

Talking the Talk

The Spoken Language Systems group strives for computerized voice-based systems that deal more naturally with human speech

It's not hard to find a computer to talk to these days: just call an airline to get arrival times, and you'll soon be navigating a menu of options by answering questions and reciting numbers. But stray too far outside the computer's repertory of stock phrases, and its fluency can fail dramatically.

The Spoken Language Systems group in the Computer Science and Artificial Intelligence Laboratory (CSAIL) is pushing the frontiers of language processing to allow people to talk naturally to computers and get helpful information in return. "There's still a big gap between what humans are able to do [with speech] and what machines are able to do, but the gap is closing," says Jim Glass, head of the group.

Consider the way an airline reservation system steers you toward words it recognizes. Are you interested in foreign or domestic flights? What city does the flight originate in? I think you said Boston; is that correct? "It's sort of a game of 20 questions the computer is playing with you," Glass says.

By contrast, the MIT group has developed several conversational interfaces that use natural spoken language. One of them, Mercury, is a prototype reservation system. When the computer asks how it can help, the caller could say, "I need a flight from Boston to San Francisco next Tuesday." The computer might then respond, "I have 30 flights leaving Boston and arriving in San Francisco on Tuesday. What time would you like to travel?"

Unlike the first system, in which the computer has simply to identify a word from a limited list, Mercury uses a variety of rules and statistical analyses not only to recognize individual words but to figure out their relation to one another. It also bases its interpretive decisions on previous exchanges. If the caller says, "What about Wednesday instead?" the system knows it's still looking for Boston to San Francisco flights next week.

Some fairly advanced speech recognition systems are already on the market, in which a computer asks an open-ended question such as "How may I help you?" and then routes a call based on the response. But Allen Sears, vice president of information systems at the nonprofit Corporation for

National Research Initiatives, says even those systems are fairly limited in the range of subject matter they can cover. "There's a whole lot of structure behind it," Sears says. "They're letting you take the lead and then mapping to that structure."

The work at MIT helps move speech recognition toward a more unstructured environment, where the number of possible responses keeps growing, Sears says. A system for providing logistics to the military, for instance, would be far more complex than one for determining whether a customer wants to dispute a bill or ask for repair service. Spoken Language Systems researchers are trying to enable such systems by learning how people actually talk to computers. The lab makes its prototypes available over toll-free numbers and analyzes data from the interactions that result.

Much of the work involves developing statistical models for speech understanding based on characteristics such as the order in which sounds arrive or the frequency with which they're uttered. The group has developed rules to accommodate the different sounds of male and female speech, different accents,

and background noise. They've even taught the software to figure out if someone is speaking Japanese or Mandarin Chinese and to respond appropriately.

Housed with the rest of CSAIL in the Stata Center, the Spoken Language Systems group consists of two dozen researchers, including faculty, staff scientists, graduate and undergraduate students, postdocs, and visitors. One alumnus, Mike Phillips, was a cofounder, in 1994, of SpeechWorks, a maker of voice recognition software that was later acquired by ScanSoft. The group collaborates with other CSAIL researchers to develop "smart rooms," in which computers interpret what users are doing both by listening to their speech and by watching their hand gestures.

The continuing challenge, Glass says, is to make the technology flexible enough that it can handle all the variations in human speech. "You shouldn't have to think about how to phrase what you want to say," he says. "You should just say it, and it's up to the technology to understand."

IN BRIEF

Name

Spoken Language Systems group

Director

Jim Glass

Contact

groups.csail.mit.edu/sls

Major projects

- » Airline reservation system
- » Weather information system
- » Restaurant guide
- » Smart rooms

Spoken Language Systems group

\$\$

