TALKING TO ROBOTS

From C-3PO to Sonny, the movies are full of robots fluent in human languages, but making a computer brain understand the intricacies of speech is proving a difficult task.

Computer sound recognition has two distinct branches: voice recognition and speech recognition. Voice recognition is a branch of biometrics, the measuring of human characteristics to confirm an individual's identity. Computers with voice recognition capabilities can listen to a person's 'voiceprint' and conclude if they are who they say they are. Speech recognition is the far more complex science of getting a computer to understand what a person is saying.

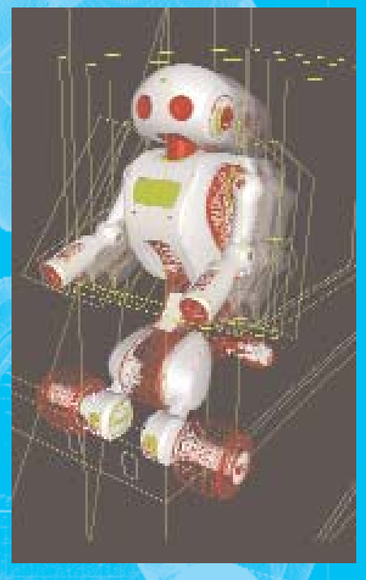
We usually take our ability to talk and understand speech for granted, but it's really a very complex process. So complex, in fact, that scientists still don't know exactly how we do it. Teaching machines to do something we don't fully understand ourselves is therefore understandably tricky.

Being able to give spoken commands, instead of operating manual controls, will make life much easier for people operating very complex machines such as the combat aircraft.

INSIDE I-ZAK

SOUND SYSTEM

I-Zak is fitted with a number of microphones and speakers and is programmed to recognise human speech and respond. He is able to seek out the source of a sound and turn towards it. Once you have his attention, you can talk to him. I-Zak can understand common command words – such as 'Stop', 'Come', 'Left' and 'Right' – and will respond by taking the appropriate action. He will also acknowledge a spoken command by saying 'OK' or lighting one of his indicator LEDs. I-Zak can also be taught to recognize his owner's voice using 'biometric analysis'.



ANALOGUE TO DIGITAL

There have been many advances in computer vocal communication, and it is now possible to program computers to understand spoken commands and recognize individual words. But computers are digital devices, and we exist in a world of analogue sound. Our ears can pick up the analogue sound waves and our brains can sort the different sounds into words and sentences, but to a computer, analogue sound is just a lot of noise. So to get a computer to hear us, we need to bridge the analogue-digital boundary. © © ©

Bridging the communication gap between human and machine will be essential if we are to have truly integrated robot workers. The first step is to turn the analogue sound waves of speech into digital data. This is a simple matter compared to the task of getting the robot to understand all the complexities of human communications.

Early speech recognition software used a system called 'discrete speech', in which each word must be spoken individually, followed by a clear pause. This system is fine for simple commands but it's no good for dictation. A newer system, 'continuous speech', allows a computer to recognize the words of a person speaking at normal speed. This enables computers to take dictation at speeds of 160 words per minute.

There are some ongoing problems, though. People have different accents and inflections – two people may say the same word in very different ways.

Computers have to be taught to recognize individual speech styles. At the moment, this means that someone with a different accent will not be understood by the computer, but when computers have faster processors and greater amounts of random access memory (RAM), it will be possible to program them with a number of different accent profiles.

FACT FILE

SPEAKING THE TRUTH

Voice recognition is used as a security measure by some banks and other services. The systems rely on customers recording their voices onto the system when setting up their accounts. Using the voiceprints, the computer can decide whether customers are who they say they are. To avoid the possibility of fraudsters using recordings to cheat the system, customers are asked to repeat random words. As this technology becomes more commonplace, call centre staff could become an endangered species.

LISTEN AND LEARN

Speech recognition is used in an increasingly wide range of applications today, and professions in which large amounts of dictated text must be typed out often use it to cut down on administration time. The medical professions, for example, with their need for reams of paperwork, find speech recognition particularly useful and specialist medical words can be programmed into the computer's vocabulary. But one of the groups to benefit the most from speech recognition has been the disabled, especially the blind. For those who find it difficult to use a keyboard, being able to speak input into a computer is a real breakthrough.

As well as being able to recognize spoken words, some computers can translate them into foreign languages. The VoxTec Phraselator P2, for example, is a hand-held device that can translate more than 15,000 English phrases into 53 different languages.

Another practical use of speech recognition is in military aircraft cockpits. The Eurofighter, for example, has a voice input system that allows the pilot to use spoken commands to control the cockpit display functions.

HEARING VOICES

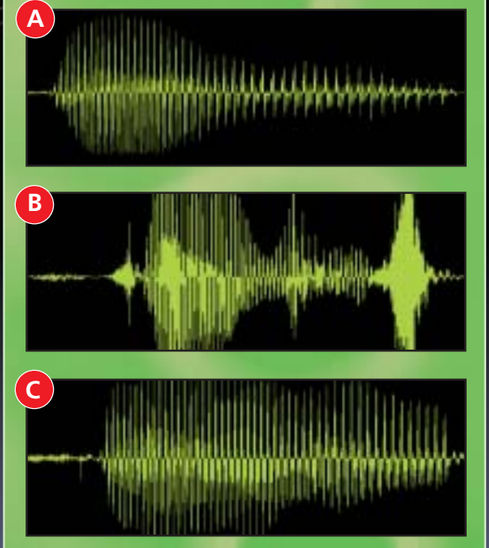
Voice recognition is a function that allows computers to analyze the characteristics of an individual's voice and so confirm their identity. This means that a person's voice can be used to prevent fraud, just as fingerprints, iris scans and face mapping can. However, unlike these other branches of what is known as biometrics, a voice can easily change and sound different. Researchers have therefore focused on those aspects of the voice that never change. Within each person's 'voiceprint', a computer can discern signals that reveal the size of the speaker's mouth and throat. These unique characteristics are very difficult to fake or disguise. For example, a person with a cold that changed the way he or she sounded would still be recognizable to the computer.

Being able to confirm a person's identity by voice alone will have great security benefits for a range of customer service-oriented professions.

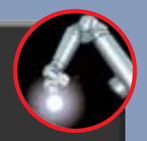
FACT FILE

COPING WITH ACCENTS

When the same word is spoken by people with different accents and intonations, it can be difficult for a computer to understand them all. While we can usually work out what people are trying to say, computers lack our language abilities. The three sound-wave patterns below show the problems that computers face: they are of same word ('forward'), but spoken by three different people. They each look very different.



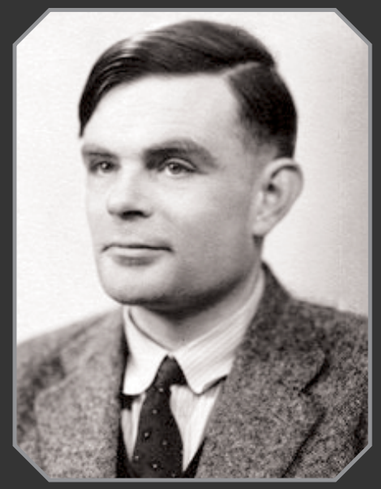
A researcher at IBM talks to a computer. Her analogue word sounds are converted into binary code that the computer can understand.



ROBOTS WITH EARS

One of the key attributes of an effective android will have to be the ability to listen and communicate like a human. Owners do not want to have to type their commands via a keyboard – they want to be able to address their robot assistants as they would humans, and have them respond immediately to their requests.

So far, we have managed to create robots that can listen and talk, but their level of understanding of what is being said is still basic. Some robots can respond to programmed-in command words, as i-Zak can, but none can really understand exactly what is being said, although they can be taught to recognize any number of words. The problem is that while robot brains can learn more words (vocabulary) than any human could ever remember, they cannot understand the intricacies of syntax (sentence structure and meaning).



Alan Turing introduced the idea of the Turing Test in an article called Computing Machinery and Intelligence, *published in 1950 in the journal* Mind.

To reach this level will require an advanced stage of artificial intelligence. In fact, the Turing Test – a goal set by the mathematician and father of computer science, Alan Turing – uses the ability to hold human-like conversation as the benchmark of whether a machine is intelligent or not.

Turing proposed that an intelligent machine could be tested by having it hold a conversation with a human being. If that human being did not realize that the other 'person' was not actually human, then the machine had reached the goal of true artificial intelligence. No computer has yet passed this test.

To pass the Turing Test, a computer must be able to hold a convincingly realistic conversation with a human being who is unaware that he or she is communicating with a machine.

INTO THE FUTURE

THE FUTURE OF ROBOTIC LISTENING

When robots are finally able to understand human language, they will have become true androids. Using their voice recognition capabilities, they will be able to recognize their owners' voices, as well as those of their owners' colleagues, families and friends.

Such robots could be used in the workplace without the need for complicated programming – their managers could simply tell them what to do. And unlike human workers, these robots will never forget the details of what they have been told.

Obedient and observant robots would also be very useful in the home, especially as helpers for people who are unable to look after themselves properly. For the disabled or elderly, robots could act as both carers and workers, responding to requests and helping to operate household appliances.