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 Speecliab 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 SpeechIJab and have the computer generate the output selected for that particular sound. The vocabulary size of Speechlab is a function of the
speech recognition 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-operated door, etc.) using a repeatable sound such as a grunt. One can use Speechlab, too, as a vocal processor to add spoken commands 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.](image)

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 F1, 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 switch S1 along with three reference voltages 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 voltage 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 multiplexer (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.](image-url)
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 1% Mylar types, and all picofarad places are CMOS types.
C1,...C16, C21, C25, C47, C60, C52, C57...0.0047 µF
C2, C31...100 pF
C3, C37, C20...270 pF
C4, C7, C8, C10, C13, C15, C27, C32, C33, C34, C35, C36, C37, C44, C53, C55, C61, C62...0.1 µF, 25 V disc
C5, C14, C18, C24, C34, C36...1011 µF
C6, C42, C45, C53, C56...2000 pF
C9, C40, C46...0.002 µF
C11, C29...47 pF
C13...15 pF, 25 V tantalum
C15, C22, C31, C59...0.005 µF
C23...0.032 µF
C25, C26, C28...1 µF
C29, C39, C46...0.047 µF
C41...1 µF
C50, C58...0.001 µF
D1, D3 through D6...1N4148 or 1N914 diode
D2...1N914 diode
E1, E4, I9, IC1, IC15, IC16...
LM300 quad amp
L2...4518 8.180 MHz ceramic resonator
L6...4018 quad analog switch
L8...LM301 comparator
R1...4024 1-stage binary counter
R11, R17...740±7% DIP OP
R12...4032 hex buffer
R13...4049 hex buffer inverter
R14...4031 s/n test dual-DIP OP
R17...74LS00 8-input NAND gate
R19...8097 three-state hex buffer
R20...8093 three-state quad buffer
R22...4001 NPN transistor
MIC...Mura DX-121 dynamic microphone (part of stereo set, Mura DX-242)
R1...12...2 µH choke
R3...4018 quad analog switch
R5,...LM301 comparator
R6...4032 hex buffer
R7...4049 hex buffer inverter
R8...4031 s/n test dual-DIP OP
R9...74LS00 8-input NAND gate
R10...8097 three-state hex buffer
R11...8093 three-state quad buffer
R12...4001 NPN transistor
MIC...Mura DX-121 dynamic microphone (part of stereo set, Mura DX-242)
R1...12...2 µH choke
R3...4018 quad analog switch
R5,...LM301 comparator
R6...4032 hex buffer
R7...4049 hex buffer inverter
R8...4031 s/n test dual-DIP OP
R9...74LS00 8-input NAND gate
R10...8097 three-state hex buffer
R11...8093 three-state quad buffer
R12...4001 NPN transistor
MIC...Mura DX-121 dynamic microphone (part of stereo set, Mura DX-242)
Fig. 4. Three bandpass filters and their associated time averagers. The encompass three ranges corresponding to frequency 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.
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 flipflop 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 test program 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 computer 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.

Calibration and Test Program.
The test program (Table 1) is a general purpose calibration, test, and diagnostic program for the SpeechLab. It 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.
Calibration and Test Program continued

The program collects four 256-byte buffers of data from four of the eight possible 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 signal generator 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 combinations 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 recognition 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 including table space.

![Flow chart](image)

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
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 (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 milliseconds 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 amplifier 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. Moreover, 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 computer still talks with a rather heavy "robotic" accent, but we have hopes that someday it will "humanize". To paraphrase Sam Johnson: "Sir, a computer 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