be by town, then by village, etc. Only at the lowest level would the item finally be sent on its way, thus ensuring the maximum reliability of the operation.

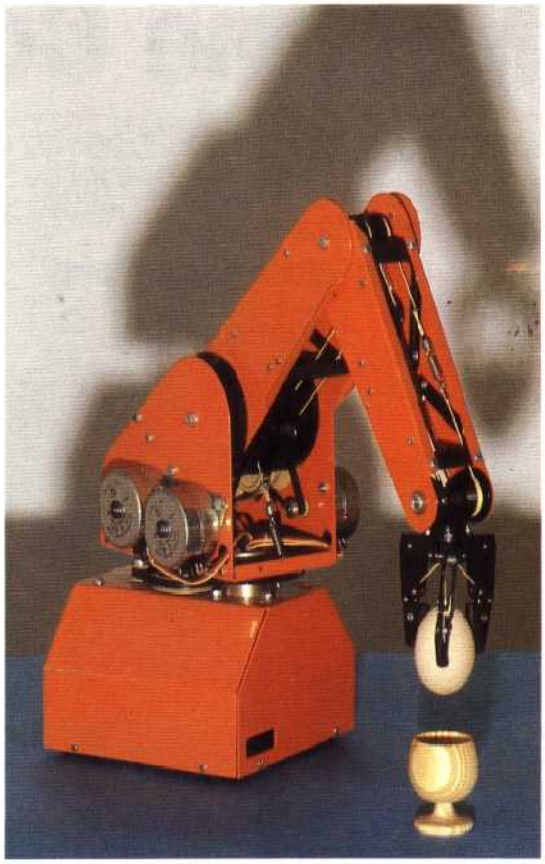
Voice Analysis

Speech recognition is usually tackled in one of two ways. The 'quick' way is simply to feed all the speech through an analogue-to-digital convertor, and use the power of the computer to perform all the analysis. Unfortunately, this method has a number of drawbacks, most notably the time taken to perform the analysis. Systems using this method can take up to two or three seconds to recognise the input. For speech recognition to be of any real use the computer must 'understand' the speech as fast as another human, and the number crunching approach rarely achieves this.

The other method is to use pre-processing. Rather than analyse the speech signal mathematically, it is possible to do much of the work with standard electronics. What is then delivered to the computer is information about the spoken input: the frequency content, pitch, energy, etc. Frequencies can be measured by filtering the signal and detecting the level in each frequency band, rather like using tone controls on a hi-fi to 'bring out' the bass drum. Because all this electronic processing is done at the same time as the original speech signal is fed to the circuits, the analysis is almost instantaneous. Performing a similar operation on the digital data from an A/D convertor would require several computers working on the numbers at once. The pre-processing method is still at the research stage — no commercial system using it has yet been marketed — but it certainly appears to have more potential.

Once the information about frequency content, pitch, energy, etc. has been extracted from the original signal (regardless of the method), the actual recognition is performed by comparing the current set of figures with a number of models stored in the computer's memory. These models are created by 'training' the recognition system. The words that are to be recognised are spoken into the system one at a time and the resulting information is stored in a digital 'library' of examples. The complete set of words is then spoken again and the computer compares the input with its current model. If they agree, the second set of information is added to the first to form a more complete version of the model. This can be a continuous process, constantly adding new information to the library for more and more speakers.

To recognise a spoken word, the computer must match the pattern of information from the input with one or more of the models stored in the current library. In many cases, several possible matches will be found as parts of other words will match the input pattern. The first two syllables of 'international', for example, are the same as those of 'interpreter'. At the end of the search, one word should stand out as being more perfectly matched than any of the other possibilities, and this is the one that the computer will interpret the input as being.



IAN McKINNELL

Environmental Control
Most recent applications of speech recognition are of an educational nature. One of these is called the 'limited environment', which involves a computer, a robot arm, and a number of simple objects that the arm can manipulate. Speaking into a microphone, the user can instruct the arm to 'PLACE THE EGG IN THE EGG CUP'. The computer will have to interpret the commands, and look up the positions of the objects in its memory

Speech recognition facilities are certain to find many applications in the future, but they are likely to be most readily used as a 'front-end' for complex software packages, such as databases, where the commands are selected from an on-screen menu. This type of application will remove the single biggest obstacle to computer usage by non-experts: the keyboard. Viewdata systems such as Prestel have reduced the input device to a simple numeric keypad, but this substantially limits the amount of interaction that a user can achieve. A speech-driven interface that can recognise a standard set of database interrogation commands, as well as numeric symbols and the letters of the alphabet, would provide a powerful facility that requires little, if any, conventional computer training to use.

There are now commercially available recognition units that can be plugged into home computers, but these are very unsophisticated devices. Systems like 'Big Ears' and Heuristic Inc's 'Speech Lab' use a lot of processing power to recognise just a few words spoken by one person. What is needed before speech recognition can become really useful is an ability to recognise words spoken by *any* person, regardless of dialect or accent. The limiting factor, at this stage, is the amount of memory available to hold the models. One interesting possibility is that of using a video disc to hold a standard set of models: this would use hardly any internal memory and the reduction in speed would be barely noticeable.