



Gary Hallgren

By JOHN R. QUAIN
Published: May 20, 2007

May 20, 2007
TECHNOLOGY

Voice Recognition Closes In on HAL

By JOHN R. QUAIN

NOT long after computers settled into cars to impose their electronic authority on engines, transmissions and safety systems, microprocessors began speaking up, nagging passengers to buckle the seat belts and issuing alerts about doors that were ajar.

It wasn't long before revelation struck: what we really wanted was not cars that talked, but listened.

Decades later, voice recognition systems are a staple of the luxury class, an alternative to buttons, joysticks and touch-screens that control onboard entertainment systems and satellite navigation devices. Voice recognition is, in theory at least, a relief from the human-machine interfaces that most drivers can only rate as frustrating.

This HAL 9000 fantasy is behind the development of voice recognition programs that enable people to converse with (or at least have their spoken commands understood by) machines. Already, desktop software like Dragon NaturallySpeaking from Nuance Communications and ViaVoice from I.B.M. have made it possible for a PC to take dictation. But so far automobiles have been limited to basic commands.

In part, speech recognition in cars has been held back by the limited amount of onboard computer memory and the processing demands of the software. Large vocabularies need to be stored on hard disk drives and processors have to work with complex language models — the combination of linguistics and statistics — to interpret what is being said.

Today, navigation systems in luxury cars like the Acura MDX can respond to perfunctory commands, like “Find the nearest gas station.”

“Voice destination entry, or V.D.E., has been a big thing these days,” John Watts, manager for product planning at Honda’s Acura division, said. Indeed, Acura’s navigation system, with voice recognition derived from I.B.M.’s ViaVoice software, can even perform concierge-in-a-car tricks like finding an address based on a phone number dictated by the driver.

While such systems do not need to be trained like the dictation programs used on desktop PCs, drivers often find they have to train themselves to use the right words in the right context. Asking for “Traffic, please” at the wrong time can flummox most systems. Did you want music from the ’70s supergroup or information about gridlock on your route?

As costs have come down and the complexity of onboard computers has ramped up, the computing limitations are largely fading away, Roberto Sicconi, director for speech technology integration at I.B.M. Research in Yorktown Heights, N.Y., said. The voice recognition software I.B.M. has developed for cars, for example, “is now as powerful or more powerful than ViaVoice was on a desktop computer” just six years ago, he said.

So why is a car not able to tell the difference between a request for a tune by Steve Winwood's former band and the need to avoid congestion on Interstate 95?

The main problem in seamlessly switching between systems — entertainment and navigation, for instance — using conversational speech has been getting the car to understand the context of what the driver is saying, according to Tom Freeman, co-founder of VoiceBox Technologies of Bellevue, Wash. VoiceBox has been working on what it calls a conversational speech engine for cars that should be ready in the next year or so, he said.

“Having consumers memorize 1,000 different commands for the right context can be distracting, even deadly” if there is confusion when you are driving 60 miles an hour, he said. So VoiceBox developed specialized software programs that overlay existing speech recognition engines like those from I.B.M. and Nuance. These digital agents are designed specifically to interpret the driver's intention and determine the correct context for a given request.

In a recent demonstration — driving around Manhattan with a prototype portable VoiceBox system — the advantage of such sophisticated software was immediately apparent. Simply saying, “I'd like to listen to jazz and blues,” turned on the XM satellite radio and tuned in the correct station while we hunted for a parking spot. Asking the system “Who's playing?” brought a response that [B.B. King](#) was the artist.

After replying to a series of questions about current basketball scores, in response to, “What about Boston?” the VoiceBox computer also correctly surmised that we wanted to know about the Celtics and not the rock band or traffic in the city. And quickly changing topics from “What's the traffic like on I-95?” to “Turn it down!” did not confuse the system.

To develop its integrated voice recognition capabilities, VoiceBox has been working with companies like I.B.M., Toyota and XM radio.

According to Mr. Freeman, the company has developed different versions of its software tailored to a variety of existing in-car systems. The first products using the software, to go on sale late this year, will be portable navigation units, according to the company, followed by factory-installed car systems next year.

Still, even such advanced voice recognition products face challenges. While constant sounds like road or wind noise no longer present much of a problem, sharp random noise can throw a speech recognition program off course. It's not just a matter of keeping the kids quiet in the back seat. The thump of a closing window or the squeak of a windshield wiper can drive a computer to distraction.

That is why voice recognition systems must be calibrated for particular cars, according to VoiceBox and Acura. Microphones, for example, must be positioned to accurately pick up the driver's voice, but be placed as far away as possible from noisy equipment like fans and turn signals.

Researchers at I.B.M. are going a step further, adding interior video cameras to improve the accuracy of speech recognition systems. Using specially developed software, the camera can locate a driver's face and lips to tell when he or she is talking. "It can also do some crude form of lip reading to complement the speech recognizer," said Dr. Sicconi of I.B.M. "Different consonants may sound similar but look different in a visual domain," he explained.

In the future, such cameras may not only eliminate miscues but also eliminate the need for the driver to push a button to talk, as one has to do with VoiceBox's system. I.B.M.'s prototype conceals the camera behind the rearview mirror and can follow the driver even in the dark.

In addition, a passenger's position could be constantly monitored by the camera. This would make it possible to determine if someone sitting in the front seat was leaning against the door and decide whether it would be safe to deploy a side air bag in an accident. Or, by judging a passenger's height, cameras could help reduce injuries from air bag deployments. Lane departure systems could also use cameras to

determine if the driver's head position indicated that he intended to change lanes or had become dangerously drowsy.

In addition, putting a camera in a car means that automakers could potentially add face and speaker identification software. So every time your teenager gets behind the wheel, the vehicle would automatically adjust the seat, pedals and mirrors without the driver saying a word. The same software could refuse to start the car should someone with an unfriendly face try to drive your vehicle without permission.

I.B.M. says such technology is available today, but cost and complexity will keep it from being used in production models anytime soon.

"The challenge is making a cost-effective system," said Paul Liao, chief technology officer for Panasonic in North America. Panasonic, which has its own research program dedicated to speech recognition, says there is a significant tradeoff between cost and performance. "You may have 90 percent accuracy with speech recognition," said Dr. Liao, "but when you push a button, you get 100 percent accuracy."

Mr. Freeman of VoiceBox said more work and computing power were needed to achieve the fantasy of HAL, the talking computer in "2001: A Space Odyssey."

"But perhaps in five years," he said, rather than having to read a manual the size of the Encyclopaedia Britannica to understand a car's features, "you'll just ask the car what it can do for you."