COMPUTER VOICE RECOGNITION

Kevin Greer and Suresh Pereira
Thousand Islands, Grade 13, Brockville, Ontario

In 1950, Bell labs attempted to develop a rudimentary voice recognition, but lacked the technology for such a feat. Since them with the advent of Artificial Intelligence and more efficient information storage, voice recognition has become much more feasible. However, to date most people are unaware of its existence, and it remains quite difficult to accomplish.

Basically, this project revolves around a voice recognition system that is both user-dependent and single word; that is, one person programs in some words, and the computer will then be able to recognize the words when repeated.

When a word is spoken into the microphone, it travels to an analog-digital converter where the analog signal of the voice is turned into a digital signal understood by the computer. This digital signal is read through the printer port. The software (designed exclusively by the entrants) then produces a dot-graph plotting amplitude vs. time, and this dot-graph is then used to check the word.

Three methods were originally developed to check the word entering the system. The first method imposed a ten by ten grid on the amplitude-time graph (A-T graph) and likened the grid to a matrix; thereafter, the values were placed in the corresponding matrix locations. This matrix was then stored beside the word. The second method first imposed ten columns on the A-T graph, and the average displacement (with all values considered positive) of the points from the Time axis was stored in a 10-length array, after that ten rows were imposed on the A-T graph, and the amount of dots within each row was stored in a 10-length array. Since it was felt that the second method, by using the row aspect, had a slight advantage, a third method was developed which only used the column aspect of the second method.

In order to compare stored words with incoming words, the three methods developed their respective matrices, which were then subtracted from the stored matrices. Each location within the resultant matrix was then squared to emphasize error. And then the locations were summed, producing a number that represented deviance from the stored word.

Experiments were then performed to determine the reliability of the system for 1) Number of words in Library, 2) Number of Lessons, lesson being the amount of times each word is stored in the library, and 3) Background Noise. These tests served a dual purpose. First, they determined what factor affected the systems, and second they determined which of the three methods was best.

However, it was found that some words were had to tell apart, cases being "one" and "four" and "left" and "right". The reason for this was suspected to be the lack of the use of frequency to check the words. To this end, two methods of frequency analysis were developed. The first was a spectro-analysis, but this method hasn't yet been used to test words, though it soon will be. The second method was similar to a spectro-analysis but it only paid attention to the major frequencies by checking for dots several units above and below the T-axis rather than checking for dots on the axis. This second method was easier to implement, and thus used in lieu of the spectro-analysis because of the lime-limitation. While comprehensive tests weren't performed to check the accuracy of the frequency analysis, the system no longer mistook "one" for "four" or "left" for "right"; and it just couldn't miss with a musical note.

Further developments include more efficient amalgamation of A-T and frequency analysis. Also, applications of voice recognition are to be both simulated on the computer and demonstrated.

Our conclusions were that your system, using A-T graphs, was accurate up to ten words, at which point it became almost useless. Further, it was concluded that any real recognition system would have to have frequency analysis of some design. It was also concluded that our second method of A-T analysis (the column then the rows) was the most reliable of the three.

Bibliography

Scientific American. (1988), "Trends in Computing", Computer Software for Intelligent Systems, Special Issue. Vol. 1.

Byte Magazine. (1984). The Texas Instruments Speech Command Systems, June.

Discover Magazine. (1988). 2001 Computers: Living inside your PC, November.

Note: Accompanying appendices showing test results and sample screen output were included in the original report.

(For this project, Kevin Greer and Suresh Pereira were awarded the CAA prize in Acoustics at the 1989 Canada-Wide Science Fair. Both students have recently completed their first year with "A" standing at the University of Waterloo in Computer Science and Physics, respectively.)