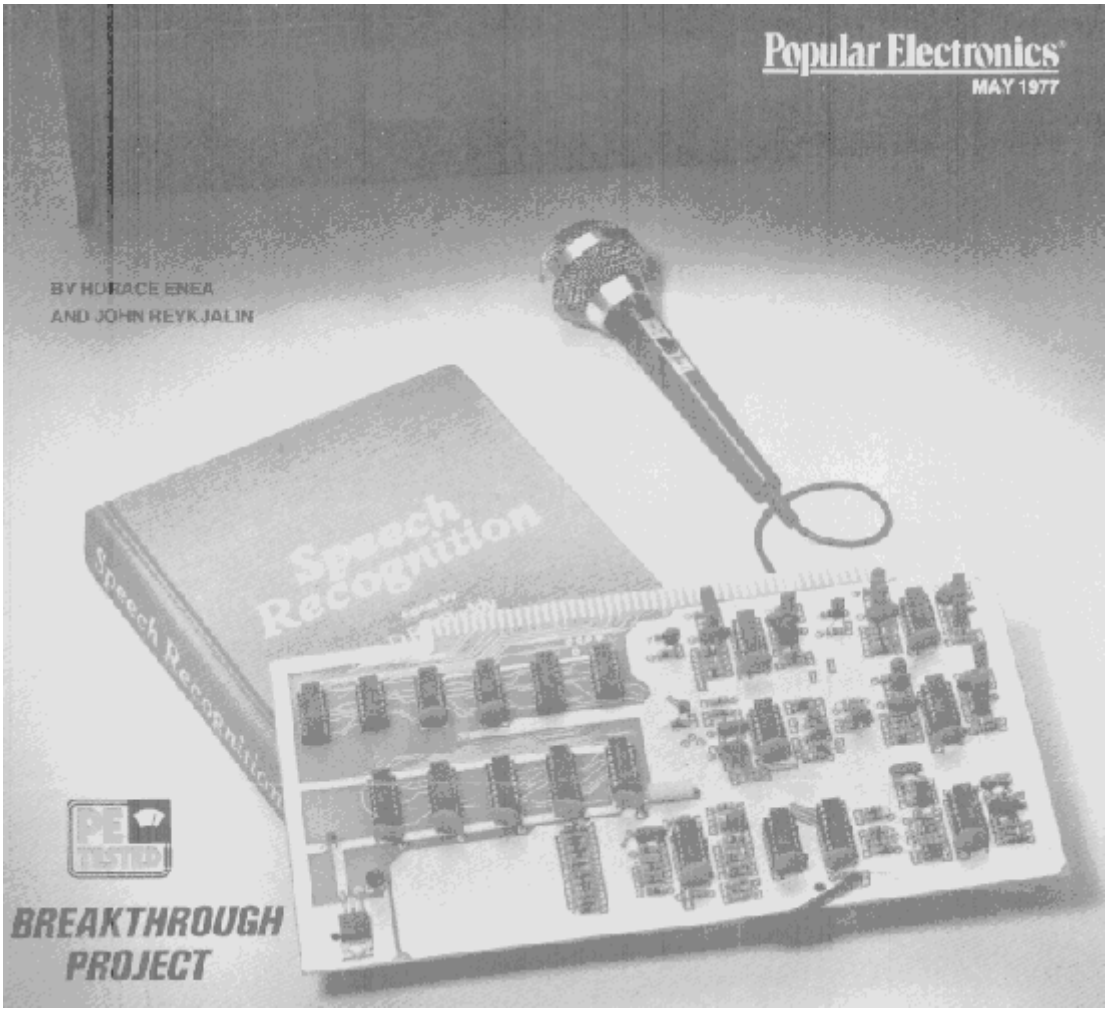
sp recog Speech Lab

Speechlab

dropbox



INTRODUCING SPEECHLAB - THE FIRST HOBBYIST VOCAL INTERFACE FOR A COMPUTER!
Now your computer can respond to vocal commands by the simple addition of a $250 single-board unit.

> IMAGINE being able to talk to your computer and have it respond by way of a hard-copy device or by activating some external appliance! Computer hobbyists can now enjoy this facility by building "Speechlab", a new, low-cost (under $250) computer peripheral. To use it, all one does is plug the single Speechiab PC board into an Aftair-bus connector (used by many microcomputer manufacturers), enter a special program, and the computer does the rest. It's a state-of-the-art approach at a moderate cost. One section of the program allows the user to "train" the computer to accept a vocal input (via a microphone), analyze the spoken word, and create a digitized version that is stored in memory. The second part of the program allows the user to speak to the SpeechIab and have the computer generate the output selected for that particular sound. The vocabulary size of Speechiab is a function of the

speech recognitlon algorithm used and the amount of memory available. For the program used in this article, it is 64 bytes per spoken word. The unique characteristics of Speechlab open many formerly closed doors. Since Speechlab will operate with any audio input (not necessarily a recognized language), a person who's vocally handicapped can operate almost any number of appliances (TV receiver, stereo system, solenoid-perated door, etc.) using a repeatable sound such as a grunt. One can use Speechlab, too, as a vocal processor to add spoken cornmands to many computer games (such as the "Star Trek" game), or enter the world of artificial intelligence and advanced programming.

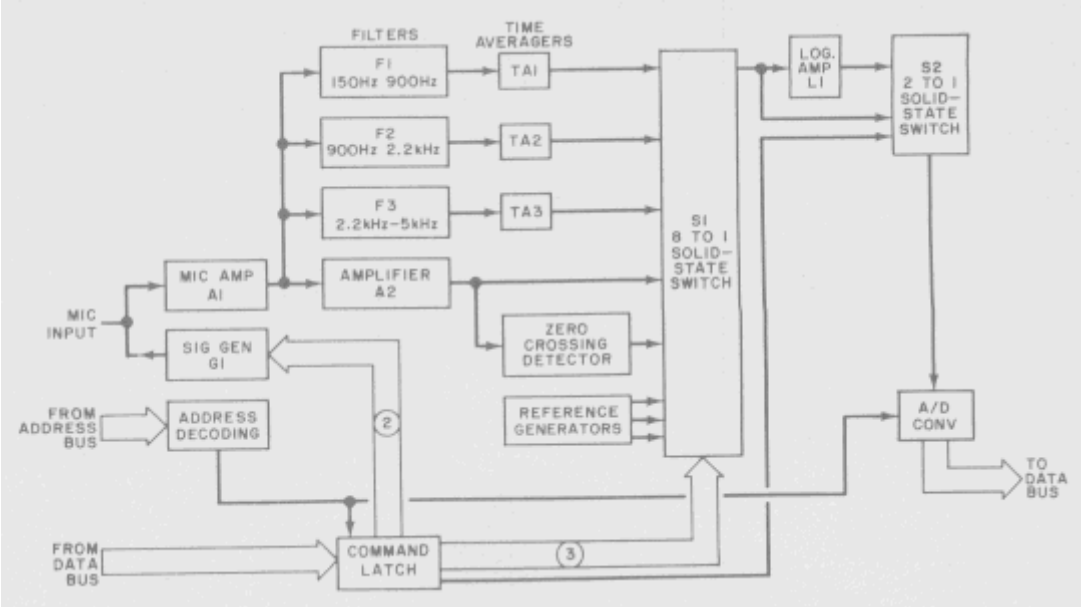


fig. 1. The mic input is amplified, filtered and applied to S1 along with raw audio, zero-crossing detection, and three reference voltages. Output of S1 is computer selected by switch S2 for digitizing.

Circuit Operation.

The basic block diagram of Speechlab is shown in **Fig. 1.** The audio input is amplified by A1 and applied to three 80-db/decade rolloff band-pass filters Fl, F2, and F3. These filters encompass the ranges of 150 to 900 Hz, 900 Hz to 2.2 kHz, and 2.2 kHz to 5 kHz, respectively. These ranges correspond to the frequency ranges of the first three resonances of the average human vocal tract.

Each filter is passed to a time averager (TA1, T2, and TA3) to generate a voltage proportional to the level of the speech waveform within each band. The amplified audio signal from A1 is further amplified by A2 to generate an unfiltered waveform that can swing approx. 2 volts about a rest level of 2 volts. This signal is also applied to a zero-crossing detector that generates a voltage proportional to the number of times the speech waveform crosses the 2-volt rest level in a given period of time, thus generating a measure of the dominant frequency in the speech signal.

These five voltages TA1, TA2, TA3,A2, and ZCD are fed to solid-state s witch S1 along with three reference voftages used for calibration and self test. A computer output command selects one of these five voltages to be passed through S1.

The selected output from S1 is passed to a second solid-state switch (S2), and to a logarithmic amplifier (L1) that emphasizes the low-level signal before being passed to S2. Switch S2 can select either the direct output from S1, or the output from L1, and pass this selected signal to a 6-bit A/D converter

where the voftage is converted to a digital value. The output of the A/D converter is fed to the computer data bus.
All operations of the Speechlab are controlled through a single I/O port (address AFhex). As shown in **Fig. 2.** , six bits are used: bit-5 disables the 8-to1 muItiplexer (S1), and is used when switching between bands; bit-4 controls signal generator G1 which is used either to drive the microphone so that it acts like a miniature loudspeaker for prompting during voice input, or to drive the filters and zero-crossing detector during calibration and test; bit-3 selects either linear or logarithmic scaling of the voltage applied to the A/D converter; while bit-2, bit-1, and bit-0 select one of the eight signals from S1 for A/D conversion.
The input data word contains the 6-bit A/D output in bits 0 through 5, blt-6 is unused and is always 0, while bit-7 is the A/D converter status with a 1 corresponding to busy, and 0 corresponding to finished. Speechlab is physically configured to occupy one slot in the Altair bus, and the complete schematic is shown in **Fig. 3. through Fig. 7.**

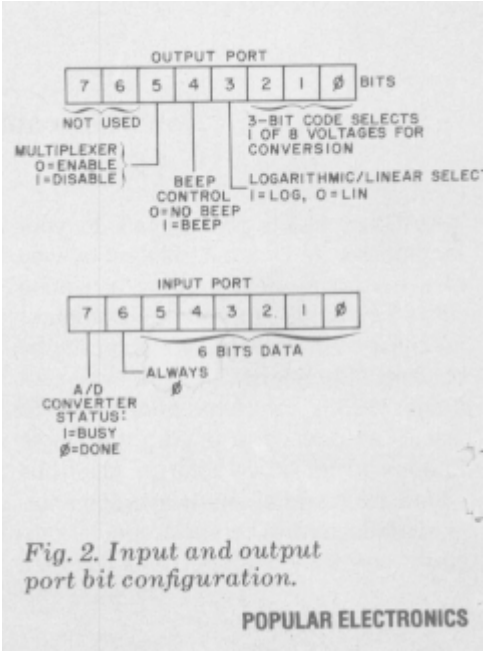


Fig. 2. Input and output port bit configuration.

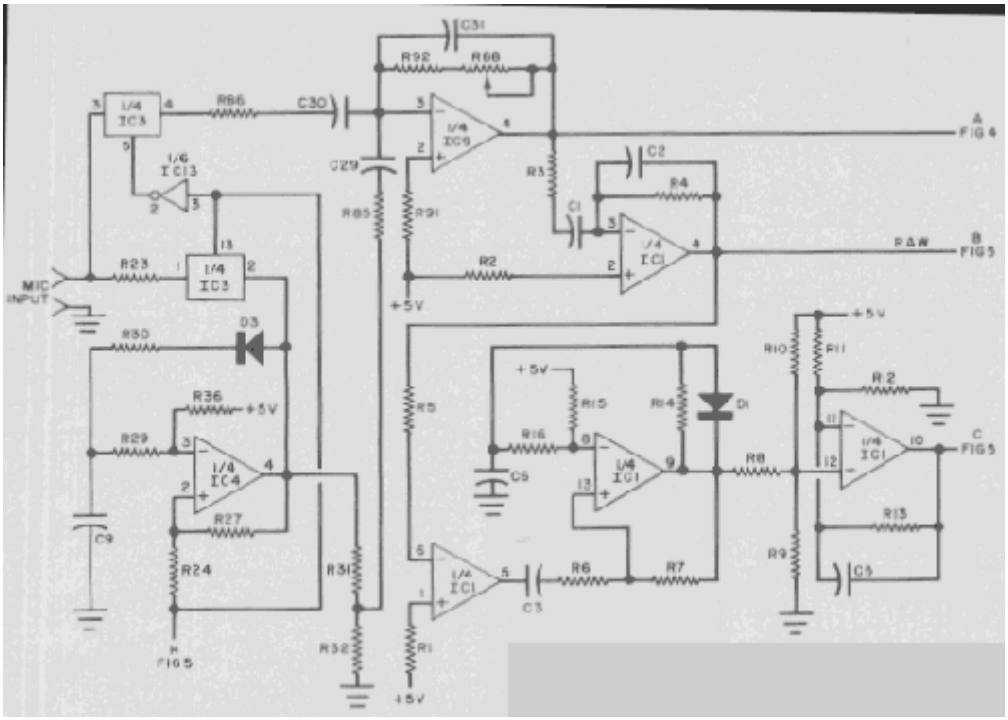


Fig. 3. Amplifier 1/4IC9 takes either audio or tone from 1/4IC4 depending on computer command. IC1 circuits are used as raw amplifier and zero-crossing detector.

PARTS LIST

Unless otherwise noted, the following capacitors are 10% Mylar types, and all picofarad sizes are CM05 types.
C1, C16, C21, C43, C47, C49, C52, C57—0.0047 µF
C2, C31—100 pF
C3, C17, C20—270 pF
C4, C7, C8, C10, C12, C19, C27, C32, C33, C34, C35, C36, C37, C44, C55, C61, C62—0.1 µF, 25-V disc
C5, C14, C18, C24, C54, C60—0.01 µF
C6, C42, C45, C53, C56—240 pF
C9, C40, C48—0.022 µF
C11, C29—47 pF
C13—15 µF, 25 V tantalum
C15, C22, C51, C59—0.0015 µF
C23—0.0022 µF
C25, C26, C28, C38—1 µF
C30, C39, C46—0.047 µF
C41—0.1 µF
C50, C58—0.001 µF
D1, D3 through D6—1N4148 or 1N914 diode
D2—1N746 diode
IC1, IC4, IC7, IC8, IC9, IC15, IC16—LM3900 quad amp
IC2—4051 8-to-1 analog multiplex
IC3—4016 quad analog switch
IC5—LM311 comparator
IC6—78M05 5-volt regulator
IC10—4024 7-stage binary counter
IC11, IC18—74C174 D flip-flop
IC12—4050 hex buffer
IC13, IC22—4049 hex buffer inverter
IC14—4013 (see text) dual-D flip-flop
IC17—74LS30 8-input NAND gate
IC19—8097 three-state hex buffer
IC20—8093 three-state quad buffer
IC21—4001 NOR gate
MIC—Mura DX-121 dynamic microphone (part of stereo set Mura DX-242)

L1—22-µH choke
Unless otherwise noted, the following resistors are ¼-W, 5%
R1—619,000 ohms, 1%
R2—1 megohm, 1%
R3—6810 ohms, 1%
R4—332,000 ohms, 1%
R5—200,000 ohms, 1%
R6,R20,R21—30,000 ohms
R7, R100—3 megohms
R8, R9, R10, R12, R14, R16, R104—1 megohm
R11—910,000 ohms
R13—2.7 megohms
R15, R48—10 megohms
R17,R18—20,000 ohms
R19, R22, R106—10,000 ohms
R23—1000 ohms
R24, R27—1.2 megohms
R25, R34, R39—470,000 ohms
R26, R38—750,000 ohms
R28, R31—100,000 ohms
R29—110,000 ohms
R30—39,000 ohms
R32—47,000 ohms
R33, R41—68,100 ohms, 1%
R35, R96,R102—75,000 ohms
R36—3.9 megohms
R37, R45—357,000 ohms, 1%
R40, R50, R52, R54, R56, R58, R60 R61—10,000 ohms, 1%
R42—12,100 ohms, 1%
R43, R49—4750 ohms, 1%
R44—4320 ohms, 1%
R45, R47—681,000 ohms, 1%
R51, R53, R55, R57, R59—4990 ohms, 1%
R62—274,000 ohms, 1%
R63—7500 ohms
R64, R66, R72, R75—160,000 ohms
R65, R71—12,000 ohms

R67, R70—300,000 ohms
R68—931,000 ohms, 1%
R69—2 megohms
R73—620,000 ohms
R74, R76, R90, R92—62,000 ohms
R77—15,000 ohms
R78, R83, R84—147,000 ohms, 1%
R79, R80, R87—51,100 ohms, 1%
R81,R82,R89—174,000 ohms, 1%
R85—330,000 ohms
R86—680 ohms
R88—100,000-ohm pc trimmer potentiometer
R91—270,000 ohms
R93—249,000 ohms, 1%
R94—4300 ohms
R95, R97, R103, R105—360,000 ohms
R98, R101—820,000 ohms
R99—845,000 ohms, 1%
R107—158,000 ohms, 1%
R108—4700 ohms
R109, R111, R117, R119—82,000 ohms
R110, R116—5100 ohms
R112, R115—180,000 ohms
R113—549,000 ohms, 1%
R114—1.6 megohms
R118—510,000 ohms
R120—6800 ohms
R121—2000 ohms
Misc.—Sockets (one 8-pin, thirteen 14-pin, seven 16-pin), regulator mounting hardware, tie-wrap etc.
Note 1: The following is available from Heuristics Inc., 900 N. San Antonio Rd. (Suite C-1), Los Altos CA 94022 (Tele: 415-948-2542): complete kit of all parts including pc board, sockets, microphone, hardware manual, and 200-page lab manual, SpeechBasic, and assembly language programs at $249. (California residents please add 6½% sales tax.)

Parts list.

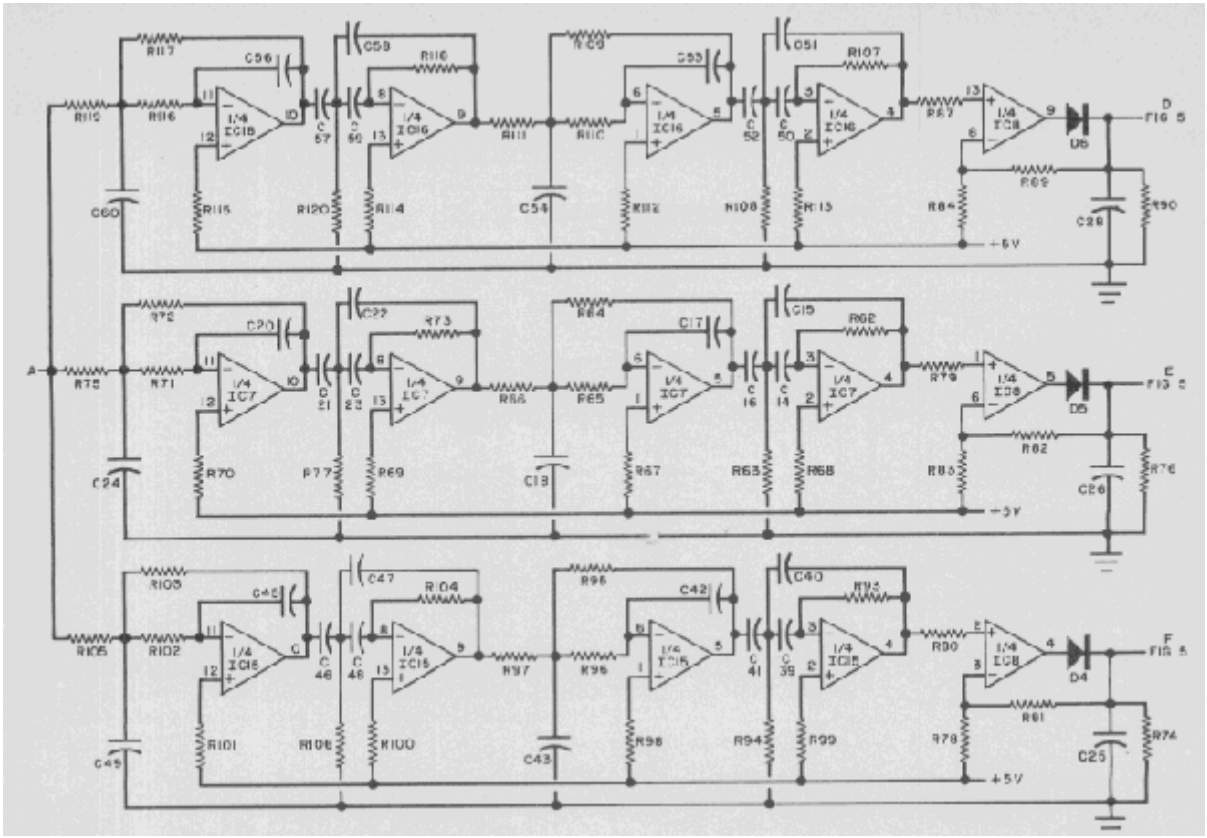


Fig. 4. Three bandpass filters and their associated time averagers. The encompass three ranges corresponding to freqnency ranges of the first three resonances of an average human vocal tract.

Construction.

The two foil patterns (Speechlab uses one double-sided PC board) are shown half-size in **Fig. 8.** (Blow up to full size on film only.) Component layout is shown in **Fig. 9.**

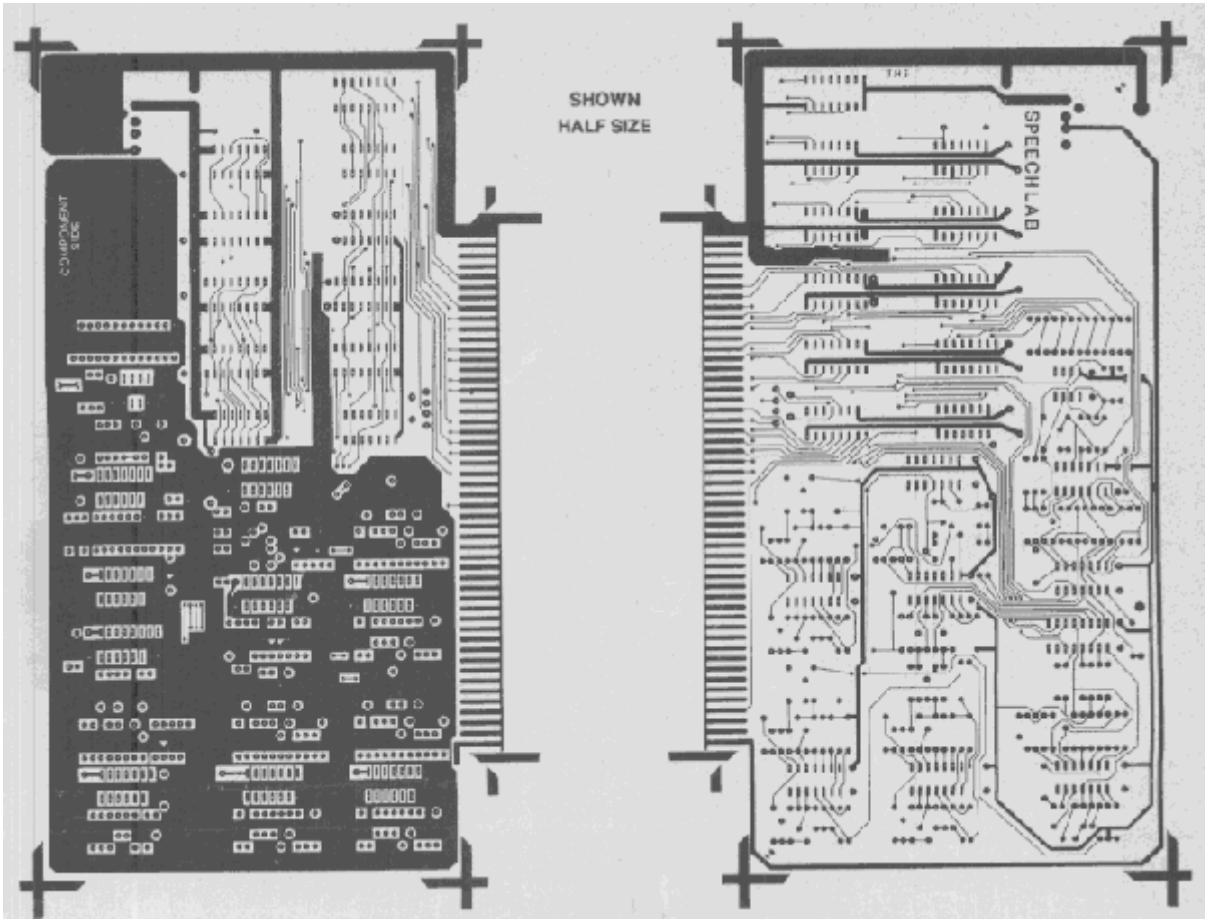


Fig. 8. Etching and drilling guides for pc board are shown half size. Guide at left is the component side. Component layout is in Fig. 9.

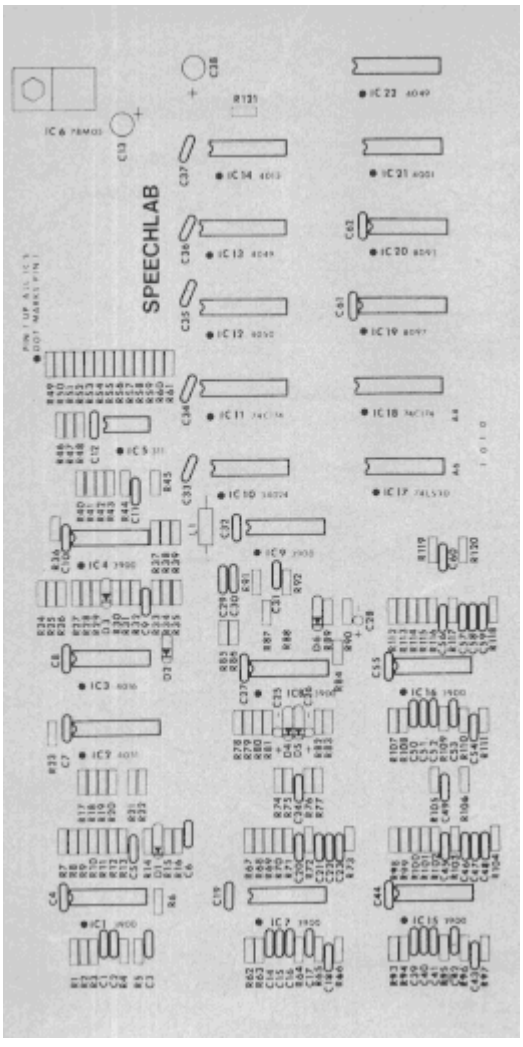


Fig. 9. Component layout for the Speechlab. See etching and drilling guide on previous page.

construction continued

All the components are mounted on one side of the board, with all the soldering done on the noncomponent side. Sockets are recommended for all IC's since most of them are MOS-types that may be damaged by improper handling. Integrated circuits IC1, IC4, IC7, IC8, IC9, IC15, and IC16 should be selected so they are capable of delivering a 4-volt output when using a 5-volt supply. Dual flipfIop IC14 can be from any manufacturer but Fairchild, as their truth table is somewhat different from the conventional table.

Start construction by installing the voltage regulator (IC6), all the discrete components, and the IC sockets do not install the IC's at this time. Check the board for correct parts installation, and to make sure that there are no solder bridges between adjacent foil traces. Mount the board in an Altair bus connector, and check for the presence of 5 volts at the output of the voltage regulator and at appropriate socket pins. Remove the board from the computer.

Install IC2 through IC5, IC10 through IC14, and IC17 through IC22. Install the board back in the Altair bus connector, and turn on the computer. Load the test Load the testprogram of Table 1 at 100 (hex). NOTE: all program data in this article is in hex.

You must jump to your monitor routine at address 0164-0165. Load address 195 with 05 and run the program. This will input the fixed reference voltage levels to the A/D converter and check the signal paths from switch S1 to the cornputer data bus.

After running this program, examine locations 200 through 20F, 300 through 30F, and 400 through 40F. Location 200 through 20F should contain 12 approx. 4, 300 through 30F should contain 24 approx. 4, and 400 through 40F should contain 36 approx. 4.

Insert the remaining IC's in their sockets, load location 195 with 10, and run the test program**(Table 1)** . This test uses the signal generator (G1) to create an input for the filters, amplifiers, and zero-crossing detector, and thereby checks the remaining signal paths on the board and calibrates the microphone preamplifier. After running the program, examine locations 200 to 20F to see if it contains 16 to 18. If not, adjust potentiometer R88 and rerun the program until these outputs occur.

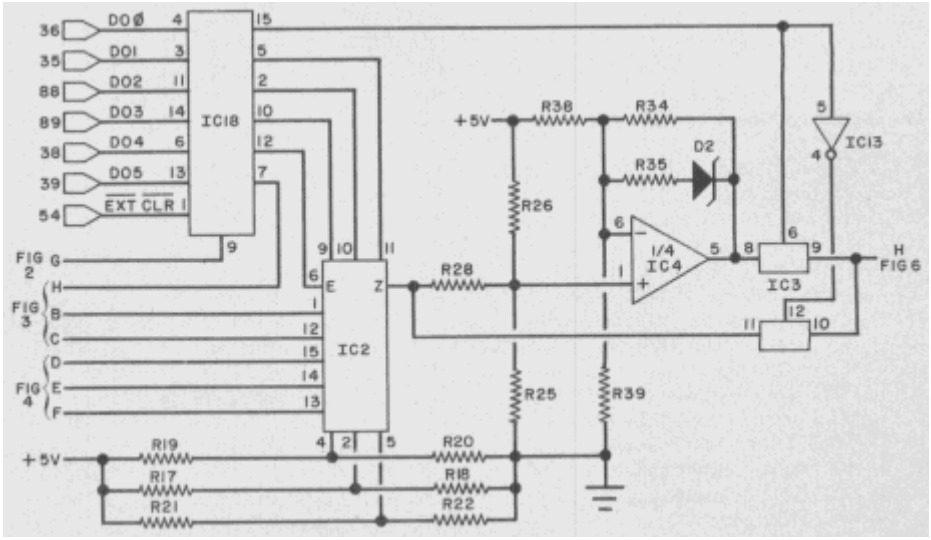


Fig. 5. Command latch (1C18) can activate tone generator and switch Si (1C2). Op amp (1/4 1C4) is logarithmic amplifier.

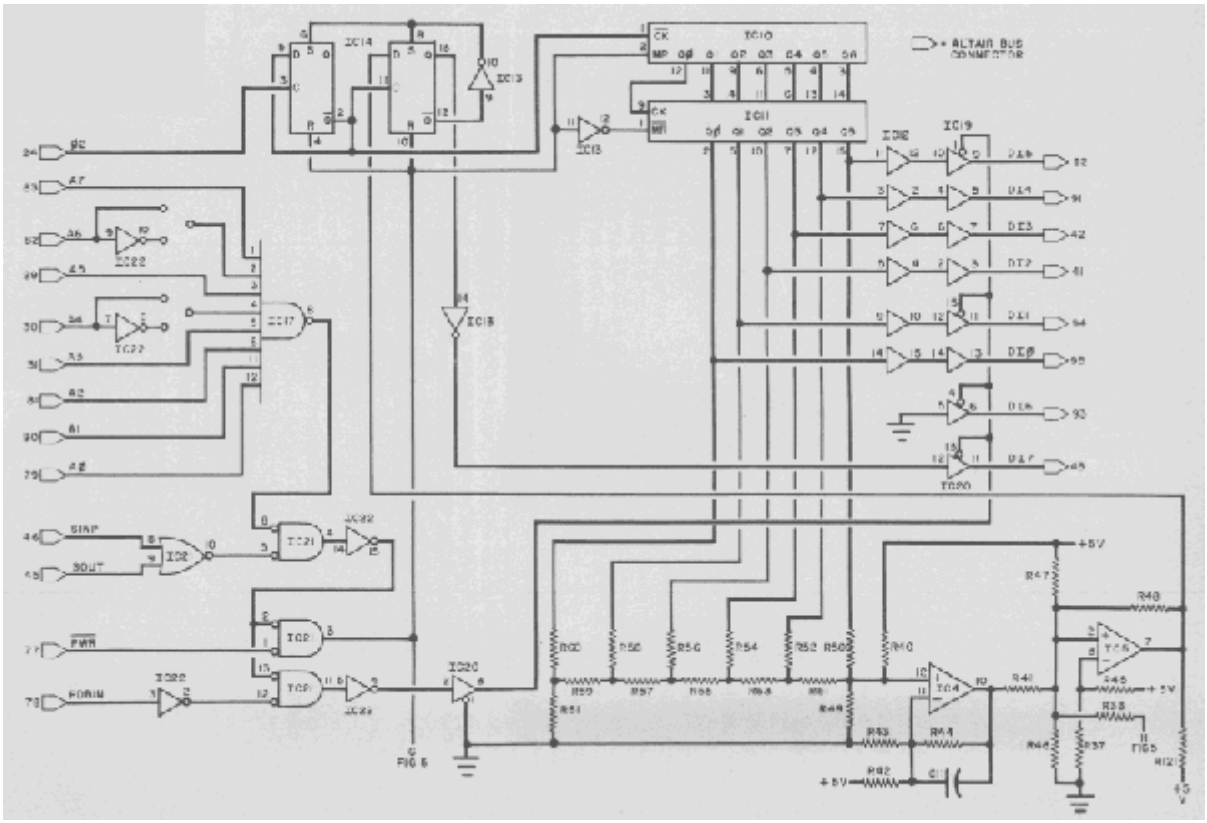


Fig. 6. IC17 circuit selects board address and IC14 forms S2. IC10 and IC11 form 6-bit A/D converter. Digitized data is then passed to computer.

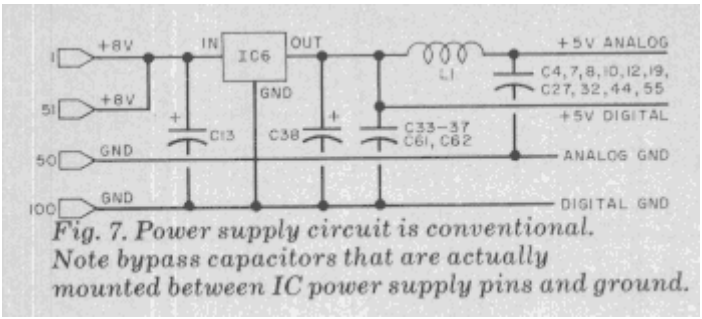


Fig. 7. Power supply circuit is conventional. Note bypass capacitors that are actually mounted between IC power supply pins and ground.

Calibration and Test Program.

The test program **(Table 1)** is a general purpose calibration, test, and diagnostic program for the SpeechIab. lt loads at location 100 and requires memory from 100 to 600 for program and data areas. Locations 163-165 should be loaded with a lump to your monitor address so that the program will return control to your monitor after execution. If you do not have a monitor, place a halt at this location.

TABLE I

```
0100                    ORG 100H
0100 =                  START EQU 100H
0100 210002             LXI H,START+100H
0103 228D01             SHLD TEMP1
0106 210003             LXI H,START+200H
0109 228F01             SHLD TEMP2
010C 210004             LXI H,START+300H
010F 229101             SHLD TEMP3
0112 210005             LXI H,START+400H
0115 229301             SHLD TEMP4
0118 3A9501     GO      LDA COMAND
011B CD6601             CALL INPUT
011E 2A8D01             LHLD TEMP1
0121 77                 MOV M,A
0122 2C                 INR L
0123 228D01             SHLD TEMP1
0126 3A9501             LDA COMAND
0129 C601               ADI 1
012B CD6601             CALL INPUT
012E 2A8F01             LHLD TEMP2
0131 77                 MOV M,A
0132 2C                 INR L
0133 228F01             SHLD TEMP2
0136 3A9501             LDA COMAND
0139 C602               ADI 2
013B CD6601             CALL INPUT
013E 2A9101             LHLD TEMP3
0141 77                 MOV M,A
0142 2C                 INR L
0143 229101             SHLD TEMP3
0146 3A9501             LDA COMAND
0149 C603               ADI 3
014B CD6601             CALL INPUT
014E 2A9301             LHLD TEMP4
0151 77                 MOV M,A
0152 2C                 INR L
0153 229301             SHLD TEMP4
0156 CA5F01             JZ STOP
0159 CD7701             CALL DELAY
015C C31801             JMP GO
015F 3E00       STOP    MVI A,0
0161 D3AF               OUT 0AFH
0163 C3XXXX             JMP SYSTEM
0166 F620       INPUT   ORI 20H
0168 D3AF               OUT 0AFH
016A E6DF               ANI 0DFH
016C D3AF               OUT 0AFH
016E DBAF       LOOP    IN 0AFH
0170 17                 RAL
0171 DA6E01             JC LOOP
0174 DBAF               IN 0AFH
0176 C9                 RET
0177 C5         DELAY   PUSH B
0178 3E05               MVI A,5
017A FE00       DEL0    CPI 0
017C CA8B01             JZ RETDEL
017F 0669               MVI B,69H
0181 00         DEL1    NOP
0182 00                 NOP
0183 05                 DCR B
0184 C28101             JNZ DEL1
0187 3D                 DCR A
0188 C37A01             JMP DEL0
018B C1         RETDEL  POP B
018C C9                 RET
C000 =          SYSTEM EQU 0C000H
018D            TEMP1 DS 2
018F            TEMP2 DS 2
0191            TEMP3 DS 2
0193            TEMP4 DS 2
0195 YY         COMAND EQU $
0100 21 00 02 22 8D 01 21 00 03 22 8F 01 21 00 04 22
0110 91 01 21 00 05 22 93 01 3A 95 01 CD 66 01 2A 8D
0120 01 77 2C 22 8D 01 3A 95 01 C6 01 CD 66 01 2A 8F
0130 01 77 2C 22 8F 01 3A 95 01 C6 02 CD 66 01 2A 91
0140 01 77 2C 22 91 01 3A 95 01 C6 03 CD 66 01 2A 93
0150 01 77 2C 22 93 01 CA 5F 01 CD 77 01 C3 18 01 3E
0160 00 D3 AF C3 53 00 FA 20 D3 AF E6 DF D3 AF DB AF
0170 17 DA 6E 01 DB AF C9 C5 3E 05 FE 00 CA 8B 01 06
0180 69 00 00 05 C2 81 01 3D C3 7A 01 C1 C9 9B F7 F0
0190 FE 57 DC 49 E2 A6
```

Calibration and Test Program continued

The program collects four 256-byte buffers of data from four of the eight possibie inputs to the A/D converter. The first of the four bands is specified by the Test Command word, which also specifies beeper on/off and linear or logarithmic scaling. The next three bands are 1, 2, and 3 greater than specified by the Test Command word. Each band is sampled every five milliseconds until 256 samples have been collected from each of the four bands. Data from the first band is stored in 200 to 2FF, the second band from 300 to 3FF, the third from 400 to 4FF, and the fourth from 500 to 5FF.

For example, if the Test Command word is set to 00, after the test program is run, the four data areas will contain numbers representing the outputs of band-0 (low frequency), band-1 (mid frequency), band-2 (high frequency), and band-3 (zerocrossing detector). Anything that was spoken into the microphone while the program was running, is filtered, converted into a binary number, and stored in the data areas.

If the Test Command word is set to 05, the first three data areas will contain constant numbers corresponding to the three reference voltage levels to the A/D converter on band 5, 6, and 7. This is useful for checking the A/D converter operation and isolating problem areas to one side or the other of the

8-to-1 analog switch S1. If the Test Command word is set to 10, signal generator G1 is enabled which begins to "beep" the microphone and connects the signalgenerator output into the microphone preamplifier A1. The four data areas contain data from bands 0, 1, 2, and 3 as when the Test Command word was 00, but the input signal comes from the signal generator rather than from the microphone. This allows calibration of the microphone preamplifier and isolates problems in one of the filter-averager chains.

Adding blt-3 to the command word will cause logarithmic rather than linear data scaling and will isolate problems to the log amplifier or either of the two analog switches comprising S2, the 2-to-1 analog switch.

Various comblnations of bits in the Test Command word will allow quick calibration and fault isolation, and also provide a quick way to look at raw data from any input through the microphone.

Software.

A simple technique for speech recognition of the digits zero through nine with a recognitlon rate of 90% or better, is shown the flowchart of **Fig. 10** . An 8080 program for this algorithm is shown in **Table II** . The program starts at memory location 0100 and requires less than 4K bytes of storage induding table space.

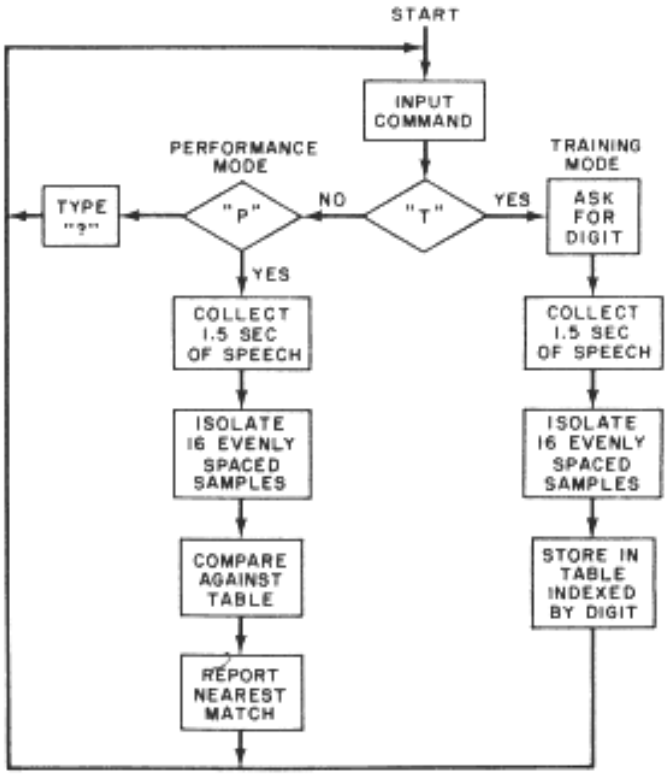


Fig. 10. Flow chart of a simple program that is used to "T" (train) and "P" (perform) a vocal operation. The program is shown in Table II

TABLE II

```
0100  31 00 10 CD 5B 03 21 35  0B CD 3A 03 CD 0F 03 FE
0110  54 C2 1A 01 CD 33 01 C3  00 01 FE 50 C2 2B 01 CD
0120  5B 03 CD 79 01 CD 81 03  C3 00 01 3E 3F CD 70 03
0130  C3 00 01 F5 C5 D5 E5 CD  5B 03 21 3D 0B CD 3A 03
0140  CD DA 03 32 6E 04 3C 47  3A 6D 04 B8 DA 2B 01 CD
0150  5B 03 CD A3 02 CD 99 01  CD 07 04 3A 6E 04 4F 16
0160  40 CD F7 02 11 13 07 CD  A3 03 60 69 0E 40 11 C8
0170  06 CD 8C 01 E1 D1 C1 F1  C9 C5 D5 E5 CD A3 02 CD
0180  99 01 CD 07 04 CD 3D 02  E1 D1 C1 C9 F5 EB 7E 23
0190  EB 77 23 0D C2 8D 01 F1  C9 0E 00 21 6F 04 CD F9
01A0  03 FE 06 D2 B2 01 79 FE  96 C2 9E 01 21 44 0B C3
01B0  9D 02 79 3D 32 08 07 47  CD F9 03 FE 06 DA CC 01
01C0  79 FE 96 C2 B8 01 21 4E  0B C3 9D 02 79 3D 3D 32
01D0  09 07 90 3C 32 C7 06 FE  0A DA A6 01 0E 01 CD F9
01E0  03 FE 06 D2 ED 01 79 FE  0A C2 DE 01 C9 3A C7 06
01F0  01 4F 3A 08 07 81 4F C3  B8 01 21 FF FF 22 13 0B
0200  3E 00 32 6E 04 3A 6D 04  3D 4F 06 00 C5 79 87 4F
0210  21 15 0B 09 46 23 4E 21  13 0B 56 23 5E CD B2 03
0220  DA 2B 02 C1 0D FA 39 02  C3 0C 02 21 13 0B 70 23
0230  71 C1 79 32 6E 04 C3 24  02 3A 6E 04 C9 CD 64 02
0240  06 00 48 16 40 3E 00 5F  21 13 07 CD D5 03 CD 75
0250  02 1C 78 FE 3F C2 48 02  04 3A 6D 04 BB C2 42 02
0260  CD FA 01 C9 3A 6D 04 B7  4F 3E 00 21 15 0B 77 23
0270  0D C2 6E 02 C9 E5 D5 C5  F5 21 C8 06 16 00 19 96
0280  F2 05 02 2F 3C 57 78 B7  21 15 0B 4F 06 00 09 23
0290  7A B6 77 D2 98 02 28 34  F1 C1 D1 E1 C9 CD 3A 03
02A0  C3 00 01 3E FA CD E3 02  CD 4B 03 3E 64 CD E3 02
02B0  21 6F 04 1E 96 16 00 7A  FE 04 CA C6 02 CD D3 02
02C0  77 23 14 C3 B7 02 3E 0A  CD E3 02 1D C2 B5 02 CD
02D0  4B 03 C9 F6 20 D3 AF E6  DF D3 AF DB AF 17 DA DB
02E0  02 1F C9 C5 FE 00 CA F5  02 06 69 00 00 05 C2 EB
02F0  02 3D C3 E4 02 C1 C9 F5  D5 06 00 1E 09 79 1F 4F
0300  1D CA 0E 03 78 D2 09 03  82 1F 47 C3 FD 02 D1 F1
0310  C9 F5 D5 1E 09 78 47 79  17 4F 1D CA 2F 03 78 17
0320  D2 27 03 92 C3 16 03 92  D2 16 03 82 C3 16 03 17
0330  5F 3E FF A9 4F 7B 1F D1  F1 C9 7E FE 00 CA 47 03
0340  CD 78 03 23 C3 3A 03 CD  5B 03 C9 F5 3E 10 D3 AF
0350  3E 64 CD E3 02 3E 00 D3  AF F1 C9 F5 3E 0D CD 70
0360  03 3E 0A CD 70 03 3E 00  CD 70 03 CD 70 03 F1 C9
0370  F5 DB 00 07 D2 71 03 F1  D3 01 C9 00 00 00 00 00
0380  00 E6 0F C6 30 FE 3A DA  70 03 C6 07 C3 70 03 DB
0390  00 E6 40 CA 8F 03 DB 01  E6 7F C3 70 03 00 00 00
03A0  00 00 00 E8 09 EB 42 4B  C9 F5 79 93 4F 78 9A 47
03B0  F1 C9 32 0A 07 78 BA CA  BE 03 3A 0A 07 C9 79 BB
03C0  3A 0A 07 C9 F5 C5 D5 CD  F7 02 09 05 6F D2 D1 03
03D0  24 D1 C1 F1 C9 CD C4 03  7E C9 CD 0F 03 E6 7F FE
03E0  30 DA 2B 01 FE 3A DA F6  03 FE 41 DA 2B 01 FE 47
03F0  3F DA 2B 01 D6 07 D6 30  C9 AF C5 06 03 06 23 05
0400  C2 FD 03 C1 23 0C C9 21  CB 06 22 0B 07 3A C7 06
0410  4F 16 0A CD F7 02 16 10  CD 11 03 26 00 69 22 11
0420  07 21 00 00 22 0F 07 1E  10 D5 2A 0F 07 44 4D 16
0430  0A CD 11 03 06 00 3A 08  07 5F 16 00 CD A3 03 16
0440  04 CD F7 02 21 6F 04 09  54 5D 2A 0B 07 0E 04 CD
0450  0C 01 22 0B 07 2A 0F 07  44 4D 2A 11 07 54 5D CD
0460  A3 03 60 69 22 0F 07 D1  1D C2 29 04 C9 10 CD EF
0B30  DA 0C FE 02 C2 54 20 4F  52 20 50 3F 00 4E 55 4D
0B40  42 45 52 00 4E 4F 20 53  50 45 45 43 48 00 4F 55
0B50  54 20 4F 46 20 57 49 4E  44 4F 57 00 00 32 5C 00
```

Software continued

There are two modes of operation training and performance. During training, speech examples of the digits are read into the microphone and the parameters of the speech input are extracted and placed in the tables. In the performance mode, an unknown utterance is presented and recognized.

To use the program, enter it into the computer starting at location 0100, and then run the program. The Teletype will respond with "T" (train) or "P" (perform). Type a "T" and the Teletype will respond with "NUMBER?" which can be between 0 and F. Type the digit you desire, and the microphone will emit a "beep" indicating that the speech window is open. When this beep occurs, vocalize the same digit you just typed in. The microphone will beep again to indicate that the speech window is now closed. The machine will then type T or P again. You answer with a T, and the process is continued as long as you want. Do not exceed 16 entries with this sample program.

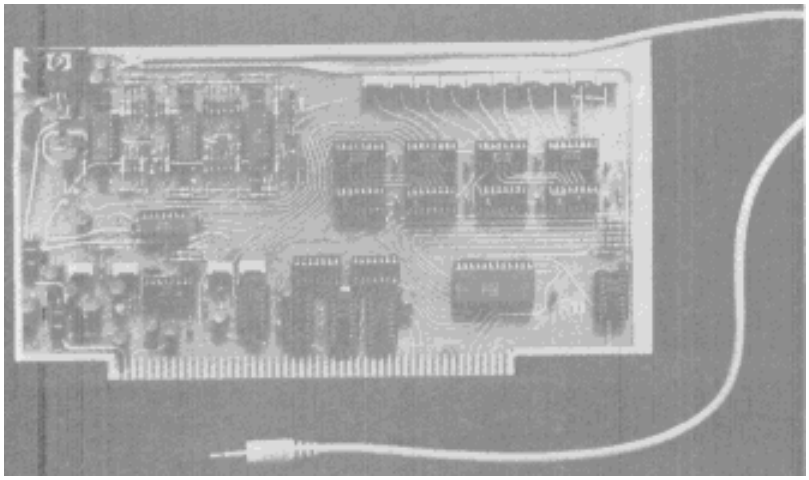
Once you have some vocalized digits in memory, run the program again. This time, when the Teletype asks T or P, answer with a P (for perform). Now, as you speak the digits into the microphone, the Teletype will respond by typing that digit. When used in a quiet room, with the same vocalization, this algorithm can be expected to have a recognition rate greater than 90%.

The program works as follows: the sampling subroutine is entered to obtain a sample of the amplitude every 10 miliseconds in each of the three frequency bands and to estimate the number of zero crossings during each time period. One hundred and fifty samples are collected, allowing up to 1.5 seconds of speech (between microphone "beeps"). A preset threshold is used to find the beginning and end of the word. The duration of the word can now be computed by a simple subtraction. Typically, this duration will be about 400-milliseconds for the digits. The duration time is divided by 16 to select 16

evenly spaced parameters from the three bands and zero crossing information.

The 64 bytes obtained (16 parameters from each of the four bands) are compared with similar parameters which were collected during the training mode. A summation (running total) of the difference between the 64 parameters of the sample and the parameters of the training "templates" is computed. The totals represent a measure of the difference between the sample and each of the previously stored templates. The template with the smallest difference from the sample is then selected as the answer (output).

The above algorithm, while relatively simple, illustrates many of the basic concepts of speech recognition. A manual supplied with the Speechlab kit contains descriptions of other approaches to speech recognition, along with sample programs to demonstrate the techniques of speech recognition.



BY LESLIE SOLOMON, Technical Editor

While testing the speechlab, we borrowed an AI Cybernetic Systems (Box 4691, University Park, NM 88003) Model1000 Speech Synthesizer ($325, assembled) to see if our microcomputer could "talk" as well as "hear." The Mode' 1000 is designed to fit into one slot of an Allair bus and delivers its output via an audio cable that can be plugged into any audio amplifler system. The output level Is 0.6 volt p-p; impedance is 1000 ohms; and frequency range is 150 to 4500 Hz. This synthesizer is phoneme-oriented. Accordingly, you can program it to say anything, as opposed to speech synthesizers that have only several words fixed in ROM. Esserwilally, the Model 1000 is a hardwired analog of the human vocal tract and various portions of the circuit emulate the vocal cords, the lungs, and the variable-frequency resonant acoustic cavity of the mouth, tongue, lips and teeth.

All the information necessary to perform the synthesis functions are located within a ROM that is accessed by the program. Words and sentences are formed by supplying a string of ASCII characters as would be done when outputting to any port, except that these strings also use some non-alphanumeric characters (i.e., the "+" is used to form "th" as in "thaw" or "earth"). Each ASCII character represents a particular phonetic sound or phoneme. If desired, you can create a program that produces simultaneous printout and "voiceout" of the same string.

The device requires very little software to implement: less than 50 bytes of assembly language or a handful of BASIC statements. The manual accompanying the synthesizer covers speech generation in detail, how it is created, and what is involved. It also illustrates how to "mechanize" speech, with several examples shown.

After working with the synthesizer for a couple of weeks, we found that we have a lot to learn about how humans create speech. After many hours of studying, experimenting, and redoing programs, we made the Model-1000 utter some recognizable sentences. It is not easy, our experience showed, even when one uses the wealth of instructions provided.

Working with a phoneme-oriented speech synthesizer is a little like learning to use a microprocessor. All the logic is there, but programming it properly is another story. like working with a processor for the first time, one must crawl frustratingly before walking. Slowly, however, the ideas start to percolate. Our com~ puter still talks with a rather heavy "robotic" accent, but we have hopes that someday it will "humanize". To paraphase Sam Johnson: "Sir, a cormputer talking is like a dog walking on its hind legs. It is not done well; but you are surprised to find it done at all." We have along road ahead to the "HAL-9000", but the first step has been taken.

Popular Electronics may 1977