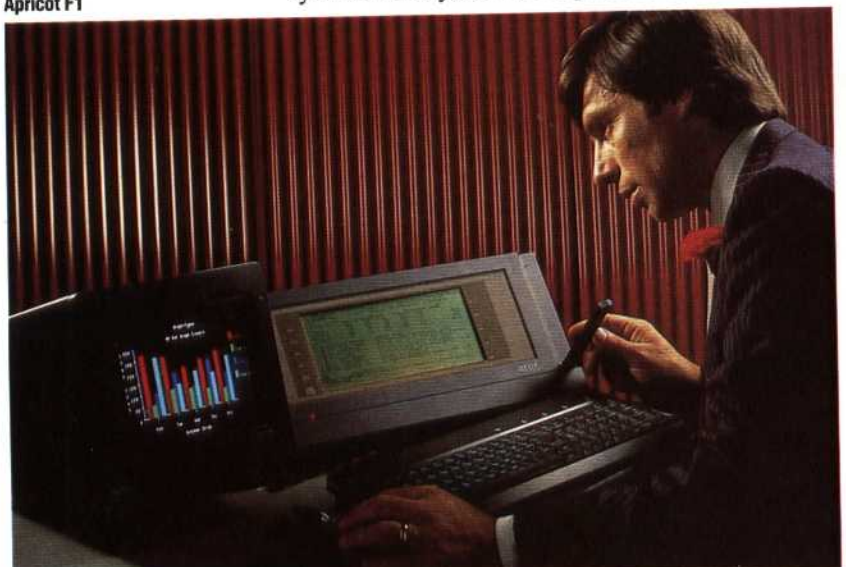
good semantic analysis, but the world it was able to make sense of was extremely simple. A robot working in the chaos of the real world would have had much greater difficulty understanding what was being said to it.

In order for robots to employ language usefully, there must be a message that is passed between you and the robot and vice versa. It is relatively easy for robots to speak, because anything that a robot would be likely to want to express would be very limited, since its knowledge is so restricted. It is much more difficult for a robot to understand what you might want to say to it, because anything that you might wish to communicate is much more difficult to analyse.

At one time it was thought that speech input to robots would be analysed by carrying out a syntactic analysis of the input and that this would reveal the meaning of the message. But recent work has shown the importance of knowledge of the surrounding world and the context in which the message is spoken. This has led to experiments in which a tentative syntactic analysis of the speech signal is made in order to make a first guess at the meaning. Then, in the light of what the robot knows about the world and the likely things that might be said in its world, the robot revises its original syntactic analysis in the hope of gradually homing in on a correct analysis of what is being said. However, this is far beyond what any commercially available robot can currently do. Here, we will look at how contemporary robot systems speak and understand speech.

Apricot F1



Big Ears

Currah

COURTESY OF WILLIAM STUART SYSTEMS LTD

Hearing And Speaking
Robot and computer speech is fairly simple to create. Speech synthesis devices, like the Currah shown here, are available for even the smallest home computers. But recognition is more difficult because of the variations in the way humans pronounce vowel sounds, and because of the amount of processing power and memory required to handle a vocabulary of more than a few words. Systems like Big Ears, and the Apricot F1, have a small range of recognisable commands built in, but too few to cover more than a minimal number of operations

SPEECH SYNTHESIS

The simplest method of speech synthesis employs a tape recorder in which a message spoken by a human being is recorded on tape and played back by the robot at an appropriate time. This might not seem to be quite what you had in mind when you first thought of robot speech — but it is the starting point of all speech synthesis systems. We will look at the limitations of this method and then see how we can improve on them.

The most obvious limitation is that a tape recorder is mechanical, expensive, bulky and liable to break down. So the next step is to take the same message and convert it into digital form so that it can be stored on a chip in the robot's memory. This is done using an analogue-to-digital converter, in which numbers are used to represent the continuously varying waveform of speech. This is exactly the same method that is used in the digital recording of music on, for instance, compact disc systems.

This method has its drawbacks, too. One of the main problems is that a digitised signal takes up a lot of room in memory. Compact disc recording samples the acoustic signal around 44,000 times per second with a resolution of approximately 16 bits (i.e. the amplitude of the waveform at any moment is stored as a 16-bit number, which enables 2^{16} levels to be discerned, where $2^{16} = 65,536$). Using this system, each second of the recording would occupy 88,000 bytes of memory. Clearly, a spoken message exceeds the storage capacity of any microcomputer. However, this sampling rate is only applicable to high fidelity sound reproduction; a simple speech system could be operated with a resolution of eight bits, and a sampling rate of 3,000 samples per second, which only uses up three Kbytes of memory!

However, in order to free the maximum amount of memory space, further economies need to be made. Linguists have found that spoken language can be conveniently broken up into units of speech called phonemes. In all, there are generally agreed to be some 40 different phonemes for most spoken languages, so it is possible to store the exact acoustic information necessary to describe each of these 40 phonemes and then use these as the foundation of robot speech. Typically, the phoneme information is held on a commercially manufactured speech synthesiser chip and all the