into a source of 120 VAC.
6. Follow the power-up instructions below.



CONNECTING VOXBOX TO YOUR EXPANSION INTERFACE

FIGURE 2

2. Powering Up and Loading the Driver Program

Okay, you've got the hardware set up now. But the VOXBOX is still only halfway installed.

That's because an essential part of the VOXBOX is the software—the driver program that communicates back and forth between your application program and the hardware. Before you can begin using the VOXBOX, you have to "install" the driver program.

This driver program is a machine-language file stored on cassette, and loaded into the highest RAM addresses in your machine. There are three different machine-language tapes, SP16, SP32, SP48, corresponding to the three possible memory sizes, 16K, 32K and 48K RAM.

1. Turn on the VOXBOX. The Power Button is on the rear of the case at the left.
2. Turn on any other peripherals you are going to use.
3. Turn on the Computer as usual. (TRSDOS users should load and execute the BASIC interpreter.)
4. Now it's time to reserve memory for the driver program. Answer MEMORY SIZE? according to the following table:

IF YOU HAVE	ANSWER MEMORY SIZE?
16K RAM	28671
32K RAM	45055
48K RAM	61439

5. Now locate the proper SPnn tape for your System. (For 16K RAM, use SP16; 32K, SP32; 48K, SP48). Prepare your recorder to play the tape.

6. TRSDOS users type CMD “T” ENTER before cassette operations.

   Type:

```
        SYSTEM  ENTER
        *? SPnn  ENTER          nn=16, 32 or 48
```

   The tape will now load, as indicated by the blinking asterisks on the Display. If there is an error, one of the blinking asterisks on the Display will be replaced by a “C.” Wait till the tape stops, adjust the volume, and repeat step 6. If the tape doesn't stop, reset the and start over at step 1. (The program is recorded twice on each side of the tape.)
7. Once the tape has loaded successfully, (the *? returns and there's no C in the upper right corner), press BREAK :

```
        *? BREAK
```

The VOXBOX is now completely “installed” and ready to run one of the applications programs.

The driver program will remain intact until you turn off the computer or replace the contents of the highest 4096 RAM bytes in your computer. Pressing RESET will not affect the driver program. (But always reserve memory as in the preceding table.)

Saving the Driver Program on Disk

With a disk system, you can save the driver program (SPnn) in a disk file. Use the TAPEDISK utility program described in the TRSDOS Reference Manual. Prepare the recorder to play the correct SPnn tape, and type:

```
        TAPEDISK   ENTER
        ? C        ENTER
        ? F SPnn/CMD:Ø aaaa bbbb 4Ø2D      ENTER
        ? E        ENTER
```

Get values for nn, aaaa and bbbb from the table below.

Memory Size	nn	aaaa	bbbb
16K	16	7000	7FFF
32K	32	B000	BFFF
48K	48	F000	FFFF

Once you've saved the driver program, you can load it in the DOS READY mode by typing:

```
        LOAD SPnn/CMD ENTER    (nn=16,32 or 48)
```

Then load BASIC and answer MEMORY SIZE as described on the preceding page.

3. Hints on Using VOXBOX

Computer Speech Recognition is still in its early stages. You cannot, for example, expect the VOXBOX to understand you **unless you make every effort to make yourself understood.** Put another way, it is easy to fool the VOXBOX; the goal is to **communicate** with your Computer through it.

To obtain the best results from your VOXBOX, always follow the following suggestions:

Environmental Conditions

Your VOXBOX, like you, will work best in a clean and **quiet** room. Background noise will be picked up by the microphone and mixed with speech, reducing the recognition accuracy of the VOXBOX.

Using the Microphone

Hold the microphone almost touching your lips. Movement of the microphone introduces additional variations into the system which increase the chance of recognition mistakes.

When VOXBOX is ready for you to talk, press the push-to-talk button on the Mic; release the button when you have finished pronouncing the word or phrase.

Speak slowly and distinctly into the Mic, using as little inflection as possible. You will need to pronounce your words consistently, so don't use a forced or unnatural pronunciation.

Choose Your Language and Words

You may use any language or words with VOXBOX, programming it in the language and words of choice. Even though the video display may "prompt" you in English, it is **not** necessary to use the same English word.

Be consistent in your speech

Once you have trained the VOXBOX on a given vocabulary, pronounce the words consistently so the VOXBOX can recognize them more reliably.

4. Running the Application Programs

Three BASIC programs are included on tape to let you get used to the VOXBOX:

- An Inventory Demonstrator that lets you specify items and quantities from an imaginary inventory.
- A Lunar Lander that simulates the landing of a space craft on the lunar surface. You control the rate of descent by commanding the craft to fire specified fuel bursts.
- A Voice Plotter that gives you a graphic representation of any sound you make.

In all these demonstration programs, the Computer will display a dot in the upper left corner of the Screen when it is listening. This means it is time for you to press the push-to-talk button and say the appropriate word or phrase.

To interrupt the Computer while it is waiting for speech input, hold down the BREAK key while you speak into the Mic.

Loading the Programs

All the programs are designed to run under LEVEL II BASIC. If you are operating under TRSDOS, type BASIC2 ENTER.

Note: These three Applications programs will not run under DISKBASIC because of differences in the USR function. However, you can write programs (or modify these) to run under DISKBASIC. (See Section 5.)

Prepare the recorder to play the desired tape program, then use the BASIC command, CLOAD, to load the program. Instructions for running each program follow.

Inventory Program

This program illustrates how data may be entered using the VOXBOX. It also illustrates use of a keyword to start a command or data sequence. Such keywords or "wake-up" words are useful to minimize interference from background noise.

These program control words are used in this demonstration:

"PART"	To begin the data entry sequence at the part number.
"QUANTITY"	To end the part number and begin entry of the quantity of that part.
"BACKSPACE"	To allow the previous entry to be corrected (re-entered).
"SCRATCH"	To cancel a complete sequence and begin over again with the word "PART".
"OK"	To accept an entry after visual verification.
"SAVE"	To signal the end of the entry program.

After training the system you may enter part numbers and quantities by entries of the following form:

PART (digit)(digit) . . . QUANTITY (digit)(digit) . . . OK

For example:

PART 123   QUANTITY 45   OK

"OK" signals the system to enter this data. If you or the system makes an error, you can say BACKSPACE to erase digits. You can also say SCRATCH to clear the entire entry. In the part field the letters A or B, or the character "–" (dash), may also be used to make up the part number. When you are through trying the Inventory demo, say SAVE instead of PART to end the program. In this demonstration, data is **not** actually saved at this point.

Lunar Program

In LUNAR, you try to land a rocket on the moon without crashing. It's a real challenge, since the amount of fuel is limited and the effect of burning a given amount of fuel varies with the mass of the ship. As you burn fuel the mass gets smaller, and the change in velocity occurring for each pound of fuel burned increases.

You can set the burn rate as large as you like. Each entry, for example "FIVE", ZERO, ZERO" (500), represents the burn rate for the next **ten** seconds in pounds of fuel per second. For example, if you say "FIVE" (5), you will burn 50 pounds of fuel. If you say "ZERO" (0̸), you will burn no fuel.

Make your entries in the form—

(digit)(digit) . . . FIRE

If you or the system makes an entry mistake, you can clear a digit by saying "BACKSPACE".

After saying "FIRE", the computer will compute the 10 second burn and display elapsed time in seconds, distance to the moon in miles per hour, and pounds of fuel remaining.

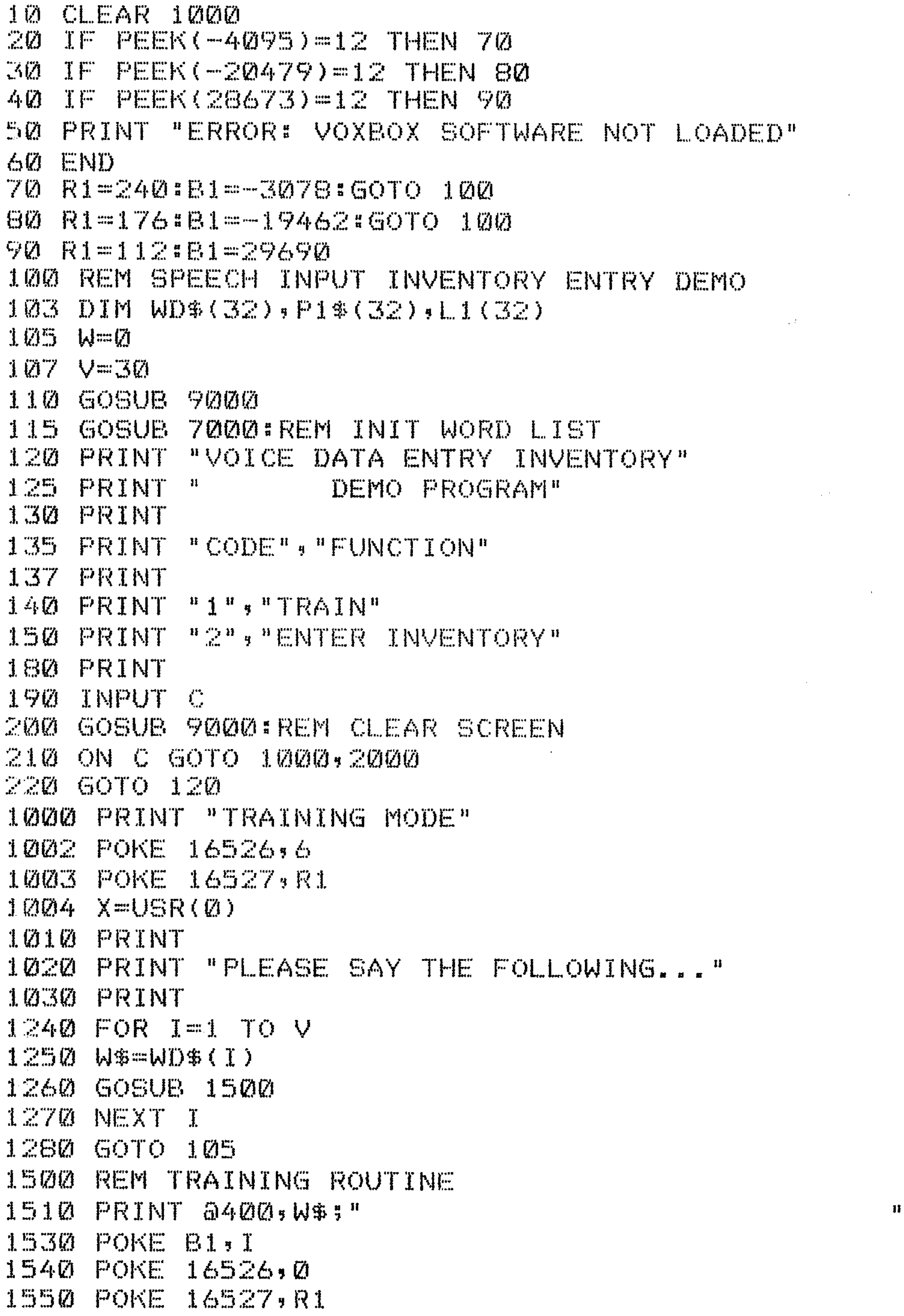
Good Luck!

Plot Program

This program shows you a plot of your voice. This is the raw data which the computer works with to recognize words. The top plot is a measure of energy in the frequency range 900 to 2200 hertz. High (trace up) means energy present. The second line is energy in the 100 to 900 hertz energy range. Notice that vowels (AH, OO, EE, etc.) have high energy in the low band while many consonants (S, F, V, etc.) don't.

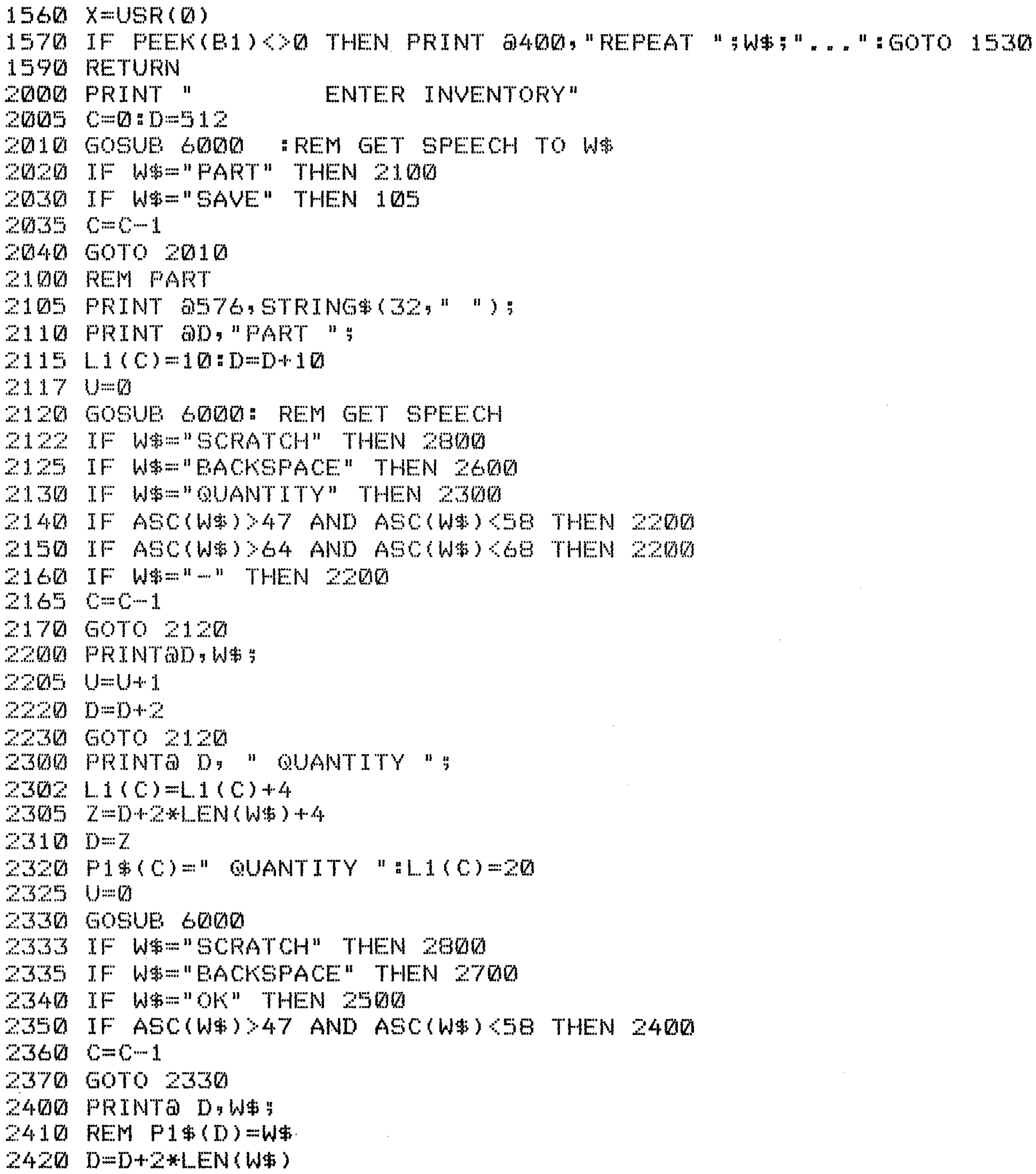
The next (lower) plot is an estimate of the dominant frequency in the band 900 to 5000 hertz. The lowest plot is an estimate of the dominant frequency in the band 200 to 900 hertz. All of these values are measured at 10 millisecond intervals.

Inventory Demo Program



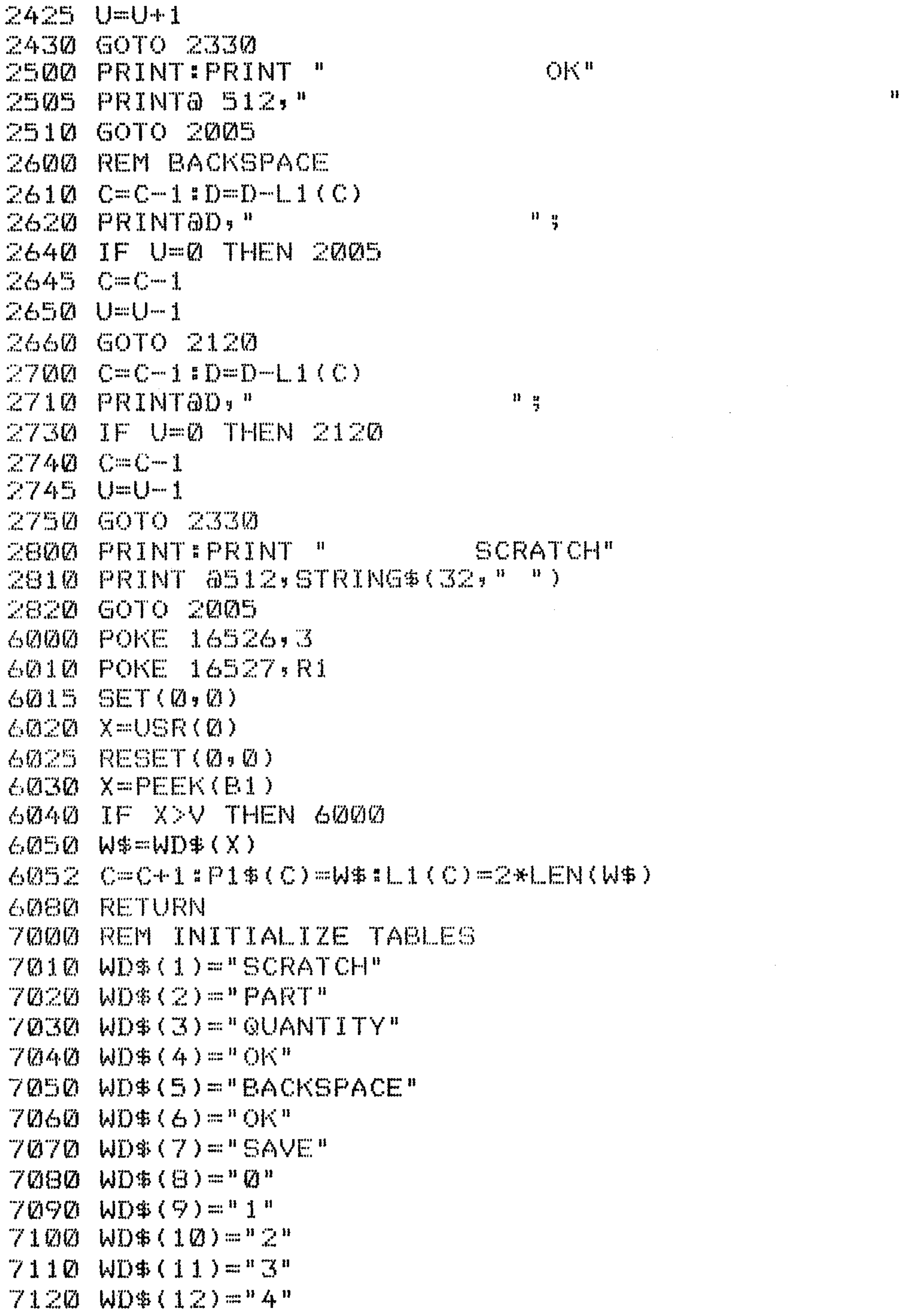
```
10 CLEAR 1000
20 IF PEEK(-4095)=12 THEN 70
30 IF PEEK(-20479)=12 THEN 80
40 IF PEEK(28673)=12 THEN 90
50 PRINT "ERROR: VOXBOX SOFTWARE NOT LOADED"
60 END
70 R1=240:B1=-3078:GOTO 100
80 R1=176:B1=-19462:GOTO 100
90 R1=112:B1=29690
100 REM SPEECH INPUT INVENTORY ENTRY DEMO
103 DIM WD$(32),P1$(32),L1(32)
105 W=0
107 V=30
110 GOSUB 9000
115 GOSUB 7000:REM INIT WORD LIST
120 PRINT "VOICE DATA ENTRY INVENTORY"
125 PRINT "        DEMO PROGRAM"
130 PRINT
135 PRINT "CODE","FUNCTION"
137 PRINT
140 PRINT "1","TRAIN"
150 PRINT "2","ENTER INVENTORY"
180 PRINT
190 INPUT C
200 GOSUB 9000:REM CLEAR SCREEN
210 ON C GOTO 1000,2000
220 GOTO 120
1000 PRINT "TRAINING MODE"
1002 POKE 16526,6
1003 POKE 16527,R1
1004 X=USR(0)
1010 PRINT
1020 PRINT "PLEASE SAY THE FOLLOWING..."
1030 PRINT
1240 FOR I=1 TO V
1250 W$=WD$(I)
1260 GOSUB 1500
1270 NEXT I
1280 GOTO 105
1500 REM TRAINING ROUTINE
1510 PRINT @400,W$;"                                        "
1530 POKE B1,I
1540 POKE 16526,0
1550 POKE 16527,R1
```

Inventory Demo Program (Continued)



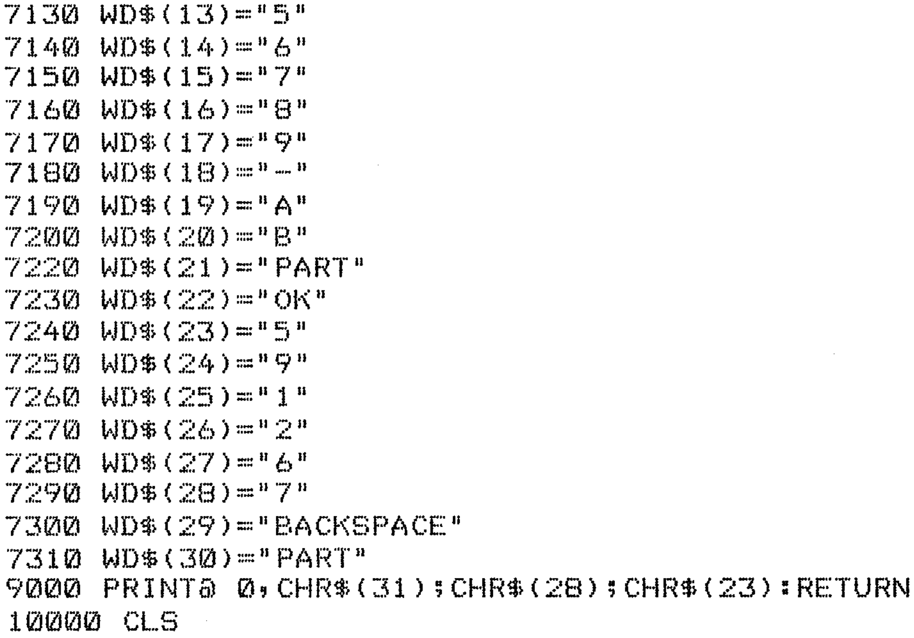
```
1560 X=USR(0)
1570 IF PEEK(B1)<>0 THEN PRINT @400,"REPEAT ";W$;"...":GOTO 1530
1590 RETURN
2000 PRINT "         ENTER INVENTORY"
2005 C=0:D=512
2010 GOSUB 6000  :REM GET SPEECH TO W$
2020 IF W$="PART" THEN 2100
2030 IF W$="SAVE" THEN 105
2035 C=C-1
2040 GOTO 2010
2100 REM PART
2105 PRINT @576,STRING$(32," ");
2110 PRINT @D,"PART ";
2115 L1(C)=10:D=D+10
2117 U=0
2120 GOSUB 6000: REM GET SPEECH
2122 IF W$="SCRATCH" THEN 2800
2125 IF W$="BACKSPACE" THEN 2600
2130 IF W$="QUANTITY" THEN 2300
2140 IF ASC(W$)>47 AND ASC(W$)<58 THEN 2200
2150 IF ASC(W$)>64 AND ASC(W$)<68 THEN 2200
2160 IF W$="-" THEN 2200
2165 C=C-1
2170 GOTO 2120
2200 PRINT@D,W$;
2205 U=U+1
2220 D=D+2
2230 GOTO 2120
2300 PRINT@ D, " QUANTITY ";
2302 L1(C)=L1(C)+4
2305 Z=D+2*LEN(W$)+4
2310 D=Z
2320 P1$(C)=" QUANTITY ":L1(C)=20
2325 U=0
2330 GOSUB 6000
2333 IF W$="SCRATCH" THEN 2800
2335 IF W$="BACKSPACE" THEN 2700
2340 IF W$="OK" THEN 2500
2350 IF ASC(W$)>47 AND ASC(W$)<58 THEN 2400
2360 C=C-1
2370 GOTO 2330
2400 PRINT@ D,W$;
2410 REM P1$(D)=W$
2420 D=D+2*LEN(W$)
```

Inventory Demo Program (Continued)



```
2425 U=U+1
2430 GOTO 2330
2500 PRINT:PRINT "            OK"
2505 PRINT@ 512,"                                "
2510 GOTO 2005
2600 REM BACKSPACE
2610 C=C-1:D=D-L1(C)
2620 PRINT@D,"               ";
2640 IF U=0 THEN 2005
2645 C=C-1
2650 U=U-1
2660 GOTO 2120
2700 C=C-1:D=D-L1(C)
2710 PRINT@D,"              ";
2730 IF U=0 THEN 2120
2740 C=C-1
2745 U=U-1
2750 GOTO 2330
2800 PRINT:PRINT "        SCRATCH"
2810 PRINT @512,STRING$(32," ")
2820 GOTO 2005
6000 POKE 16526,3
6010 POKE 16527,R1
6015 SET(0,0)
6020 X=USR(0)
6025 RESET(0,0)
6030 X=PEEK(B1)
6040 IF X>V THEN 6000
6050 W$=WD$(X)
6052 C=C+1:P1$(C)=W$:L1(C)=2*LEN(W$)
6080 RETURN
7000 REM INITIALIZE TABLES
7010 WD$(1)="SCRATCH"
7020 WD$(2)="PART"
7030 WD$(3)="QUANTITY"
7040 WD$(4)="OK"
7050 WD$(5)="BACKSPACE"
7060 WD$(6)="OK"
7070 WD$(7)="SAVE"
7080 WD$(8)="0"
7090 WD$(9)="1"
7100 WD$(10)="2"
7110 WD$(11)="3"
7120 WD$(12)="4"
```

Inventory Demo Program (Continued)



```
7130 WD$(13)="5"
7140 WD$(14)="6"
7150 WD$(15)="7"
7160 WD$(16)="8"
7170 WD$(17)="9"
7180 WD$(18)="-"
7190 WD$(19)="A"
7200 WD$(20)="B"
7220 WD$(21)="PART"
7230 WD$(22)="OK"
7240 WD$(23)="5"
7250 WD$(24)="9"
7260 WD$(25)="1"
7270 WD$(26)="2"
7280 WD$(27)="6"
7290 WD$(28)="7"
7300 WD$(29)="BACKSPACE"
7310 WD$(30)="PART"
9000 PRINT@ 0,CHR$(31);CHR$(28);CHR$(23):RETURN
10000 CLS
```

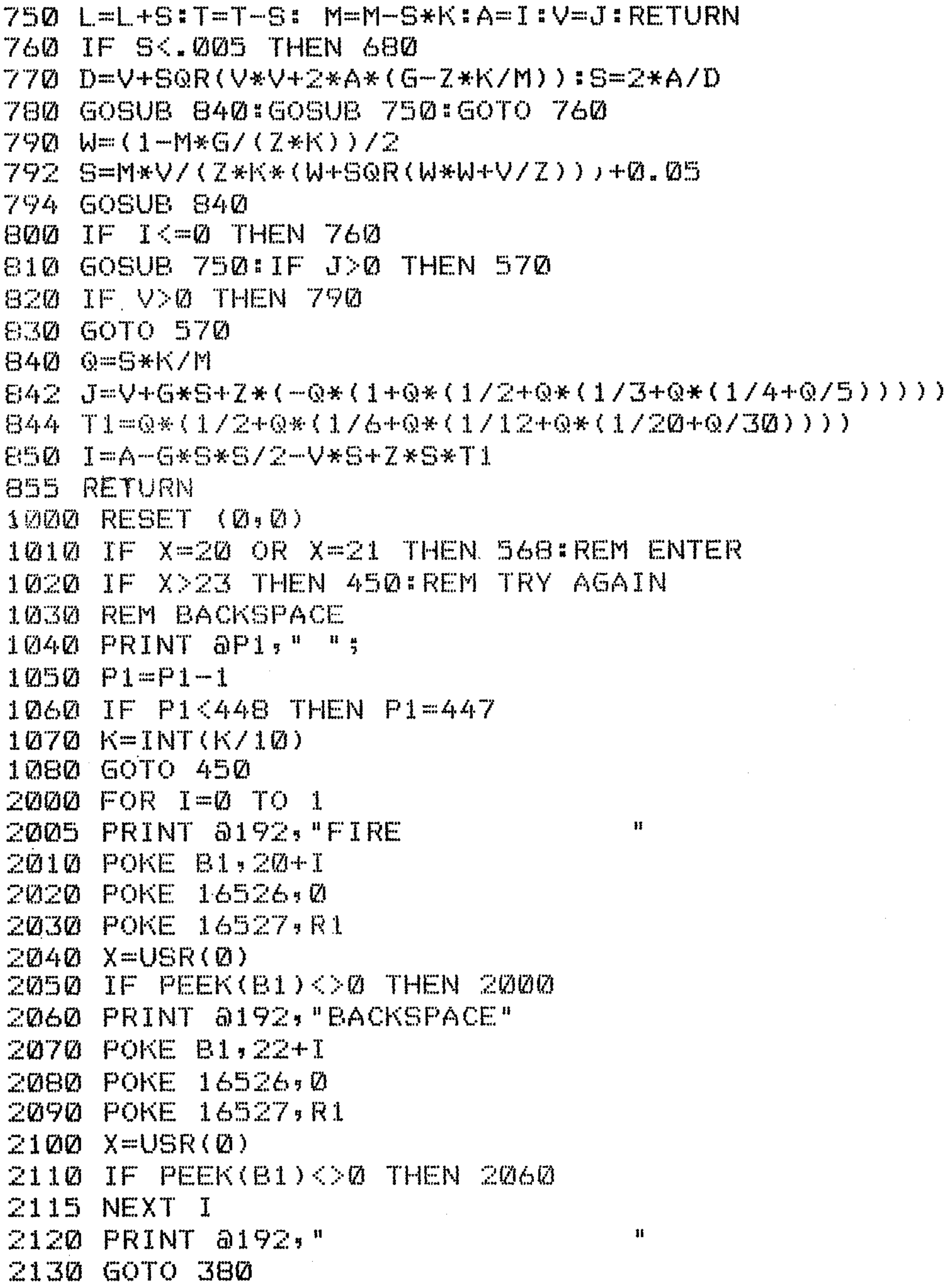
Lunar Program

```
10 CLS
20 IF PEEK(-4095)=12 THEN 70
30 IF PEEK(-20479)=12 THEN 80
40 IF PEEK(28673)=12 THEN 90
50 PRINT "ERROR: VOXBOX SOFTWARE NOT LOADED"
60 END
70 R1=240:B1=-3078:GOTO 100
80 R1=176:B1=-19462:GOTO 100
90 R1=112:B1=29690
100 REM SPEECH INPUT LUNAR
105 DIM W$(1)
110 REM INITIALIZE
120 POKE 16526,6
130 POKE 16527,R1
140 X=USR(0)
160 PRINT "THIS IS A SIMULATION OF AN APOLLO LUNAR"
170 PRINT "LANDER. THE ON BOARD     COMPUTER"
180 PRINT "HAS FAILED, SO YOU HAVE TO LAND MANUALLY."
190 PRINT "SET BURN RATE OF ROCKET TO ANY VALUE"
200 PRINT "BY ENTERING ONE DIGIT AT A TIME 'TIL YOU"
205 PRINT "GET THE VALUE YOU WANT, THEN SAY 'FIRE'."
208 PRINT "SAY 'BACKSPACE' TO ERASE THE LAST DIGIT."
209 PRINT "YOU'LL SEE THE NEW POSITION, WEIGHT, ETC."
210 PRINT "UPDATED AFTER EACH BURN. FIRST"
220 PRINT "WE TRAIN THE SYSTEM ON THE DIGITS. TYPE"
230 PRINT "ANY KEY TO GET STARTED."
250 A$=INKEY$:IF A$="" THEN 250
260 CLS
270 PRINT "PLEASE SAY THE FOLLOWING..."
280 FOR J=0 TO 1
290 FOR I=0 TO 9
300 PRINT@ 192, I
310 POKE B1,J*10+I
320 POKE 16526,0
330 POKE 16527,R1
340 X=USR(0)
350 IF PEEK(B1)<>0 THEN 300
360 NEXT I
370 NEXT J
375 GOTO 2000
380 PRINT "TRAINED..."
390 FOR I=0 TO 1000:NEXT I
400 CLS
```

Lunar Program (Continued)

```
410 L=0
420 PRINT @128,"SEC","MI+FT","MPH","LB FUEL"
425 PRINT @384,"BURN RATE"
430 A=120:V=1:M=33000:N=16500:G=0.001:Z=1.8
440 PRINT @256,L,INT(A);INT(5280*(A-INT(A))),3600*V,M-N,
445 P1=447:K=0
450 REM RECOGNIZE
460 SET(0,0)
470 POKE 16526,3
480 POKE 16527,R1
490 X=USR(0)
500 X=PEEK(B1)
510 IF X>19 THEN 1000
520 RESET (0,0)
530 T=10
540 X=(X-INT(X/10)*10)
545 W$=CHR$(X+ASC("0"))
550 P1=P1+1
555 K=10*K+X
560 PRINT @P1,W$;
565 GOTO 450
568 PRINT @448,"                              ";:P1=447
570 IF M-N<0.001 THEN 650
580 IF T<.001 THEN 440
590 S=T:IF M>=N+S*K THEN 610
600 S=(M-N)/K
610 GOSUB 840:IF I<=0 THEN 760
620 IF VV=0 THEN 640
630 IF J<0 THEN 790
640 GOSUB 750: GOTO 570
650 PRINT@704, "FUEL OUT AT ";L;"SECONDS"
660 S=(-V+SQR(V*V+2*A*G))/G
670 V=V+G*S:L=L+S
680 W=3600*V
690 PRINT@ 768,"ON MOON AT ";L;" SECONDS-- IMPACT VELOCITY ";W;" MPH"
700 IF W<=1.2 THEN PRINT@832 "PERFECT LANDING!!!":GOTO 390
710 IF W<=10 THEN PRINT@832 "CRAFT DAMAGED. YOU'RE STRANDED":GOTO 390
720 PRINT@832,"SORRY THERE WERE NO SURVIVORS, YOU BLEW IT"
730 PRINT@896,"IN FACT, YOU BLASTED A NEW CRATER ";W*.227;" FEET DEEP"
735 FOR I=0 TO 2000:NEXT I
740 GOTO 390
```

Lunar Program (Continued)



```
750 L=L+S:T=T-S: M=M-S*K:A=I:V=J:RETURN
760 IF S<.ØØ5 THEN 68Ø
77Ø D=V+SQR(V*V+2*A*(G-Z*K/M)):S=2*A/D
78Ø GOSUB 84Ø:GOSUB 75Ø:GOTO 76Ø
79Ø W=(1-M*G/(Z*K))/2
792 S=M*V/(Z*K*(W+SQR(W*W+V/Z)))+Ø.Ø5
794 GOSUB 84Ø
8ØØ IF I<=Ø THEN 76Ø
81Ø GOSUB 75Ø:IF J>Ø THEN 57Ø
82Ø IF V>Ø THEN 79Ø
83Ø GOTO 57Ø
84Ø Q=S*K/M
842 J=V+G*S+Z*(-Q*(1+Q*(1/2+Q*(1/3+Q*(1/4+Q/5)))))
844 T1=Q*(1/2+Q*(1/6+Q*(1/12+Q*(1/2Ø+Q/3Ø))))
85Ø I=A-G*S*S/2-V*S+Z*S*T1
855 RETURN
1ØØØ RESET (Ø,Ø)
1Ø1Ø IF X=2Ø OR X=21 THEN 568:REM ENTER
1Ø2Ø IF X>23 THEN 45Ø:REM TRY AGAIN
1Ø3Ø REM BACKSPACE
1Ø4Ø PRINT @P1," ";
1Ø5Ø P1=P1-1
1Ø6Ø IF P1<448 THEN P1=447
1Ø7Ø K=INT(K/1Ø)
1Ø8Ø GOTO 45Ø
2ØØØ FOR I=Ø TO 1
2ØØ5 PRINT @192,"FIRE                "
2Ø1Ø POKE B1,2Ø+I
2Ø2Ø POKE 16526,Ø
2Ø3Ø POKE 16527,R1
2Ø4Ø X=USR(Ø)
2Ø5Ø IF PEEK(B1)<>Ø THEN 2ØØØ
2Ø6Ø PRINT @192,"BACKSPACE"
2Ø7Ø POKE B1,22+I
2Ø8Ø POKE 16526,Ø
2Ø9Ø POKE 16527,R1
21ØØ X=USR(Ø)
211Ø IF PEEK(B1)<>Ø THEN 2Ø6Ø
2115 NEXT I
212Ø PRINT @192,"                        "
213Ø GOTO 38Ø
```

Plot Program

```
10 IF PEEK(-4095)=12 THEN 60
20 IF PEEK(-20479)=12 THEN 70
30 IF PEEK(28673)=12 THEN 80
40 PRINT "ERROR: VOX BOX SOFTWARE NOT LOADED"
50 END
60 R1=240:B1=-3078:L1=-2411:B2=-3041:GOTO 90
70 R1=176:B1=-19462:L1=-18795:B2=-19425:GOTO 90
80 R1=112:B1=29690:L1=30357:B2=29727:GOTO 90
90 CLS
100 PRINT @50,"SPEAK...";
110 POKE B1,0
120 POKE 16526,0
130 POKE 16527,R1
140 X=USR(0)
150 CLS
160 L=PEEK(L1)
165 PRINT@114,"LEN=";L;
170 B=B2
180 FOR I=10 TO L+10
190 IF PEEK(B)=0 THEN X=1 ELSE X=0
200 SET(I,X)
210 IF PEEK(B+1)=0 THEN X=1 ELSE X=0
220 SET(I,X+4)
230 SET(I,47-PEEK(B+2)/4)
240 SET(I,47-PEEK(B+3)/4)
250 B=B+4
260 NEXT I
270 GOTO 100
280 END
```

5. How to Write Your Own Application Programs

General Information

The VOXBOX allows you to control the TRS-80 system or enter data into it with spoken words. Up to 32 different words may be used. You select what the words are to do and what words are to be used. What the words do (the action they cause or the data they provide) is written into the program. The actual words spoken are selected (programmed) by the user at the time he runs the program. For convenience they are usually the same words, but they need not be. One user may refer to the digit "1" as "ONE" and another user as "UNO" and to the word "YES" as "YES" or "SI".

As is true with human listeners, your VOXBOX will sometimes make mistakes when listening to you. In general, it should correctly recognize 85-95% of the words you use, provided you take reasonable care in speaking clearly and distinctly (see Section 3).

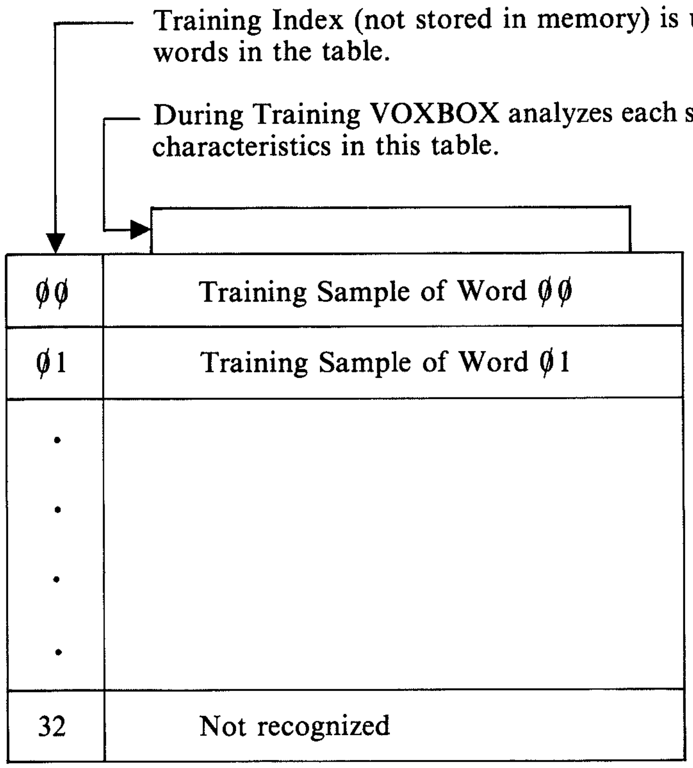
In writing programs, you should **assume** that some misunderstanding will occur and provide for checks on the data or verification of what was said at selected points in your program, and have the program allow for correction if that is important and required (as is done in the INVENTORY demo).

Technically speaking, VOXBOX is a speaker-trained, isolated-word, speech recognition system with a 32-word vocabulary. "Speaker-trained" means that the user must enter one or more examples of each word or phrase in the vocabulary. It also means that there are two operating modes for the device. First, a **TRAINING** mode in which the spoken examples of each word are provided, and then a **RECOGNITION** mode in which previously trained words are recognized. Isolated word means that each word or phrase must have a .1 second period of silence before and after the word. This is necessary to allow the computer to determine when a word has started and when it has finished.

The current vocabulary is stored in a table in high memory (as part of the driver program). Each table entry consists of the code for that word **as you pronounced it.** See figure 3 below.

VOXBOX Speech Files

(See Figure 3.)



Training Index (not stored in memory) is used to specify individual words in the table.

During Training VOXBOX analyzes each spoken word and stores its characteristics in this table.

ØØ	Training Sample of Word ØØ
Ø1	Training Sample of Word Ø1
• • • •	
32	Not recognized

FIGURE 3

VOXBOX Use

The VOXBOX has three routines you can call from BASIC:

1) Initialization
2) Training
3) Recognition

They are usually used in this order. Each routine is called by the USR function of Level II (or USRØ, USR1, USR2 for DISK BASIC). These routines create the speech data files or match unknown speech data against these files to find the corresponding entry and thus the desired word that was spoken.

Initialization clears the speech files, and resets the speech sample table to all Ø's. Previously stored speech data is lost. Use Initialization at the beginning of the program once; it must be followed by training.

Training accepts known spoken words and extracts features from these words which are stored away for future comparison with unknown words during recognition. During training, a training index number (TR=Ø to 31) is given to identify the specific word sample.

Recognition identifies unknown words spoken to VOXBOX. It may be called as frequently as desired after training. Training samples are preserved during recognition. During recognition, speech is presented and the training index number corresponding to the set of stored characteristics closest to the unknown word is returned. TR=32 is reserved for "no such word in my list."

A detailed discussion of the use of each routine follows below; and a typical program is developed step-by-step.

Initialization (Entry at Relocation +6)

Note: The term "Relocation" refers to the actual first address of the driver program in your system. If you loaded driver program SP16, then Relocation=28672; if SP32, then Relocation=45056; if SP48, Relocation=61440.

The Initialization routine clears all tables before use. Previously stored speech samples will be lost. Initialization is usually run once just before VOXBOX is trained. First set up the USR function by POKEing (Level II) or DEFUSR (DISK BASIC). The entry address equals Relocation +6. Then execute the statement, X=USR(Ø); the computer will branch to "Initialization," using (in Level II) the data in memory locations 16526 and 16527.

For example (See Appendix B) for DISK BASIC equivalents):

```
10 REM   SET UP CALLING PARAMETERS
20 REM   MS=MEMORY SIZE IN K-BYTES
30 PRINT"IS THIS A 16K, 32K OR 48K SYSTEM ";
35 INPUT" (TYPE 16, 32 OR 48) ";MS
40 IF MS=16 THEN AD =112:  GOTO 80
50 IF MS=32 THEN AD =176:  GOTO 80
60 IF MS=48 THEN AD =240:  GOTO 80
70 GOTO 35
80 POKE 16526,6
90 POKE 16527,AD
100 X=USR(0)
```

Training (Entry at Relocation +Ø)

Before you can recognize words you must train the unit on your (or the intended user's) speech patterns. The training subroutines do this The training subroutine must be called once for each word, or sample of a word, to be later recognized.

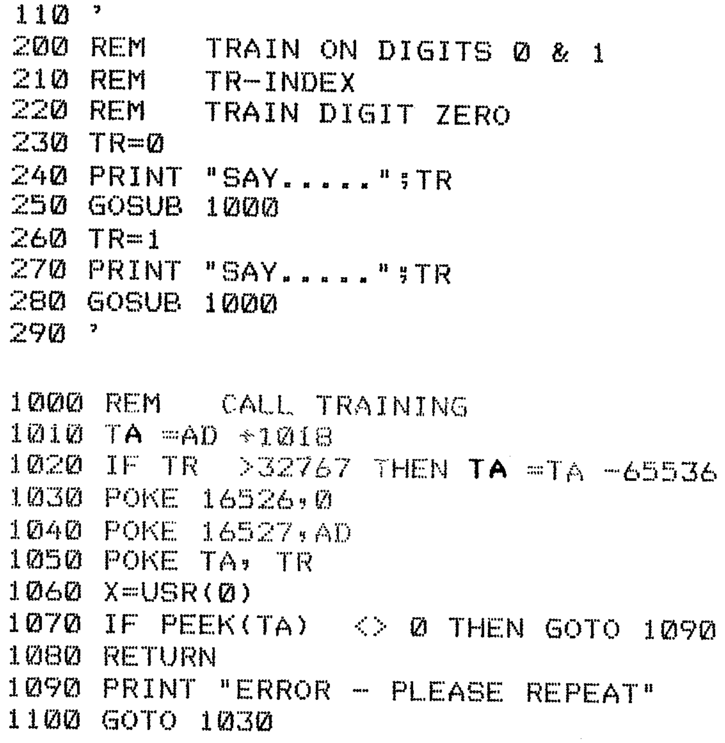
Usually, you will want to tell the speaker which word he's to say through the video display by, for example, printing out a statement that says "SAY... START" when "start" is the intended word, and then immediately calling the training routine for that word.

There are 32 slots in the VOXBOX table for samples of words to be recognized. They are numbered Ø to 31. The number (index) of the "slot" which is to receive the training parameters is passed using a POKE command, as are the addresses of the subroutine.

First set up a USR routine with entry at the Relocation Address. Then call the routine each time you want the user to input a word or phrase for training.

Frequently a small subroutine is helpful to organize the flow of your program.

For example, the following program lines let you "train" the VOXBOX on the words "ZERO" and "ONE." (This assumes ADR contains the proper address for your system from the Initialization routine, lines 10-100 above.) See Appendix B for DISK BASIC equivalents.



```
11Ø '
2ØØ REM   TRAIN ON DIGITS Ø & 1
21Ø REM   TR-INDEX
22Ø REM   TRAIN DIGIT ZERO
23Ø TR=Ø
24Ø PRINT "SAY....."; TR
25Ø GOSUB 1ØØØ
26Ø TR=1
27Ø PRINT "SAY....."; TR
28Ø GOSUB 1ØØØ
29Ø '

1ØØØ REM    CALL TRAINING
1Ø1Ø TA =AD +1Ø18
1Ø2Ø IF TR  >32767 THEN TA =TA -65536
1Ø3Ø POKE 16526,Ø
1Ø4Ø POKE 16527,AD
1Ø5Ø POKE TA, TR
1Ø6Ø X=USR(Ø)
1Ø7Ø IF PEEK(TA)  <> Ø THEN GOTO 1Ø9Ø
1Ø8Ø RETURN
1Ø9Ø PRINT "ERROR - PLEASE REPEAT"
11ØØ GOTO 1Ø3Ø
```

After executing X=USR(Ø) (line #1060) the VOXBOX will be listening. Parameters from that word will be stored away, associated with the training index, TR. The word or phrase spoken may be up to 1.5 seconds long with one or more embedded silences not exceeding 0.1 seconds. A silence of 0.1 second or longer signals the end of the word.

After each training word is spoken, an **error code** is returned at the same location (TRA) where the training index, TR, was POKE'd. If the speech presented to VOXBOX was longer than 1.5 seconds the contents of the location will be **non-zero.** If no error occurred, the contents of the location will be **zero.** This can be checked as in line #1070 above.

Recognition (Entry at Relocation +3)

Now you are ready to recognize words. When you want a verbal input, it helps to instruct the user by printing a cue on the Video Display, such as "SPEAK..." or "YOUR MOVE..." or "CONTINUE." If the words are easily forgotten, it may help to print on the display a partial list of those that the VOXBOX can recognize as a reminder.

In recognition, the user function instructs the VOXBOX to listen. When a word is detected, the number (index) of this word recognized is returned to you in a specific location in memory. You will use PEEK to get that number.

If no trained word is recognized but a word-like sound was heard, the VOXBOX will return a "reject" code (Code 32 or 20_H)|in this location. The same code is returned if the word or phrase detected is longer than 1.5 sec. The "speech longer than 1.5 secs." error can result from using the unit in a noisy environment.

To interrupt the system (return to BASIC monitor) while waiting for a speech input, press down the "BREAK" key and say a word into the microphone (such as "break"). This will cause a "break" to occur.

To use the recognition routine, first set up a USR routine with entry at Relocation +3. Then call the USR routine when you want the user say the word or phrase.

Here is an example you might follow for recognizing the digits θ and 1 as they were trained above. The addresses ADR and TRA were computed in the previous routines, and do not have to be computed again here.

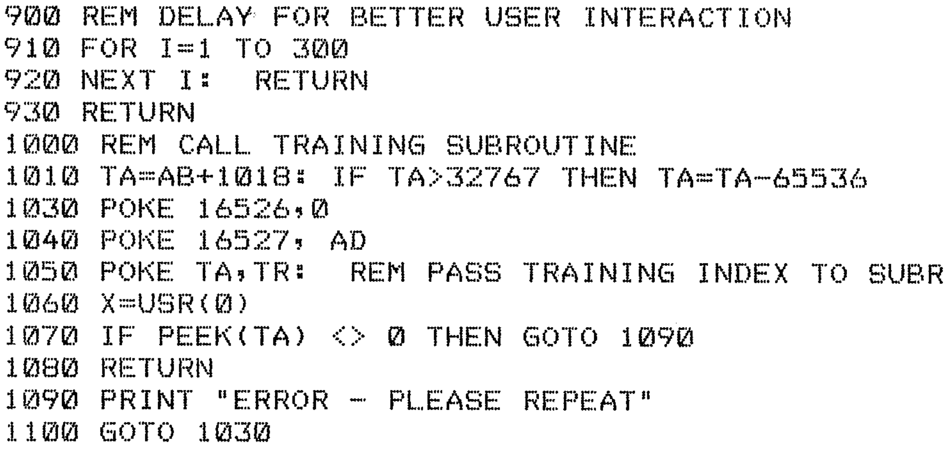
```
400 REM   GET VOICE INPUT
410 POKE 16526,3
420 POKE 16527,AD
430 PRINT "SPEAK...."
440 X = USR(Ø)
450 W = PEEK(TA)
460 IF W > 31 THEN GOTO 430
470 PRINT "YOU SAID";
480 GOSUB 900:  REM DELAY
490 GOTO 300: REM GET NEXT WORD
900 REM   DELAY
910 FOR I=1 TO 300
920 NEXT I: RETURN
930 '
```

The above examples of the three elements of speech recognition can be put together (and slightly expanded) to yield a demonstration program which will recognize 6 digits and 5 words.

Here is a complete program:

```
10 REM   SET UP CALLING PARAMETERS
20 REM   MS=MEMORY SIZE IN K-BYTES
30 PRINT" IS THIS A 16K, 32K OR 48K SYSTEM ";
35 INPUT" (TYPE 16, 32, OR 48)";MS
40 IF MS=16 THEN AD =112:  GOTO 80
50 IF MS=32 THEN AD =176:  GOTO 80
60 IF MS=48 THEN AD =240:  GOTO 80
70 GOTO 35
80 POKE 16526,6
90 POKE 16527,AD
95 AB=AD *256
100 X=USR(0)
200 REM
210 REM TR=INDEX; W$=STRING ARRAY
220 DIM W$(32)
230 W$(0)="0"
240 W$(1)="1"
250 W$(2)="2"
260 W$(3)="3"
270 W$(4)="4"
280 W$(5)="5"
290 W$(6)="YES"
300 W$(7)="NO"
310 W$(8)="PERHAPS"
320 W$(9)="FASTER"
330 W$(10)="SLOWER"
340 FOR TR=0 TO 10
350 PRINT  "SAY...";W$(TR)
360 GOSUB 1000: REM TRAIN
370 NEXT TR
400 REM GET VOICE INPUT
410 POKE 16526,3
420 POKE 16527, AD
430 PRINT "SPEAK...."
440 X = USR(0):   REM CALL RECOGNITION
450 W = PEEK(TA)
460 IF W > 31 THEN GOTO 500
470 PRINT "YOU SAID "; W$(W)
480 GOSUB 900: REM DELAY
490 GOTO 400: REM GET NEXT WORD
500 REM  ERROR
510 PRINT "PLEASE REPEAT THAT"
520 GOSUB 900:  REM DELAY
530 GOTO 400: GET WORD
```

Complete Program (Continued)



```
900 REM DELAY FOR BETTER USER INTERACTION
910 FOR I=1 TO 300
920 NEXT I:  RETURN
930 RETURN
1000 REM CALL TRAINING SUBROUTINE
1010 TA=AB+1018: IF TA>32767 THEN TA=TA-65536
1030 POKE 16526,0
1040 POKE 16527, AD
1050 POKE TA,TR:  REM PASS TRAINING INDEX TO SUBR
1060 X=USR(0)
1070 IF PEEK(TA) <> 0 THEN GOTO 1090
1080 RETURN
1090 PRINT "ERROR - PLEASE REPEAT"
1100 GOTO 1030
```

Programming Hints

Multiple Training Samples

When the vocabulary size is less than 32 words, significant improvements in recognition accuracy are possible through the use of multiple training samples. By associating 2 or more speech table entries with the same word the system can better handle the normal variations of speech. Coding with multiple training samples is simplified by use of a string array. For example:

```
DIM WD$ (32)
WD$(1)="ENTER"
WD$(2)="ENTER"
```

would set both indexes 1 and 2 of WD$ equal to "ENTER." If you trained the speech recognition system so that words 1 and 2 of its tables were both "ENTER" you would now have 2 training samples of the word "ENTER." In Recognition, you might get either (1) or (2), but both would be interpreted to mean "ENTER."

Articulation

People have a tendency to drop some sounds out of speech. For example, "A" and "8" sound alike if the "T" sound is dropped from "8". Be careful to articulate clearly, or choose words where such confusion is not possible.

Words That Sound Alike

Many speech sounds are hard to distinguish. For example, M and N, L and R, or E, B, D, and P. Avoid words that only differ by these sounds, for example, MINE and NINE.

Use of String Arrays

Coding is simplified if you set up a string array with your vocabulary words assigned the same index as they are given in the speech tables. This is especially true when you use multiple training samples.

Appendix

A. Modifying Driver Program Parameters

The closeness with which an unknown speech sample must match one of the samples of known words before a word is "recognized" (properly or improperly) is controlled by a number stored in the driver program. This "rejection parameter" may be made larger to allow for a looser fit between spoken and trained words, or it may be made smaller to require a tighter fit between spoken words and the trained words.

When the number is made larger, fewer spoken words will be rejected (that is, not recognized at all) but more mistakes where the wrong word is identified may occur.

If the number is made smaller, fewer mistakes where the wrong word is identified should occur, but in general more words will be rejected as unidentifiable.

This parameter has been carefully set at a value which provides an acceptable tradeoff of these characteristics. The current value of the parameter is 350 DEC (Ø15E HEX).

This parameter is stored in two bytes at relocation +254 (least significant byte) and +255 (most significant byte). These locations in absolute address are as follows:

User Memory	Locations (DEC)	Locations (HEX)
16K	28926-28927	7ØFE,7ØFF
32K	45310-45311	BØFE,BØFF
48K	61694-61695	FØFE,FØFF

The least significant byte of the value is stored in the lowest location. The most significant byte of the value is stored in the higher location. Remember to follow instructions for Decimal to Hex to Decimal conversion given in the TRS-80 User Manual for POKEs. Use POKEs to store new data as in the example below:

```
REM INPUT LOWER LOCATION, L
IF L > 32767 THEN L=L-65536
REM SET VALUE TO 360 [0168 HEX] FOR LOOSER MATCH
REM 68 HEX = 104 DEC
POKE L, 104
REM 1 HEX = 1 DEC
POKE L+1, 1
```

B. Summary of Important Addresses and USR Calls

	USER MEMORY SIZE		
	16K	32K	48K
Answer Memory Prompt for Available Memory Limit	28671	45055	61439
LOAD TAPE (NAME)	SP16	SP32	SP48
To Initialize VOXBOX (clears Tables, Training, resets everything)	POKE 16526,6 POKE 16527,112 X=USR(Ø) or DEFUSR= 28678 X=USR (Ø) (initialize entered at 7ØØ6$_H$)	POKE 16526,6 POKE 16527,176 X=USR (Ø) or DEFUSR= 45062 X=USR (Ø) (initialize entered at BØØ6$_H$)	POKE 16526,6 POKE 16527,24Ø X=USR (Ø) or DEFUSR= 61446 X=USR (Ø) (initialize entered at FØØ6$_H$)
To Train VOXBOX	POKE 16526,Ø POKE 16527,112 POKE 29690, TR X=USR1(Ø) or DEFUSR1=28672 POKE 29690, TR X=USR1(Ø) where TR=the training index, a decimal number from Ø to 31.	POKE 16526,Ø POKE 16527,176 POKE-19462, TR X=USR1(Ø) or DEFUSR1=45056 POKE 19462, TR X=USR1(Ø) where TR=the training index, a decimal number from Ø to 31.	POKE 16526,Ø POKE 16527,24Ø POKE -3078, TR X=USR1(Ø) or DEFUSR1=61440 POKE 3078, TR X=USR1(Ø) where TR=the training index, a decimal number from Ø to 31.

To have VOXBOX Listen and Recognize	POKE 16526,3
	POKE 16527,112
	X=USR2(Ø)
	W=PEEK(29690)
	or
	DEFUSR2=28675
	X=USR2(Ø)
	W=PEEK(29690)
	where W is the index of the word recognized
		POKE 16526,3
		POKE 16527,176
		X=USR2(Ø)
		W=PEEK(−19462)
		or
		DEFUSR2=45059
		X=USR2(Ø)
		W=PEEK(−19462)
		where W is the index of the word recognized
			POKE 16526,3
			POKE 16527,24Ø
			X=USR2(Ø)
			W=PEEK(−3078)
			or
			DEFUSR2=61443
			X=USR2(Ø)
			W=PEEK(−3078)
			where W is the index of the word recognized

C. VOXBOX Technical Specifications

General

Function:	Word Recognition for Data Input and Control.
Type:	Isolated-Word, Speaker Trained
Vocabulary Size:	32 Words
Connects to:	TRS-80, or TRS-80 Expansion Interface
Method of Connection:	Cable provided for connection.
Word Definition:	Word or phrase length 0.1 to 1.5 sec. Beginning of word identified by 0.1 sec. of continuous speech. End of word identified by first interval of silence at least 0.1 sec. which follows detection of beginning of word.
Equipment Required:	TRS-80 with Level II Basic and at least 16K RAM. May be used with other peripherals and more capable versions of TRS-80 system.
Memory Requirements:	4096 bytes RAM reserved at the high end of user memory.
Peripheral Device Address Used:	I/O PORT AF HEX (175 DEC)

Inputs:

Power:	120 VAC, 50-60 Hz, 0.1 Amp via separate 15 VDC power pack (included).
Logic Signals:	TRS-80 peripheral bus
Voice:	Push-to-Talk microphone (included)

Physical Specifications

Temperature:	32° F-110° F Operating –40° F-160° F Storage
Humidity:	0-95% Non-condensing
Size:	1¾″H × 7 ″W × 5½″D (5 cm × 20 cm × 14 cm)
Weight (shipping):	2 lbs (1 kg)

NOTES

LIMITED WARRANTY

For a period of 90 days from the date of delivery, Radio Shack warrants to the original purchaser that the computer hardware described herein shall be free from defects in material and workmanship under normal use and service. This warranty is only applicable to purchases from Radio Shack company-owned retail outlets and through duly authorized franchisees and dealers. The warranty shall be void if this unit's case or cabinet is opened or if the unit is altered or modified. During this period, if a defect should occur, the product must be returned to a Radio Shack store or dealer for repair, and proof of purchase must be presented. Purchaser's sole and exclusive remedy in the event of defect is expressly limited to the correction of the defect by adjustment, repair or replacement at Radio Shack's election and sole expense, except there shall be no obligation to replace or repair items which by their nature are expendable. No representation or other affirmation of fact, including, but not limited to, statements regarding capacity, suitability for use, or performance of the equipment, shall be or be deemed to be a warranty or representation by Radio Shack, for any purpose, nor give rise to any liability or obligation of Radio Shack whatsoever.

EXCEPT AS SPECIFICALLY PROVIDED IN THIS AGREEMENT, THERE ARE NO OTHER WARRANTIES, EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE AND IN NO EVENT SHALL RADIO SHACK BE LIABLE FOR LOSS OF PROFITS OR BENEFITS, INDIRECT, SPECIAL, CONSEQUENTIAL OR OTHER SIMILAR DAMAGES ARISING OUT OF ANY BREACH OF THIS WARRANTY OR OTHERWISE.

RADIO SHACK A DIVISION OF TANDY CORPORATION

U.S.A.: FORT WORTH, TEXAS 76102
CANADA: BARRIE, ONTARIO L4M 4W5

TANDY CORPORATION

AUSTRALIA
280-316 VICTORIA ROAD
RYDALMERE, N.S.W. 2116

BELGIUM
PARC INDUSTRIEL DE NANINNE
5140 NANINNE

U. K.
BILSTON ROAD WEDNESBURY
WEST MIDLANDS WS10 7JN

PRINTED IN U.S.A.
679