IDEAS

TECHNOLOGY

Voice-Recognition Systems Boom

Computers that respond to the spoken word are gaining in speed, accuracy, and applicability

By Simson L. Garfinkel
Staff writer of The Christian Science Monitor
BOSTON

PHONES you can dial simply by repeating someone's name and typewriters that type what you say to them may be commonplace by the middle of the next decade, thanks to recent developments in computerized speech-recognition technology.

Word processors that can listen to a person talk and produce a printed transcript have been on the market for almost a year. Although they are expensive and require the speaker to pause between each word, the machines have made dramatic differences in the lives of many people with physical disabilities and promise to save time and money in clerically intensive fields such as commodities trading and medical reporting.

"I [can] dictate a letter . . . and send it over a network to another computer and have it printed," says Frank Whitney, a computer programmer at the United States Department of Defense who is able to use only one finger. "If I were doing it using my finger, I would be halfway through the first paragraph," in the same amount of time, he says.

"I have spoken to half-a-dozen handicapped folks, paraplegics, and others who are using these systems," says Christopher R. Seelbach, an analyst at Probe Research, a market-research firm that follows the voice-processing industry. "For those who can afford the $10,000 to $15,000 for a system, it basically changes their lives."

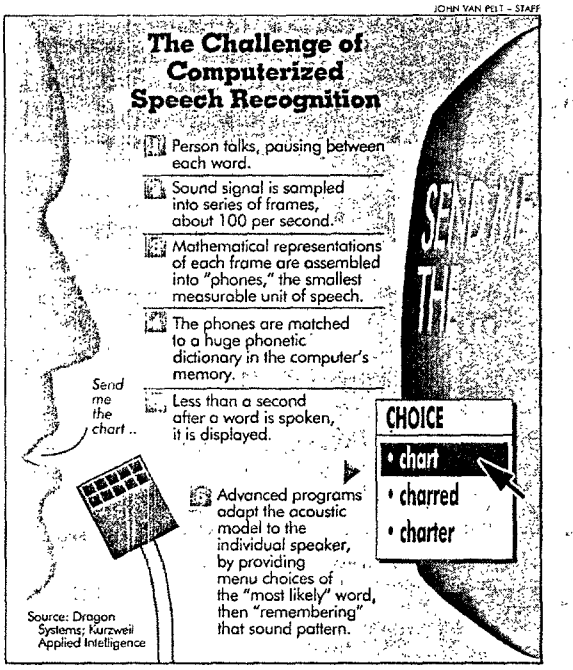
The concept is not new. Computers designed to recognize 50 to 100 spoken words have been around for nearly 15 years, says Janet M. Baker, president of Dragon Systems, a Boston-area company that sells voice-recognition equipment and software. "The early systems didn't work very well," often making mistakes, and they were unable to tell the difference between background noise and speech, says Dr. Baker.

By the mid-1980s, however, the accuracy of these small-vocabulary systems had improved. Companies started using them for inventory and quality control.

"Three years ago, Xerox Corporation was able to conduct a cost-effective, 100 percent audit of 2.2 million parts in two months [using such a system]," Baker says.

Small-vocabulary systems are speaker-dependent. They must be "trained" to recognize the user's voice in a 10-minute session, during which the computer flashes words on the screen and the user repeats them. Both Dragon and Kurzweil Applied Intelligence, another Boston-area firm, have recently developed large-vocabulary, speaker-independent systems that do not require training for each new user. Dragon sells a system that Baker says can recognize 30,000 spoken words. In March, Kurzweil plans to introduce a system for medical dictation that will recognize up to 10,000 words, says Vladimir Sejnoha, a research engineer with the company.

JOHN VAN PELT – STAFF



Although the specific recognition techniques employed by Dragon and Kurzweil are different, basic speech recognition involves converting sound picked up by a microphone into a series of acoustic frames or segments, each 1/100th of a second long. Each frame is analyzed and a set of mathematical constants representing tone and change in volume is extracted. The constants are in turn translated into "phones", the smallest distinctive element of spoken language. Silences between phones are used to signify breaks between the words. The phones are matched against a phonetic dictionary and then changed into standard English spelling.

The actual systems are much more complicated, Baker stresses. "Just doing a phone identification, and then doing a look-up on that, does not work. . . . You need to make use of many kinds of information simultaneously." For example, the software considers the context of the spoken word in the sentence to determine probability of a match against words in the phonetic dictionary. Such techniques also help the system decide between homonyms like "through" and "threw."

Less than a second after the word is spoken, the computer displays it on the screen. If the computer is not positive about its decision, it also displays a box of similar-sounding words; by saying "take-two" or "take-three," the speaker can correct the computer and substitute the second or third choice for the computer's first. Depending on the system, the speaker, and the background noise, the computer's first choice is correct anywhere from 80 to 95 percent of the time.

"One of the problems with speech recognition is the high degree of variability among individuals," says Mr. Sejnoha. Kurzweil's system features a special enrollment procedure in which a person speaks 400 representative words. These are used "to produce a model for how you say the rest of the vocabulary in the system," explains Bernard Bradstreet, the company's president.

Dragon Systems uses an adaptive algorithm or procedure in which the computer updates its internal model on the basis of each word that is correctly matched.

Although having to pause between each spoken word is a drawback, most people can dictate at 15 words per minute the first time they use the system, Baker says.

"My 12-year-old son did all of his social studies reports on it last year," she adds. "It was faster for him than typing. Most 12-year-olds are not skilled typists."

Neither are the "majority of people in the professional and business community," she says. As an added benefit, every word is perfectly spelled. "We think [widespread use of speech-recog-

MASTER VOICES

■ According to Christopher Seelbach of Probe Research, a market-research firm that follows the voice-processing industry, applications of speech recognition fall into six main categories:

1. Very-small-vocabulary recognition to get information by telephone. Such a system might use only 0 through 9 and "yes" and "no."
2. Dialing for cellular telephones. "[Car] phones offer a real hazard" when people dial and drive at the same time, says Mr. Seelbach.
3. Voice-activated locks, where a voice print is used to verify an individual's identity.
4. Small-vocabulary-recognition systems for hands-free data entry in quality control and inventory.
5. Large-vocabulary systems for transaction processing. "Financial trading is one particular market," Seelbach says. Shearson Lehman Hutton, a New York investment firm, is already using such a system.
6. Speech typewriters for creating letters, reports, and other documents. "There are real questions about how quickly that is going to develop," he says.

– S.L.G.

Talk Devices Train Air-Traffic Controllers

BOSTON

COMPUTERIZED aircraft simulators have saved dollars and lives by training pilots in a wide range of flight situations. But air-traffic controllers have not had similar success with computer-based training because of their profession's reliance on oral communication.

Although computers are used in air-traffic control training today, the systems require an operator called a pseudo-pilot to listen to the trainee's commands and tell the computer how to update the heading of each aircraft.

"That's manpower intensive," says Robert Wesson, a researcher who studies the application of artificial intelligence to air-traffic control. "You have to take a trained controller off the line and put him in the back room [to run the trainer]. They're already dramatically understaffed. That translates directly into time, dollars, and safety."

Because of developments in computerized speech processing, the operator can now be replaced with a voice-recognition system. "What you want is a stand-alone training device, where a controller can turn on the machine, strap on a headset, and go," says Dr. Wesson.

In 1988, Wesson was selling a $49.95 computer game that let a person step into the shoes of an air-traffic controller, route flights around the country, and deal with pilots.

"The controllers discovered it and commented, 'if you put voice recognition on the thing . . . it could be used as a professional-level training device,' " says Wesson. With a little work, he got a Dragon Systems speech recognizer to operate with his game.

Then Wesson International, Wesson's software development firm in Austin, Texas, was given a US Air Force Small Business Innovative Research Contract to develop the professional simulator. Ten systems have already been sold for research and evaluation.

– S.L.G.

nition technology] would dramatically improve the accessibility of computers and the information available through them."

Dictation systems for physicians based on both Dragon's and Kurzweil's technology were shown recently at the annual meeting of the Radiological Society of North America in Chicago a few weeks ago.

"This is the future," says Melvyn Conrad, a radiologist at the Nan Travis Memorial Hospital in Jacksonville, Texas. "One of the major advantages is you can see the words appear in front of you, so you don't have to review the words later." Most radiologists today record their reports on tape, and sometimes have to wait as long as a week to review the typed copies. In addition to the delay, "the secretary doesn't always type what you say," says Dr. Conrad.

Medical reporting has been one of the first applications "to really take off," because of the limited vocabulary and the tremendous clerical backlog, says Mr. Seelbach of Probe Research. Kurzweil has 137 medical dictation systems already in the field, says Mr. Bradstreet.

Nevertheless, researchers are working hard to develop systems that are more accurate and able to recognize speech without pauses between the words. In its Yorktown Heights research laboratory in upstate New York, IBM has developed a speech-recognition system that averages between 90 and 95 percent accuracy on discrete words. Even that, say some researchers, isn't good enough. "It puts too much burden on the user to have to correct every 10th word," says Lalit Bahl, manager of natural-language speech recognition at the laboratory.

Besides increased accuracy, says Dr. Bahl, speech-recognition systems need to be able to process continuous speech, so that users don't have to pause between each word. IBM has such a system, but "it's about 50 to 100 times slower than the real-time systems," Bahl says. Running on an IBM 3090 mainframe, it takes the computer more than an hour to analyze a minute-long recording of a person speaking in a normal voice.

"If you look toward the future, the area of speech recognition that we are working in will clearly lead to [a system with] a functionally unlimited vocabulary, complete speaker independence, and continuous speech recognition," says Kurzweil's Bradstreet.

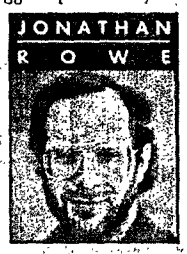
Voice-recognition researchers expect that it will be three to five years before desktop computers have enough power to do the job.

COMMENTARY

The Rolling Stones Come Home

THE Rolling Stones don't travel light anymore. They rehearsed for their recent and much-publicized tour at a lavish estate in Connecticut. On the road their tastes ran in the same direction.

At Shea Stadium in New York, for example, the Stones turned the baseball Mets' locker room into what the New York Times called a "luxury suite," with leather chairs and drapes. The office of Mets' manager Davey Johnson became Mick Jagger's private tea room, with "love seats, Oriental rugs, elaborate tables, lamps, mirrors, and a keyboard."



By one estimate, ticket sales alone from the tour brought the group more than $10 million per member, not counting sales of related paraphernalia, such as $500 leather motorcycle jackets with the title of the Stones' new album on the back.

On stage the Rolling Stones are still rapscallion and rowdy. But their act has become a kind of cultural lip-sync, like the canned backup that – according to one report – was dubbed into their show. The Stones aren't dangerous or even contrary these days. With their rich tastes and corporate sponsors – they're shilling for a beer company on TV – the group is absolutely mainstream, Reagan rock.

The tomcats have become house cats, and some fans are dismayed. Weren't the Stones, as one writer put it, "the embodiment of 1960s-style anti-establishment anti-materialism"?

No, they weren't. And there's nothing surprising about what they are now. It's where they've been headed all along.

What's called today "The Sixties" was really two totally different things. On the one hand was the idealism of the civil rights movement and the Peace Corps, and to some extent the campaign to end the Vietnam War.

This Sixties rejected the consumerist frenzy into which the nation had plunged after World War II. It organized the first Earth Day and sought to live in more self-reliant and environmentally conscious ways. (These efforts were ridiculed then, but given today's trash crisis, they seem prophetic.)

The other Sixties was the "Youth Culture": acid and rock, the Doors and the Stones. It wasn't so much anti-materialist as countermaterialist, preferring pot to martinis and tie-dyes to suits. Its main concern was "lifestyle": Leave me alone, let me enjoy.

The line between the two Sixties was never airtight. Bob Dylan, for example, bridged both camps, as did the antiwar movement. (The draft was a strong stimulant to geopolitical concern.) But the tendencies were as distinct as Eugene McCarthy and Timothy Leary, Mick Jagger and Joan Baez.

When the Sixties mentality petered out during the Carter presidency, it was the youth culture strand that proved most resilient; as Rolling Stone magazine perceived, the kids could become a market.

For all his preachy ineptness, Jimmy Carter himself was a Sixties idealist. Since leaving office, he has been fixing tenements for poor people and helping resolve international disputes – the Peace Corps and VISTA both at once. In office, President Carter tried to tell us that we couldn't have everything. Energy was short, the budget was out of balance, self-gratification was not the most important thing.

This sounded like Sunday School, and the nation didn't like it. So along came Ronald Reagan, who assured us there was no need for hardship, that we could just enjoy ourselves. It is understandable that many from the Woodstock generation – yuppies now – could feel so comfortable in the Reagan camp.

Reagan appealed to the same romantic individualism, the same yearning for self-gratification. Reaganite economics was the hedonism of the '60s drug culture, transferred to the economic realm.

It is precisely here, at the intersection between Woodstock and Wall Street, that the Rolling Stones come in. Most big commercial acts take their fans with them. Frank Sinatra, Elvis – their crowds became grayer as the years passed. The Stones, by contrast, are generational crossovers, gathering new fans as they go.

In the context of the times, this is not surprising. What better soundtrack for the Reagan years than "I Can't Get No Satisfaction"? What better symbol than the leather chairs in the locker room? The Stones haven't changed. They have come home.

Sentimental Memoir – With a Dash of Lemon

By Pamela Marsh

SIT down with a Rumer Godden novel and reality disappears in a most satisfying way. Or does it? It's clear from the two books she has written about her own life ("A Time to Dance, No Time to Weep" and "A House With Four Rooms") that her fiction has always been firmly rooted in her experience.

The first volume, "A Time to Dance, No Time to Weep," is mostly an account of the days of her childhood and of her first marriage. Set in India, it is permeated with her love of that country, a cruelly tried but enduring love.

This second volume opens with her arrival in 1945 England with two small daughters, little money, but abundant talent and determination. She finds homes, schools, friends, her place in the literary world, goes back to India for the filming of her novel "The River," does a stint in Hollywood, and dabbles in farming ("it's all petrol and killing").

Quoting from "A House With Four Rooms" is irresistible; it satisfies the reviewer's hunger for a companion who can be nudged in the ribs and begged to "listen to this, listen to this."

For instance, I like her description of a woman speaking Portuguese, ". . . it sounds like a little horse trotting," and her account of the birth of a pig, "a neat little pink parcel. Never could anything have been more perfectly packed; forelegs and backlegs were folded against the chest and stomach, ears folded back against the head, tail tucked in. Mrs. Pig gave the package a thwart; ears, tail, legs were shaken out. Next second a tiny complete piglet was standing upright. . ."

BOOKS

A HOUSE WITH FOUR ROOMS
by Rumer Godden
New York: William Morrow
320 pp., $18.95

There's a Christmas story too that cries out to be noticed at this time of year. Just after V-E Day, Rumer Godden invited two young German prisoners of war to supper on Christmas Day. Mrs. Phelps, her housekeeper, disapproved. Adamantly. She and her little boy, Patrick, would keep to their room:

"Josef and Hans arrived spruce and clean. They had carved, with their penknives, a little wooden angel for each child, including Patrick."

After supper, they sang carols. First Patrick stole in to listen, then his mother. Josef, due to return to Germany, had a son he had never seen, so Rumer Godden gave him some of her young nephew's outgrown clothes, "at which Josef burst into tears. He had had nothing to take back to his unknown boy. Mrs. Phelps went out. She came back and silently laid something on Josef's lap, something for which I knew she had saved her coupons and scant money for a long time, Patrick's new pair of shoes."

Rumer Godden's sister, friend, and fellow-writer, Jon Godden, called her stories a mixture of "children, animals, flowers, houses, a little sentimentality and piety all written and done as you only can do it."

Jon's assessment fits this memoir too, though mention should also be made of Rumer's gift for telling an absorbing story, for tapping exactly the right word

squarely on the head, and for adding just the right touch of lemon to cut the sentimentality – except of course when it comes to the Christmas spirit.

■ *Pamela Marsh is a former Monitor book editor.*