Recognizing the Melody of Speech

JULIA B. HIRSCHBERG

Professor of Computer Science

Anyone who has ever navigated an interactive voice recognition system to make a reservation or review a charge knows that anger and sarcasm change nothing. But one day they might, thanks to research by Professor Julia Hirschberg.

Hirschberg studies prosody, the intonation and melody of speech. Often, it conveys subtle differences in meaning. For example, "I like cats" may sound like a statement, but raising the pitch at the end turns it into a question.

"During deceptive speech, you experience emotions like fear if you think you'll be detected or elation if you're getting away with it. This shows up in the prosody of your speech. The best people at judging liars are criminals. Police were worse than average, and parole officers the worst of all, because they assume people are always lying," she said.

Hirschberg's goal is to teach computers to understand such subtle variations and reproduce them in natural sounding speech. This involves understanding how prosody changes under different circumstances.

"When I was at Bell Labs, we did lots of experiments that looked at people's speech, and tried to predict what words he or she would emphasize," Hirschberg related. "We looked at syntax, context, the part of speech being uttered – you use whatever information you have, and usually that's not a whole lot."

At Columbia, she has analyzed the prosody of charismatic and deceptive speech. "Much of the perception of charisma is not about what people say, but how they say it," she explained. "In English, charismatic speakers are very expressive, vary their pitch contour a lot, and speak more rapidly."

She has also conducted extensive experiments in which people either lied or told the truth. In these experiments, the speakers told the truth about 61 percent of the time. Her automated computer system labeled identified truth tellers and liars about 70 percent of the time. Humans got it right about 58 percent of the time, worse than if they had just guessed "truth" every time, Hirschberg said.

She is also working on teaching interactive voice response systems a technique called entrainment. This occurs when one speaker mirrors back the same vocabulary, pitch and speed as another. "People like people who entrain to them more than those who do not," Hirschberg said. "We want to teach computers to change their pitch, intensity, speaking rate and other factors to sound more like the user."

If that doesn't mollify the next generation of callers, at least the computer will recognize their anger when they express it.

B.A., Eckert College, 1968; Ph.D., Michigan, 1976; MSEE, Pennsylvania, 1982; Ph.D., 1985