



Then and NOW

ROBOTS WHO LISTEN

by Tom Carroll

Speech recognition systems applicable to robot use have dramatically dropped in price over the last few years. Rather than going through the myriad of computer-based solutions that we can use to make our robots listen to us, I'd like to talk a bit about why it has been so difficult to implement this listening ability.

The understanding of human speech seems to top the list of desires for robot experimenter's projects, sometimes even more than basic mobility. "Now, a robot who can listen to my voice and obey my commands; that is the starting point for an intelligent machine," you say to yourself. Speech is a human's way to communicate with others so it stands to reason that speech recognition is the most natural way for us to communicate with our robots.

This is the subject that I always enjoy talking about with other experimenters. I still like to go back to Isaac Asimov's *Robot* series of short stories and the story — Robbie. Young Gloria Weston was given a robot — Robbie — as a babysitter and companion. This mute robot would quietly sit next to Gloria as she told him stories. She frequently would give him numerous verbal commands in her childish ways, which he would quickly obey. Her mother quickly tired of the robot and had her father send it back to the factory without Gloria's knowledge of what they had done. One day, while visiting a museum, Gloria was transfixed by the world's first talking robot and began to ask it "Mr. Robot, Sir. Have you seen Robbie?" Of course this robot had no clue about what she was asking and was about to 'blow a circuit' when the

operator ran up and told the gathering crowd that they could not talk to the robot without an attendant.

For those of you who know Asimov's stories, this is where young Susan Calvin is introduced as a student, quietly taking notes on the robot and the spectators. She later became Dr. Calvin, a robopsychologist for US Robots and Mechanical Men, Inc. Later in the story, Robbie would save Gloria's life and gain acceptance by all the Weston family members.

What I've always found to be interesting is how robot speech came after speech understanding for robots in Asimov's stories. When he wrote the first story about Robbie in 1940, it was entitled *Strange Playfellow* and published in *Super Science Stories*. Crude "robot" speech was already being investigated, yet true machine speech understanding was still a long ways in the future.

I assume this speech understanding was a must for Asimov's robots to be able to react to his "second law of robotics" — "A robot must obey the orders given it by human beings except where such orders would conflict with the first law." (The first law prevented a robot from injuring a human or allowing a human to be injured.)

What is Speech Recognition?

The terms, speech recognition or voice recognition have always bothered me as they really do not imply the features we actually desire for our robots to possess, but that's the screwy English

language that we use. As I've mentioned in other articles, I can "recognize" Russian speech and, in fact, actually recognize that it is the language that Russians use, but I don't have a clue what Russians are saying. A dog can recognize its master's voice and come running with tail wagging when it hears him say "hey, boy, ya wanna go to the vet and get neutered?"

We really cannot use the term that I use sometimes — speech understanding or cognition. Well, enough of that. We cannot even come to a unified definition of a robot, so how are experimenters supposed to decide on how to describe how robots listen?

A speech recognition system installed on a computer can identify each word through a complex set of algorithms and print them out in a sentence, but few computers available to experimenters actually "understand" what the line of words mean or the context in which they are used. We just program into the microprocessor or stand-alone speech recognition board that the words "go right" triggers another line of code to make the left motor (in a differentially-driven robot) turn more revolutions than the right motor. Well, it's really not quite that easy, but that's the basic principle.

The bottom line is: Speech recognition sounds a bit more applicable to a computer understanding commands given to it, as "speech" refers to a series of words that imply an idea, command, or meaning. "Voice" recognition can refer to just the sound of a person's voice or a single word triggering the computer. Though many magazines and companies



Figure 1. Butler in a Box.

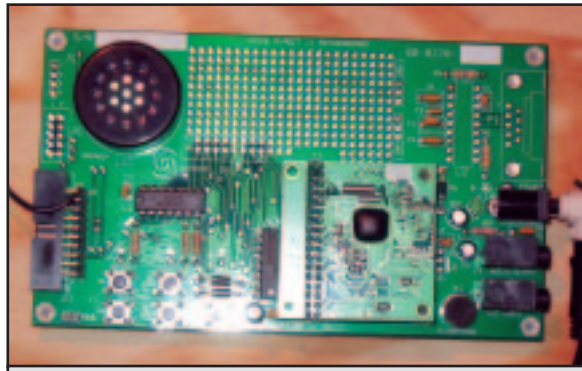


Figure 2. Voice Direct II.

use the two terms interchangeably, I'll go along with speech recognition for this article. Besides, speech recognition had 19 million hits on Google, vs. 11 million for voice recognition.

The Nuts and Bolts of Speech Recognition

Speech recognition is the ability of a machine or program to receive and interpret words spoken to the system, or to understand and carry out spoken commands. Applications can vary anywhere from dictation (office steno, voice input word processing, court reporting, etc.), control of machines (industrial, medical, battlefield, etc.), telecommunications (cell phones, telephone verification messages, etc.), and, of course, robots (home, industrial, etc.). There are numerous other applications and many are applicable to physically disabled persons to assist them in daily living.

The bane of all speech recognition systems has always been background noise (machines, others talking, barking dogs, etc.) and variations in the human voice. Directional or closely held microphones and filters help clarify the input voice signals to the computer. For use with computers, the audio signal from a microphone or audio source must be converted into digital signals by an analog-to-digital converter.

If a computer is to understand the speech input, it must have a database or vocabulary of words or phonemes and a rapid way of comparing this data with the input. It is at this point where we can divide the speech recognition systems into speaker independent and speaker dependent systems.

Higher-end speaker independent

systems can be so designed that anyone can enter a voice command, word, or phrase and the computer can understand them. A speaker dependent system requires a voice template (or templates) of a single person's voice speaking the required vocabulary words. It is sometimes advantageous to have a computer or robot that can verify the speaker and be controlled only by that person. This variation is called speaker verification.

The most complex system is continuous listening of speech without having to press a button to start a period of listening. The reference speech patterns can be stored on a hard drive for a computer-based program, or static RAM or Flash memory for a stand-alone board. A comparator checks these stored word or phoneme patterns against the output of the A/D converter and makes a determination of what word was spoken into the microphone.

Early Speech Recognition Systems for the Experimenter

The earliest "speech recognition" that I remember for home experimenters was not really speech recognition at all, but was just "syllable counting." One of my early robots used this method for fairly crude control. Quite a few toy vehicles have also used this system that converts spoken word syllables into pulses to drive relays for control.

I happened to have a bunch of 12V relays and some telephone office stepping relays that my brother had given me. The huge telephone relay, the amplifier and microphone, and other relays almost doubled the size of the robot and

I had to almost shout the commands: "Stop! Now go right! Go left!" etc. If I managed to screw up the three syllables "Now go right" and said instead, "Now go left," the robot would count the three syllables and still go right. My friends quickly found that watching paint dry was more exciting than watching my stupid robot go in the wrong direction.

Microprocessor Based Speech Recognition

Since there have been so many manufacturers of speech recognition systems, including main frame, personal computer based, and stand-alone board level units, I'll concentrate on the systems available to the experimenter. The government, military, and many universities have long been experimenting with systems for their particular purposes, yet the greatest breakthroughs came with inexpensive experimenter's units.

Back in the early '80s, many of the new 6502, Z-80, and 8080-based personal computers had manufacturers designing speech synthesis and recognition stand-alone boards for experimenters. Steve Ciarcia featured the Lis'ner 1000 speech recognition system in his popular *Ciarcia's Circuit Cellar* column in the November 1984 *Byte Magazine*. The Lis'ner 1000 was a low-cost (\$150), high-performance speech-recognition system for the Apple II or any 6502-based system.

After experimenting with the Votrax SC-01 speech synthesizer for a while, I was given a Lis'ner 1000 by my friend, Dave Freeman, co-founder of Advanced Computer Products in Santa Ana, CA. It was a bit cantankerous to program and use but it finally made my robots seem a bit more human, at least more human than the telephone relay cheating thing. I first used a KIM-1 single board computer and later, an AIM-65 board as the controller. Now, instead of my robots leaping to their death from my workbench when I turned on the power, at least they learned to wait for my verbal command before taking the unexpected death dive. (I later discovered that the set of "H" bridges that I designed myself occasionally went from zero output to full output when a signal was applied. Using a friend's design corrected the problem).

ViaVoice from IBM and Dragon System's Naturally Speaking are two software-based systems that became available to experimenters a bit later that improve over time as the computer learns and adapts to individual speaker's speech patterns. These two companies have further refined their products into Dragon Dictate and "VoiceType" from IBM, and Apple Computer also has a speech recognition program called Voiceprint, all for speech-to-typed-word applications. Say goodbye to the stenographer.

Butler in a Box

One of the first out-of-the-box fully integrated systems that I ever got a chance to experiment with was the Butler In A Box from Mastervoice, now called AVSI, Inc. Butler In A Box was created by founder and company president Gus Searcy, a professional magician. I invited Gus to one of our Robotics Society of Southern California meetings in the mid '80s to demonstrate his new speech-controlled home automation device. All of the RSSC membership voted this device the coolest thing that they'd ever seen. "A magician shouldn't have to get up to turn the light on," he said. "As a result, I decided to create the illusion of an invisible magic genie."

Figure 1 shows the large dictionary-sized standalone automation computer that uses a smart controller and X-10 compatible devices. You can program the system by using a keypad on the box. There are several Butler In A Box models ranging in price from \$1,795 to \$3,995. These aren't cheap puppies but the 'cool factor' is way up there.

Sensory Speech Recognition Products

To really get a handle on the latest speech recognition systems available to robot experimenters, I spent quite a while on Google checking out manufacturers, comments, about systems, and which ones were more applicable to robot experimenters. I was somewhat familiar with Sensory's Voice Direct II \$49.95 speech recognition board that features continuous listening and recognition technology and allows

almost any device to be controlled with just the sound of one or two key words or a short phrase (see Figure 2).

It then listens for up to three seconds to recognize one of up to 15 additional "command" words or phrases lasting up to 2.5 seconds each. When a command word is recognized, the Voice Direct II will raise one or two output pins high for one second, which can be used to control external devices. All the trigger and command words are speaker dependent, so the recognition technology will work for any language.

It can also be configured to have one to three different continuous listening trigger words or phrases, each with up to five speaker-dependent command words. The University of Vermont's College of Engineering and Mathematical Sciences used this system in an experimental van for the physically disabled where the driver could verbally command various functions such as windshield wipers, lights, etc., without removing a hand from the steering wheel.

The VR Stamp

Recently, the Dallas Personal Robotics Group had a series of responses on its website to a discussion on Sensory's VR Stamp, available for \$39.99 at Digi-Key and other places. Figures 3 and 4 show the 40 pin DIP configuration and the block diagram of the VR stamp and the on-board microcontroller. After reading the different replies, I decided that I had to have their VR Stamp Toolkit to develop a reliable speech recognition system for a robot I'm working on. The kit comes with two VR Stamp modules and a programmer

board with a built-in microphone. Also included is a VR Stamp Toolkit CD-ROM with a lot of useful documentation and speech tools, a serial-USB cable, speaker, and a wall-wart power supply.

I was a bit disappointed that there was no printed documentation with the kit as I had to search back and forth on the CD-ROM and print out what I needed, but there is an amazing amount of information on the CD needed to set up the system and for other applications. Besides basic information guides in Adobe Acrobat format, there are four software installations that you can use for different applications and the needed drivers.

A quick connection to the 9 VDC wall-wart power supply, the included speaker and the USB cable to the computer and the system was up and running. The 40 DIP zero insertion force socket on the VR Stamp Programmer board allows a quick interchange of two programmed VR Stamps. The 3-1/4" x 4" programmer board size is convenient for mounting on a robot, though the 40 pin DIP Stamp and an associated microcontroller is all you'll really need after programming the Stamp. I found the VR Stamp to be quite accurate, though pronunciation and word speed and separation helped the accuracy a bit.

Figure 3. 40 pin DIP configuration of the VR Stamp.

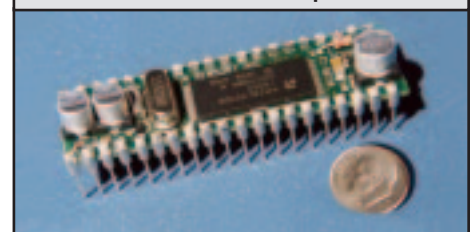
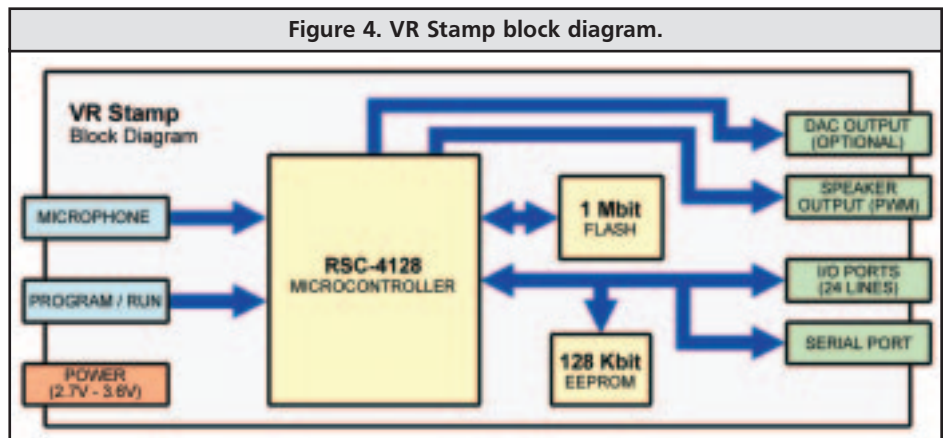


Figure 4. VR Stamp block diagram.



RESOURCES

Sensory Systems
www.sensoryinc.com

AVSI Automated Voice Systems
www.mastervoice.com

Dragon Systems
www.speechtechnology.com/dragon/

IBM (and Dragon)
www.voicerecognition.com

Another company associated with Sensory and their VR Stamp is a company based in Belgrade (formerly Yugoslavia) – mikroElektronika. It was established in 1997 as a publishing firm specializing in electronics, and has become well known for PIC, AVR, 8501, and other microcontroller development tools. They also make the Easy-VRStamp development system board for the VR Stamp voice recognition modules. It was designed for students and engineers to explore the capabilities of VR Stamp voice recognition modules. The development board's \$129.95

price includes a \$40 VR Stamp – a real bargain for those who want to start experimenting with speech recognition.

I have just touched upon this complex and exciting part of the new robotics age. Speech recognition, like most areas of electronics, is making rapid advances. One day soon we'll be able to tell our robot "Robbie, go deep and catch this forward pass." Of course, we'll need to work a bit on the mechanics, but true speech cognition is definitely on the near horizon for robot experimenters. Enjoy talking with your new friend who will gladly listen to you. **SV**