robot has to do is to string together those phonemes to generate the required message. This message is usually held as a string of phoneme numbers in the computer's memory.

Most of the speech synthesisers in use can be programmed by writing out the message that the robot is to speak in a phonetic version of English. Thus, the message 'Can you come here?' might be written as 'kan yew kum heah' and this would be sufficient for the synthesiser chip to produce the correct string of sounds. This is not exactly the same notation that linguists use when describing phonemes — they have their own specialised alphabet — but it suffices for robots.

At this point you will notice that the robot is no longer using a pre-recorded message — it is actually generating messages of its own. Because of this it is possible to make the robot say anything we wish without the need for having the whole message stored beforehand.

So, if we wanted to we could try programming in some of the rules of grammar in an attempt to make the robot say quite original things. But, as already mentioned, the number of different things that a robot might want to say is fairly limited so there is no need for too much complexity unless we happen to be feeling either adventurous, or curious to see what can be done.

If you have ever heard a speech synthesiser on a robot you will know that the quality of the speech, although usually comprehensible, is by no means perfect. This is due to two factors. The first is that the form that a phoneme takes when used by a human speaker varies considerably depending on the phonemes that precede and follow it. The second is that the overall sound of human speech varies depending on the meaning that we wish to convey. 'Will you sit down?' and 'Will you sit *down*?' are two identical written messages, but they will sound quite different if the first is said by a courteous host to his guest and the second is uttered by an exasperated schoolteacher. Some attempts have been made to capture this intonation in speech synthesis systems, but it is difficult to apply as a robot has no knowledge of the meaning of the words it is speaking.

SPEECH RECOGNITION

The inherent problem to solve when devising a speech recognition system is that the things we may wish to say to a robot, and the different ways in which we might express them, are many and varied. The problem could be approached by using a tape recording of everything that we might want the robot to understand. When we spoke, it could then simply scan through all of its tape recordings and look for the one most like the message it just heard — and that, in principle, is how many robots do recognise speech. They store internal 'templates' of spoken messages and, on being spoken to, simply look for the template that offers the best match. These templates are usually obtained by training the robot — repeating a word or phrase several times — until it has an 'average' template of what we have said. This method works well if you only have a small number of things to say to the robot and are going to say them in roughly the same way every time. It is used for robots that respond to simple commands such as 'forward', 'turn left', and so on.

However, this is a comparatively simple problem and is known as 'discrete speech recognition' because each spoken item is 'discrete' — that is to say, it is separated from other messages by a slight pause during which nothing is said.

The real problem emerges when we wish to speak to the robot using 'continuous speech', which is the type of speech we normally use when speaking to each other. Try saying 'It's a nice summer day' and listen closely to what you said. You will find that it comes out as something like 'Itssan ice ummerday' with the words and sounds running into each other.

The way that people tackle this problem when they are listening to others speaking is by guessing what it is that the speaker means to say — not usually a hard task — and using this guessing to decode the message. But for a robot to do this it would have to know a great deal about what was likely to be said and what it was likely to mean — a very complex task.

In general, speech synthesis by robots is becoming quite common, although there is still room for improvement in the quality of their speech. Speech recognition is a much more difficult task and, currently, the best that can easily be achieved is to endow the robot with an understanding of speech equivalent to a well-trained dog that responds to spoken commands, as long as there are not too many of them. However, there is a tremendous interest in solving all of the problems of robot speech and the next few years are likely to see substantial advances.

Hear Me, Feel Me
The Voicemate is a voice-controlled robot arm developed for laboratory and industrial use by the science engineering department at Newcastle Polytechnic

COURTESY OF NEWCASTLE POLYTECHNIC