TechTalk

TALK TO ME

Like everything else, they needed the survey yesterday—eighty-four questions, a five-point Likert scale angled at 90 degrees, and clear directions with examples. They said, "It has to fit on two pages, and we want the 'fold and staple' design, with the return address preprinted on the back. Oh yeah, and use plenty of white space."

Right! So there I was trying to polish off the last few design changes. Let's see—to insert the graphic, "click this, press Control-Alt Y, and...." The survey disappeared!

I yelled at my computer, "No! You're supposed to insert the graphic!"

I clicked the "undo" button. Oops, it couldn't undo.

I sighed and shut my eyes. Well, they said they wanted a lot of white space.

Luckily I had a back-up copy, so the story ended well. But wouldn't it be great if you really could *tell* your computer what to do? Forget the keyboard commands. Forget the icons. Just say, "Insert graphic, please," and there it is. And while we're dreaming, why not have your computer take dictation, too?

You say, "Take a letter for William." And your computer pulls up the letter format, inserts William's address, and displays on the screen whatever comments you utter.

You say, "Great. Now fax it." The computer opens the fax program, initiates the dial sequence for William's number, and sends the letter on its way.

Speech-recognition systems. If that scenario piques your interest, I've got some information for you. Right now, today, thousands of computers are doing what their owners *tell* them to do. That's right. Thousands of computers are responding to verbal commands just as if the commands had been entered from a keyboard or mouse.

You need at least three things to put speech recognition, or SR, to work for you:

▶ a relatively fast computer (at least a 386 or equivalent)
▶ a digital audio card (with microphone and headset)
▶ compatible speech-recognition software.

Together, the audio card and SR software can set you back anywhere from $200 for low-end systems to about $5,000 for top-of-the-line PC systems. The faster your computer and the more memory (RAM) you have, the better.

In general, SR systems work by digitizing incoming sound in small segments (about 10 milliseconds) and comparing the digital patterns to prerecorded speech patterns (words or phonemes) called "contour maps." To help match the digital patterns, the systems employ such sophisticated statistical analysis methods as dynamic time warping (a process that identifies temporal differences in live and prerecorded sounds) and the hidden Markov model (which among other things helps to identify and filter ambient noise).

Once the system identifies a word, the computer takes a particular pre-programmed action. The action can range from simply displaying the word on the computer screen (for example, if you are creating a letter) to activating your modem and dialing a phone number. The point is that the computer can act on spoken words just as it would on any other data it receives from the keyboard or mouse.

What's in it for me? I believe an understanding of SR will be important for

The premier edition of T&D's new quarterly column looks at speech-recognition technology.

those of us working with CBT, multimedia, expert systems, and electronic performance-support systems—especially for use in hands-busy, eyes-busy environments (for example, for pilots, technicians, mechanics, and equipment operators).

SR will almost certainly become part of multimedia authoring packages. And, with the proliferation of high-quality, low-cost sound cards, SR is likely to affect the rest of us who use computers simply to get our daily tasks done. Sound cards are already part of the multimedia personal-computer standards (the MPC 2 standards). In fact, when it comes to IBMs, Apples, and many other types of systems that offer SR systems, we may soon find it difficult to buy a computer that doesn't have SR capability.

Corporate applications. In the late 1980s, SR began finding its way into corporate America in large-scale projects. In a 1992 article in the *Wall Street Journal*, John Keller reported that American Express (Lehman Brothers), GM (Saturn), and AT&T had implemented speech-recognition systems. Lehman Brothers is reported to be using SR to increase efficiency in initiating trades. Saturn is reportedly using SR in quality control. And AT&T is said to be working on an SR system that will handle incoming calls—not by touch tone, but by voice.

I've spoken with officials from Boeing who said they use SR-equipped computers to display wiring diagrams by voice command. With the SR system, technicians can continue to wire the aircraft while verbally requesting the computer to display needed wiring diagrams. Emergency-room doctors and radiologists are using SR systems in more than 400 locations across the United States. Even the U.S. post office is using SR systems in some locations to help new employees sort bulk mail.

Implementation challenges. Before you rush out to buy an SR system, you should know the systems aren't perfect. First of all, recognition rates are never 100 percent, and they vary with different systems and applications.

The first time I used an SR system for wordprocessing, I came up with about a 50 percent recognition rate, initially. After about 15 minutes, the system learned to recognize my southern drawl, and the rate increased to about 95 percent when I used words it had "heard" me say before.

Speaker-independent systems, such as the one I used, come with relatively large, generic vocabularies (thousands of words stored as digital patterns). They have the ability to adapt to multiple users on the fly. You can add your own words too, and the more you use the system, the better the recognition gets.

Speaker-dependent systems, on the other hand, typically handle small vocabularies (dozens or hundreds of words). They require each intended user to give them hours of "training" (saying the words you want a system to digitize and store in its vocabulary) before they can recognize word one.

In addition, SR systems are not well suited for environments with a lot of ambient noise—especially random noise. To an SR system, noise becomes part of the words you speak, making them more difficult to identify. Constant background noise—such as the hum of a fan—can be filtered out. But random noise will generally reduce recognition rates.

Due to PC speed limitations, most SR systems can handle only discrete speech. In other words, to have the computer understand your commands, you—have—to—say—them—like—this, using small pauses (about a quarter-second minimum) between words. This is fine for individual commands, but some find it tedious for dictation. True continuous-speech systems are in prototype development now, and are likely to become available as PC processing speeds increase.

Finally, positioning the cursor and making corrections in documents using voice commands can be frustrating (for example, "cursor—down—5, left—2, highlight—5, delete," and so forth). Current SR systems seem better suited for applications that do not require detailed cursor positioning.

For those of you who are interested in learning more about SR systems, I've listed a few sources in the box.

"TechTalk" *is a quarterly column written by* **Bart Dahmer,** *a technical analyst in HRD for Federal Express, 3035 Director's Row, Suite 501, Memphis, TN 38131.*