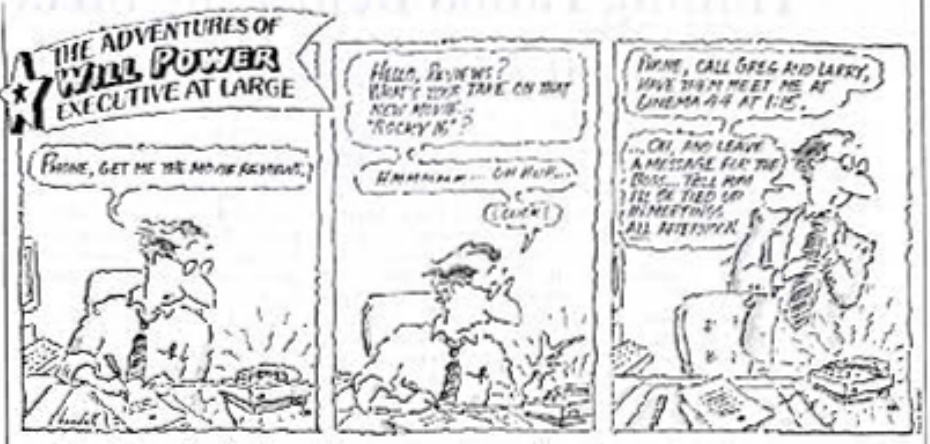
Speech Recognition Moves From Single Words to Smooth Talk

BY LOREN TALLEY

Murray Hill, N.J. — At the 1991 Telecom Conference, AT&T Bell Laboratories President Emeritus Ian Ross addressed AT&T's rapid advancements in speech recognition technology, and predicted that by 1995, some machines will have 20,000 word vocabularies — the ability to understand continuous speech. Today, Bishnu Atal and his colleagues in speech research have taken that prediction a step further.

"The focus of speech technology is shifting from just the recognition of words to understanding their meaning. That is what researchers are grappling with right now," says Atal, head of the Speech Processing Research Department here. Atal predicts that in the next five years, speech recognition will become a universally applied technology allowing people to speak almost as freely with computers as they would their best friends.



Not Just Research Anymore

Evolving from the laboratory to development, many AT&T products and services feature voice recognition and are on their way to achieving more natural interaction with people. Consumer Communications Services (CCS) is one of several business units committed to pushing speech recognition across its product line.

1987

Please say collect or calling card...

Voice recognizers spot isolated words and digits. Automated operator service is one example.

"Going forward, we see speech as a key differentiator in our services," says Carroll Creswell, technical manager of the Consumer Lab in Murray Hill. "People are beginning to use their voices to interact with a wide variety of devices, such as programming their VCRs (video cassette recorders) by voice. It's not universal yet, but we're getting there. The challenge is for the machine to learn about human contact and adapt to that."

Creswell's group is currently building and testing a prototype service for CCS that enables callers to speak their choice of applications from a main menu that includes messaging, voice dialing with names or numbers, movie reviews, and directory assistance.

A movie review service asks callers to say the name of a current movie and then reads the corresponding review from the Internet using text-to-speech synthesis.

In addition, users can create their own directories of automatically dialed numbers so when placing calls, they simply say the name of the person they want to contact.

This "label dialing" feature is incorporated in a related CCS offering called AT&T Voice Line, a speech-dialing service that uses AT&T Conversant® technology, and is accessed by a calling card. Voice Line is currently available to limited markets. Plans are to introduce the service to widespread markets later this year.

Speaking from a consumer's viewpoint, Creswell notes that people generally enjoy using today's voice recognition features in their telephones instead of the touch-tone or rotary dial, but they become frustrated with some of the mechanics, such as having to dial a number first to access voice dialing.

So why can't we simply pick up the phone and say, "Call Aunt Minnie?"

"You can do this now if you live where a local exchange carrier is testing voice dialing, or if you have speech recognition built into your terminal equipment," explains Creswell. The drawbacks, he points out, are that you can't usually use such systems when you're away from home, or when you're not using a speech-recognition based telephone.

According to Creswell, the opening of the local telephone market for competition and deployment of advanced intelligent network features should allow more "anytime, anywhere" speech-based services. For now, he adds, voice-activated features will continue to be integrated into AT&T's long-distance network.

1990

Did you say 'one' or 'fun?'

When one recognizer verify what you say, the first choice is not always correct, but the second choice might be.

Beyond One-Word Responses

Besides network applications, Global Business Communications Systems (GBCS) developers are fine-tuning products like Intuity™ Conversant and Intuity Audix® voice mail to make them more "voice-friendly" for the user.

"Our objective is to get systems to go from sequences of 20 questions to more open-ended requests," says Bob Perdue, technical manager, Core Technology Group, Voice Transactions Systems Department, Columbus, Ohio. Perdue explains that, to create natural sounding conversations between humans and computers, the system must recognize connected words and strings of important phrases.

Perdue gives as example of a future-generation Intuity Conversant system for banking: "Ultimately, if a person calls the bank and says something like, 'I'd like to transfer $100 from savings to checking,' the computer would pick out the key words and phrases — 'transfer', '$100' and the 'savings to checking'."

GBCS recently announced a similar feature called FlexWord™. It enables businesses to build their own speech recognition vocabularies of up to 2,000 words for Intuity Conversant menus. Perdue's group is working to expand FlexWord's vocabulary and grammar, so it can recognize multiple key words or phrases.

Expanded voice capability is also planned for a future version of Intuity Audix, which is expected to expand from voice mail to include multimedia applications. "You'll be able to say commands like 'play next message' and 'delete all messages.' That will make the service more user friendly and allow you to walk away from your keypad," says Perdue, adding that name dialing is also likely to be added.

1995

Where's Forrest Gump playing in Naperville?

Today's speech technology understands parts of fluent speech within a specific context like movies or airline reservations.

Intelligence in the Network

The notion of open-ended questions and natural response is what propels speech recognition technology from research trials to network services. Imagine calling an operator and hearing an automated voice ask, "How may I help you?" Services like collect calls, direct dialing and directory assistance would be at the tip of your tongue. Technology is headed in that direction, according to David Thomson, technical manager, Speech Processing Group at AT&T Network Systems (AT&T:NS). Indian Hill Developers at AT&T:NS are building switching equipment that incorporates speech recognition functions more and more.

"Ten years ago, we didn't have speech recognition anywhere in the AT&T network. We had vocabulary systems in the lab, but deployment was in the distance. Now, we're building dialogue systems and the technology is at a point where it will work for the customer," Thompson says.

An experiment recently completed by Thomson's department found that a large amount of what we say is decipherable using today's speech technology. The experiment involved a movie locator system that provided callers information about movies playing at local theaters. When users called the computerized service, they could ask questions like, "What science-fiction movies are playing in Naperville?" Collecting key words and phrases representing the movie category and city, the voice recognizer would process the appropriate response.

Thomson calls the process "concept spotting." "Since we weren't telling people what to say, we tried to build a recognizer that understood words likely to come up in conversation such as the name of the movie, town or time of day," he explains, adding that 90 percent of callers who used the service reported they received the information they wanted most or all of the time. That capability may be applied to the AT&T network, where customers can tell an automated operator who to call, when to make the call, and how to pay for it.

Concept spotting has proven to be a major step forward for Bell Labs efforts in achieving natural language understanding. Last year, Bell Labs researchers were among the leaders in a contest sponsored by the federal

continued on page 3

SPEECH

continued from page 2

government's Advanced Research Projects Agency in which they successfully created a voice-activated airline reservation system. People could casually ask questions like "What are the next flights leaving from uh, let's say, Philadelphia and Chicago?" The system responded correctly 91 percent of the time.

"It was a difficult test because in natural speech, people don't know what they're going to say right off the bat, so there is some speech that is grammatically incorrect," says David Roe, head of the Applied Speech Research Department here.

Roe explains researchers created statistical models for concepts — such as destination, city or time of day — and plugged in various ways of expressing those questions. The computer had to learn those concepts as people would speak their requests.

The Speech Research Department is currently working to build machines that figure out the meaning behind words and phrases. "While the airline system could not carry on a two-way dialogue with the caller, it did demonstrate that a computer can understand part of what a person is saying. But we ultimately want systems to learn semantics and be able to ask for more information," says Atal.

He observes that another big challenge for researchers is dealing with speech recognition in noisier environments. "When you talk to a person at a party, you focus on that person's speech even though there are others talking around you. We want speech recognizers to do the same thing. We've overcome the problem somewhat with acoustics, but it's not the final answer."

Roe predicts that by the end of the decade, you'll see systems deployed that will perform more complex functions such as airline reservations or directory assistance by voice. "We're at the brink of having computers understand us, and we're moving this capability into AT&T products and services." ▩