THE USE OF COMPUTERS FOR RESEARCH IN PHONETICS

P. DENES

More and more often these days we find that the person engaged in the experimental investigation of speech is surrounded by a complicated array of electronic instruments and only too frequently his immediate problems are centred around what I call the "soldering iron", the electronic performance of a valve, and so on, rather than on the question he originally set out to investigate: the acoustic or physiological characteristics associated with the units of linguistic organisation, the phoneme, the word, etc.

The reason for this trend is that the basic relationships between the acoustic, physiological and linguistic levels of the speech event are highly complex and are strongly influenced by acoustic and linguistic context and therefore any phonetic instrument that is to take these diverse factors into account in its operation must be correspondingly complicated. Let us take just one brief example: the work of the Haskins Laboratories has shown us that the plosive consonants are associated with certain variations of the second formant of preceding or following vowels. These variations are the so-called formant transitions. The actual transitions for any one consonant are however not invariant but depend on the nature of the associated vowels. If we wanted to recognise such a consonant automatically, therefore, the device to be constructed would have to be able to find the second formant of the speech wave, memorise its variation over a period and then make a decision based on these variations. More generally, if the instruments we use for either speech analysis or speech synthesis are to incorporate what we know or surmise above the functioning of the speech mechanism, then they must have considerable capacity for memorising acoustic and linguistic rules and data and the ability to use this information in fairly complex ways. Only by building such complex models can we try some of our hypotheses about the operation of the speech mechanism or alternatively try the prototypes of some of the proposed speech transmission systems.

Modern electronics enables us to construct such complicated devices but they require extensive design followed by lengthy test procedures and this is the reason why the experimental phonetician has to spend so much of his time with the soldering iron.

During the last few years, however, digital computers have become commercially available and are likely to change this situation radically. These computers offer extensive facilities for processing data in a great variety of often extremely complex ways; they also have large memories for storing data and rules for handling them. In fact, they offer just the kind of facilities that we have seen are frequently required for the experimental study of speech. What is more, these facilities are available without the need for any electronic design on the part of the experimenter: that has been taken care of by the designers of the computer. All the phonetician has to do is to write down a series of instructions that embody the logical sequence of steps that he wants to have carried out: the computer will do the rest.

As an example of the set of instructions, or program as it is called, required to carry out a specific task, let us assume that a speech wave has been applied to a bank of filters, that the identity and value of each filter output is available to the computer and that we wish to find the spectral peak so as to determine a formant frequency; in other words, the task to be performed is the one that is often referred to as "peak picking". The computer program for finding the filter with the biggest output will take the following form: look at the output of the first two filters, the computer is told, compare them and select the greater one. Next compare this output with that of the third filter and again select the greater of the two. Carry on in this way for all the filters in turn. The last selection will give the identity of the filter with the greatest output. The result is however still held inside the computer and a further instruction is needed to have the result printed on the printer attached to the computer If this type of operation is likely to be needed often, then arrangements can be made so that in response to only one instruction, such as "find the greatest of a set of inputs" the computer will select and carry out the whole of the above sequence of operations.

This simple example also shows an additional reason for the great power and flexibility of the computer: the program does not simply follow a rigid sequence of operations, but can alter its course depending on the result of some previous operation.

There is one difficulty in using computers for speech research that should be mentioned here. The computers are designed to operate solely on numbers and any speech data must be converted into numerical form before the computer can handle them. Let us see how this can be done. In experiments involving speech analysis it is the sound wave to be examined that has to be converted into numerical form. As we know, the sound wave consists of variations of air pressure. The varying pressure is sampled at frequent intervals and each time a number is produced that is proportional to the size of the sample. The sound wave is thereby represented by a sequence of discrete numbers. The instrument for carrying out the conversion, the so-called analogue-digital converter, is available commercially.

In the case of speech synthesis, the programme will generate a numerical sequence representing the synthesised sound wave and this can be converted into an actual sound wave by a digital-analogue converter, a reasonably simple instrument.

If linguistic data are to be used, the identity of the phonemes, etc. can easily be

coded in numerical form. The facilities for analogue-digital and digital-analogue conversion, once established, can be used without modification for a large range of experiments.

As against the drawback of requiring this initial instrumentation the use of computers can greatly reduce the amount of individual instrumentation required for many experiments. Often, electronic design and construction that could easily extend over a number of years is replaced by programming which can be completed in a number of months. Perhaps equally important is the fact that the computer programmer is largely concerned with the logic of the problem and his attention therefore is more closely focussed on the real substance of the experiment in hand. The electronic designer on the other hand has to pay so much attention to the peculiarities and limitations of the circuits he uses that he may easily be distracted from the basic problems. A further point to remember is that the ready-made facilities for complex data handling and memory afforded by modern computers are so large that many experiments that could not have been attempted by conventional electronic means because of their complexity have become a possibility. This applies particularly to the range of experiments where information about acoustic and linguistic context and organisation is used to recognise or synthesise speech waves.

There is every indication, therefore, that the availability of computers is about to produce a profound change in the way in which the experimental phonetician approaches his problems, in the range of experiments open to him and in the kind of training he requires to enable him to carry out his work.

In view of their obvious advantages, computers have already been used in speech research. The Massachusetts Institute of Technology,[1] the Air Force Cambridge Research laboratories[2] and the Bell Telephone Laboratories[3] in the United States and the Universities of Tokyo[4] and Kyoto in Japan are among the places where computers have been used for such work. When I first planned this paper I intended to review the work that has already been done but unfortunately there is no time for this and, rather selfishly perhaps, I shall only describe our efforts to use computers at University College London. The general field in which we are interested is the automatic recognition of speech. The recognition processes are to take account of the acoustic characteristics of the speech wave as well as of information derived from linguistic structure. Our previous work, using electronic circuits of our own construction for this purpose, has already shown that further capacity for data processing and memory is desirable for achieving satisfactory results, and building

[1] C. G. Bell, F. Poza, and K. N. Stevens, "Automatic resolution of speech spectra into elemental spectra", *Proc. Seminar on Speech Compression and Proceeding*, Vol. 1 (Air Force Cambridge Research Center, Mass., 1959).

[2] C. P. Smith, "An approach to speech bandwidth compression," *Ibid.*

[3] M. V. Mathews, "The effective use of digital simulation for speech processing," *Ibid.*

[4] S. Inomata, T. Shinoda, M. Kumada, T. Sawada, and E. Masahata, *Progress Report: Project Logos, Dec., 1960. Vowel Recognising Program, SNCS-1* (Tokyo, 1960).

the additional circuits required would be extremely laborious and time-consuming. The advantage of using computers is therefore obvious.

The first question was: which of the several available computers to use? It was important to select the computer which was likely to be suited to the widest variety of experiments planned, because for certain technical reasons changing from one computer to another is not an easy matter. Convenience of use and of programming, size of memory and cost were some of the points of view from which available computers were compared. It the end it was decided to use the largest machine available, the I.B.M. 7090. The large memory, the high speed and the convenient programming facilities were some of the reasons for this decision. It can easily be shown that speech processing requires a large memory, and even if all the memory of the 7090 is not needed for a particular problem, it greatly simplifies programming. Surprisingly enough even the cost consideration is not unfavourable to the 7090, despite the fact that it costs 25 times as much per hour of use than the computer that was rejected in the final choice. The explanation is that for many speech problems most of the computer time is required not for the processing of the data but for transmitting information to and from the computer. The 7090, although 25 times more expensive, is 100 times faster than its competitor in its input and output operations and is therefore likely to be 4 or 5 times cheaper for any one job than its "cheaper" competitor.

Having decided on the computer to be used, preparations were made to acquire experience in handling both acoustic and linguistic data in the computer. On the acoustic side, a system is on the point of being completed that makes possible the recording in digital form of either the wave shape or the spectrum of a speech wave. The recorded version of the acoustic data is then in a suitable form to be used in any speech recognition program. This system of recording does not require the purchase of a digital recorder but uses the 7090 and the digital tape recorders attached to it in a suitable program. A digital recorder costs about £10,000 while the proposed method requires an expenditure of only about £5 per minute of speech recorded. Fig. 1 gives a block diagram of the scheme. The speech wave is either applied directly to the analogue-digital recorder, for recording the wave shape, or to a bank of filters for spectral analysis. The output of the filter channels is sampled by an electronic switch, the output of which is sent to the analogue-digital converter. Each filter channel consists of a band pass filter, a rectifier and a low pass filter. There are 70 of these filters covering the frequency band 100 c/s to 10,000 c/s. The whole operation is under the control of a computer program, which calls for the speech wave to be applied to the analogue-digital converter, for the digital information to be sent continuously into the computer and from there to be sent to and recorded on one of the digital tape recorders attached to the 7090. The program also edits the data into convenient form for future speech processing experiments, before actually recording them.

On the linguistic side, programmes have been tried or are being planned for obtaining further data about phoneme distributions in English and for using linguistic

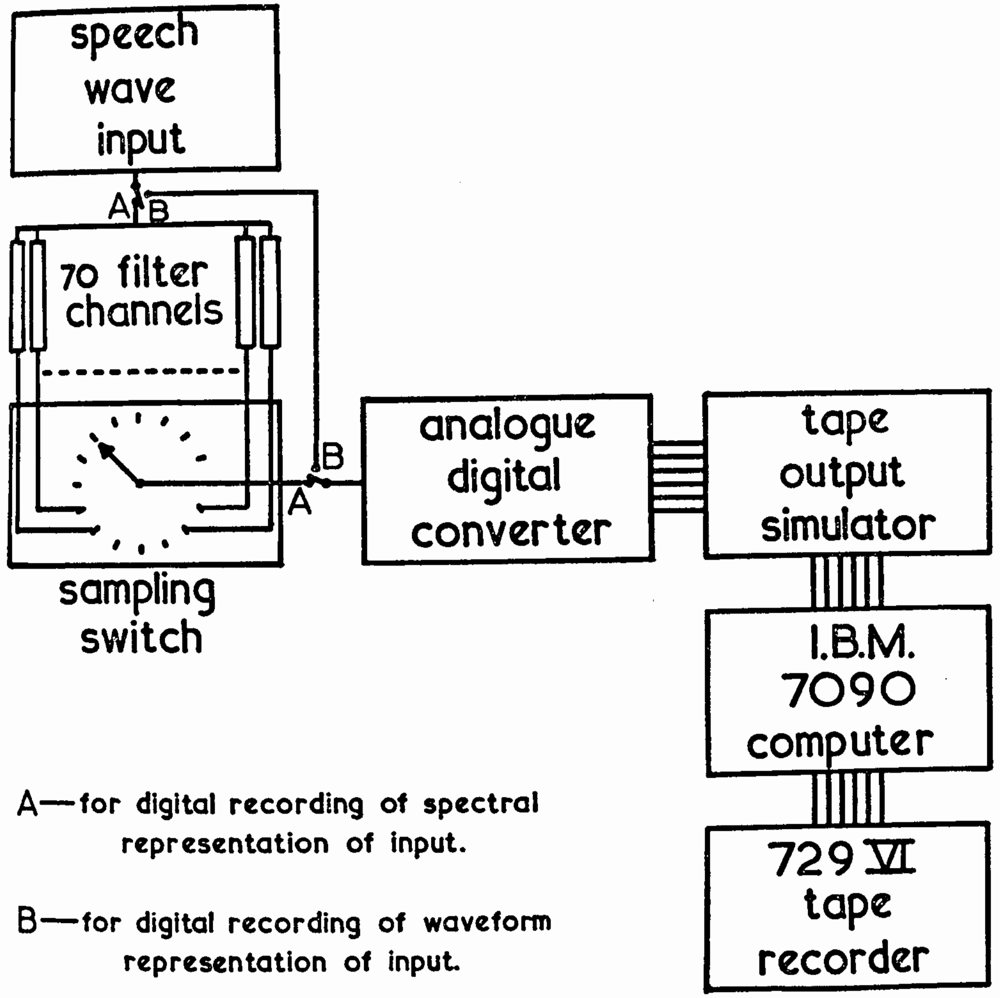
speech wave input
A B
70 filter channels
sampling switch
B
A
analogue digital converter
tape output simulator
I.B.M. 7090 computer
729 VI tape recorder
A—for digital recording of spectral representation of input.
B—for digital recording of waveform representation of input.

Fig. 1.

structure in automatic speech recognition and similar applications. Results have been obtained on the frequency of occurrence and digram frequencies of English phonemes, on the relationship of stress, word and syllable boundaries to each other and to phoneme distributions. As a preparation for full-scale automatic speech recognition experiments, programmes are being evolved to test how far errors in recognition can be corrected by information about linguistic structure. The computer is given a list of all the words that can occur in the input sequence and this information is used in the recognition of an input which includes deliberately introduced mistakes. It is also intended to write a program for the automatic transcription of English, using Daniel Jones's *English Pronouncing Dictionary* and a suitable set of structural and semantic rules.

I hope that what I have said in this short paper will not have given the impression that I believe all "good" speech experimentation will in future use a computer. Quite obviously there are many classes of experiment where the use of computers is not suitable. At the same time, I hope my paper will have given at least a partial explanation of why some of us have such high hopes of the impact of computers on speech research.

It only remains to acknowledge that the major part of the work referred to in this paper was carried out under contract with the Air Research and Development Command of the United States Air Force.

University College, London