[From: *Information retrieval and machine translation*, ed. Allen Kent. International conference… September 6-12, 1959, Cleveland, Ohio; part 2 (New York: Interscience Publishers, 1961)]

CHAPTER 37

The Current Status of Direct Input in Great Britain

ANDREW D. BOOTH

Department of Numerical Automation, Birkbeck College,
University of London, England

This short account of the current state of character and spoken word recognition in England is divided into two sections: the first dealing with written and printed characters, the second with the spoken word.

As far as printed or written characters are concerned, the pioneers in this field in England were the Solartron Electronic Group, who demonstrated late in 1956 a prototype model of their electronic reading automaton, shortly called ERA (1, 2). This device can recognize something like sixteen symbols comprising the numbers 0, 1,… 9 and a few alphabetic symbols which are required in the interpretation of printing from the till rolls used by cash registers in retail stores. Speed of recognition can be as high as 1000 characters per second, but the restricted format which this machine uses makes it unlikely that in its present shape it would be suitable for the direct input of textual material to a translating machine. The principles upon which ERA operates are quite simple. Two scanning operations take place on each character, the first to actuate servo-mechanisms which centre the scan upon the character, the second to sample various portions of the character and apply the results of this sampling to logical circuits which, by their interaction, produce an unique characterization and thus encodement in binary form of the character being sensed. The cost of the scanning device in its present form suitable for till roll appears to be of the order of £20.000, that is, about 56,000 dollars.

The second commercially available scanning device is so far applicable only to numbers. This is the EMI character reading apparatus called FRED (3, 4), or Figure Reading Electronic Device. The principle of FRED is a very simple one. Each of the numerical symbols has a special format in which it is arranged that in any of six vertical sections the total amount of blackening is either considerably greater or considerably less than some value which we arbitrarily designate as one half. Any section containing a blackening greater than one half gives rise to a code digit one, any section containing less blackening than one half to a code digit nought. The first of the sections is arranged always to contain unity as far as blackening is concerned, and thus to trigger the remaining circuits of the recognizer. The five sections which follow are scanned in sequence and give rise to a one-nought sequence which is characteristic of the number concerned. Thus, for example, in one of the type founts which has been proposed the code numbers are as follows:

0	10111	6	10011
1	11000	7	11110
2	11111	8	11011
3	11100	9	11001
4	10110	10	10101
5	11101	11	10010

There seems to be no difficulty in principle in extending FRED to the recognition of a full range of alpha-numeric symbols, but the approach to the problem which is involved, namely the use of an idealized type fount which is easily recognizable, while thoroughly praiseworthy in the production of low cost, highly accurate business machines, does not seem at all suitable for the production of inputs to a translating machine, since a wide variety of text founts will be involved and certainly the rather exotic type faces required by FRED will not find universal typographical acceptance on the part either of readers or probably of the editors of journals.

So much for the character reading machines which are commercially available in Great Britain. Next are discussed the three projects which are still in the experimental stage. The first of these, which is of only remote interest to machine translation, is the work of W. K. Taylor and J. Z. Young at University College, London on the simulation of animal pattern recognition by electronic analogues of a neuron net (5). Taylor's recognizing device consists of a matrix of nine photocells upon which are incident the characters concerned. The output of these photocells is then passed to what Taylor calls a "detail filter." This is in effect a band width compression device which produces from the nine quantized inputs from the original photocells nine outputs which are modified from the original inputs by the fact that each output takes account of the neighbouring inputs about a given place on the photocell matrix. The physiological justification for this device is based on the observation that when one examines an object, considerable detail is available at the actual point of examination, but that peripheral detail becomes less and less as this point of attention is deviated from. After the detail filter the outputs are passed to "majority units." These are simply devices which take various combinations of the detail filter outputs and produce binary quantized outputs according to the number of inputs which are different from zero. The output from these majority units is used not only to feed successive circuits but also to feed a centering servomechanism for the photocell matrix. The output of the majority units is also taken to a pair of maximum amplitude detectors combined in such a way that the ultimate output is on two lines and is quantized at each of three levels, 0, 1, and 2. It is not to be hoped that with a device having so coarse a detecting surface, practical recognition of diverse type founts can be obtained, and furthermore it is rather doubtful whether it would be commercially attractive to increase the size of the photocell matrix to any considerable and worthwhile extent. The device is, however, interesting in that, unlike ERA and FRED, it is able to recognise characters whose positions with respect to vertical and horizontal axes have suffered from a rotation, and also characters in various states of degradation. In this respect it is far more like the animal brain upon which it is based than are any of the preceding devices. Should it prove possible to construct a large scale recognizer of this type, the device would have the attractions of considerable speed, potentially at least in the mega-cycle range, and of greater flexibility than any of the purely logical schemes which have hitherto been thought of, so that further developments are awaited with interest.

The remaining direct input projects in the character field are those at the University of Manchester and in my own laboratory. At Manchester Kilburn and his collaborators (6) have constructed a machine programme which makes what may be described as topological examination of input characters. Their recognition scheme is simply a multiply branched programme which dissects an incident character into various standard units consisting of straight lines, V-shaped unions, rectangular joins and curved segments, and then, as a result of this dissection, follows a logical course which leads ultimately to the recognition of the character. The present programme contains some 4000 instructions and takes about 60 seconds to recognise any alpha-numeric symbol. The programme size might perhaps be reduced, but its complexity is likely to remain and, since the machine upon which it is run is one of the fastest which is available in Great Britain at the present time, it seems unlikely that this recognition program in its existing form will be of much practical utility. On the other hand, the method contains numerous pointers to the way in which an ultimate recognising device might take account both of malformation of characters and of malpositioning under the scanning device.

The work in The Department of Numerical Automation has been divided into two phases. The first was the construction of GPC or General Purpose Coder. This is simply a device which scans an area of incident text, quantizes the intensity of blackening into two levels, 0 and 1, and punches the result of this scan in the form of 1024 punchings on paper tape. The tape is then taken to a computing machine for ultimate recognition. The scanner itself is of course in no way novel. Flying spot scanners have been made in the past and, although not common in the United Kingdom, are probably everyday devices in electronic laboratories in the United States. The idea of coupling such a device to a tape punch by means of relatively simple circuits appears, however, to be new, and it has certainly borne fruit in my laboratory. The second phase of the programme with this device has so far proceeded as follows. In the first place a noise-removing programme has been written. This takes a tape prepared by the original scanner, examines it and uses criteria of connectivity in order to identify and later remove punchings which are due to noise generated by dirt or surface imperfections on the paper being scanned. The second programme takes the character concerned and normalizes it, that is, moves it (actually by generating two comparison indices to identify the relevant portions) so that the top of the character touches the Y axis and the left hand side of the character touches the Y axis. In addition two scale factors are generated which arrange that the height and width of the character are enlarged or reduced so that it fills a standard interval. When the normalizing and noise removing operations have been performed, a variety of courses of action are possible; for example, one of the simplest recognition procedures which were suggested originally in 1952 (7) is to generate a characteristic number which is the aggregate of the number of intersections of scan lines with the character concerned. These numbers characterize uniquely type of a given fount and it is easy to identify any particular character by comparing the number generated by scanning it with a dictionary of such numbers held in the store. On the other hand, such numerical procedures are not ideally suited to high speed operation since the generation of characteristic numbers is a time-consuming process. We have therefore experimented both with the typological approach in which we have obtained substantially identical although less extensive results than Kilburn, and also with the moment generation procedure which was first suggested by Franz Alt. In this method the successive moments of the character about a pair of axes passing through its centre of gravity are generated. Now, it is well known mathematically that the infinite series of moments about both axes characterises completely any given spatial distribution. The interest arises in seeing how few moments are adequate to characterise symbols arising in different founts. This work is at present in progress, but it rather appears that six moments are required, each specified by a precision of one part in 100 if alpha-numeric symbols in the type founts usually encountered on typewriters are to be identified.

The recognition of tapes produced by such general purpose recognisers on a computing machine is, of course, prohibitively slow, but the attraction of using a scheme of this sort is that it involves building the minimum of hardware and enables the maximum number of experiments to be carried out. It is thus hoped that when this work has been in process for some time more, to be able definitively to specify the ideal character recognition scheme either in terms of hardware or of speed or possibly both. So much then for the direct recognition of characters. The production of a satisfactory device of this sort is of paramount importance in the further development of machine translation, particularly for Russian and for the oriental languages.

A second mode of input to translating machines, which seems to be of peculiar appeal to the popular press, is that of the direct input of the spoken word. Here the subject is of considerable antiquity, since, although it is perhaps unnecessary to recognise the spoken word, it is certainly desirable to compress the information contained in it into the smallest numerical compass if the best possible use is to be made of such things as transatlantic telephone cables. Thus the various telephone laboratories, from the Bell Laboratories in the United States to the G.P.O. in England, have been working for something like twenty years on devices which would find ready application in the direct input of the spoken word. Neither of these groups seems, however, to have been very interested in producing a suitable device for translating machine input. There does exist, however, one group of workers in England who have not only done pioneer work in the field, but have also produced a working device which will type in phonetic script from a spoken input. The inspiration in this group is derived from D. B. Fry at University College, London, and, working under his direction, P. Denes has constructed a phonetic recogniser.[8] The device consists of five main sections excluding input and output. The input speech waveform, usually in the form of a tape recording, is passed to a set of filter circuits. The outputs of these filters are then fed in pairs to a set of phoneme analysers. These analysers are two input devices having the characteristic of multipliers. The inputs from the filters are connected in a manner which is dictated by previous experiments in the field, but are such that both inputs to any given phoneme analyser are generally only large when that phoneme is uniquely characterized. Since the inputs to any given analyser are effectively multiplied together, if either is small, then the output from the analyser is also small. The group of analysers is followed by a maximum detector and the outputs from this maximum detector in early forms of the machine were used to feed the various coding devices required either to produce a punched paper tape or directly to operate a typewriter. It so happens, however, that in real speech with all its imperfections, several of the phoneme recognising multipliers may produce an appreciable output. An ambiguity thus results. To overcome this defect, Fry and Denes have added two additional features to their machine. The first is a memory which effectively stores the last output from the maximum detector during the input of the next, and this memory device is followed by a store of linguistic knowledge, the effect of which is to multiply the output from each of the phoneme recognising devices by a number which is effectively proportional to the probability of that phoneme occurring after the phoneme which has just been analysed. It is clear that the device in a sense makes use of context. The disadvantage of this particular scheme lies in the fact that improbable combinations of phonemes will always be incorrectly recognised, whereas a more sophisticated version of the device could be constructed in which the store for linguistic knowledge operates upon the phoneme recognisers in such a way as to produce the most probable output in the event that several equally probable ones are present, but to be inoperative if one particular output had a far stronger a priori probability than any of the others. The accuracy which is obtainable with this recogniser depends upon whether a single speaker is involved and the machine is correctly adjusted for his idiosyncrasies, or whether several speakers are to use it. For example, for a single speaker about 70% reliability was obtained in one series of trials, but this dropped to about 45% when a second and third speaker were added. At the present time only 13 phonemes are contained in the repertory of the machine, but is is hoped to extend its scope in the near future and also to improve its performance by inserting into the linguistic store not only digram frequencies, but also trigram ones.

Finally in the field of mechanical speech recognition a small amount of work has been carried out in the Department of Numerical Automation. This sprung from the idea which came some years ago of reducing speech waveforms to a sequence of numbers. This was to be done by passing the incident waveform first to a stardardising circuit, second to three filter circuits selecting low, medium and high frequencies, and third to a set of three counters which evaluated the number of axis crossings of the resulting waveforms in equal intervals of time. This triad of numbers generated for each of the time intervals into which a sound was partitioned was shown under certain controlled conditions to characterise the sound itself. Unfortunately the device was extremely sensitive both to the position of the speaker and to room acoustics. Ahmed and Fatehchand, working in the Department of Numerical Automation, then commenced a fundamental study of the possible ways in which this recognition process could be made more reliable. The results of their work have recently been published (9, 10). Suffice it to say that the opinion of Fatehchand, expressed recently, was that a recogniser based upon the principles which he and Ahmed advocate could be constructed during the next two or three years and would afford an accuracy of recognition on the phoneme basis of about 95%. On the other hand it is only fair to remark (a) that this would apply only to a single user of the machine and (b) that as yet no really satisfactory method has been devised which could analyse running text. This is because normal speakers tend to run words into one another. The presence of a pause between words appears to be essential to the proper working of any of the analysers which have hitherto been considered and although one might proceed on the basis of the sentence rather than the word, this would involve a dictionary of sound sampling numbers of the same sort that one encounters in machine translation of words. Since most workers in the field of machine translation are unanimous in suggesting that the storage of anything like a complete dictionary of sentences is impracticable, presumably the same argument applies to the recognition of spoken sounds by similar techniques, so that, short of an accurate standardization on the part of the user of such a machine, both of his phraseology and of his precision of articulation, it does not seem likely that the device will be very useful in the near future. The reason for pursuing this development is in no sense its utility or possible utility in the field of machine translation. It is rather because a simple recording device, with recognition facilities either separate or combined with a computing machine, has very considerable application in store keeping and other recording operations, where first of all only a limited number of sounds are to be recorded, and secondly precise articulation is easily attained.

REFERENCES

1. E.R.A., "Engineer," 203 (1957), p. 414.
2. E.R.A., "Electronic Engineering," 29 (April 1957), pp. 189-190.
3. F.R.E.D.—A Figure Reading Electronic Device, "Electronic Engineering," 31 (Jan. 1959), p. 45.
4. F.R.E.D.—A Figure Reading Electronic Device, "Brit. Commun. and Electronics," 6 (March 1959), p. 185.
5. Taylor, W. K., Pattern Recognition by Means of Automatic Analogue Apparatus, "Proc.I.E.E.," 106B (March 1959), pp. 198-209.
6. Grimsdale, R. L., Sumner, F. H., Tunis, C. J., and Kilburn, T., A System for the Automatic Recognition of Patterns, "Proc. I.E.E.," 106B (March 1959), pp. 210-221.
7. Booth, A. D., On the Recognition of Spoken Sounds by Means of a Computer, "Computers and Automation," 4:2 (1955), p. 9.
8. Denes, P., The Design and Operation of the Mechanical Speech Recogniser at University College, London, "J. Brit. I.R.E.," 19 (April 1959), pp. 219-234.
9. Ahmed, R., and Fatehchand, R., An Electronic Speech Sampler for Studying the Effect of Sample Duration on Articulation," J. Inst. Telecomm. Engineers," 5:2 (1959), pp. 86-88.
10. Ahmed, R., and Fatehchand, R., "J. Acoust. Soc. Amer.," 31:7 (1959).