

# ALL EARS

**With the increasing popularity of intelligent personal assistants such as Alexa, we're becoming accustomed to talking to machines – and this is just the beginning. Matt Burgess reports on the growing capabilities of voice-recognition technology**

Victor Collins' body was found face down in his friend James Bates' hot tub in Benton County, Arkansas one morning in 2015. The previous evening, the 47-year-old former police officer had gone to Bates' house to watch football, and Bates maintained he'd left Collins in the tub when he went to bed. His death must have been an accident, he said. But the local police charged Bates with murder, alleging Collins had been strangled.

A possible key 'witness' in the case was a nine-inch tall cylinder of metal going by the name of Alexa. It turned out that music had been playing on Bates' Amazon Echo speaker that night, and because it would need to hear the voice command "Alexa stop" to stop playing, its seven microphones would have been listening. That meant a recording of what had happened might lie somewhere on Amazon's servers.

When police sent a warrant to Amazon ordering it to release anything that had been recorded, the tech giant refused, citing privacy grounds. However, Bates allowed the recording to be handed over. He claimed he hadn't done anything wrong, so there couldn't be anything that could incriminate him.

The case was the first example of a smart speaker providing evidence in a court case – and a sign that voice-recognition technology is increasingly becoming part of our daily lives. Apple, Microsoft, Google and Samsung have introduced their technologies into similar home-based products. And voice-recognition systems aren't just confined to smart home speakers: watches, lightbulbs, home thermostats, fridges, TVs, cars and more are being equipped with the ability to interpret speech. All around us, gadgets are listening to what we say...

7  
MINS

ILLUSTRATION: MAGIC TORCH





## THE IMPACT

"The way we interact with the internet is moving completely away from buttons and screens to being something verbal and almost atmospheric," says Lucie Greene, the worldwide director of J Walter Thompson's Innovation Group.

Greene's organisation, along with Mindshare Futures, recently produced a report, *Speak Easy*, outlining how voice recognition will impact companies and consumers, concluding that, "The opportunity is ripe for businesses to get involved in the voice-activated world."

And it's a world that's growing fast. Research company Ovum reckons there will be more digital assistants than humans on the planet by 2021. And those who already have them are showing a strong emotional attachment to their new virtual helpers. The *Speak Easy* report found that almost half of regular voice technology users love their voice assistant so much they wish it was a real person. Remarkably, more than a quarter say they have had a sexual fantasy about their voice assistant.

Bearing in mind this cosy relationship, it's clear that the old-fashioned interruptive style of advertising won't cut it in the brave new voice-activated world. One early approach to the problem has seen brands develop services for personal assistants. Amazon's Echo currently provides more than 7,000 'skills' for users, which range from the super-practical (learning first aid) to more brand-driven information. The Johnnie Walker 'skill', for example, provides history, recommendations and drink recipes – as well as information on the nearest shop selling the whisky. "Now our customers can entertain at home in a unique way with the hands-free convenience of Alexa," said Amazon Alexa director Rob Pulciani on launch.

However, a report by VoiceLabs reveals that less than a third of Alexa's 'skills' have more than one customer review, suggesting take-up is limited. What this means is that

in future, companies will have to fight with each other to make their brand the one that a voice assistant suggests when a user asks for recommendations – which, of course, is good news for the manufacturers of voice assistants.

Beyond the world of personal speakers, brands will need to think about how to make their physical assets voice enabled – and think about what that voice will be and what personality it has. This includes packaging that talks: two-thirds of global smartphone users in the *Speak Easy* report agreed with the statement: "I like the idea of being able to ask my products questions about their provenance."

"There's an opportunity for brands to come across as more human and almost like sentient beings," Greene says.

Remarkably, more than a quarter say they have had a sexual fantasy about their voice assistant

## THE RISKS

On a Wednesday afternoon in April, owners of the Google Home – the search giant's rival to the Amazon Echo – were startled when their device started speaking of its own

volition. Nobody in the room had uttered the wake words of "OK, Google", but the voice assistant had been triggered. The reason was that on televisions around the country, fast food retailer Burger King had shown its new 15-second advertisement.

Within the short commercial an unnamed employee says that 15 seconds isn't enough time to explain what all the ingredients are in one of its burgers. "OK, Google, what is the Whopper burger?" the staff member says, as the camera moves to a close-up. The result? Google Home reads out the Wikipedia entry – cleverly edited to relay Burger King's message – to unsuspecting owners. The incident was one of the first

cases of companies exploiting voice-recognition devices in people's homes.

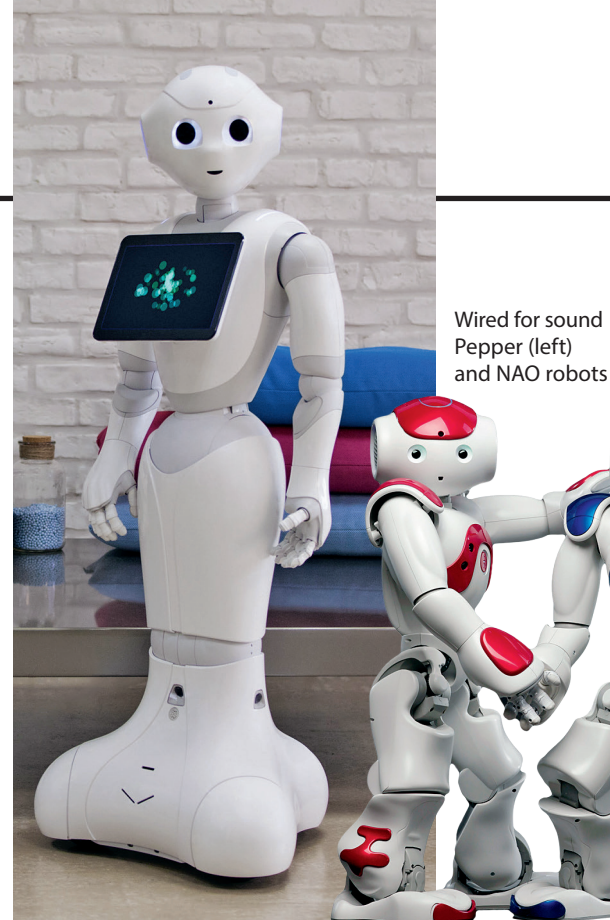
Google promptly stopped Burger King's advert from working but the example, like the murder case involving Collins and Bates, highlights the potential intrusiveness of voice recognition and AI systems within the home.

"The use of biometrics such as voice recognition in everyday life gives rise to real concerns about our society's changing attitudes to privacy and normalises what is, essentially, an intrusive tracking capability," says Silkie Carlo, a policy officer at human rights group Liberty. She believes that potential owners of voice-controlled systems need to question whether they are comfortable to let a product into their home that has the potential to be listening at all times, as well as providing voice information to the firms that create them.

The systems are also not infallible. When the banking firm HSBC turned on voice recognition for its online customer services, the twin of a BBC reporter mimicked his sibling's voice to gain access to his bank account. He was able to access balances and recent transactions, and transfer money between internal accounts. Separately, researchers from Princeton University in the USA were able to take control of Android phones and the Amazon Echo by playing ultrasonic sounds, inaudible to the human ear.

## THE FUTURE

Voice technologies have the potential to radically alter how we interact with the items around us. The future offers a



Wired for sound Pepper (left) and NAO robots

recognition systems to create a safer world for humans. Kelly Davis, the manager of Mozilla's machine learning group, says "home hubs" could be used to provide safety and health monitoring for the ill and elderly, as well as being used in cars

to stop drivers from taking their eyes off the road. In one glimpse of what is to come, Toronto-based company WinterLight Labs has developed voice-recognition systems that are able to deconstruct a person's voice and determine within 45 seconds whether he or she is living with Alzheimer's disease. There's also the potential to help those with disabilities improve their lives.

"Voice interfaces for the web could enhance browsing experiences for people with visual and

potential world where it will be possible to tell a television what to show, instruct a car where to take us, and converse freely with robots. In fact, any object that's able to have a chip installed and be connected to the internet could feasibly understand what we are saying.

"Speech recognition and computer vision will redefine how we interact with technology," says Xuedong Huang, a technical fellow at Microsoft who is the key figure in the company's spoken language processing technologies. He believes a state of 'ambient computing' will be achieved in which humans don't have to interact with physical screens and devices to get digital tasks done.

At the moment, the technologies are still being adopted.

"It is on our phones and available in IoT applications at home but is not used as much as such a highly skilled achievement should be," says Alastair Harvey, chief solutions officer at AI firm Cortextica. He believes voice-recognition technologies will be most useful when paired with other forms of artificial intelligence. One example is manufacturing robots, which can work alongside humans and build products. These will be transformed by their human boss being able to order them to complete certain tasks and act as an additional employee, just by talking to them.

There is also a huge potential for voice-

physical limitations, giving them the option to talk to applications instead of having to type, read or move a mouse," Davis says.

At the moment, robots are being given the ability to act in more human ways. Machine vision is giving them the ability to see objects around them, and with the addition of voice recognition and pronunciation they will be more human than ever. Softbank Robotics' NAO and Pepper humanoid robots are currently able to listen to humans and answer queries, but in the coming decades their voice

capabilities and other technologies may make them proactive.

Instead of waiting for a human to ask a question, the robot may be able to use its experience to predict what will be asked and instantly provide information.

A major development for voice technology will come when

systems are able to understand what a human is feeling, not just the words they are

saying. Computer science researchers are developing emotional intelligence, which can then be applied to voice-recognition technologies. Future systems will be able to understand when a person is angry or elated and provide responses to their questions with a knowledge of the type of response they want. We're heading for a world where we can talk to everything – and everything can talk right back. ■

"Speech recognition and computer vision will redefine how we interact with technology"

## WHO'S LISTENING? A SPOTTER'S GUIDE TO HOME ASSISTANTS

From left: Apple HomePod; Harman/Kardon Invoke with Cortana by Microsoft; Google Home; Amazon Echo; and from China, Baidu's Xiaoyu Zaijia ('Little Fish') and the brilliantly named LingLong DingDong

