U.S.News & WORLD REPORT

Science & Technology; News You Can Use

Let's Talk

Janet Rae-Dupree
1,255 words
12 May 2003
U.S. News & World Report
58
English

I wrote the first draft of this story by talking at my computer. I thought it would be easy. After all, technical writer Karl Barksdale has composed the last 30 of his 43 books that way. "I got hooked after I went from 70 words per minute typing to 140 words per minute talking," says Barksdale, who also teaches eighth grade in Provo, Utah. These days, he says, better software and faster hardware mean that most of those words come out right.

But Barksdale has been gabbing away at his computer since 1997, and several of his books are training manuals for speech recognition software. For me, experimenting with two of the latest programs was an exercise in frustration. On a four-year-old laptop, neither the operating system nor the processor was up to the task, resulting in slow response and spotty accuracy. After I switched to a state-of-the-art desktop, the programs galloped along at nearly the speed of my voice and accuracy improved dramatically.

Yet faithfully capturing spoken words is not enough. For a generation used to composing on the keyboard, reciting fully formed sentences doesn't come naturally, and the software does not make it easy to backtrack and reword your utterances. In the end, I shut my mouth and plied the keyboard. Bill Meisel, president of speech industry consulting firm TMA Associates, says that's a common reaction among users over the age of 20.

Many others, however, are happy to switch from typing to talking. Sales of speech recognition software and hardware--from natural-voice dictation systems for PCs to commercial telephone systems--are expected to rise from $680 million last year to $2.2 billion in 2006. Disabled people and those with repetitive motion injuries have embraced the technology. In fact, Barksdale first got interested in speech recognition when he bought IBM's ViaVoice dictation program for a student with cerebral palsy. Within weeks, he says, the boy was entering text at twice the speed of the school's fastest student typist. And a whole new generation of users is on the way. Last year, the National Business Education Association recommended that all students from fourth through 12th grade learn how to use both speech and handwriting recognition software (as well as learn to type). So far, hundreds of high schools and community colleges around the country have instituted programs.

You're the boss. Two other categories of speech recognition already have begun to work their way into our day-to-day lives: the perky computer-driven personas that take your information on phone service sites from Amtrak to Moviefone (story, Page 60), and the "embedded" speech recognition built into chips in cars' navigation systems, hand-held computers, and even toys. Bossing around your car or ordering movie tickets, however, is a very different proposition from chatting to your desktop computer. Telephone personas and embedded speech devices only have to recognize a limited number of words and phrases--a few hundred to a few thousand--although they must be able to cope with a wide range of accents and mispronunciations.

PC dictation systems, though, have to recognize 250,000 words or more. The latest versions gradually tailor themselves to a single user, learning from mistakes and adapting to the user's speech patterns as they parse raw, digitized sound into words (box, Page 59). Because they often have to distinguish words that sound alike, the systems also pay attention to context in deciding what to put up on the screen.

Hardware improvements have helped, too. In 2000, Intel redesigned the Pentium processor to make it better able to decipher audio input. At the same time, the price of computer memory plummeted, enabling cheaper computers to handle the huge vocabulary database needed for speech. And headset maker Plantronics introduced some of the first noise-canceling headsets intended specifically for speech recognition, improving the software's accuracy.

Don't try running the software on an older machine. I first loaded ScanSoft's Dragon NaturallySpeaking 7 Preferred software and IBM's ViaVoice 10 Pro USB Edition (around $170 for either package, including a microphone) onto a four-year-old Windows 98 desktop. Long delays between my spoken words and the computer's written ones indicated that the software makers have set their minimum system requirements far too low. But on a brand-new Windows XP desktop, both programs trotted through their paces with aplomb. With a bit of practice and about 20 minutes of "training" the software to my voice, I could dictate text from a book or magazine with scarcely a glance at the computer screen to verify accuracy. Dragon seemed slightly more accurate than ViaVoice, missing only one or two words per hundred rather than three or four.

When it came time to make corrections, however, Dragon stumbled. It often misunderstood the words or short phrases used as editing commands, such as "define," "select," or "delete that." Ask it to define text and it's as likely to type "the find text" as it is to highlight your selection. ViaVoice is better at following the commands, but talking out revisions is still laborious.

And most users will need to make plenty of corrections. The stop/start nature of thought and the inevitable "ums" and "ahs" of extemporaneous speech throw both the machine and your concentration for a loop. "It's a learned skill to say something the way you want it to appear on paper," says Meisel. "Most people have no idea how to dictate." You may just decide it's easier to type.

Look, Ma, no hands! The up-and-coming younger generation is a different story. Raised on a steady diet of Star Trek thinking computers and fast-moving video games, youths expect computers to automate what has not been automated before. High school typing teachers aren't discouraging them. In 2000, Barksdale and three other teachers set up speech-recognition seminars for about 300 middle and high school business teachers--what a previous generation knew as typing teachers--from 11 states. By the end of 2002, his informal group had trained more than 4,000 teachers in all 50 states.

After eight to 12 hours, most students and teachers are able to dictate more than twice as fast as they can type, says Cindy Agnew, program developer for technical education at Washington State's department of education. The benefits go beyond fast text entry. "The kids are learning how to talk, how to read more clearly, enunciate, and speak up," she says. That's especially valuable for non-native English speakers, special-education students, and very young children, she says. Her 9-year-old son has learned to create stories using Dragon NaturallySpeaking. "Talking and seeing his words come out on the screen is just amazing to him."

And what college student, late on a 2,000-word essay, wouldn't rather talk than type? Last semester Seton Hall University in New Jersey, in an effort to keep undergraduates on the cutting edge, started distributing IBM ViaVoice software to its entire freshman class. Says Chief Information Officer Stephen Landry, "They're the ones who will carry it into the business world. They're the ones who will make all of us want this and need this."

The software is capable, but dictating to your PC is harder than it looks

Drawing: No caption (ILLUSTRATION BY CATHY GENDRON FOR USN&WR)

Document usn0000020030506dz5c0001v