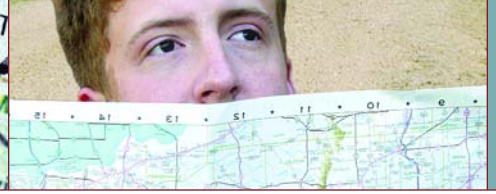
RESEARCH TO THE

rescue



Speech research helps a lost soul stay on course.

Anyone who's been lost late at night on the back roads of Mississippi knows that finding an isolated residence is not for the faint of heart. Rural roads and homes often are poorly marked, and even popular Internet map sites provide limited coverage of extremely remote locations. What's a lost soul to do?

Research at Mississippi State University may be coming to the rescue.

A team led by Joseph Picone, professor of electrical and computer engineering, is developing speech recognition technology for in-vehicle applications. Head of the Institute for Signal and Information Processing (ISIP) at MSU, Picone leads an effort to develop speech recognition software for applications ranging from workforce training to national security. He currently is collaborating with MSU's Center for Advanced Vehicular Systems to develop the in-vehicle dialogue system for automobile navigation in a project led by Julia A. Baca, a research professor at CAVS.

While a limited number of high-end vehicles now have some speech-recognition capabilities, the MSU goal is more far-reaching. The team's vision is a system that lets drivers have a "dialogue" with a computer to retrieve navigation information as needed. Picone predicts that within five years such systems will be significantly less expensive and much more flexible both in the type of information provided and in the free-form structure of the dialogue.

"Instead of being constrained to a fixed dialogue structure in which you answer a sequence of questions, you'll be able to ask 'How do I drive from the football stadium to the nearest McDonald's?'" he said.

"Conversational language often defies the rules of grammar," he noted. "While machines now are approaching human performance on very limited tasks, it's a daunting undertaking to achieve high-performance speech recognition."

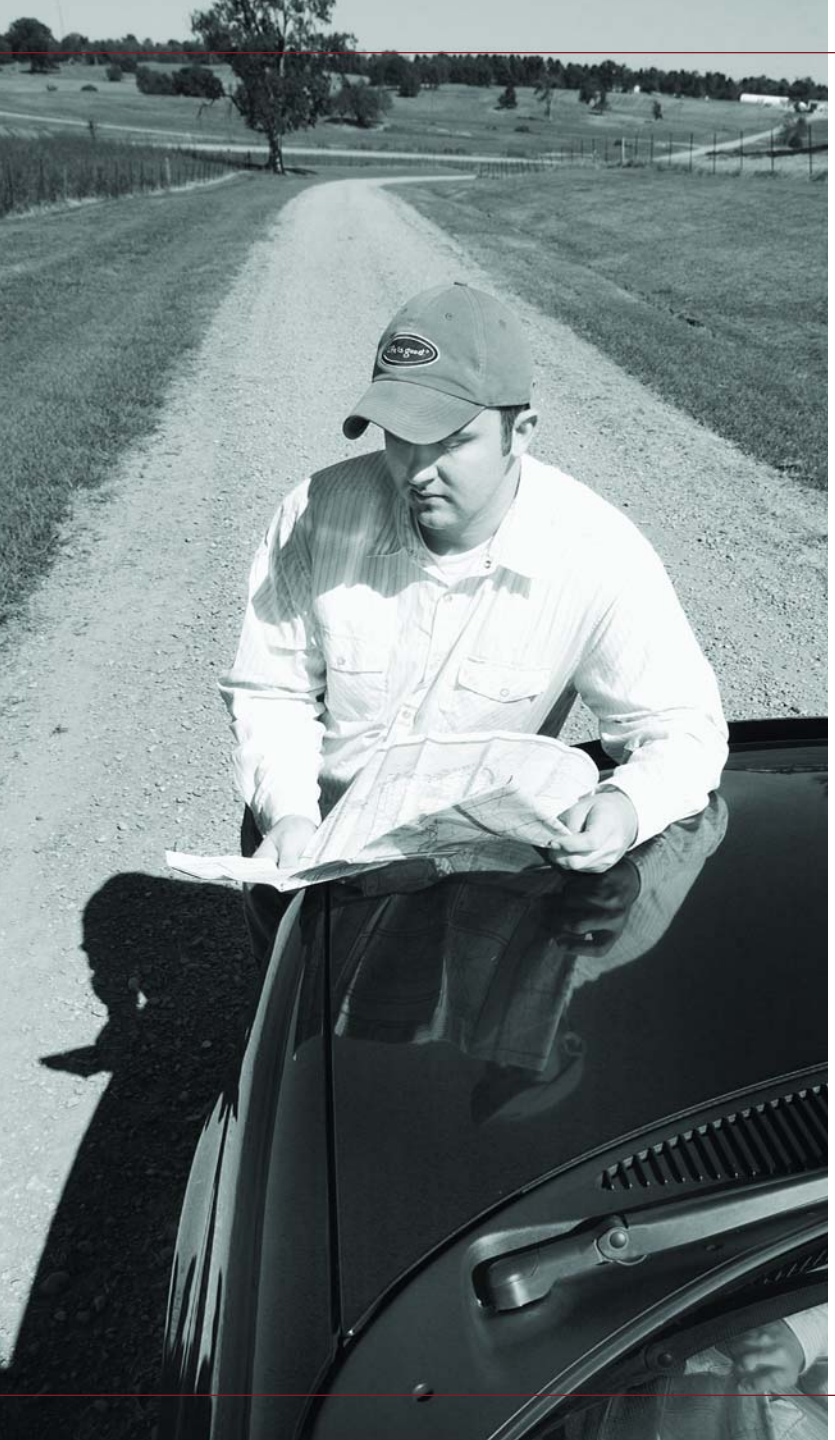
"For example, it is natural to ask a question: 'Are there any healthy food options available adjacent to gas stations located halfway between Jackson and Starkville?' Such requests would be extremely difficult to type into a standard Internet search engine," he explained. He adds that such dialogues are complicated by the fact that in certain applications, such as automobiles, users have a limited display area and can take their eyes off the road for only brief periods to manipulate a computer screen. "Voice input and output becomes a natural and enabling technology," Picone said.

Accurately interpreting complex nuances such as pronunciation, context and emotional tone, as well as filtering out background noises, presents difficult challenges for researchers. Current approaches involve complex statistical models that estimate the average values of measurements performed by electronics that convert the analog sound pressure wave generated by the human voice to a digital representation. The statistical models are trained in such a way that the values of the parameters are continually adjusted until the models predict—with the highest likelihood possible—the words that have been spoken.

"This approach was pioneered in speech recognition research, but now is finding its way into many areas of science, including the life sciences, where it's being used for applications such as genome sequences," Picone said.

But the approach has difficulties. "Because the representations are invariably based on average behavior, there are problems with the richness and variety of dialects in the United States, particularly Southern-accented speech such as that spoken in Mississippi," Picone said.

Despite the fact that it sounds like science fiction, speech recognition technology has practical, compelling uses.



ISIP is pioneering a new approach that draws heavily on the classical concept of risk minimization and goes beyond today's standard modeling techniques.

Consider, for instance, the problems investment counselors face in suggesting an individual's optimal investment strategy. The professional must take into account a client's willingness to balance risk and return. Because individuals have different tolerances for risk, the optimal individual strategy must vary accordingly.

"Investment decisions thus are based not just on daily stock values, but many intangible factors such as risk aversion, a factor we refer to as prior information," Picone said.

As illustration, he cites a hypothetical problem developed by the University of Cambridge's D.J. MacKay. "Consider the observation that tossing a coin 250 times resulted in 140 heads and 110 tails. One would quickly conclude that the coin is biased, since over such a large number of trials one would expect a more balanced distribution. What is the probability that such an event could happen? What are the chances the coin is biased based on this one set of experimental data?" Picone asks.

"If one assumes that the only possible prior evidence is that the coin is fair, then this outcome would have a fairly small probability of occurring. However, if one averages over all possible

prior distributions for the coin, MacKay shows one can conclude the evidence from this experiment points (weakly) in favor of an unbiased coin. Hence," Picone concludes, "depending on what prior assumptions you make about the data, the inferences you draw from the data can vary wildly."

Picone's group is extending this analysis to the speech recognition problem. In ISIP's approach, parameters are evaluated for their impact on the overall solution to the problem.

In this way, if a parameter of the model is not relevant to the problem, it automatically will be discarded, an approach known as automatic relevance determination.

"Our first attempts at this approach have resulted in speech recognition systems that use more than an order of magnitude fewer computing resources to reach a solution," Picone said. "Significantly, such systems could now be realized on much smaller amounts of hardware, making the technology more cost effective and pervasive."

An added benefit of systems based on risk minimization, he said, is that they directly learn to discriminate. Current technology learns to represent the average manner in which a sound is articulated, and decides which words were spoken by choosing the nearest average value (in a probabilistic sense). Often, accented speakers perform poorly because their pronunciations fall too far from this median behavior.

Risk minimization approaches directly model differences between sounds, learning to classify sounds based on a decision surface that separates these sounds in a mathematically-abstract space used to represent the measurements of the incoming signal. "It is our hope that such an approach will result in systems that are able to solve new problems without being exposed to examples of the data, thereby significantly reducing the cost to develop new applications," Picone explained.

ISIP has been developing software that puts these ideas into practice. Since Picone joined Mississippi State in 1994, ISIP has developed and released the first state-of-the-art public domain speech recognition system. ISIP maintains one of the most extensive Web sites in the industry, with public access available at www.isip.msstate.edu/projects/speech. More than 175 sites around the world have used the software, with others joining daily.

Despite the fact that it sounds like science fiction, speech recognition technology has practical, compelling uses, Picone explained. "It has military applications for antiterrorism and homeland security, as well as industrial and educational

"This is an excellent example of bringing research to bear on economic development issues with direct impact to Mississippi."

applications," he said. "Homeland security is an area in which we expect to see a significant increase in research activity at MSU over the next few years."

Picone credits a dedicated team of students for helping make the successes possible. "Our students have spent many hours developing software that is functional, well-designed, well-documented, and easy to use. In the process, they gain valuable skills in areas such as software engineering, which makes them extremely marketable after graduation."

One problem the MSU team is tackling has direct implications for Mississippi, he noted.

"We're very early in our research of using interactive dialogue to promote workforce training on an automotive production line," he said. Conducted under the auspices of the Center for Advanced Vehicular Systems, the project has applications for Nissan as well as other automotive manufacturers.

"Because cars are so customized, we envision workers getting instructions about a particular vehicle in real time, with the ability to ask questions and get responses," he explained. Especially in a work environment in which assembly personnel often are rotated, the system would increase efficiency and safety, Picone said, adding, "It's a challenging research project, because of the noise on the production line.

"This is an excellent example of bringing research to bear on economic development issues with direct impact in Mississippi," Picone continued. "It also is an example of collaboration among several MSU departments—electrical and computer engineering, industrial engineering and computer science—to build powerful and complex systems."

For more information about the speech recognition research or about the Institute for Signal and Information Processing, telephone Picone at 662-325-3149 or visit www.isip.msstate.edu.