



Voice Activated

One minute Kees Vander Putten of Kenmore, Washington, was a healthy 20-year-old art student working toward a degree in 3-D animation. The next minute, a spinal cord injury had paralyzed him from the shoulders down. Something went awry when he attempted a back flip while clowning around with friends.

Now 25, Vander Putten works as a freelance animator for game companies, doing his job from his basement apartment, which is lined with art posters. An aide comes daily to help with personal care and housekeeping, but Vander Putten can do most other things himself.

The apartment is wired so that everything electronic—from the lights to the television to the computer to the cell phone—responds to his voice commands. “Once I’m out of bed, I’m pretty independent,” he says. “Anything that has an on-off switch, I can operate with my voice.”

Vander Putten positions his motorized chair close to a microphone attached to a stand beside his bed and uses the mic to operate HAL2000, a computer system produced by Home Automated Living Inc. that puts his world within his control. The microphone is wired to a computer, which sends signals across the power line to communicate with electrical devices and other equipment.

Homeowners who have the use of their hands may choose to “talk to HAL” via in-home telephones—including portable cordless phones—connected to the system. They press a certain phone key, such as the pound key, to signal HAL to listen to a command. Vander Putten uses a basic HAL setup. A whole-house open-air microphone system is more complicated and requires additional hardware and wiring. It’s easier to set up while a home is being constructed, and generally requires the help of an audio technician.

But whether you use microphones or a regular phone, you can use your voice to turn on the coffeemaker, start the popcorn popper, adjust the thermostat, log on to the Internet and turn on the TV. You can even use a program to operate a home theater. The voice command, “Prepare the family room for a DVD” can prompt HAL to turn on the TV, switch it to the right channel for playing a DVD, turn on the stereo and configure it for theater sound, set the DVD to play and dim the room lights as the movie begins.

What’s more, if you’re on vacation and decide to stay a little longer, you can call HAL from your hotel and tell it to keep the

thermostat low or have the VCR record a game you want to watch.

“You can get started with some pretty cool home automation for less than \$1,000,” says Bill Stevenson, vice president of product distribution for Home Automated Living. Costs rise, of course, with the sophistication and complexity of the system, but “you don’t have to be Bill Gates to have a smart home,” Stevenson says.

HAL—which has been featured on several *Extreme Makeover: Home Edition* episodes—is an example of speech recognition, a technology that over the next decade may change how we operate appliances, open and start our cars, make phone calls and surf the Internet.

Already another product, Microsoft’s Voice Command—released in August 2004—can be used to voice-activate car stereos and call familiar numbers from mobile phones, according to Rob Chambers, a software architect for Microsoft.

Products such as Voice Command also let you use your voice to play slower-moving computer games such as chess.

“Anything you can do with your hands, you can in theory do with your voice,” says Dan Comden, a consultant at the University of Washington Access Technology Lab in Seattle. As part of its mission of improving access to computing resources at the university, the lab provides speech-recognition software as a writing tool.

Before too long, we may be a speech-driven versus button-punching society. You might, for instance, turn on a computer with your voice while making a dinner salad, ask the system to bring up your favorite florist site and direct the site to “Send Mom flowers.” Then you could call out to the TV to change the channel to *Jeopardy!* so that you don’t have to pick up the remote and rewash your hands.

A fringe benefit of speech recognition might be reducing the chances of carpal tunnel syndrome in those of us whose work and home lives

revolve around computers—although voice strain might then become an issue.

Despite the potential of speech recognition, there are some barriers to widespread use. For one thing, manufacturers of TVs, coffeemakers and other appliances haven’t seen enough demand to warrant putting sound chips in their products, and until they do, home-control speech recognition will require talking into a telephone or microphone, with the requisite wiring.

“Home or office environmental-control systems will only reach a critical mass when voice-recognition chips become routinely imbedded in appliances,” says Marty Tibor, a nationally recognized computer expert who is president of Synapse Adaptive, a San Rafael, California, company that helps employers meet accommodation requirements for people with disabilities.

Speech-
recognition
software
may soon
transform
how we
make calls,
surf the
Web and
watch TV



By Scott Driscoll

The technology for ubiquitous voice chips is available right now, he says, but the chips are a standard feature only in most cell phones, where they have become increasingly common since 2004, thanks to improvements in voice-recognition technology and to increasing realization of the benefits of hands-free phone use. Even so, many cell phone users don't know about or don't use their speech-recognition feature, he says.

Lack of knowledge about speech recognition is one of the main reasons it's not in greater use, says Ed. (yes, his legal name is spelled with a period after the "Ed") Rosenthal, president and CEO of Next Generation Technologies Inc., a technology-consulting firm in Lynnwood, Washington, north of Seattle. "People still don't understand how competent and easy-to-use speech recognition really is."

While industry experts agree on the potential for the technology, there isn't consistency about what to call it. Most consumers use the terms "speech recognition" and "voice recognition" interchangeably, as does this story. Some techies prefer that "voice recognition" be used for programs that, to be accurate, need to be trained to recognize the peculiarities of a specific speaker's voice. Dictation and composition programs fall into this category. Other techies use "voice recognition" as the general category but divide it into subcategories of trained-to-voice "speaker recognition" and no-training-required "speech recognition." Speech recognition includes the voice-response telephone menus consumers get when they call various companies.

Voice-recognition software developers know that voice recognition got off to a rocky start in the early 1990s with first-generation programs such as Dragon Dictate, developed by Dragon Systems, a company whose assets have since been folded into Massachusetts-based Nuance Communications via a series of acquisitions. Capable of recognizing only words spoken slowly and separately, the programs were the antithesis of user-friendliness. "Training was laborious," says Comden from the UW, "and recognition was not as accurate."

But in 1997, Dragon NaturallySpeaking—"continuous-speech" software that let users speak at a normal conversational pace, with full sentences, and that dealt better with accents—impressed users. The software has gotten better and better ever since.

"The last two years saw big changes that remarkably improved speed and

bumped accuracy up to as much as 99 percent," Rosenthal says. "Today's improved microprocessors run the code better, and today's software makes better probability guesses as to what you're really saying."

Microsoft's new Windows operating system, Vista—which the company expects to introduce to consumers early next year—will include fully integrated voice recognition that Microsoft claims will provide "unparalleled voice-recognition accuracy."

"Our prediction is that by 2011, our software will be better at recognizing human speech than humans themselves," predicts Chambers, the Microsoft software architect. The system will be so easy to use that many people with disabilities will choose 100 percent hands-free operation for running their homes and using their computers, he says.

Right now, even quadriplegics whose computers have voice recognition often choose, instead, a joystick, which is actually a strawlike device attached to the wheelchair. The joystick responds to sipping and blowing, with mouselike clicks on a virtual keyboard shown on a screen. It eliminates the need to memorize voice commands.

Microsoft is also part of a group called the Speech Application Language Tags (SALT) Forum, formed in 2001, that also includes Intel, Cisco, Comverse, Philips and SpeechWorks. The forum's goal is to use Internet code such as HTML—hyper-text markup language—to create speech-recognition software that allows the user to browse the Net and interact with Websites without ever clicking a button.

This could make cybersurfing so easy that some industry analysts predict it will be the catalyst for mainstream use of speech-recognition technology. The group's progress so far has included making it easier to create Websites that give spoken commands and spoken descriptions, and that have "hot spots" that respond to spoken commands.

"The staying power of these technological solutions is unassailable," says Dan Miller, head of Opus Research, a "conversational-access technologies" advisory company in San Francisco. "The problem is, will people use them? Changing people's behavior takes longer than changing the guts of a device."

Even Vander Putten has reservations. He still uses his joystick for various tasks that could be controlled by voice, because, he says, "I began to realize I don't like talk-

ing to nonconscious objects."

Despite consumer resistance to change and consumer lack of familiarity with speech-based products, the voice-recognition revolution is already making headway in at least three major areas: service by phone, dictation and assistance for people with learning or physical disabilities.

Service by Phone

Nowadays, the first voice you hear when you call a company is often that of a speech-recognition attendant who understands your answers to her questions and speeds you to the right place for what you need. That doesn't always happen, of course, especially if your last name has lots of syllables or your request is complex, but the improvement in speech-recognition technology over the past two years has led more and more companies to boost productivity, and the bottom line, via "natural-language" voice response.

Opus Research estimates that voice-activated answering systems will handle more than a fifth of automated calls this year. That number is expected to grow rapidly because it takes \$6 to \$15 per call to have a live agent answer the phone, according to a 2003 *Wired News* report, and that's with less-expensive overseas labor factored in. Costs for putting in a speech-recognition system start at around \$10,000 for a small company, and annual operating costs are minimal, according to Opus Research. High-volume call centers may recoup the cost in just a few months.

For these companies, a speech-recognition system can mean faster customer service and fewer abandoned calls by frustrated customers, without the company's having to add more staff. Software developers say customers also like speech recognition better than touch-tone systems.

Software maker TuVox of Cupertino, California, reports that its speech-recognition applications improve the rate of successful call completion without live-agent involvement—for services ranging from looking up fares to ordering tickets to returning products—from 25 percent to as much as 45 percent over the success rate for touch-tone applications.

In an October 2005 *Call Center Magazine* story, an IBM spokesman reported 10 percent to 15 percent better completion rates for speech recognition than for touch-tone.

In June 2005, Alaska Airlines and Horizon Air joined the growing list of companies that greet callers with a friendly natural voice programmed to respond to

their needs. The warm recorded voice callers hear is that of Theresa Miller, an Alaska Airlines customer care supervisor.

"People dialing in actually call her 'Sweetie' and thank her for her help," says Karen Wells, manager of reservations technology for Alaska and Horizon.

"We receive about 30,000 calls a month," Wells says, noting that it would take 18 to 20 agents working full time to handle those calls, 5 percent of which are merely to reconfirm reservations. "This frees up those agents to handle more complicated transactions," she says.

In terms of labor costs saved, she estimates that within the first six months, the airlines began to see a return on the cost to install the service. They started small to test how effective the system would be. Speech-recognition software handles only reconfirmations and check-ins 24 hours prior to flight. So far, the airlines have had virtually no complaints, and only about 30 percent of the callers who talk to the voice-response system need to route back to an agent, Wells says.

Dictation

Members of the health-care industry also are benefiting from voice recognition, says Rosenthal of Next Generation Technologies. Hospital departments such as the Pathology Lab at El Camino Hospital in Mountain View, California, and private practitioners such as Dr. Philip Ranheim in Lake Stevens, north of Seattle, are saving time and money by using voice recognition to dictate patient records.

"A typical savings could be up to \$25,000 per year in transcription costs," Rosenthal says. "Turnaround time can be reduced from a wait of several days for the return of transcribed records to immediate delivery."

Nuance, the maker of Dragon NaturallySpeaking, states that in North America alone the manual transcription of patient data costs the health-care industry on average \$10 billion per year. "We're seeing explosive growth in all markets, but the most exceptional increase has certainly been in the medical field," says Matt Revis, Nuance's director of product management.

The company's latest medical software can process up to 160 words a minute with as much as 99 percent accuracy, he says.

Judith Maddox, manager of the Pathology Department at El Camino Hospital, says her department switched to Dragon three years ago. "We used to have to stop what we were doing to type in everything

we observed while looking at a specimen," she says. "Using Dragon frees up our pathologists' hands to keep busy with a tissue sample."

Since the switch, the department has cut its response time to doctors in half, from two days to one. It saves \$240 per day in transcription costs, and it's become so much more efficient that when a worker left, the department didn't need to replace those hours.

"With six to eight weeks of practice with an individual's voice, the software builds to 90 to 95 percent accuracy," Maddox says. Since the five members of her department share computers, each accesses his or her own user file when it's time for transcription. "We're now captains of our own ship. The pathologists have control over their own reports, which makes sense," she says. "We know better if it's correct. This is the best thing since mashed potatoes and peas."

Dr. Ranheim, an allergy and environmental specialist, was paying about \$9,000 a year for transcription. Now, at the end of each day, he sits at his desk, where he keeps his Pentium IV laptop, and dons the headset microphone, making sure the mouthpiece is no more than two-fingers' length away from his mouth. He dictates 12 to 15 pages of notes into patient charts via Dragon's medical software.

Subtracting an investment of \$4,000, which includes the computer, the software and training, Ranheim estimates that in the three years he's been using Dragon, he's saved at least \$23,000 in transcription costs, as well as time. "It used to take five to seven days to get notes back," he says. "I can now have them the minute I finish dictating."

To demonstrate how easy it is, he sits at his desk and powers up the program. Saying the patient name activates the chart. He reads from notes scribbled in consultation with the patient, a 40-year-old woman with severe environmentally induced allergies. Occasionally he adds voice commands, such as "Tab key" or "New paragraph." The screen nearly fills with notes before the first mistake is made.

"The highest being for wheat," comes out as "highest being four weeks."

Ranheim instructs the program to make a change: "Select 'four weeks.' Correct that. Select to train. Go. 'For wheat.' 'Four weeks.' Done."

In almost less time than it would take to back up and retype, the error is fixed. Ranheim says the words "select to train" so that the correction will become part of the

software's standing dictionary and he won't have to make this fix again.

Thanks in part to voice recognition's success in the medical field, educators are taking note of the up-and-coming technology as one that may one day permeate many industries.

High school business classes are catching on to the value of voice and speech recognition, says Julia Fallon, program developer for technical education for Washington state's Office of Superintendent of Public Instruction. "DigiTools," a pilot program tested in Washington between 2000 and 2001, added speech-recognition training as part of the model business curriculum. "We can't mandate curriculum choices, of course," Fallon says, "but we do feel students today need to learn speech and voice recognition to have a head start in the business world."

She doesn't know how many of the 235 districts in the state that have high schools are actually using DigiTools. She hopes all high schools will recognize its importance. "Speech recognition is definitely a technology that's evolving into everyday use."

Assistance

Another education-related use for voice recognition may be assistance for children with learning disabilities. Some students are cognitively quite competent, but they either have dysgraphia, a condition in which they struggle to express their thoughts in written language, or they struggle with spelling, says Jerry Connolly, director of the Special Education Technology Center at Central Washington University in Ellensburg. Many children with learning disabilities also struggle with reading. For these children, voice-recognition software, when complemented by text-to-speech software that allows the computer to read text out loud, can be beneficial. Essentially, the students use voice to put text in, and then they hear the result rather than read it on the screen.

A report from the National Center to Improve Practice in Special Education notes that the discrete speech requirement in the clunky Dragon Dictate that frustrated everyone in the early 1990s had an unforeseen advantage for many people with learning disabilities. Discrete speech slows the dictation process. Difficulties in pronunciation and in error monitoring are more easily handled at this slow pace. Continuous speech requires an ability to comprehend and monitor large blocks of text, a skill many with disabilities don't have.

In addition, voice recognition may work better for homework than in the classroom unless quiet corners can be found. "You need a quiet environment without background noise," says Margaret Poore, assistive technology specialist with Seattle Public Schools. "Ever notice what an elementary school classroom sounds like? Besides, students feel self-conscious talking to a machine out loud when other kids can hear."

Difficulty weeding out background noise is one of the two main limitations software makers still haven't worked out, says Revis from Nuance. The other is that software for desktop users is speaker-dependent. It trains to recognize one voice. While phone-response software can usually understand limited responses such as "yes," "no," numbers, dates and names regardless of who is speaking, the software used to create documents must respond to many more words and also must be able to handle many potential differences in the way something is said. Those requirements make it challenging to create a program that immediately recognizes the sentences of all speakers, although developers are striving for this.

Getting the software to weed out background noise may be an even bigger problem. Competing sounds may not only affect your document, but also your appliances, especially if you're using a hands-free system with a microphone versus a regular phone. For instance, after the system turns on the CD player, the music in the room might keep it from hearing your later command to turn the CD player off.

Kees Vander Putten remembers the time he asked HAL to turn on the lights and someone upstairs started a vacuum. That noise caused HAL to say, "Do you want me to turn the lights off?" Vander Putten repeated his request, and the second time, HAL heard it correctly.

Despite these kinds of challenges, voice-commanded computers are in mainstream use for people who are affected by a disability. Paul Jacobson, a San Diego resident and former marketer for a computer-hardware company, remembers well the day in November 2003 when he left work to drive to a dental appointment and ended up in a car accident. The ensuing collision injured Jacobson's spinal cord, making him a quadriplegic consigned to life in a motorized wheelchair.

"My biggest concern after the accident was not being able to communicate," Jacobson says. Two years ago, he installed Dragon on his computer at home. With

only 20 minutes' worth of training the software to recognize his voice, Jacobson could use voice commands to send and receive e-mail, shop online, browse the Internet, keep in touch with friends via instant messaging, and also go back to school.

"I'm not the kind of person to slow down," he says. Now 42 years old and otherwise in good health, Jacobson uses his voice to write papers as he works toward a master's degree in education from the University of San Diego.

Brian Peterson, a student at the University of Washington, tells a similar story. A quadriplegic as the result of a car accident in 2001, he uses Dragon to write papers and to keep in touch with friends. "The biggest thing with the accident is that it took away the use of my fingers," he says. "I was used to relying on my hands and a keyboard to do most of my communication. Now with my voice, I'm able to be independent again."

Janet Blackstone of Seattle, a former computer programmer, had to give up her line of work because of a repetitive-strain injury, which required surgery on her elbow. Voice recognition allowed her to go back to school, but Blackstone doesn't feel it's ideal for everyone.

"When you get to a certain pain level, you'll put up with the frustration. People who don't have that motivation give up before they get used to it," she says. "I can dictate using Dragon as fast as I could type, but I don't think mainstream people will use it anytime soon."

SOFTWARE DEVELOPERS are working hard to change that opinion. "We're focused on bringing speech recognition to all customers," says Chambers, the Microsoft software architect.

"Just as the mouse improved the effectiveness of the average PC user, voice control will one day take us to the next level by adding an even easier way to input messages and access information," says Rosenthal from Next Generation Technologies.

"People used to think the mouse was weird, too," comments Miller from Opus Research. "But now everyone uses it."

As with the mouse, so with the voice? Or will it just seem too weird to talk to machines? Only time will tell. Either way, voice control is no longer a science-fair project. It's real. It's here. Maybe one day we'll look back and ask ourselves—and our homes, cars, phones and computers—how we ever got by without it. ■

Scott Driscoll is a Seattle writer.