# Communicating *with* MACHINES

## Researcher presents his vision of the next generation of speech recognizers.

by JOHN TOON

*"Open the pod bay doors please, HAL."*
*"I'm sorry Dave, I'm afraid I can't do that."*

When the motion picture "2001: A Space Odyssey" opened in 1968, that conversation between a stranded astronaut and a malevolent computer named HAL seemed plausible for the year 2001 – then more than three decades in the future.

But as any user of today's automatic speech recognition technology can attest, that future hasn't quite arrived yet.

As a scientist at AT&T Bell Labs, B.H. "Fred" Juang helped create the current generation of speech recognition technology that routinely handles "operator-assisted" calls and a host of other simple tasks, including accessing credit card information. Proud of that pioneering work, Juang today is working to help create the next generation of speech technology – one that would facilitate natural communication between humans and machines.

Now a professor in the School of Electrical and Computer Engineering at the Georgia Institute of Technology, Juang presented his vision of next-generation speech systems at the annual meeting of the American Association for the Advancement of Science (AAAS) in February 2004.

"If we want to communicate with a machine as we would with a human, the basic assumptions underlying today's automated speech recognition systems are wrong," he says. "To have real human-



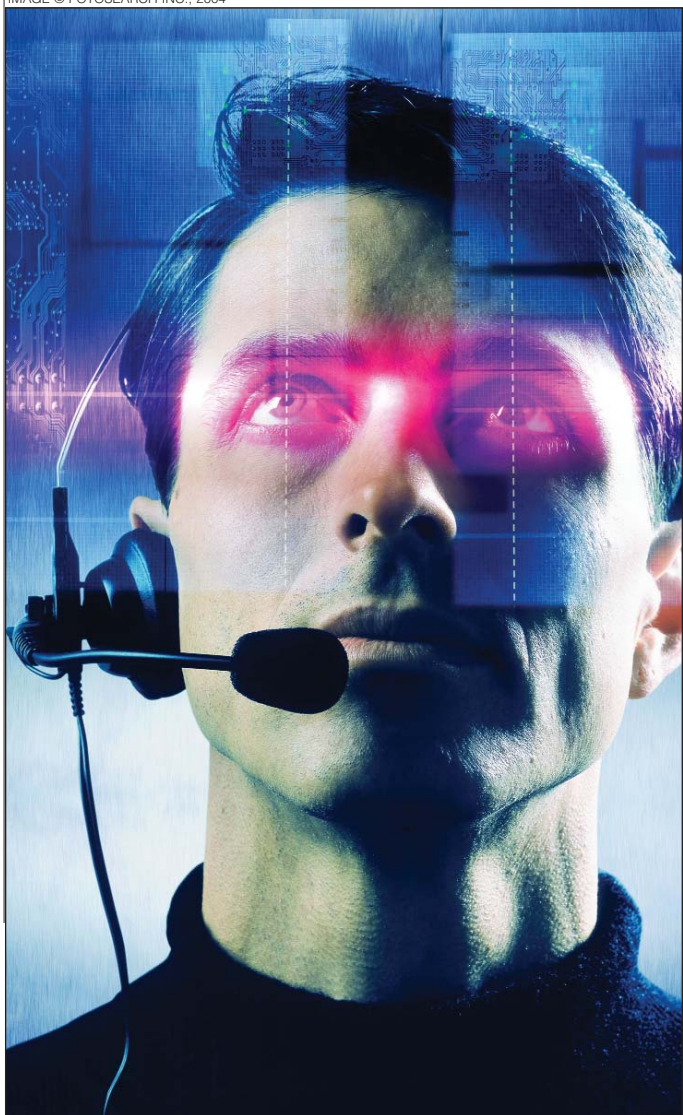

PHOTO BY BILLY HOWARD

**B.H. "Fred" Juang helped create the current generation of speech recognition technology that routinely handles "operator-assisted" calls. Today he is working to create the next generation of speech technology.**

machine communication, the machine must be able to detect the intention of the speaker by compiling all the linguistic cues in the acoustic wave. That's much more difficult than what the existing technology was designed to do: convert speech to text."

To make the speech recognition problem solvable in the 1970s, researchers made certain assumptions. For instance, they assumed that all the sounds coming to the recognizer would be human speech – from just one speaker. They also assumed the output would be text, and that recognizer algorithms could acceptably match speech signals to the "closest" word in a stored database.

But in the real world, human speech mixes with noise – which may include the speech of another person. Speaking pace varies, and people group words in unpredictable ways while peppering their conversations with "ums" and "ahs."

Speech researchers chose mathematical algorithms known as Hidden Markov Models to match sounds to words and place them into grammatical outlines. That system has performed well for simple tasks, but often produces errors that make the result of speech-to-text conversion difficult for humans to understand – and even worse for natural human-machine communication.

"It doesn't matter what you give the system, it just picks the closest sounding word and gives that to you as text," explains Juang, who holds the Motorola Foundation Chair at Georgia Tech and is a Georgia Research Alliance Eminent Scholar in Advanced Communications. "But that's quite wrong if you are interested in general communications. When you talk, a lot of information is lost if you use the current methods."

In addition, current machines cannot understand "reference," a linguistic shorthand people use to communicate. When discussing a technical issue such as electrical resistance, for instance, a group of engineers may use the word "it" in referring to Ohm's Law. Humans easily understand that, but machines don't.

"If every time we began to discuss one term, we had to define it, conversation would be very awkward," Juang notes. "Being able to understand reference is very important for natural communication. If we can create a system to do that, the machine would behave much more like a human and communicate more like a human."

The next generation of speech recognizers, he says, will have to go beyond conversion to text.

"Unlike the existing technology which gives you the closest word in a database, the new framework will consist of information detectors that provide information the machine can digest," he says. "This will involve a fusion of information, beyond the simple words."

And like humans, it will occasionally have to say "I don't understand" if it has doubts about what it's heard. Like humans, it will also be able to learn from its experiences to communicate better in the future.

The next generation of speech communications technology will require new mathematical algorithms that will go beyond the Hidden Markov Models. Researchers at university and corporate research labs worldwide have already begun working on the problem.

"We need to reformulate the problem in a different way and we will need some new mathematical tools to tackle the much broader problem of human-to-machine and machine-to-human speech," Juang says. "We are just at the beginning of developing this new paradigm, but I would say that we have perhaps 60 percent of the framework we need. There are some interesting steps and challenges ahead, but this is not an insurmountable problem."

Development of the new system will proceed in parallel at multiple institutions, each contributing its own skills and fitting them into the overall framework. Researchers will also benefit from new understanding of human cognition and linguistics that will allow machines to act more like humans.

Juang senses increasing agreement among researchers about the need to produce a new generation of speech communications able to do more than help route long-distance calls and accept credit card numbers.

"With a new system, we will be able to automate many things," he says. "We are now talking about realizing the original dream of automatic speech recognition."

■ Contact Fred Juang at 404-894-6618 or juang@ece.gatech.edu.

If we want to communicate with a machine as we would with a human, the basic assumptions underlying today's automated speech recognition systems are wrong.