

F32

Talk to me

Artificial intelligence and voice recognition will change how we design everything.

by Alan S. Brown



Gregory Abowd owns one of the first Tesla cars, built before they were capable of autonomous driving. Abowd may not have buyer's remorse, but since he is a distinguished professor at the Georgia Institute of Technology's School of Interactive Computing and an expert in human-computer interfaces, he's been giving serious thought to how he wants his next Tesla—one that presumably will be able to drive on its own—to handle.

"One other thing I learned when I took an autonomous Tesla for a drive—I would like it to mimic my way of braking," Abowd said.

"Its braking style is much too abrupt for me."

Today, teaching a smart car such tricks might require some serious programming, or perhaps a lengthy tour through multiple app screens and drop-down menus. But Abowd has a different vision. Within a few short years, he believes, we will be able to talk with our cars and tell them what we want them to do. Their voice recognition systems will not only translate our words, but apply artificial intelligence to understand our intentions as well.

Most of us may never learn how to program a car's braking performance, but soon we may have a simple way to reach deep into the heart of its control system and customize its behavior. Indeed, we may have the power to program any smart device in our homes, offices, and factories in ways that were previously impractical or impossible for all but the most sophisticated technophile.

That sounds radical, but new technology has been simplifying interfaces for decades. In the 1980s, personal computers transitioned from command lines to graphical interfaces that we could access by clicking a button on a mouse. Less than 10 years ago, the iPhone's touchscreen and accelerometers revolutionized how we operated handheld devices.

Voice recognition, its proponents argue, has the same potential to change what we expect from the everyday products. As those products grow smarter and more capable, voice promises to simplify how we communicate with smart cars, smart

homes, smart offices, and smart factories. Instead of mastering one new app after another, voice could make it simpler to command them all.

Incorporating voice interfaces will transform product design.

"The job of the mechanical engineer will be to harness those capabilities," said Henry Lieberman, a pioneer of human-computer interaction at MIT's Media Lab. "People want to have to understand the details of how things work.

"Language will become a means—not to help users understand a product more easily, but to have the product understand its users."

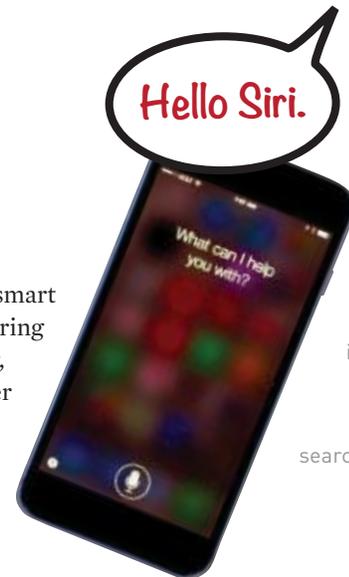
Machines that hear us

Anyone who hung up in frustration on voice-activated virtual assistants such as Apple's Siri or on voice-driven customer service centers and never went back has missed the advances in voice recognition. Today it is fast, accurate, and smart enough to understand everyday speech—and consumers are increasingly taking to it. Two years ago, spotty performance discouraged most people from using speech to run Google searches on their phones. This year, 20 percent of queries handled by Android phones were spoken, according to Google. That's 20 billion spoken queries daily.

Voice recognition is also expanding its beachhead in physical products. Many new cars use voice to place calls, set the GPS, write and receive texts, change radio stations, and adjust the temperature. The Eurofighter Typhoon military jet has a speech recognition system capable of controlling communications and allowing pilots to assign targets.

This is only the start, Lieberman said. Speech is not only convenient, but also much richer than typing or flicking an app.

"Think about it," Lieberman said. "We only speak to other human beings. So when we speak



Voice interfaces already handle billions of search queries each day.



GE appliances are among hundreds of devices that have the ability to communicate with Amazon's Alexa.

to a computer, we treat it as another human being. It's like talking to a dog. You know it doesn't really understand you, but you express yourself as if it does. That's the synergy you get from voice recognition that you don't get from typing."

The real sea change won't come from products responding to clearly enunciated commands. Rather, it will happen when they wade through the torrent of half-finished sentences, parenthetical remarks, and place-holding "ums"—and figure out what we really mean.

Artificial intelligence connected to the Internet makes that possible. The machine learning software behind voice recognition analyzes data from actual interactions to improve its performance. By analyzing the words used in searches, for instance, voice systems know which words are likely to go together, and those inferred relationships help them make sense of complex sentences.

In the connected world, machine learning software can draw on a billion interactions a day. That learning shows. Voice recognition can now easily navigate accents or pick out a single voice in a crowded room. Most voice systems are more than 97 percent accurate in identifying individual words. And while virtual assistants may not "know" the meaning of our words, their ability to link words helps them figure out what we want.

Not only is voice recognition more capable, it is also easier for engineers to use. There are many

large vendors—Amazon, Apple, Google, Microsoft, Nuance, and Baidu—and several offer free software to developers. And semiconductor firms such as ARM Holdings, Intel, and Sensory have introduced new chips optimized for voice. These chips provide fast, reliable voice recognition, even when devices are not linked to the Internet.

Nouns, verbs, and beyond

The graphical interfaces we've used for a generation make it straightforward for systems to interpret the commands they receive. A touchscreen may have clearly marked buttons for each input, or specialized apps access different operations. That clarity makes it easy for a device to understand what a user wants.

With voice recognition, the same input is used for initiating everything, from setting a thermostat to making a phone call. An always-on virtual assistant in a device that sits on a kitchen counter or desk, such as Amazon's Alexa, must field seemingly random requests and figure out whether to access a grocery list or a music library when someone asks for "some Red Hot Chili Peppers."

Vendors that want to use Alexa's voice interface to control their products must first bridge this gap. Wink is one company that has done this. It makes hubs that work with a broad range of home automation products from many different vendors,

each with its own capabilities and commands.

Wink brings order to this profusion of interfaces by creating a common model for each class of product, Matt Bornski, Wink's chief architect of enterprise services, said. Its lighting model, for example, supports every feature found in smart lightbulbs, from simple actions like "turn off" or "dim" to less common ones, such as "change colors." Each light uses a subset of these commands.

The common interface also makes it easier to link different devices with Alexa. Bornski does this by creating a framework, or domain, for each common model. The domain relates the words we might use to the actions a product can take. This enables Alexa to understand what we mean when we talk to our lights.

Wink has been so successful with its common model approach that it recently signed a deal to link the Alexa home automation system with Ford's voice-activated car consoles. The resulting system will let customers check the gas tank before the morning commute or turn on their porch light from the car.

Creating voice interfaces requires building in safeguards that might not be obvious to those used to tangible controls. For example, Alexa will activate but not disarm a security system. "You don't want a burglar to yell 'Turn off the alarm' through the back window," Bornski said.

The system also needs to anticipate that it won't work perfectly, given the limits of the equipment and requests from fallible humans.

"If I tell one light to turn red and it can't, I'll get an error message," Bornski explained. "But if I tell all my lights to turn red and only some of them can do it, I would feel frustrated if I got an error message. So our system does what a human would do, and changes all lights that accept the command."

Other companies are designing voice interfaces that take into account that speech conveys not just nouns and verbs—but also emotion.

IBM, for example, infers the emotional content of words by using its Watson deep learning technology, said Rama Akkiraju, a distinguished engineer at IBM Research in Almaden, Calif. And



You don't want a burglar to yell "Turn off the alarm" through the back window.

— Matt Bornski, Chief architect of enterprise services, Wink

IPsoft's Amelia "cognitive assistant" can tell when customers are losing patience with automated transactions and call for a live agent.

IPsoft got its start developing "virtual engineers" to automate routine IT tasks. Still, it takes experts to use the virtual engineers. Amelia uses voice recognition so anyone can ask these engineers for help.

"I can tell Amelia I want to install a new speakerphone in a conference room," said Jonathan Crane, IPsoft's chief commercial officer. "Amelia will check if the room can support the phone, whether the phone is available, and if I have the authority to order it. It fills out all the paperwork. Instead of me speaking IT, I can speak to Amelia in English and it just does it."

Such performance impressed two global consulting firms, Accenture and Deloitte. They recently signed deals to use Amelia to automate business processes and IT center engineering and

Instead of me speaking IT,
I can speak to Amelia in English
and it just does it.

— Jonathan Crane, Chief Commercial
Officer, IPsoft.

Amelia's avatar
as it appears to
users.



administration. Marc Carrel-Billiard, Accenture's global managing director for technology R&D, believes Amelia can help technicians maintain products. He points to air-conditioning repair as an example.

"We could feed a user guide into Amelia so she understands how it works. Instead of looking for information in a manual or on a tablet, a technician could explain what he or she tried and Amelia would give advice like, 'If you did this and it didn't work, try that.' Over time, Amelia would learn more about how the system worked, and one day might apply what it learned about one model of air conditioner to another."

Meanwhile, a few manufacturers have approached Crane about capturing the hard-won knowledge of an experienced but aging workforce. Amelia, Crane said, could act like an intelligent apprentice. It could look over a technician's shoulder, recording and transcribing explanations and abstracting it for later analysis.

"These conversations are giving us a strong sense of how we might solve these problems," Crane said.

Working together

Other groups are harnessing voice recognition and artificial intelligence to forge new models for human-machine collaboration.

Companies like Rethink Robotics and Universal Robotics already make collaborative robots. While they learn new tasks easily, they cannot really change collaboration strategies on the fly. But the collaborative robot built at Georgia Tech by doc-

toral student Crystal Chao, now with Google, and her advisor, Andrea Thomaz, now a professor at University of Texas, adjusts to its human partners by simply talking with—and listening to—them.

To show how this works, Chao and Thomaz created a task: building a Lego tower. They outfitted the robot not only with mechanical hands and vision sensors, but also with microphones and speakers. Then they gave the robot and its human partner different goals.

"We might tell the robot to use a red door and the human to make the tower six blocks high," Thomaz said.

Sometimes, the robot followed the human's lead, placing like-colored blocks the way one child might copy another. Other times, rather than wait for a command, the robot took the initiative. It might, for example, simply add the red door or ask if the color was okay.

The conversation flowed naturally. The robot reacted to human commands, and also to half-formed phrases, laughter, and verbal shortcuts like "uh-huh" or "uh-uh" that humans take for granted. Sometimes, the robot even interrupted with a suggestion or a question.

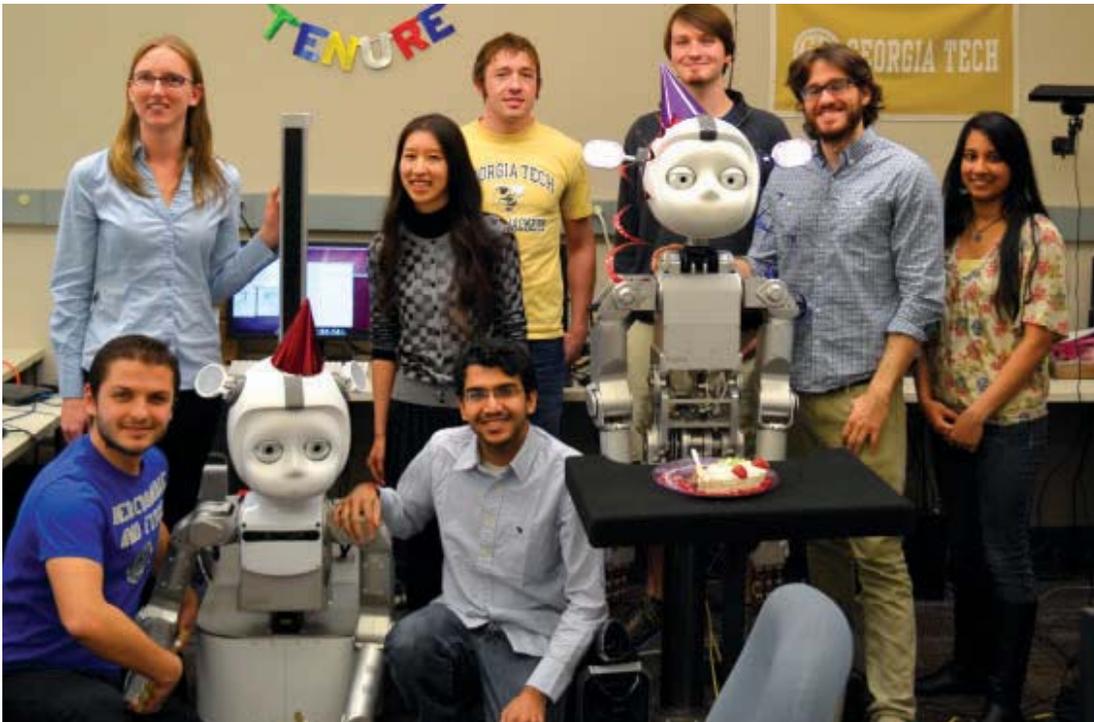
The interactions looked very much like the way humans collaborate with one another.

"In this type of collaborative dialogue, we're not leaning anything, we're just substantiating what we already know," Thomaz said.

The results were far from perfect. Humans are much better than robots at inferring what a partner is trying to do, and to reacting to dialogue that is outside the domain created by the robot's developers. Still, this robot's flexibility is anything but robotic.

It is a glimpse of how AI-driven voice recognition might soon change the way we work with machines.

Clearly, voice recognition has a way to go. It still gets simple searches wrong, and nobody is about to use it to control sophisticated machinery. But remember, this is a self-correcting technology that learns from every mistake. It will only get better and better.



Andrea Thomaz and Crystal Chao (top left) along with research students at Georgia Institute of Technology's School of Interactive Computing in 2013.

By coupling natural language requests to the deepest workings of the operating system, we may soon have new types of products that will give anyone access to features that only a professional could manipulate today. Instead of pouring through a manual to find the proper technique for an in-camera effect, one could simply tell the camera, "Focus on the faces, and make the background blurry," and the system would produce the image. A microwave would ask you what you were cooking and then apply a sequence of power cycles to crisp it to perfection.

Or the autonomous driving system of a Tesla could respond to the critiques of Georgia Tech's

Abowd and adjust its brakes—or cornering performance or acceleration—to his liking.

It is certainly not hard to imagine technicians working with flexible robots capable of reacting to their motions and commands on the fly. More powerfully, systems may one day provide advice to engineers looking to boost factory performance, or help designers work through difficult problems when they are not sure how to explain what they are want.

Language is a rich enough medium to do all that. And so much more. **ME**

ALAN S. BROWN is associate editor at *Mechanical Engineering*.

The Grammar of Privacy

Google and other online companies know a great deal about those of us who use their search engine, email, maps, calendar, news, travel sites, and other services. They know where we live, what we buy, where we travel, which subjects interest us, and much, much more.

Voice recognition systems will know even more.

"Voice allows us to tap into far more detailed analytics than text," said Sam Vasisht, chief marketing officer for MindMeld, which builds natural language systems for consumer websites. "The way we speak can reveal if we are happy, sad, or angry, and whether we're just looking or interested in buying. It can show our gender and age, and our

economic and educational status."

Voice systems will analyze this data to understand us better, so they can have more meaningful dialogues with us, said Werner Goertz, research director for personal technologies at global IT consultant Gartner. They will also use it to target advertisements at us.

"The big driver for voice recognition is its commercial use," Goertz said. "The more they understand our preferences, behavior, and proclivity to make buying decisions, the better they can predict ways to maximize the value of the ads they show us."

A world of always-on virtual agents might

feel even more intrusive, Dan Roberts, CEO of Scout Alarm, which makes Alexa-connected home automation systems, said. Alexa, he notes, is always on, but right now, it listens only for the keywords in commands and ignores everything else.

"When we talk about Alexa becoming more proactive, it will be listening for more than those keywords," Roberts said. "This will bring privacy issues to the fore. When we install any device that is always on and connected to the cloud, it raises privacy concerns for some consumers. Not everyone is going to want to stream a picture of their home." **ME**