Computers That Hear

Basic problems appear to have been overcome in the search for computers that recognize human speech.

by Carl Behrens

Ever since the first electronic computer was built, computer engineers have dreamed of a machine that would translate speech into something that a vacuum tube or transister could understand. Now at last, some promising hardware is being developed.

Engineers started trying to develop a speech recognizer at least 15 years ago. They have been been hampered by two opposing problems: the wild variability of human speech, even its most elementary spoken words, and the contrasting simple-mindedness of computer circuits which can think only in terms of black and white. Speech is mostly gray, and to a computer gray is unintellible.

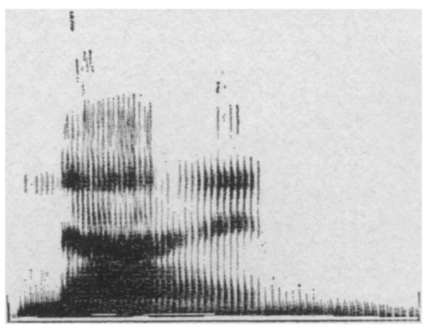
In the past few years, several major electronic firms have built machines designed to eliminate the gray areas in computer recognition of speech. One of the most recent was developed by Genung L. Clapper of the International Business Machine Corp. Clapper's machine brings a relativistic flavor to voice analysis, which allows him to eliminate many of the problems connected with absolute measurement of sound.

Both absolute and relative voice analyzers deal with two sound variables, loudness and pitch. Unlike the pure sound of a tuning fork, which has only a single pitch, speech sounds at any one instant are combinations of many pitches of varying loudness. The analyzer has to chart the pattern of different pitches that make up any given word.

A typical absolute analyzer chops a spoken word into very short segments, and for each segment draws a vertical line which represents how the loudness of the segment varies at different pitches. The louder the sound at any pitch, the darker the line is drawn at that point. When lines are drawn for all the segments of the word, a pattern representing the word can be seen.

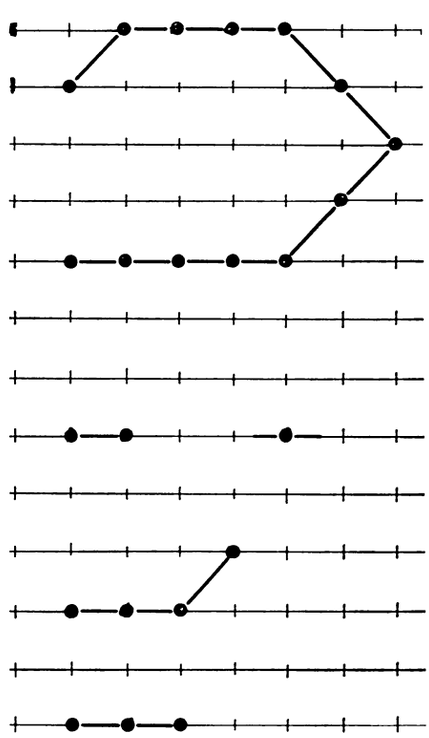
There are two drawbacks to this system. First, it is slow and clumsy, using a rotating drum and stylus to trace out a picture of a single word. Secondly, shades of gray representing loudness within a sound are charted, but the distinctions are too subtle for a computer to digest and interpret.

These problems can be eliminated by abandoning the attempt to chart the absolute loudness of each pitch in a sound segment. Instead, Clapper's machine merely indicates the pitches that



IBM

The spoken word NINE: subtle grays.



IBM

NINE: fit for computer consumption.

are louder than their neighbors. These pitches are called local maxima, and each sound segment will have several of them.

As in the absolute analyzer, putting the segments together gives a pattern representing the word, but here the pattern is a series of points, representing the local maxima, rather than a subtle gradation of lines.

A pattern of points can be digested by a computer, if the number of points is kept within reason. IBM engineers developed patterns representing the spoken numbers "one" to "nine" and fed them into a computer; the machine recognized the spoken digits 99 percent of the time, better than humans do.

Another advantage of the simplified recognition system is reduced size: the prototype machine built by IBM engineers is not much bigger than an oscilloscope. Other machines which have been built take up several racks of electronic cabinetry.

The word patterns used for computer inputs are the bare bones of the spoken word: they are the features that are always present when the word is said, no matter who says it, so a computer doesn't have to be tailored to regional dialects as it would for absolute analysis.

Beyond improving communications with computers, a speech analyzer might find application in personal identification or speech therapy. But for those a reverse set of abilities are needed. Therapy and identification require the machine to locate features which belong to the word as spoken by one person, and not by another.

Clapper says his analyzer can do just that by analyzing each segment more closely and marking more points. And because it can come up with a reading immediately, it might have real potential for speech therapists or as an identification tool for police and banks.

Successful as the machine may be, it is still a long way from the kind of science fiction computer that can understand sentences or long speeches. Words may seem to be separate and independent when written down, but in speech they are run together in a way that defies separation by any simple rules of logic. Pronunciation of many words even changes depending on the words that come before or after.

So some kind of code will still be necessary to feed information into a computer. What may be ended, however, is the dreary necessity of punching long series of numbers by hand machine, a task which occupies a huge force of keypunch operators today.

At any rate, a big step has been taken toward a computer system that can understand human speech, or at least a few words of it. What about one that talks? That is being worked on too. But that's another machine.

Science Service, Inc. is collaborating with JSTOR to digitize, preserve, and extend access to *Science News*. www.jstor.org