The Journal of

THE BRITISH INSTITUTION OF RADIO ENGINEERS

FOUNDED 1925 INCORPORATED 1932

"To promote the advancement of radio, electronics and kindred subjects by the exchange of information in these branches of engineering."

VOLUME 19 APRIL 1959 NUMBER 4

PLANNING THE CONVENTION

WHEN the Council of the Institution decided last year that there would be a Television Convention in 1959 at Cambridge, the selection of an appropriate theme was determined by the conviction that the potential applications of television were not widely appreciated. The 1951 Convention devoted to television engineering dealt almost solely with the application of these techniques to entertainment. The past eight years have seen an enormous extension of television for this purpose throughout the world, but the other uses of television may still be regarded as being in their infancy.

One of the first tasks of the Convention Committee therefore was to recommend to the Council a theme which would make clear the forward-looking nature of the Convention. Accordingly "Television Engineering in Science, Industry and Broadcasting" was selected. The Committee believes that its final programme will meet the conflicting requirements of the engineer with a general interest and those of the specialist.

There will be a total of eight sessions in the course of the four days. The opening session will consist of general papers typical of the various applications of television engineering, whilst subsequent sessions will deal successively with Receiver Techniques including Manufacture, Display Topics, Cameras and Video Engineering, Television Recording, Colour Television, Industrial and Scientific Television, and Transmission and Transmission Equipment. About fifty papers will be presented and it is intended that they will all be preprinted so that the maximum time can be devoted to discussion.

Past experience has shown that much of the value of a Convention lies in this opportunity for discussion, and arrangements have been made for special small discussion groups, which will be able to explore the implications of authors' papers after the main discussion has taken place. Another way in which it is hoped to increase still further the value of the Convention is in the provision of a continuous exhibition of working and static demonstration equipment.

The Clerk Maxwell Memorial Lecture has customarily been presented on one of the evenings of an Institution Convention, and this year we are delighted to learn that Dr. Vladimir Zworykin has accepted an invitation to give the lecture. It is particularly appropriate that as a pioneer of electronic television he should participate in a Convention on Television Engineering.

Offers of papers have been received from Denmark, France and the U.S.A. and it is hoped that one or more members of the group of Russian engineers who are expected to attend the Convention will also present a contribution. A preliminary list of the papers to be presented at the Convention will be published in the May issue of the *Journal*.

V. J. C.

INSTITUTION NOTICES

Institution Meeting in Canada

A meeting of members in Ontario is being arranged to explore the possibility of setting up a Section of the Institution in Canada. The meeting, which will be held at the Royal York Hotel, Toronto, on Monday, 1st June, will be under the chairmanship of Mr. G. A. Marriott, Immediate Past President, who is visiting North America during the early part of the summer. Arrangements for the meeting are being made by the Council's representative in Canada, Mr. L. H. Paddle, and separate advice is being sent to all members giving fuller details.

Institution Meetings in May

The following meetings will conclude the 1958-59 programme:—

LONDON

Tuesday, 5th May.

"An Experimental Diode Parametric Amplifier and its Properties," by Dr. I. M. Ross, C. P. Lea-Wilson, A. J. Monk and A. F. H. Thomson.

Wednesday, 13th May.

"Improving Communications Techniques—what have Engineers to learn from information theory?" by Professor D. Gabor, F.R.S.

The above meetings will be held at the London School of Hygiene and Tropical Medicine, Keppel Street, W.C.1, and start at 6.30 p.m.

SOUTH MIDLANDS

Friday, 1st May.

"Transistor Amplifiers," by F. Butler, B.Sc. (Member). At the North Gloucestershire Technical College, Cheltenham, at 7 p.m.

Back Copies of the Journal

The Institution has received requests for the following issues of the *Journal* which are now out of print:—

January 1953
December 1953.

Members who have copies for disposal of these issues, in good condition, are invited to send them to the Brit.I.R.E. Publications Department, 9, Bedford Square, W.C.1; a payment of 5s. per copy will be made.

Insignia Award in Technology

One of the special annual Insignia Awards of the City and Guilds of London Institute for 1959 has been conferred upon Air Commodore William Charles Cooper, C.B.E., M.A. (Member). These Awards are granted to engineers and technologists of distinction in recognition of their outstanding achievements; they are also intended to encourage younger engineers who gain the C.G.I.A. through the usual method of thesis backed up by responsible industrial experience. The latter include five Associate Members of the Institution.

Air Commodore Cooper is at present Chairman and Managing Director of Manlove, Alliott & Co. Ltd., Nottingham. From 1946 to 1956 he was Factory Manager at Ericsson Telephones Ltd, a position he took up on his retirement from the R.A.F. During a distinguished service career Air Commodore Cooper held the appointment of Director of Communications (Research and Development), Ministry of Aircraft Production. A Member of the Institution since 1945, he serves on a number of educational bodies in the East Midlands.

M.Sc. Course in Information Engineering

A graduate course in Information Engineering will be conducted in the Electrical Engineering Department, University of Birmingham, during 1959-60. Suitably qualified students (preferably with some industrial experience) can qualify for the degree of M.Sc. on satisfactory completion of this 12-month course, of which the next session runs from 1st October, 1959, to 30th September, 1960. Subjects available cover Communications, Radar, Computers and Control Systems, with some degree of choice to suit individual requirements.

It is sometimes possible to admit to courses of lectures on individual topics, engineers who can be seconded from industry for a period of five weeks only. Further information may be obtained from Dr. D. A. Bell, Supervisor of Graduate Courses, Electrical Engineering Department, The University, Edgbaston, Birmingham 15.

Computers—The Next Ten Years†

by

ANDREW D. BOOTH, D.SC., PH.D., MEMBER‡

An Address given at the Inaugural Meeting of the Computer Group in London on 4th February 1959.

In the Chair: The President, Professor E. E. Zepler.

Before I start on the provocative subject of the lecture, I want to say one or two words about the Computer Group which the Institution has just formed and of which this is the Inaugural Meeting. When a new group is formed, its function and purpose in life must be carefully determined so that I wondered just what purpose we did have in life apart from the ever-popular one of getting people together to talk—sometimes regarded as the Academic vice! One purpose is to disseminate new ideas between people working in this field, another to provide a means of introducing new recruits to the old ideas which many of us find so obvious that we do not bother to explain.

I feel that one of the rocks upon which the computing machine industry is likely to founder is the inadequacy of existing technicians both to build and to maintain the machines. I speak with a certain bitterness since one of the troubles which besets Universities is the lack of competent technical assistance in the field of electronics. This is because electronics is a popular profession at the present time and rightly commands large salaries, whereas the University technicians' salary scale is geared to the idea of the glass-blower of the 1890's. Now, although the glass-blower has very great manual skill, he does not need the logical skill and knowledge of fundamentals which is essential to modern electronics. For this reason, the Universities, which have in the past been the birthplaces of computing machines, have lost their ascendancy which has passed to large development groups in industry, where technicians and scientists of the required ability can be gathered together and can be paid adequate salaries. This conclusion is not mine alone; it is also that of Howard Aiken, whose work at Harvard I shall mention later. Thus, one of the functions of this Group is to assist in recruiting and training young people for this new and expanding industry and I can say, from very extensive knowledge of the industry, that well trained and able men find the incentive of a wide choice of well paid jobs awaiting them.

Now to the subject of the discourse: the next ten years of computers. Before talking about the next decade, I want to talk about the last ten years which I find a source of disappointment. One of the things which prompted me to give the lecture its title was that a number of people who should know better have, during the last year, given tongue and pen to the statement that "new, second generation computers are with us" and that these machines are better than anything which has been thought of before. The first part of this lecture will attempt to disabuse you of this idea for, far from thinking that any second generation computer exists, I think that we are only seeing the growing up of first generation computers. To justify this statement I remark that, in 1946-1947, I had the great good fortune to work with the late John von Neumann on the logical and physical design of computing machines. At that time there were prepared at the Institute for Advanced Study at Princeton two reports[1, 2] on aims and objectives to which we may look to see the type of computer which was envisaged at that time. This computer was in fact a machine having, in retrospect, certain rather interesting characteristics. The most obvious of these is speed of

† Manuscript received 17th February 1959. (Address No. 17.)

‡ Department of Numerical Automation, Birkbeck College University of London.

U.D.C. No. 681.14

operation and this was desired to be such that a 40-bit addition or subtraction would take about 10 microseconds. There is no computing machine commercially available in this country which achieves this addition time. The multiplication time of the von Neumann machine varied from 400 microseconds for the crudest scheme to 50 microseconds for a more sophisticated logical device which still made no use of the steam-roller electronics to be seen in at least one machine of the present day, which achieves a rather worse performance. So much for second generation speed!

Next the store. We were thinking in 1947 of a high-speed store for 4096 words of 40 bits. I remark that it is only in the last couple of years that machines have been produced with storage organs whose total capacity is of that order of magnitude and many machines which are currently manufactured have high-speed storage for only one hundredth of this number of data. Yet again, early in 1947, we were developing a highly integrated backing storage using magnetic wire. You may say, "Well, wire! That was abandoned years ago!" And although this is true, the important thing is that the system itself was an integrated one, well suited for handling the large quantities of information required in the solution of non-linear partial differential equations. Had it been put into service, it would have been equally suitable for "data processing."

The main reason for the comparative failure of the Princeton machine was that it relied for storage on a secondary emission device called the Selectron.[3] Unfortunately the Selectron, which was to have a storage capacity of 4,096 bits in a tube the size of a milk bottle and an access time of a couple of microseconds, did not materialize and although about four years later I saw a demonstration model, it stored only 16 bits in about the same space and was considerably slower.

A binary parallel adder with the required speed of operation was actually constructed at Princeton in 1947 by Richard Snyder. Unfortunately it contained certain analogue elements which performed the operation of obtaining the sum digit at each stage by means of the algorithm:

Sum digit = Sum of incident digits—twice carry.

The analogue elements required precise setting by means of potentiometers and, for this reason, the scheme was not adopted. It is interesting to speculate on the effect which modern high stability components might have had on this development.

The mercury delay line, which was the first form of high-speed, erasible, storage to be suggested, arose in a non-computer context, that of detecting moving targets by radar. It was first applied to computing machines by von Neumann in the EDVAC[4] and got off to a very bad start. This machine was used as a basis by Wilkes at Cambridge who produced the EDSAC[5]. The EDSAC was the first machine to do signficant calculations, although it was pre-dated in actual operation by the old Manchester Mark I machine, which solved a carefully selected problem in prime number theory. The EDSAC was an immensely creditable achievement for British technology and completely outdistanced American efforts in the same field. Nevertheless the mercury delay line is now a thing which has passed into the limbo of the forgotten as far as new machine design is concerned. There are a number of reasons for this: weight, cost, size and tendency in the long run to get out of order. Despite these defects it has the great charm that an appreciable proportion of the input energy is recovered at the other end, and this does not occur with most other storage devices.

The second of the war-horses is the cathode-ray tube, or "Williams tube"[6]. The late Douglas Hartree expressed to me in 1948 considerable doubt as to whether it was going to be a long-term stayer, although he thought it might be a stop-gap. History has shown the truth of Hartree's judgement, the stop-gap caused great trouble to a lot of people who had to use it and it has now vanished.

The magnetic drum[7, 8] is still with us, it was invented at Birkbeck College as a store for medium-speed machines and is not a high-speed store in any sense of the word, although fast-access storage, at the 100 kc/s rate, can easily be provided either by "revolvers" or by multi-head tracks[9].

The overall picture of present computing machines which have developed over the last ten years is as follows: First, the single storage organ, which I mentioned earlier as the idea of von Neumann, has, except for medium-speed magnetic drum machines been forced into abeyance. The reason is that the only high-speed element that has become available is the magnetic core. The magnetic core matrix contravenes a fundamental principle laid down in 1947: that no large-scale store for a computing machine was practicable if it used physically distinct elements. This *volta face* has been forced upon designers by the march (or rather the backward march) of progress. Fortunately the position is not quite as bad as it appears, since a number of programming studies seem to show that a high-speed store, which some people put as low as 10 words, and none higher than 1,000 words, plus a large magnetic drum, produces a net slowing down of the machine of from 1 per cent. to 20 per cent. according to the optimism and the particular problems under investigation. Thus, present machines consist, not of one store, one arithmetic unit and so on, but of two stores, a high-speed store of limited size and a backing store of between 8,000 and 30,000 words.

In addition to cores and drums there is magnetic tape which is just appearing on British machines but which has been on American machines for some time. The history of magnetic tape in this country is one of muddled thinking about the way in which magnetic tape might improve the performance of a computing machine, and of muddled experimentation on the way in which errors arise from magnetic tape. Magnetic tape has two possible functions in a machine, the first as an input and output medium, and the second as a very large capacity low-speed backing store. As a direct input and output the problem, both here and in America, is that no simple and adequate mechanism for direct input from a keyboard onto magnetic tape has been developed and, with the exception of the Xerographic printer a like remark applies to output. This means that information must be punched on paper tape or cards and processed either through the computer itself or by a special magnetic tape preparing device. Two rivals to magnetic tape are already visible: the file drum—a large device for storing a few million bits on a slow speed but otherwise classical magnetic drum—and the so-called RAMAC or "juke-box memory." As a magnetic drum man, I should obviously favour the file drum and, indeed, to some extent, I do. On the other hand, I cannot help admiring the immense technical virtuosity of a device like a stack of gramophone records which in about a third of a second will move a recording head up to a given level, rotate it into position and drop it on to a prescribed track on a disc. Despite this admiration, however, I cannot help saying that I do not think it will stay in the field for very long. In the same class is the device for shooting magnetic material at a wire matrix, supposedly for storing information for machine translation. It is based on a complete misconception of what is wanted and is, I think, technically unsound.

The file drum is an interesting device which runs at a very slow speed and has an information density of about 1,000 bits/inch. A few years ago I would have said that it was impossible, practically it has turned out to be possible but somewhat difficult and has shown a number of curious and unexpected technical imperfections. I mention only one of these: that apparently the oil in which the device has to run because it uses in-contact heads, develops an electric charge. This produces transient discharges which can give rise to errors. I would not have predicted this since the idea of an oil charging up seems somewhat improbable, but indeed it shows how the pursuit of the solution of one problem sometimes leads to even more difficult ones.

I have mentioned paper tapes and punched cards. Which are best? Both have certain advantages. For example, card handling mechanism is more reliable than that for paper tape at the present time: this is perhaps due to the fact that numerical information has, until recently, been non-redundant, so that, when a punched-card calculator manufacturer produced a machine, he did not put in checking digits, he assumed that his machine was going to work most of the time. With paper tape, in the telegraph industry, the situation was rather different. Probably without realizing it, manufacturers of paper-tape equipment made use of the great redundancy of language, and this meant that they did not have to transmit messages very accurately by numerical standards. I once put this point to a firm of telegraph equipment manufacturers who were horrified at the mention of one error in 10^9 as an acceptable fault rate for computer equipment, since they considered 1:5,000 as good and 1: 50,000 as phenomenal! Paper-tape producing equipment will improve in the future and its cheapness and small size will make it a serious rival to punched cards.

So much for the overall picture of machines. What of the constituent components? There has been a transition from large thermionic valves which were in common use twelve years ago to miniature valves. There was also a temporary excursion into sub-miniatures which seem now to have been abandoned, at least in the computer field. Magnetic cores were next introduced and I was one of the first people to work on these in this country. The conclusion which was reached after about three years' work was that magnetic cores required associated diodes to prevent back circuits and since these diodes could themselves perform the required logical operations, the cores were redundant. My present feeling is that whilst cores have a definite place in computing machines, it is not in doing logic. I think that the function of the core is in the storage of information.

The transistor is the latest component to arrive from America. It is not very recent and has gone through a period of a number of vicissitudes and transatlantic factors. Only two years ago I tested some 400 Mc/s transistors from the United States, but unfortunately for a reason I never discovered, in crossing the Atlantic they turned into 4 Mc/s transistors of the same sort as we have over here!

Semi-conductor diodes are the staple diet of the computing machine designer. They require no heater supply, they are well enough understood so that circuits can be designed without much experiment, they can be easily and reliably manufactured. I think that the diode will eventually replace most other passive elements in future machine. Transistors will be used for power amplification, diodes for computing. Both the diode and the transistor suffer from hole storage effects, but these are likely to be better understood and controlled in the future, and are unlikely to inhibit progress towards faster machines. There is one important way in which junction diodes and transistors differ from thermionic valves: their impedance levels are from 10 to 100 times lower. Low impedance is very important since it means that a computing machine can be built which will be relatively free from pick-up of extraneous noise pulses. The elimination of heater supplies is so obvious an advantage that it hardly requires mention, but I believe that the main impact of solid-state deviceson computer technology will lie in improvement of reliability and not in mere reduction in size. And, although a sprawling monster the size of a room is undesirable, one much smaller than an office desk is equally awkward both to service and to manufacture. The only place where a small computer is really justified is in a moon rocket.

The assessment of reliability is a thorny subject. I could tell you how reliable my machine is—I am sure you would not believe me. You could tell me how reliable your machines are—I am quite sure that I would not believe you either. The literature contains many ambiguous statements concerning reliability and quoted figures vary from 60 per cent. to 99 per cent. I think that these figures show the danger of statistics because, investigating the 99 per cent. reliability one, we find that out of an eight-hour day three-and-a-half hours are spent in "routine" maintenance and I would say that a machine that needs three-and-a-half hours' routine maintenance per day is only about 62 per cent. reliable. It is quite clear that if all the unreliable time is removed, the machine is very reliable for the rest of it! To show the futility of unexplained statistics I mention that I went to see one of the American machines in 1951. It was labouring under a cloud at the time, but I was told that it was working for 50 per cent. of the time. I looked at a unit which clearly was not working at all, because some of the pulses were missing, and the engineer quite glibly said, "Oh, it's working for 50 per cent. of the time—every other pulse!"

To conclude this brief summary of present "fast" machines, I mention that their speeds lie in the range: addition and subtraction 10-200 microseconds, multiplication and division 100 to 10,000 microseconds, and pass on to consider some of the problem which these machines have been used to solve. For example π and e have been calculated to 3,000 places. You may think that this was because of a request by mathematicians, in fact it was not. The first 2,000 digits of π were calculated on ENIAC by the maintenance engineers during a slack weekend. They produced the statistics on the digit distribution in π shown in Table 1: These were so interesting that the mathematicians insisted on the calculation of another thousand places, shown in the second line of Table 1. Next there was the odd affair of the Riemann-Zeta function[11]. Many of you will not have seen the paper where this particular work was published, so it is worth telling the story as recorded by the late Alan Turing, one of the pioneers of ACE[12]. Turing says, in effect: "I have been working for the past few years on the Manchester Mark I computer. Towards the end of this period I had an unusual run of error-free time. I calculated some zeros of the Riemann-Zeta function. Unfortunately my programme was wrong and therefore the results are worthless, but I give them to you for your information." This is an unexpected side-light on machine reliability and demonstrates the honesty and näivety of mathematicians. Calculations with large numbers have also interested computer mathematicians and it has, for example, been shown recently[13] that $2^{2281} - 1$ is composite.

In Engineering, computers have been used for control system stability in reactors. They have also been used to solve the electric traction problem[14] which is that of investigating the performance of a new railway engine from theoretical performance curves. The classical work is that of Gilmour who has devised a programme which simulates engine performance in a manner which would require a trial run at a speed of 6,000 miles/hr to equal it. An interesting by-product of this work is the production, by computer, of optimum running schedules. One example, for a return journey from Liverpool Street to Shenfield has shown how, by replacing an all-out run by a coasting one, the operating cost can be reduced by 40 per cent. for an increase in running time of only 14 per cent.

One aspect of machine use which is becoming more and more clear as time goes on, is summed up in the dictum: "Don't use the machine to produce numbers on paper when in fact it can be used to produce, directly, the results you want." Recently there have been several papers on cathode-ray tube display systems for computing machines. This idea is very old, the first "Numeroscope" or number display on a cathode-ray tube was invented in 1948 and the first use of a cathode-ray tube to display an actual computer output for a physical problem in an acceptable form was, I think, done in my own laboratory in 1951 during work on the crystal structure of Oxalic Acid. The calculations which lead to values of electron density from which a contour map is to be drawn take a computer about two weeks instead of the same number of years for human computation. The actual reduction of these data to a form suitable for plotting takes about one week but, by using the machine to generate contours directly on a cathode-ray tube, the actual plotting was completed in 1/50 second.

Other technical problems solved on computers include the design of optical systems, the improvement of the Tessar lens system, for example, where the aberrations of this system were reduced by 50 per cent. by the use of an electronic computer. Stellar dynamics and the

Table 1

Digit Distribution in π

No. of digits	0	1	2	3	4	5	6	7	8	9
2036	184	213	210	191	198	211	204	200	207	218
3090	269	315	314	276	322	326	311	297	318	342

theory of continuous creation involve an investigation of the way in which stars grow, rise to maturity and then decay. These problems are difficult and occupy some of the fastest modern computers for periods of the order of two or three weeks.

Finally from the sublime to the ridiculous, computers have been used to do P.A.Y.E. They have also been used on a small scale for production control. Accountants, however, are slow to support and quick to criticize. They *may* agree to purchase a machine developed by the efforts of engineers, but usually suggest trivial modifications which earlier participation would have avoided. One American business group bought a computing machine known to be in the early development stage and which did not work before it was delivered. Eventually I saw it in a gleaming glass and chromium room in a Texas factory. Questioned on the uses to which the machine was put, the owners said, "Yes, we are very proud of it. It's not wired up. We only have it to show the customer." I suppose that is as good a use for a computer as any.

So much for the past. To attempt a prediction of the future we look at what is in the laboratories at the present time and to try to see which laboratory devices are likely to be practicable in computing machines. The spin-echo effect of nuclear-magnetic resonance[16] appears to be in the impracticable category, an exciting idea, but not one which I think is likely to come into general use in computing machines. Next the transistor. Here the problems are those of eliminating the hole storage effects and increasing the effective cut off frequency. There are in existence techniques for making "good" transistors but we have been testing transistors for about the last five years and still find that goodness is a very relative term, you can get *a* very good transistor; you may, if you are very lucky, get two very good transistors whose characteristics resemble one another; but to get many identical and good transistors is rather more difficult. This, I think, is the crux of the whole matter and the next ten year period will be one of consolidation in transistors.

Ferrite cores will supply the high-speed storage of computers for probably five years whilst other things are being developed. Ferrite cores are getting smaller, easier to handle, and more reproduceable in production. They form a reliable storage medium in the speed range of 0·5-5 microseconds.

Ferroelectrics[17] have a less rosy future, those known at present suffer from the defect that no field strength exists below which switching never takes place. This implies that co-incident voltage selection techniques cannot be used without periodic regeneration of *all* stored data and this, in turn, makes the ferro-electric matrix far less attractive than the core one. In fairness it should perhaps be said that in gates and shifting registers there may be some sphere of application for these devices.

A more promising new element is the thin magnetic film. Experiments have already shown[18] that information can be written, read and regenerated in times of the order of 20 millimicroseconds. This is faster, by a factor of 25, than the speed attainable with ferrite cores. Perhaps even more important, the speed is such that the "light barrier" is looming up. The light barrier occurs when the time of transit of a signal from one side to the other of the computer becomes appreciable in terms of the arithmetical speed. For example, propagation across a distance of 30 cm takes one millimicrosecond, and several modern computers have dimensions which are from 10 to 100 times as great as the figure quoted. This effect will introduce considerable complexities in computer design and probably means that the ultimate arithmetical speed will not exceed one millimicrosecond for addition. The problems which still remain to be solved for thin films are, firstly, those of reproduceable production of matrices, a technological problem which should not present much difficulty, and secondly, the problem of designing an efficient access system. With speeds of the order mentioned, it becomes difficult to conceive, let alone develop, physical mechanisms for selecting, writing and reading. One possible mechanism involves the use of transmission line techniques and this may have the advantage of resolving the impedance matching problem at the same time. One interesting aspect of thin film techniques is that, despite the small quantity of magnetic material present in any storage element the output produced is still quite large (20 mV in some cases).

This is a direct consequence of the fast rise times of the interrogating pulses. It may be argued that thin magnetic film techniques will go the same way as ferroelectrics, but in my opinion this is not likely to occur because the basic physical phenomena of thin films are well understood and are known to be sound.

For the second quinquennium two other things appear on the horizon and are, I think, significant. They are the "Twistor"[19] invented by Andrew Bobeck at the Bell Telephone Laboratories and the "Tensor"[20]. Both belong to the class of devices which make use of certain physical properties of continuous media. One form of Twistor is shown in Fig. 1. A nickel wire, AB, a few thousandths of an inch in diameter, is twisted about its axis by means of a permanently applied external couple. The resulting maximum of compressive stress will then lie at an angle of 45 deg to the wire axis in a screw sense which is the same as that of the applied couple. For a strain-sensitive material, such as unannealed nickel, the preferred direction of magnetization will follow this direction of maximum compression so that the flux path

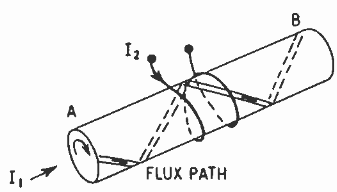


Fig. 1. The "Twistor."

is as shown. When there is sufficient difference between the ease of magnetization axially and along the circumference of the spiral the analogue of co-incident current core storage can be obtained. To do this simultaneous currents, I_1 and I_2, are applied along the wire and to an auxiliary coil as shown, and at coincidence currents of suitable magnitude will switch the wire from one state to the other; separately little or no change results. An alternative co-incident current system uses two auxiliary coils and no passage of current through the wire.

Read-out can be obtained either at the auxiliary coil by switching the flux by means of current applied to the wire itself, or from the wire by pulsing the auxiliary coil. With two auxiliary coils co-incident current read out is obtained. Results already obtained suggest storage capacities of 10 bits/inch on 0·001 in. molybdenum-permalloy wire and a switching time of 0·2 microsecond.

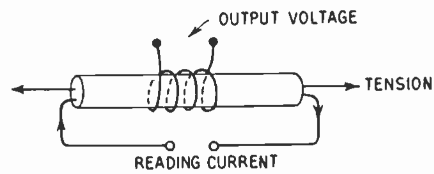


Fig. 2. The "Tensor."

The Tensor store of Gianola, shown in Fig. 2, depends for its action upon the fact that any magnetic memory depending upon remanence can be interrogated non-destructively by means of an orthogonal perturbing field. With isotropic storage materials, however, there is considerable irreversible flux change for the first few tests. The Tensor introduces anisotropy by stretching the wire. Information is introduced via the auxiliary sensing coil and the same coil generates the output voltage during interrogation, the perturbing field being introduced by passing a current pulse through the wire itself. Coincident current techniques can also be used for writing-in although these would appear to need wires of special magnetic materials. The preliminary experiments were made using 4-79 permalloy tape and here signals of 1-4 millivolts per turn were recovered for read-out times of about 1 microsecond. It is suggested that the information density will be 50 wire diameters per bit.

Another of the more exciting things which magnetics holds for the future has appeared quite recently in the physical literature. It arises from work done by Ludwig Mayer[21] in the United States on the direct reading of magnetically recorded information by means of electron-mirror microscopy. In this work we may be seeing the future equivalent of the magnetic drum. The reading process is a static one which depends upon the deflection of the electron beam in an electron microscope by the magnetic field of the data recorded on a magnetic film of conventional type. One advantage of the process is that the inherently great magnification of the electron microscope makes the reading of closely-packed data easy, another advantage is that the reading element is an electron beam which can be deflected electrically and thus facilitates high speeds of operation. Naturally the advantages of non-destructive read-out are still present as with the magnetic drum.

Even if this were the only feature of Mayer's work it would be exciting, but he has also shown[22] how information may be written onto a magnetic medium by means of an electron beam. This is achieved by magnetizing the medium initially by means of an external field, and then heating small regions above the Curie point by means of a sharply focused electron beam. When the material cools below the Curie point again an inversion effect takes place and this enables the electron mirror microscope to detect the previously heated spots. Mayer suggests that information densities of 10^5 bits/cm^2 will be possible and that the writing speed might be 10^5-10^6 bits/sec.

Even more exotic elements for storage and arithmetical units of the future depend upon superconductivity. The first of these to be described was the Cryotron of Dudley Buck[23]. This device makes use of the fact that in the presence of an applied magnetic field the temperature of incidence of superconductivity of metals is reduced.

It happens that Niobium is less sensitive to the destruction of its superconductivity by applied fields than is Tantalum and this makes possible the Cryotron flip-flop shown in Fig. 3.

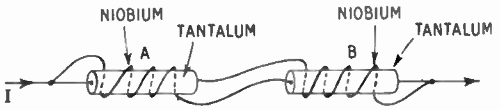


Fig. 3. Cryotron flip-flop.

The operation of this device is as follows. The circuit is cooled to 4·2° K is boiling liquid helium. In the absence of current all of the materials are superconducting but, as the applied current, I, is increased, a point is reached at which the field generated by the Niobium forces the Tantalum wire to assume its normal, ohmic, resistance. Because of small asymmetries in the layout this will occur in one side of the circuit before the other. If this is assumed to be side A in the diagram, the increased resistance of the associated Tantalum wire will keep the current in the Niobium wire of B below the de-superconductivity point. The arrangement is thus stable. From symmetry the state in which B is superconducting and A is normal is also stable so that the device behaves as a flip-flop.

It is clear that the flip-flop can be set and reset by means of gates constructed in the same manner from Niobium coils on a Tantalum wire, so that all the pre-requisites for a computing machine are present.

Unfortunately the speed of operation of the original Cryotron is relatively slow (10^{-3} sec) but by vacuum deposition of thin film elements it is claimed that this figure can be decreased to 10^{-6} sec or less.

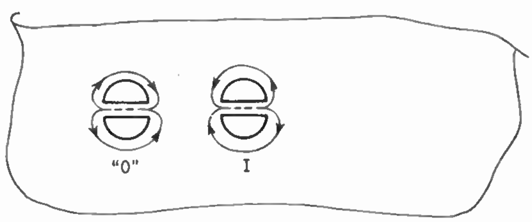


Fig. 4. Superconductive store.

The Cryotron, in this form, is not directly suitable for a large-capacity store but two variants have been proposed for this purpose by I.B.M., the Ramo-Wooldridge Corporation and Duke University. One of these is shown in Fig. 4.

A thin superconducting film is vacuum-deposited upon an insulating substrate. In this film are "D" shaped openings separated by a narrow bridge. Data are inserted and read by means of wires along the bridges and to which current pulses are applied. It is clear that a current can flow in one of two directions along a bridge and that such circulating currents can be produced by the magnetic fields generated by the wires just mentioned.

It is claimed that operating times of 10^{-8} sec can be attained by the "Persistor," as it is sometimes called.

The importance of superconducting elements is likely to increase as satellites and spaceprobes become more common, because of the natural insulating properties of the interplanetary vacuum. Even without this incentive, however, superconductivity is an attractive element for large computers because of the relatively small power dissipation. It is worth noting that the helium which is used can be recovered, liquified, and used over and over again.

A depressing commentary on so-called progress is that von Neumann and the author discussed a superconducting storage element, almost identical with that shown in Fig. 4, at an evening party in 1947. Unfortunately, by morning our enthusiasm, like the liquid helium, had evaporated!

I have talked a lot about future computer elements but I have been deliberately vague about circuitry and logic because I think it unlikely that extensive advances will be made in these fields. There has been some talk lately about a supposedly new idea of computers which will interrogate various priority stations and break off from doing a bread-and-butter job in favour of a priority job. In fact this has been envisaged for many years and involves nothing new. No special instructions are really needed since all existing computers can be programmed to do just this by the use of branch instructions and the normal input or inputs. It may be that it is better to provide special facilities on big installations, but this is not by any means necessarily the case and no new idea is involved.

In the general field of input/output equipment clearly magnetic tape will come into general use. Xerography, too, is likely to be applied both to printing and to replacement of punched paper or film. A second form of recording "Digitape" based on "Teledeltos" paper is, I think, less likely to be perpetuated because of its expensive basic material. Devices for the direct reading of printed characters are, at present, moderately unreliable. The imperfections of print, and especially of typewritten print, present quite considerable technical difficulties to the designer of a machine which, independently of context, will recognize any given printed character with great accuracy. By using context the problem is rendered easier and such an extension of the basic character reader is almost inevitable. Direct recognition of the spoken word is on the horizon. One possible application is to stocktaking where a pocket recorder and lapel microphone would leave hands free for inspecting the stock. This appears to save two operators, one to check and one to write. However, on suggesting this particular application to a firm, I was told that the second operator, although he is apparently redundant, is actually learning the job, so that the advantage of spoken word recognition is largely illusory.

Machine translation of language is another application of the future. The Russians are devoting considerable effort to it and I suspect that if they produce a working translator, the Americans will not be long to follow. I am not quite sure who will get there first, but it is comforting to know that we have done all of the preliminary work and have shown that translations of about 90 per cent. accuracy from technical French are possible and that the same principles apply, with greater or less complexity, to any other language.

There is much talk of auto-codes nowadays and there is no doubt that they are a useful adjunct to a computing machine. Personally, I have always been a little bit sceptical and find them unnecessary. Psychological experiments have been performed which show that, although you can teach a student an autocode, he is almost as inaccurate in it as he is with a normal computer code. However, autocodes remove some of the burden of mathematical formulation from the programmer and, to this extent, they speed up the work and will find an ever increasing place in the computing laboratory of the future.

Finally, computing applications of the future. I think that Science and Technology will occupy a smaller *proportion* of available computing time as the art develops: the exciting avenues of progress are going to be in non-numerical applications. Machine translation is one of these and may be developed on a large scale to cope with the vast masses of technical information which are accumulating. "Cradle to grave" economic planning is another large scale prospect and preliminary surveys have already been made at the National Physical Laboratory. Newly-born children will be given a number, this will be traced throughout life: if the person moves or tries to escape by altering his name, the computer will be able to re-construct it—the 1984 of George Orwell is at hand!

It is not worth describing other probable applications in detail since they are readily predictable from present trends, but one general principle forms a worthy note upon which to conclude: whatever the application, the machine should always produce the result which is required and not a mass of paper for human consumption. One example of this is already to hand in the computer-controlled machine tool, others, such as the automatic patent office and the computer-controlled airport, are likely soon to be with us.

References

1. A. Burks, H. H. Goldstine and J. von Neumann, "Preliminary Discussion of the Logical Design of an Electronics Instrument" (Princeton, 1946).
2. A. D. Booth and K. H. V. Britten, "General Considerations in the Design of an All Purpose Electronic Digital Computer" (Princeton, 1947).
3. A. D. Booth and K. H. V. Booth, "Automatic Digital Calculators," p. 145, 2nd Ed., (Butterworths, London, 1956).
4. J. von Neumann, "First Draft of a Report on the EDVAC," (Moore School, Philadelphia, 1945).
5. M. V. Wilkes and W. Renwick, "An ultrasonic memory unit for the EDSAC," *Electronic Engineering*, **20**, pp. 208-213, July 1948.
6. F. C. Williams and T. Kilburn, "Storage system for use with binary-digital computing machines," *Proc. Instn Elect. Engrs*, **96**, pp. 81-100, 183, March 1949.
7. A. D. Booth, "A magnetic digital storage system," *Electronic Engineering*, **21**, pp. 234-238, July 1949.
8. Ref. 3, pp. 122-132.
9. A. D. Booth, "The M.2. Computer," *Madrid Conf. Proc.*, 1958, (in press).
10. "Some comments on a NORC computation of π," *Math. Tables Aids Comp.*, **9**, pp. 162-164, 1955 and "An ENIAC determination of π and e to more than 2,000 decimal places," *Math. Tables Aids Comp.*, **4**, pp. 11-15, 1950.
11. A. M. Turing, "Some calculations of the Riemann zeta-function," *Proc. London Math. Soc.* (3). **3**, pp. 99-117, 1953.
12. J. H. Wilkinson. "A discussion on computing machines," *Proc. Roy. Soc., A.*, **195**, p. 265, 1948.
13. R. M. Robinson, "Progress in finding large primes," *Math. Tables Aids Comp.*, **11**, pp. 265-268.
14. A. Gilmour, "The application of digital computers to electric traction problems," *Proc. Instn Elect. Engrs*, **103B**, (Supplements), pp. 59-67, 1956.
15. Ref. 3, pp. 216-218.
16. A. G. Anderson *et alii*, "Spin echo serial storage memory," *J. Appl. Phys.*, **26**, pp. 1,324-1,338, November 1955.
17. M. Prutton, "Ferroelectrics and computer storage," *J. Brit.I.R.E.*, **19**, 93-100, February 1959.
18. E. M. Bradley, Private communication.
19. A. H. Bobeck, "A new storage element suitable for large size memory arrays," *Bell Syst. Tech. J.*, **36**, pp. 1,319-1,340, November 1957.
20. V. F. Gianola, "Non-destructive memory employing a domain oriented steel wire," *J. Appl. Phys.*, **29**, pp. 849-853, 1958.
21. L. Mayer, "Electron mirror microscopy of patterns recorded on magnetic tape," *J. Appl. Phys.*, **29**, pp. 658-660, April 1958.
22. L. Mayer, "Magnetic writing with an electron beam," *J. Appl. Phys.*, **29**, pp. 1,454-1,456, 1958.
23. D. A. Buck, "The cryotron—a super-conductive computer component," *Proc. Inst. Radio Engrs*, **44**, pp. 482-493, April 1956.

ATTENDING THE 1959 CONVENTION . . .

Dr. Vladimir K. Zworykin will give the Clerk Maxwell Memorial Lecture in Cambridge during the Convention. Dr. Zworykin received the Electrical Engineering Degree of Petrograd Institute of Technology in 1912, and subsequently carried out post-graduate work in Physics at the College de France with Paul Langevin. After spending a short time in England he settled in the United States, and in 1929 joined the Radio Corporation of America as head of its electronic research laboratory. He has been responsible for original and significant work in the perfection of television as well as for devices such as the secondary emission multiplier and the electron microscope. Since 1954 Dr. Zworykin has directed the Medical Electronics Centre at the Rockefeller Institute of Medical Research, and has been particularly concerned with the extension of electronic methods in medicine and biology: he took a leading part in the establishment of the International Conference on Medical Electronics, which meets for the second time in Paris just before the Brit.I.R.E. Convention. Dr. Zworykin's work has received many awards from Governments, academic bodies, and technical institutions all over the world.

Mr. Victor J. Cooper (Member) is the Chairman of the 1959 Convention Committee. He has held the position of Chief Television Engineer with Marconi's Wireless Telegraph Company since 1956 and has indeed been with that Company for 23 years. Since 1947 Mr. Cooper has been closely concerned with the advanced development of professional television equipment, including the 50 kW B.B.C. transmitter at Holme Moss and the Crystal Palace transmitter. He has also worked on studio equipment both for black and white and colour television.

Mr. Cooper, who was educated at Sir Walter St. John's School and at City and Guilds College, Imperial College of Science and Technology, was elected a Member of the Institution in 1953. He holds over 40 patents and is the author of 15 papers in technical journals. These include a paper in the *Brit.I.R.E. Journal* on "New Amplifier Techniques" which was read at the 1951 Convention and subsequently received the Marconi Premium.

Professor Sisir Kumar Mitra, D.Sc., F.R.S., is Emeritus Professor of Physics of Calcutta University; he occupied the University's Khaira Chair of Physics from 1923 to 1935 and the Sir Rashbehary Ghose Chair from 1935 until his retirement four years ago. Professor Mitra introduced the study of radio as a post-graduate subject in Calcutta University nearly thirty years ago, and he was largely responsible for the foundation of the Institute of Radiophysics and Electronics at the University. He has initiated and led much of the work on ionospheric research in India and is author of "The Upper Atmosphere"—the first treatise on this subject. In 1958 his contributions to this field of knowledge were recognized by election as a Fellow of the Royal Society.

The many official appointments held by Professor Mitra include chairmanship of the Radio Research Committee (Council of Scientific and Industrial Research), and presidency of the Indian Science Congress Association in 1955, and of the National Institute of Science of India for this year. He has been a Member of the Brit.I.R.E. since 1952 and serves on the Institution's Indian Advisory Committee. Professor Mitra has been invited to represent Indian members at the Convention and will himself give a special lecture.

APPLICANTS FOR ELECTION AND TRANSFER

As a result of its March meeting the Membership Committee recommended to the Council the following elections and transfers.

In accordance with a resolution of Council, and in the absence of any objections, the election and transfer of the candidates to the class indicated will be confirmed fourteen days after the date of circulation of this list. Any objections or communications concerning these elections should be addressed to the General Secretary for submission to the Council.

Direct Election to Full Member

ATHERTON, Leonard, B.Sc.(Hons.). *West Wickham.*
GRAY, Bernard Francis, B.Sc.(Eng.). *Hatfield.*

Transfer from Associate Member to Member

BENT, Ronald Albert. *Rugeley, Staffs.*
MORGAN, Group Captain Sydney George, R.A.F. *London, S.W.3.*
NORMINGTON, Alfred Charles, B.Sc.(Eng.). *Manchester.*
SIMS, Hubert Vincent. *Evesham.*

Direct Election to Associate Member

ADHIKARI, Birendra Mohan, M.Sc., B.Sc. *London, W.2.*
ARMSTRONG, Sqdn. Ldr. Robert Leslie, R.A.F. *New Malden.*
COX, Flt. Lt. Ronald Mochrie, R.A.F. *High Wycombe.*
GAYLER, Sqdn. Ldr. Leslie Ward, R.A.F. *Woodbridge.*
GREEN, Flt. Lt. Frederick Henry, R.A.F. *Chester.*
GRIFFITHS, Major Clifford, St. John, R.Aust.Sigs. *New Milton.*
HALL, Robert James. *High Wycombe.*
HARKER, Major Robert Kenneth, Sommerville, R.A. *Anglesey.*
HART, Francis. *Sidcup.*
HULME, Vernon Boyce, M.A.(Cantab). *Ashford, Middlesex.*
KERRY, Lieut-Com. Herbert Sidney, R.N. *Sheerness.*
MIDDLETON, Frank Martin. *Crewe.**
PASCOE, Sqdn. Ldr. Donald Thomas Charles, R.A.F. *Wallington.*
PECK, Lt. Col. Ivor William, R.A. *Swindon.*
PRIDHAM, Gordon John, B.Sc.(Eng.) Hons. *Cheshunt.*
RICHARDSON, Colin Clark, B.Sc.(Eng.). *Reading.*
ROSENBROCK, Howard Harry, B.Sc., Ph.D. *Leatherhead.*
SHOTTON, Alan Carr, B.Sc. *Upminster.*
SMITH, Thomas. *Pendine.*
TEAGUE, Bryan Thomas, B.Sc.(Hons.). *Birmingham.*
TWEEDIE, Sqdn. Ldr. Kenneth Andrew, R.A.F. *Ewell.*
TULLEY, Kenneth Roland, B.A.(Hons.). *Rugby.*

Transfer from Graduate to Associate Member

BELCHER, John Charles. *Shipley.*
DAY, Lieut. Geoffrey, R.N. *Winchester.*
LOB, Gideon Kurt. *Ramataim, Israel.*
NEWLING, Robert Jefferson Grant. *Cambridge.*
THORN, Roger. *Newbury.*
WARD, Reginald Havelock. *Crawley.*

Transfer from Student to Associate Member

DAVIS, Frank. *Wolverhampton.*

Direct Election to Associate

BARLOW, Harry Campbell. *Basingstoke.*
BERMINGHAM, Thomas. *Karachi.*
CLARK, Sidney Joseph, B.Sc. *Harlington.*
HEREDGE, Ronald William Vincent. *Loughton.*
INGHAM, Sidney Alexander. *Peterborough.*
LEE, John. *Cheltenham.*
MOORES, Captain Peter Roderick, R. Sigs. *Catterick.*
OSBORNE, Mostyn Arnold. *Sierra Leone.*
THOMAS, Cyril Robert. *Loughborough.*
TOOKE, Peter Edward. *Selsdon.*
VALE, Dennis William. *Ilford.*
WHITLEY, Richard Harold. *Singapore**.

Transfer from Student to Associate

BURKILL, Arthur Herbert. *Lagos.*

Direct Election to Graduate

ABRAM, Lieut. Thomas Richard, B.Sc., R.N. *Hants.*
AGRAWAL, Plt. Off. Prem Prakash, M.Sc., I.A.F. *Madras.*
BALDWIN, David Arthur. *Hounslow.*
CASE, Raymond Ernest. *Reading.*
CHESTERS, John Lhind. *Liverpool, 13.*
COOMBS, Norman Wycherley. *Edgware.*
DUXBURY, Frederick Grunwell, B.Sc. *Liverpool, 14.*
EMMETT, Stanley James. *Reading.*
GURR, Major Arthur Roy, R.A. *Westerham.*
KAY, Timothy Roger. *Littleborough.*
LITTLE, James *Liverpool, 6.*
MASON, Christopher James Woodville. *Ewell.*
MILTON, Robert Leonard. *London, S.E.10.*
MONGER, Kenneth Sydney. *Weston-Super-Mare.*
MOODY, Ronald Reginald. *Waltham Cross.*
PARAMITHAS, Christos. *New Haven, Conn.*
PRICE, Peter John, B.Sc. *Radlett.*
ROSS, Alan, B.Sc. *Auckland.*
SAINS, Harry, B.Sc.(Elec. Eng.). *Bishop Auckland.*
TAYLOR, Derek, B.Eng. *Wigan.*
THURBIN, Patrick John. *Hanworth.*
WHITEMAN, Donald Arthur. *Bexhill-on-Sea.*
WILLIAMS, Roger Morley. *Taplow.*
WYNDHAM, Brian Alexander. *Malvern, Worcs.*

Transfer from Student to Graduate

BHATTACHERJEE, Amal Kumar, B.Sc. *Bangalore.*
BROOKS, William Gilbert Ernest. *Ontario.*
CHESTER, Michael William. *Beccles.*
COLLINS, James Edward. *Basildon.*
DUNNETT, Paul Wesley. *Aldershot.*
HERLEKAR, Balvant Vishnu, I.A.F. *Agra.*
HOWES, Bentley Arthur. *Thornton Heath.*
HUBBARD, Raymond Thomas. *Oranjemund, S.W. Africa.*
KAMALJIT, Plt. Off. Singh, B.A., I.A.F. *Bangalore.*
KANAAR, John Christopher Norris. *Bromley.*
NARASIMHAN, Villiambakkam Venkatachari. *Madras State.*
ROY, Biman Bihari, I.A.F. *Kharagpur.*
SOOD, Omkar Nath. *Dehra Dun.*
TEWARI, Man Haran Kumar, M.Sc. *Jaipur.*

STUDENTSHIP REGISTRATIONS

The following 33 Students were registered at the February meeting; the names of the 40 Students registered at the March meeting, will be published later.

BAKER, Richard Sahagun. *Ilford.*
BANERJEE, Asoke Kumar, B.Sc. *Howrah.*
BEAN, Arthur Derek. *London, S.W.18.*
CHAKRAVARTY, Amarendra Nath, M.Sc., B.Sc. *Calcutta.*
CHONG KWONG FAYE. *Singapore.*
COIA, Henry Joseph. *Montreal.*
DANE, Alan Arthur. *Dagenham.*
DIXIT, Ganesh Vishwanath. *Poona.*
DRIVER, Desmond John. *Salisbury, Southern Rhodesia.*
EVANS, Ronald H. J. P. *Shrewsbury.*
GOGIA, Mahinder-Pratap. *London, S.E.27.*
GOVINDAYYA, Jilla, B.E. *Poona.*
HARVEY, Edwin. *Cheltenham.*
KENNETT, Barrington George. *Paignton.*
McKEON, Donald George. *Kuala Lumpur.*
MALTBY, David. *Welwyn Garden City.*
MARCHBANK, Phillip Campbell. *Vernon, British Columbia.*
MUKERJEE, Jyoti Kumar. *London, N.W.3.**
MULCAY, Michael. *London, N.9.*
NOAKES, Michael Ernest. *Eastbourne.*
OAKES-JONES, Clifford. *Wellington.*
POULSON, Barrie Kenneth. *Watford.*
PURCHASE, Colin Gordon. *Hockley.*
PURI, Devindra Nath, B.Sc. *London, S.E.27.*
REINER, Joseph. *Tel-Aviv.*
ROBINSON, John David. *Hornsby, N.S.W.*
SAMPSON, Peter Charles Frederick. *London, S.E.22.*
SARATSIOTIS, Panayiotis. *Athens.*
SREEDHARA, Mijar Kanakabettu, B.A. *Bangalore.**
STERRY, Peter John. *Basingstoke.*
TOWNSEND, David Bryan Kevin. *Wolverton.*
TRUESDALE, William R. *Newport, I.O.W.*
WILD, William George. *Sidcup.*

* Reinstatement.

Theoretical Aspects of Mechanical Speech Recognition†

by

Professor D. B. FRY, PH.D.‡

A paper read before the Institution in London on 28th January 1959.

In the Chair: Dr. A. D. Booth (Member).

Summary: The human listener is able to take in speech wave-motions and to perform a variety of operations based on them. As a device for transforming speech wave-motions into typescript, the listener is virtually free of errors in a great diversity of conditions. His recognition of speech sounds is based on the use of a language system, that is a system of linguistic units (phonemes, morphemes, words and sentences). In taking in speech, he makes use of his knowledge of the constraints operating in his language, and thus resolves uncertainties and corrects errors arising at the level of acoustic recognition. The redundancy of speech is also exploited at the level of primary or acoustic recognition. A mechanical speech recognizer needs to simulate these two features of the human mechanism if it is to achieve even a small fraction of the flexibility and accuracy of the latter; it must carry out acoustic recognition by inspecting wave-motions in a variety of ways and must then apply statistical knowledge to the results of acoustic recognition.

1. Introduction

Speech communication fulfils a great diversity of functions in human life generally; it may form a direct expression of the speaker's mental and emotional state, may reveal much about the history of the individual speaker and may be a powerful factor in regulating the behaviour of the speaker himself and of other people. It is the last of these functions that is of major interest when we begin to think about the mechanical recognition of speech. A human listener can take in long trains of speech wave-motions and is able, as a result, to perform a great variety of operations, some of them concerned with the message itself and others that consist in carrying out instructions embodied in the message. The listener may, for example, speak back the message, he may write or type it or produce a phonetic transcription of it. In all these cases, he has to decode and to re-code the message received in the form of sound-waves. When he carries out instructions conveyed in the message, he again has to decode it but here any re-coding that occurs is carried out in terms of behavioural patterns not directly connected with speech.

Mechanical speech recognition has applications of both these types. It may be of service in transmission systems, particularly in analysis-synthesis telephony; it might for some purposes be useful to produce a typed output mechanically from a speech input and it would certainly be valuable as an element in some control systems, for example in automatic telephone dialling and in feeding instructions into computers. This paper is concerned with the theory of mechanical recognition and a consideration of these functional applications will therefore be left to the supporting paper§. For the present purpose, a single example of the use of a mechanical speech recognizer will serve throughout. This is the case in which a mechanical recognizer is used to convert speech input into a typed transcript. It must be stressed that this application is chosen for the sake of clarifying the discussion and not because it represents the most immediate objec-

† Manuscript received 17th January 1959. (Paper No. 495.)

‡ Department of Phonetics, University College, London.

U.D.C. No. 534.782

§ P. Denes, "The design and operation of the mechanical speech recognizer at University College, London," *J. Brit.I.R.E.*, **19**, pp. 220-230, April 1959.

tive or the most profitable direction for present research. Experience has made it clear, indeed, that this use of the recognizer is probably the most remote and possibly the least fruitful of all.

A human listener, then, can take in speech wave-motions and can, in suitable conditions, type what he hears. A mechanical speech recognizer would perform an analogous function if it accepted the sound-waves of speech, processed the incoming information in some way and yielded as the output a set of instructions that would operate typewriter keys in the appropriate sequence. The success of such a device could be measured by counting the correspondences between the letter sequence that it gave and the sequence given by a human operator in response to the same input. It will be well, at this stage, to leave aside a consideration of the difficulties that are peculiar to English. The English listener recognizes sequences of English sounds; when he types the message he re-codes the information in accordance with the rules of English spelling. The question whether these rules could be built into a mechanical recognizer is a separate one from that of making correct recognitions of the sounds. The present discussion will be limited to the problem of correct sound recognition so that the objective may now be re-stated as the design of a mechanical recognizer that will give a typed phonetic transcript of the speech input, that is to say a sequence of letters having a one-to-one correlation with the sequence of phonemes constituting the input.

2. The Variability of the Speech Input

The most important feature of the human listener as a device for transforming sound waves into phoneme recognitions is his very great reliability in the most diverse conditions of reception. Speech is taken so much for granted in everyday life that we notice it only when something goes wrong. In fact the listener is virtually free from errors even in the face of very unfavourable signal/noise ratios or of grave distortion of the speech, and, most important of all, despite the extreme variability of the speech input. This order of reliability is at present very far from being possible in a mechanical speech recognizer, largely on account of the variability of the input. If we take the case of a single English speaker and examine the wave-motions he sends out when uttering the same word on different occasions, for example, the word *no*, we see that the physical output will show great variation. The mean intensity level will be different for different utterances and so too will the variation in intensity within the word. The frequency of the vocal cord vibrations, that is the fundamental frequency of the acoustic output, will vary more or less continuously during an utterance and the range of these variations will differ widely on different occasions. As a result of these changes in overall intensity and in fundamental frequency, the time pattern of spectral distribution of energy will vary with the utterance: the frequency and relative amplitude of harmonics, the position and relative amplitude of the formants will be different for different utterances. Yet in spite of this variability, the listener is able on each occasion to recognize the same word, *no*.

If we now extend the scope of our examination to cases where the listener takes in the speech of many different speakers on different occasions, the variability of the acoustic stimulus for the same word, *no*, is now many times greater still. We have not only the range of variation already indicated for one speaker but, in addition, the changes introduced by the speaker's personal voice quality and above all by his particular pronunciation of the word. This may cover a very wide range indeed from the wide diphthong of Cockney pronunciation to the close vowel of Scottish speech. Again, and even more remarkably in the face of these variations, the listener is still able to recognize the same word. It appears then that he is able to take in a great range of stimuli which are acoustically different and to assign them to a single class.

In other cases, and particularly when dealing with the speech of different speakers, the listener produces the opposite effect: he takes sounds which are acoustically similar and nonetheless assigns them to different classes. A sequence that we may transcribe phonetically as /lain/ may in one speaker be the word *line* and in another speaker the word *lane*. The listener is able to assign these similar stimuli to different classes in the two samples of speech. Or again in the speech of different speakers, the range of sounds used in distinguishing *men* and *man* may show considerable overlap: the listener may encounter two very similar vowel sounds, one of which occurs in the singular form for one speaker and the other in the plural form for the second speaker. Once more the listener's recognitions are free of errors and in this second case it is quite obvious that he cannot be operating on a purely acoustic basis.

It is the fact that for successful speech recognition many different acoustic inputs have to be classed together and that on occasions similar inputs have to be classed separately that constitutes the central difficulty of mechanical speech recognition. In order to discover whether this difficulty can be overcome, at least to the extent of making mechanical speech recognition a practical possibility, we must now look more closely at the way in which the listener processes the incoming sound waves of speech. In order to decode the incoming message, the listener classifies the sounds that he hears. What system of categories does he use and how does he carry out the task of classifying the acoustic inputs?

3. Linguistic Systems

The categories are those which form the linguistic system of the particular language used for communication. Any language is essentially a system of units of different magnitude which combine together according to a set of laws peculiar to that language. In spoken English (and it must be remembered that in the whole of this discussion we are concerned only with the audible form of the language), the smallest unit that need concern us is the *phoneme*, which corresponds in magnitude to the single letter of the printed language. Thus the word /kat/ consists of three phonemes, the word /stiks/, of five, and the word /bro:dka:st/, of eight. The phonemes have no other function than to act as substitution counters in the language system: each phoneme has no meaning in the everyday sense of the word, but it has a function in the system. If in an English sequence of phonemes we replace one member of the system by another, then we introduce a change in the meaning of the whole sequence, and by applying this criterion we can arrive at an estimate of the total number of units in the system. We may have, for example, the following three-phoneme sequences which mean different things in English:

hi:d
hid
hed
had
ha:d
hod
ho:d etc.

In all these, the phonemes occupying positions 1 and 3 are the same in each case, and there is a commutation in position 2. Since the change in position 2 involves a change of meaning in all these sequences, we can state that these are different phonemic units. The middle units in this type of English three-phoneme sequence are the vowel phonemes and by continuing the series of commutations (not necessarily with the context /h – d/ only) we can arrive at an exhaustive list of English vowel phonemes, of which there are about 20. The actual number depends upon the particular regional variety of English pronunciation that we are dealing with.

By a similar process involving series like:

pig	and	pip
big		pit
dig		pik
fig		pin
rig		piŋ
wig etc.		pil etc.

we can obtain an exhaustive list of English consonant phonemes, of which there are about 24. The total number of English phonemes is therefore of the order of 40 and any sequence uttered in English can be represented as a succession of phonemes from this total repertory.

The next linguistic unit in order of size is the *morpheme*; phonemes combine together to make morphemes, which have mainly a grammatical function in the language. For example, the phoneme sequence /stik/ makes up a morpheme with a verbal function [stik]. To this we add another morpheme [iŋ] if the present participle is required, or [s] for the third person of the present tense, [i] to make an adjective, and a third morpheme [li] to make an adverb from the adjective, or [nes] to make a noun from it.

The morphemes are combined together to make *words* and the words join together to make *sentences*. Both of these are units with which the ordinary speaker is quite familiar; it is only necessary to note that in the spoken language any complete remark forms a sentence, and we may have sentences consisting simply of "No," or "Yes," "Not very," "All right" and so on.

4. The Redundancy of Speech

These four levels, that of the phoneme, the morpheme, the word and the sentence, constitute the whole linguistic system of English. We have seen that the total number of phonemes in the system is about 40; the total number of possible morphemes and words is very much greater, of the order of tens of thousands, and of possible sentences, very much larger still—hundreds of thousands. The whole of speech communication is a stochastic process: at any position in the sequence, the probability of occurrence of any unit is dependent upon the units which precede that position. At the phoneme level, for example, in a sequence where n phonemes have already occurred, the probability that any phoneme will appear at position $(n+1)$ is not equal for all phonemes in the repertory and is not independent of the order of the n phonemes. In information theory terms, speech as a communication process exhibits redundancy, and the degree of redundancy is in fact very high. What is true at the phoneme level is true also at the other levels; just as the occurrence of a phoneme is constrained by the phonemes that precede it, so the appearance of a morpheme, word or sentence is constrained by the preceding morphemes, words or sentences.

When a listener takes in a spoken message, he has available a complete knowledge of the linguistic system whose general character we have just outlined. He knows all the phonemic units that make up the system; he knows virtually all the morphemes and words that can occur and all the possible sentences and he also knows the laws governing the combination of units at each level. This knowledge is so thorough that the listener knows a great deal about the sequential probabilities that obtain at any point in the sequence and at any level. All this information is collected in the process of learning to speak a language and in general we remain quite unaware of its existence. Shannon[1], in his earlier work on information theory, evolved a method of demonstrating the existence of this knowledge which he applied to printed English; it can be used equally well with regard to the spoken language. The experiment[2] consists in selecting a spoken English sentence and asking an English listener to guess the succession of phonemes that occur in the sentence, without knowing anything about the subject-matter or the context of the sentence. He begins by suggesting one after another the phonemes that might fill the first position, and he is told as soon as he suggests the right one.

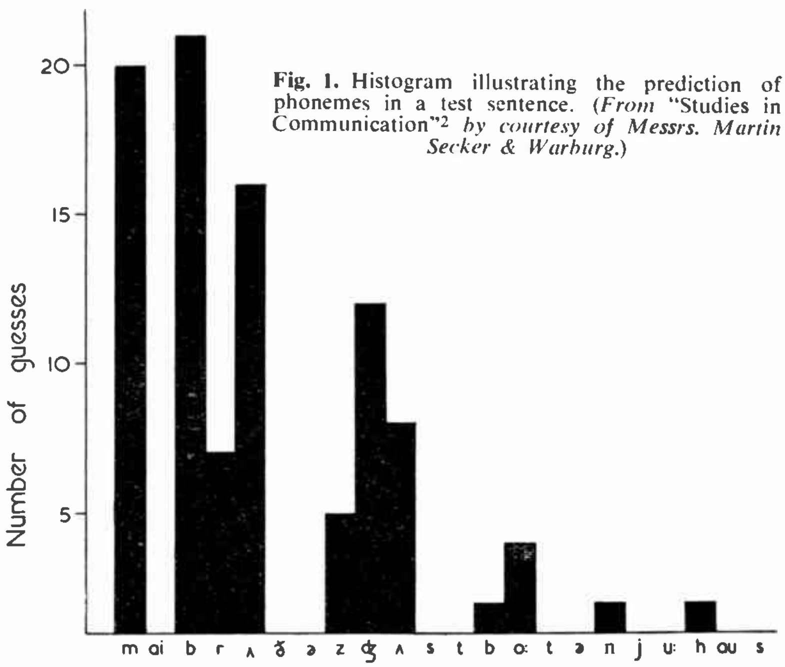


Fig. 1. Histogram illustrating the prediction of phonemes in a test sentence. (*From* "Studies in Communication"[2] *by courtesy of Messrs. Martin Secker & Warburg.*)

He then goes on to guess the phoneme in the second position and so on throughout the sentence. The results of such an experiment are shown in the histogram of Fig. 1. The selected sentence was "My brother's just bought a new house"; a transcription of the phonemic sequence is given on the horizontal axis and the vertical axis shows the number of guesses needed by the listener at each position. It is clear that since the listener knows the complete phoneme repertory, he could not need more than about 40 guesses at any position. But in fact constraints operate at all positions and the greatest number ever needed was 22. Even in the first position, where the uncertainty is obviously quite high, the phonemes are not all equally likely to appear and the listener required 20 guesses to arrive at /m/. Having done this, he immediately made use of his knowledge of constraints and guessed that the first morpheme was [mai] and supplied /ai/ with his first guess. Thereafter, there are major fluctuations; the high values generally occur at the morpheme or word boundaries, but there is an overall tendency for the number of guesses to decrease as the listener proceeds through the sentence. In other words, he becomes more and more certain what the sound must be, the constraints become stronger and stronger, the uncertainty less and less. It is worth noting that of the 22 phonemes that make up the sequence, 11 were supplied correctly at the first guess. If we accept this as a method of estimating the redundancy of spoken English, this particular example would suggest that this redundancy is of the order of 50 per cent.

In producing this result the listener is not only making use of his knowledge of phonemic constraints. It seems clear from the distribution of guesses that information is fed back from the word to the morpheme, from the morpheme to the phoneme level, and the constraints at one level are used to resolve doubts at a lower one. This effect can be seen more specifically in the reception of speech on the rare occasions when the listener makes an error. We are all, as listeners, familiar with the case in which we take in a long phoneme sequence and at a certain point in the reception we have to go back and mentally make a correction at the phoneme level. We do this because at the morpheme or word level, or even at the sentence level, the sequence does not make "sense" and we detect the presence of an error.

5. Primary or Acoustic Recognition

Linguistic knowledge of the kind we have just described is essential for the reception of speech. When we say that someone "knows" a language, we mean that he carries in his brain this particular kind of information with respect to that language; it constitutes the *a priori* information without which there can be no reception of speech and it has been treated here at some length because it forms by far the most important factor in the recognition of speech sounds. The listener in taking in a spoken message combines incoming information derived from the acoustic wave with *a priori* information of a linguistic nature, so that recognition comprises two stages. In the first stage, the listener carries out *primary* or acoustic recognition and in the second stage he modifies the results of primary recognition by the application of statistical knowledge and so resolves uncertainties and corrects errors that arise in the first stage.

In the stage of primary recognition the listener is already operating within a linguistic framework, since all incoming waves have to be classified within the system of 40 phonemic units. How does the listener reduce the widely varying acoustic inputs to such a relatively small number of categories?

There are two important features of the process by which this is achieved. First, the phonemic system is represented at the acoustic level by a system of relations and not by a set of absolute values.[3] To go back to an example given above, the listener recognizes correctly the vowel of the word *line* not as an absolute acoustic value but because of its relation to the vowels of *lane* and *loin* pronounced by the same speaker. Similarly, the vowel of *men* is recognized by reference to the vowels of *man* and *min* and to all the other vowels in the system of that speaker. In the whole process of primary recognition, the listener is dealing with relations of intensity, of fundamental frequency, of formant frequency, of duration and not with absolute values.

The second feature of primary recognition is that here too the redundancy of speech is made use of. At this level, speech exhibits redundancy in the sense that articulatory and hence acoustic constraints are operating; every articulatory movement is conditioned by the movements that have preceded it, and hence any section of the acoustic train is dependent upon the preceding sections. In a less rigorous sense, however, redundancy plays a great part in primary recognition because the listener has at his disposal a number of cues for the recognition of any given phoneme; he uses any combination of cues as occasion demands or as the conditions of reception allow but he never needs all the cues at the same time.

The whole field of the acoustic cues used in primary recognition is now being thoroughly explored, largely through the medium of synthesized speech sounds.[4-7] It will be possible here to mention only some of the results of this work. In order to make a single distinction, such as that between /p/ and /t/, the listener can make use of several cues. First, each consonant gives rise to a short burst of noise which appears in a different frequency region for the two phonemes; for /t/ the peak of the noise energy is generally between 3,000 and 4,000 cycles/sec and for /p/ between 300 and 600 c/s. Second, the intensity of this noise burst is greater for /t/ than for /p/ and third, the noise lasts longer in /t/ than in /p/. Then with each of these consonants there is a change of frequency (transition) in the second formant of the associated vowel section. For /t/, the second formant transition is positive, i.e. the formant frequency falls from a higher value if the consonant precedes the vowel, and rises to a higher value when the consonant follows the vowel; in the case of /p/, the second formant transition is negative.

Here then the listener has at least four different acoustic cues which enable him to distinguish between /p/ and /t/. Changing conditions of reception make it necessary for him to use sometimes one, sometimes another combination of these cues in order to resolve uncertainties.

Another set of cues can be used for the distinction between /s/ and /z/. Here the noise spectrum is approximately common to the two sounds but in the case of /z/ there is sometimes a low frequency component provided by the vocal cord vibrations, and in this case there is the added cue of strong periodicity in the acoustic wave. Again in these two sounds there is a difference in overall intensity, /s/ having greater intensity than /z/, a difference in duration, the noise lasting longer in the case of /s/ than in /z/. When these consonants follow a vowel, an important cue is provided by the duration relations of the vowel and succeeding noise; for /s/, the noise is long compared with the vowel, for /z/ it is relatively short.[8]

We see from these examples that identification of phonemic units from the acoustic input depends largely upon weighing up several cues and deciding upon the evidence that the unit is /p/ and not /t/ or /k/, that it is /z/ and not /s/ or /t/ and so on. In receiving speech, the listener carries out this operation of primary recognition first and then uses his knowledge of constraints to correct errors and resolve uncertainties. It is only in this way that the ordinary listener is able to deal successfully with messages from hundreds of different speakers and in varying conditions of reception.

6. Speech Recognition in the Mechanical Analogue

It is clear that to achieve the flexibility and accuracy displayed by the human mechanism in the reception of speech would require a machine almost as complex as the human brain. There are, however, practical uses for an analogue that would obtain even a small measure of success in this direction and the foregoing theoretical discussion suggests several lines upon which to proceed. The listener succeeds in recognizing speech sounds because he has at his command a vast store of *a priori* information. The success of a mechanical recognizer will be very closely correlated with the extent of the information that we can afford to build into it. This information will be used both at the stage of acoustic recognition and in a second stage where statistical information is applied to the results of acoustic recognition.

The mechanical speech recognizer, like the listener, has first to process the incoming acoustic wave. The classical method of doing this, and still the most useful, is by means of frequency analysis, that is by inspecting the spectral distribution of energy and its variations with time. The listener relies only in part on this method, however, and the analogue is likely to be more successful if it combines with the results of frequency analysis information derived from inspecting the acoustic input with respect to other dimensions. The supporting paper will deal in detail with these possible methods; it will be enough to mention here that relations of intensity between different parts of the acoustic input and relations of duration are acoustic cues that are valuable and probably essential to the successful working of the analogue.

We have seen that a great part of the listener's success depends upon his exploiting the redundancy of the linguistic system and we may be certain that a mechanical recognizer is not likely to be very successful if it relies solely on the processing of acoustic information, no matter how sophisticated or refined its methods of doing so. It needs also a store of statistical information that will enable it to take advantage of linguistic constraints. The human brain, with its very large storage capacity, is able to make use of the constraints at all linguistic levels and to feed back information from one level to another, thus getting rid of practically all errors in its reception of speech. In a mechanical system it is practically necessary to work with a much smaller storage capacity and it is doubtful whether it will ever be feasible to simulate the multiple levels of linguistic operation that characterize the reception of speech by a listener. If not, we have to be prepared for a considerable proportion of errors in the output of the mechanical recognizer and further, we are faced with the problem of selecting the level at which the machine is to operate, that is of choosing the recognition units.

7. The Choice of Recognition Units

In theory any of the four linguistic levels might be chosen as the basis of operation for the mechanical recognizer. We might design the machine to recognize phonemes, morphemes, words or complete sentences. Recognition involves the matching of an incoming pattern with patterns stored in the machine and one important factor in deciding the choice of recognition units will obviously be the amount of storage required in any given case. In the general discussion of linguistic systems, it was said that there is a very great increase in the number of possible units as we go from phonemes to morphemes, words and sentences. A machine designed to recognize English phonemes would need a store of only about 40 patterns for comparison with incoming patterns in order to deal with the whole of English speech. If it were designed to recognize morphemes or words, this number would have to be raised to tens of thousands, or in the case of sentences, hundreds of thousands. The storage problem does not, however, end there since the numbers just indicated refer only to the store required for purposes of acoustic recognition. In addition to this, as we have seen, the machine needs a stock of statistical information concerning sequential probabilities. Since the number of possible phonemes is only 40, information concerning sequential probabilities to two places (digram frequencies) would require the storage of 1,600 items and to three places (trigram frequencies) would still need only 64,000 items. Supposing, however, that the machine operated on a word basis, the initial store might well be 10,000 word patterns, each of which would be much longer and thus require more storage space than a single phoneme. In order to use knowledge of word sequential probabilities, a store of 10^8 items would be required for sequences of two words and 10^{12} for sequences of three.

The choice of a unit as small as the phoneme is therefore attractive in theory because the size of the initial store is small, both in number of items and in the length of each item, and because the size of store that will afford useful statistical information is also relatively small. We must set against this, however, the relative effectiveness of acoustic and of statistical constraints in reducing errors. The longer the unit, the stronger are the acoustic constraints and hence, presumably, the smaller the probability of error. A machine would in fact be more likely to recognize correctly an acoustic train corresponding to a complete sentence than one corresponding to a single phoneme. It cannot yet be said with certainty whether the additional accuracy gained by carrying out primary recognition with longer acoustic trains would outweigh the capacity for error correction provided by additional statistical information stored in the machine. In a practical system, the need for accuracy would in fact have to be balanced against the storage requirements.

8. Segmentation of the Acoustic Input

The question of choosing suitable recognition units leads to a final theoretical consideration that concerns the segmentation of the acoustic input. If a given unit, e.g. phoneme or word, is selected as the recognition unit, then the mechanical recognizer must be able to de-limit the phonemes or the words in the acoustic input. The latter is of course continuous and it can be cut up into discrete segments only by adopting one of two methods: either by introducing a standard integration time, i.e. by repeatedly cutting the acoustic train into equal time segments, or by allowing time segmentation to depend upon magnitude of change in some other dimension. The first of these methods is unsatisfactory because, no matter what the chosen unit, the duration of the acoustic trains corresponding to these units is not constant. All the phonemes, or words or sentences, correspond to acoustic stretches of different durations and the adoption of a standard integration time could only make successful recognition more difficult.

In the case of the listener, the de-limiting of phonemes appears to depend largely upon change in sound quality.[3] This dimension in the acoustic sphere corresponds to change in the spectral distribution of energy and it will be shown in the supporting paper that change in this dimension may provide a suitable basis for time segmentation of the acoustic input to the recognizer.

At levels above the phonemic level, the listener de-limits on a purely linguistic basis and not on an acoustic one. If larger units such as words or sentences were adopted as the basis for recognition, there would be additional difficulty in time segmentation. A possible solution would be the introduction of artificial boundary signals by the speaker, such as a pause at the end of each word or sentence. Another possible method would be the continual scanning of the stored repertory of word or sentence patterns and the triggering of the recognition mechanism as soon as a match was obtained. This would lead to difficulty and possible errors particularly in the case of words comprising several morphemes. A word such as *trustworthiness*, for instance, would produce a recognition decision after *trust;* the next part of the word might be adequately dealt with as a separate word, *worthy*, but this would leave the suffix *-ness* unattached and unrecognized. Such difficulties of segmentation can probably be overcome only by the aid of a very considerable statistical store in the machine, and it would appear that both from the point of view of time segmentation as well as from that of economy of storage, the phoneme is probably the best choice as a recognition unit.

9. Conclusion

We may summarize the working of the human recognition mechanism in the following way: the continuous wave-motions that form the acoustic input are first segmented, mainly on the basis of changes in sound quality: the acoustic features of the wave yield a variety of cues for recognition and the segments are sorted into phoneme categories with the aid of these cues. Phonemes are combined to form higher order units—morphemes, words and sentences—and at each level recognition is determined by the aid of statistical criteria. Information is continually fed back from level to level in order to establish the changing probabilities and to correct errors.

The success of a mechanical speech recognizer will depend upon the extent to which the chief features of the human mechanism can be simulated. It will need to operate with a variety of acoustic cues and to use statistical information of some kind. The success of the machine, in terms of the proportion of errors that it makes, is directly correlated with the storage capacity which is available and in any practical system it will be necessary to balance the requirements of accuracy against economy in storage space.

10. References

The references quoted in this paper are included in the list at the end of the supporting paper. (Page 229.)

The Design and Operation of the Mechanical Speech Recognizer at University College London†

by

P. DENES, M.SC.‡

A paper read before the Institution in London on 28th January 1959.

In the Chair: Dr. A. D. Booth (Member).

Summary: The design of a practical speech recognizer, incorporating many of the principles discussed in a companion paper, is described. The principal parts of the recognizer are as follows:—acoustic spectrum analyser; spectral pattern matcher; store of information about linguistic statistics; phoneme sequence memory; the computer; output circuits (a) the storage of computer decisions, (b) the control of the automatic typewriter. The speech material used with the completed recognizer is described, together with the results obtained. Future developments in speech typewriter research and possible uses of a successful recognizer are indicated.

1. Introduction

An automatic speech recognizer, or speech typewriter as it is sometimes called, is a device which converts the sound waves produced by a speaker into a series of typewritten digits, corresponding to the sequence of speech sounds, or phonemes, uttered by the speaker. When the design of such a device was first considered, it seemed a fairly simple task. After all, the argument ran, when the speaker articulates one sound after the other, he changes the position of his vocal organs, tongue, lips, teeth, etc., and as a result sound waves of different characteristics are produced; simple spectral analysis with a bank of filters should make possible the automatic identification of the various speech sounds. Unfortunately this simple approach does not yield the desired results. The Bell Telephone Laboratory's book "Visible Speech"[9] gives the spectrographic analysis of numerous samples of speech and shows quite clearly that no simple relationship exists between the spectral patterns and the speech sound units. Later work, mainly at the Haskins Laboratories, New York[7], showed that more sophisticated examination of the sound wave yields more consistent relations between the acoustic characteristics and the speech sound units, but even so the relationship is not a very constant one and devices for detecting these intricate characteristics, even if they could be instrumented successfully, would not make the automatic speech recognizer very reliable. As Professor Fry has explained, in his paper, it seems likely that the human listener in recognizing speech does not rely solely on acoustic characteristics, and it is unlikely that any single acoustic characteristic or combination of characteristics uniquely identifying any speech sound does in fact exist in the acoustic wave. It is much more likely that the acoustic cues are only sufficient to identify the sound as one of a number of possibilities and that this ambiguity is then resolved by the linguistic cues that are at the disposal of the listener.

As in the case of the human listener, the automatic speech recognizer to be described in this paper uses linguistic as well as acoustic information in its operation. The principal aim in constructing the recognizer was, in fact, to see what improvements, if any, could be achieved when certain forms of linguistic information were used to modify the output of an acoustic recognizer, rather than to pursue the design of the acoustic detector to a great degree of refinement. Figure 1 shows the basic mode of operation of our speech typewriter. The acoustic recognizer examines the characteristics of the speech sound wave and makes a preliminary phoneme recognition. The linguistic

† Manuscript received 29th January 1959. (Paper No. 496.)

‡ Department of Phonetics, University College, London.

U.D.C. No. 534.782

store contains information about linguistic statistics and supplies information to the "computer" about which phonemes are most likely to occur next in the context of the words being recognized. The computer combines the information derived from the acoustic recognizer

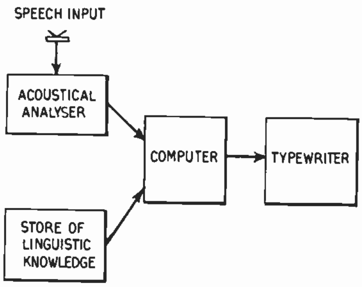


Fig. 1. Schematic diagram of the automatic speech recognizer.

with that from the linguistic store and selects that phoneme which is most likely in the light of both acoustic and linguistic information. The computer indicates the phoneme recognized by operating the appropriate key of the typewriter.

The constituent parts of the recognizer will now be described in greater detail.

2. The Acoustic Input

The speech material to be dealt with by the machine was first recorded on tape and the speech sound wave was obtained from playing this tape rather than from live speech via a microphone. The tape recorder used is an Ampex 600, which was modified to run at 15 in./sec. At this speed, with full track heads, it gave a signal/noise ratio of 60 db. This was an important feature because of the well-known wide dynamic range of speech sound waves. The output of the tape recorder was first given pre-emphasis and then peak-clipped.

Pre-emphasis is used to compensate for the reduced high frequency content of the average speech spectrum and to ensure that the peak output of all the filters to be used in the recognition process is the same. The pre-emphasis amounts to an increased gain of about 5 db/octave from 500 c/s to 5,000 c/s. Apart from this pre-emphasis the frequency response is flat from 60 c/s to 10,000 c/s. Peak clipping of about 20 db from the highest peaks reached through a long speech passage is used as a simple form of volume compression. Compression is necessary to reduce the very wide dynamic range of ordinary speech to the requirements of the circuits used in the recognizer. Spectrographic investigation showed that, at least as far as the detection processes to be utilized in the recognizer were concerned, the distortion produced by this degree of peak clipping is of no importance.

The output of the peak clipper is connected to a power amplifier. All the circuits of the acoustic recognizer derive their inputs from this power amplifier.

3. The Acoustic Recognizer

A number of different methods are used to detect the acoustic features relevant to phoneme recognition. Some of these methods are based on spectral analysis of the sound wave, and this analysis is achieved by applying the speech voltage to a bank of 18 band-pass filters covering the frequency range 160 c/s to 8,000 c/s in adjacent third octave wide bands. These filters were chosen, not because they were the most suitable, but because they were available at the time when construction of the recognizer was started; filters with equal linear bandwidths up to about 1,000 c/s and with equal logarithmic bandwidths above 1,000 c/s would probably be preferable. The output of each filter is rectified and integrated. The time constants of the integrating circuits vary from about 10 msec for the low frequency filters to about 1 msec for higher frequencies as a compromise between the requirements that they should be long enough to smooth out variations due to the individual cycles of the filter output voltage and short enough to preserve the variations of the overall energy level in the filters.

At this stage, as a preliminary to further design, a number of test words were applied to the analyser and the rectified output of all the filters was recorded simultaneously, using a number of pen recorders. The resulting spectral patterns were inspected, and it was observed that many of the sounds were characterized by peak outputs in two of the filters. For instance, for the sound /i/ the 250 c/s and the 3,200 c/s filters gave large outputs, whilst for /s/ it was the 6,000 c/s and 8,000 c/s filters. The pairs of filters that gave characteristically large outputs for the different phonemes were selected by visual inspection of these spectrographic records. In this way a set of spectral patterns with double peaks was obtained, where each pattern was characteristic of one of the phonemes in the repertory of the machine. Automatic recognition of the phonemes is then carried out by comparing the speech spectra with these previously selected characteristic patterns and indicating instant by instant which of the characteristic patterns is most similar to the speech spectrum appearing at the filter outputs. This spectrum-matching process is achieved, as shown in Fig. 2, by first multiplying the outputs of the pairs of filters that were found to be characteristic of particular phonemes. This means that as many simultaneous multiplications have to be carried out as there are phonemes to be recognized in this way. Then, all the multiplication products are compared and the largest selected. The particular multiplication that gives the largest product indicates the phoneme to be recognized. For instance, it has already been stated that the phoneme /i/ is characterized by greatest outputs in the 250 c/s and 3,200 c/s filters, whilst the phoneme /s/ is characterized by greatest outputs in the 6,000 c/s and 8,000 c/s filters. The output voltages of these pairs of filters are continuously multiplied and the products compared. Whenever the former of these products is greater, the machine will indicate that the sound /i/ has occurred, while whenever the latter multiplication product is larger, the phoneme /s/ will be recognized.

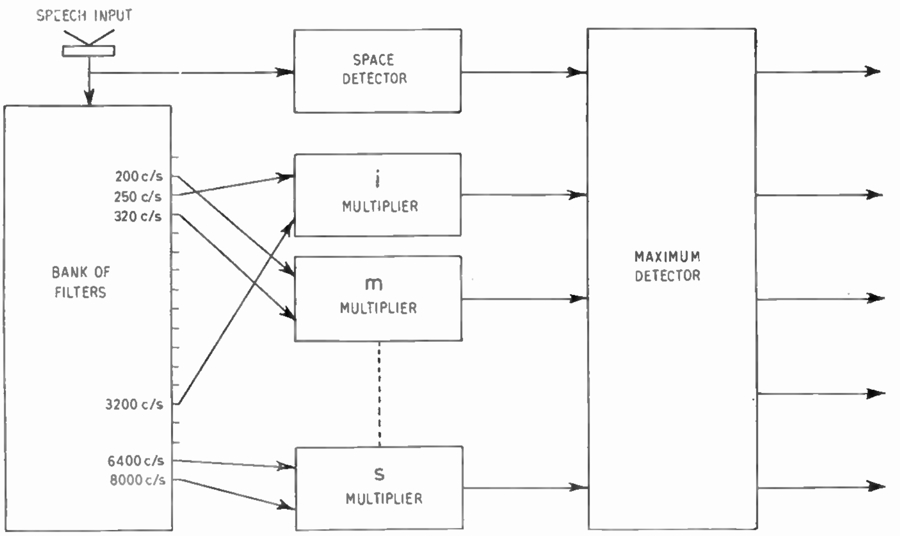


Fig. 2. Schematic diagram of the spectrum matching process.

Multiplication, rather than addition, is necessary for the pattern-matching process, in order to ensure that both spectral areas contribute proportionately to the resultant voltage. If addition were used instead, then excessive output from one of the filters would serve as a substitute for lack of output from the second of the characteristic pair. The multiplier circuits used are of the variable mark–space ratio type and the principle of operation is shown in Fig. 3. An oscillator common to all the multipliers

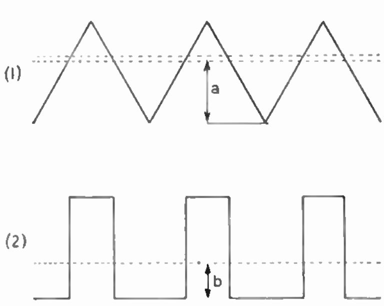


Fig. 3. Principle of operation of multiplier circuits.

generates a voltage of triangular shape. In each multiplier a narrow slice is cut out of this triangular wave, and this slice constitutes a square wave in which the mark–space ratio is determined by the level at which the slicing is carried out. The square wave is amplified and then peak-clipped. One of the two voltages to be multiplied (a) determines the level at which the triangular voltage is sliced and the other voltage (b) sets the peak-clipping level. In this way one of the voltages to be multiplied determines the mark–space ratio and the other the amplitude of the resulting square wave. The square wave is integrated and the resulting voltage is proportional to the product of a and b. The frequency of the triangular wave is relatively high, so that the time-constant of the integrating circuit can be made long relative to the duration of the triangle and at the same time short compared with the rate of variation of the voltages to be multiplied.

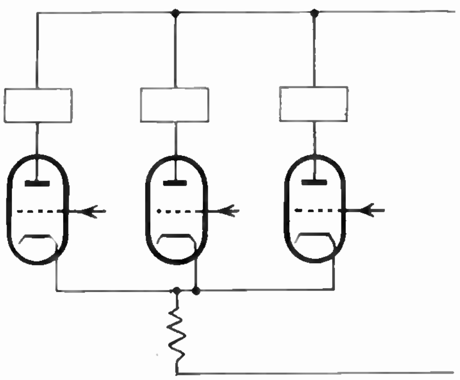


Fig. 4. The maximum detector circuit.

The maximum selector circuit is shown in Fig. 4. The multiplier output voltages are first amplified about 10 times by simple d.c. amplifier circuits and then applied each to the grid of a separate triode. The number of triodes is equal to the number of voltages to be compared. The common cathode resistance is given such a value that the current going through it is sufficient to operate one of the relays in series with the triode anodes, but not enough to activate two of them. Whichever grid is the most positive will make that valve conduct, and the relay in the anode circuit of this valve operates, indicating the particular multiplication producing the largest product. The others will cut off as their cathodes are too positive, relative to their grids, because of the current going to the valve with the most positive grid. The circuit detects voltage differences of about $1\frac{1}{2}$ volts and the peak input voltage is about 200. All the vowels and some of the consonants that the recognizer can deal with are identified by such a spectrum-matching process. Other consonants are recognized by detecting some other feature of the associated acoustic wave.

The consonant /t/ is characterized by the presence of high frequency energy and its spectrum is very similar to /s/. The duration of the sound associated with /t/, however, is much shorter than that of /s/ and therefore measurement of the duration of high frequency energy forms the basis of the /t/ detector circuit. The basic circuit is shown in Fig. 5. The output of the 8,000 c/s filter operates valve V1 and relay A. Relay A triggers the time-measuring flip-flop V2 and thus de-energizes relay B. While V2 is operated the contacts of relay B are in position 1. If the output of the 8,000 c/s filter disappears before the flip-flop returns to normal, then relay C operates and this indicates the occurrence of /t/. If the flip-flop returns to normal before the output of the filter has disappeared, then relay C will not operate. In other words, as soon as the 8,000 c/s filter shows an output the flip-flop operates and measures off a fixed time interval. If the output of the filter disappears in less than this time, the occurrence of /t/ is indicated. If the filter output is still present when the flip-flop returns to normal, then the /t/ indicator cannot operate. It must be remembered that the same filter output that initiates the action of the /t/ detector is

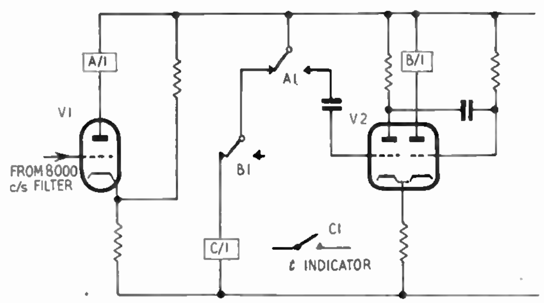


Fig. 5. The /t/ detector circuit.

also connected to the /s/ multiplier circuit and will produce an /s/ output before the /t/ circuit has had time to decide whether the appearance of the high frequency energy indicated a /t/ or an /s/. The operation of the recognizer is therefore inhibited for the duration of the operation of the /t/ detector. This is achieved by having an additional triode in the maximum detector circuit, connected in the same way as all the other valves, except that there is no relay in series with its anode. While the /t/ circuit is operated a large voltage is connected to the grid of this extra triode in the maximum detector so that all the current goes to it and no output can appear until the voltage is removed from this grid; this happens as soon as the /t/ circuit has made its decision.

The consonant /k/ is detected with a similar arrangement, but with the input derived from the 1,600-c/s filter.

The detection of the consonant /z/ is also based on duration measurement. It is known that usually the spectrum of /z/ is very similar to that of /s/, but the amplitude and the duration of the sound is less for /z/ than for /s/[8]. Whenever the /s/ spectrum is recognized a flip-flop is triggered and the output of the filters is channelled to the /z/ or /s/ circuits depending on whether the relevant filter outputs cease before or after the end of the flip-flop cycle respectively. Neither the /z/ nor the /s/ circuit can operate while the /t/ circuit is making its decision. In effect therefore the same filter outputs will actuate the /t/, /z/ and /s/ circuits, depending on whether the spectrum configuration lasts for a short time only (/t/), for a medium duration (/z/) or for a long time (/s/).

The spectrum of the consonant /f/ is in many ways similar to that of /s/ but its amplitude is considerably less. A simple Schmidt trigger is used to perform the necessary amplitude discrimination and switch the filter outputs to the /f/ or the /s/ circuits.

The proper operation of the computer section of the recognizer requires that the incidence of intervals between words be known. A conventional voice-operated relay is used to indicate the occurrence of the silent intervals between words. This circuit will be referred to as the space detector in the rest of this paper.

Although several other methods for more successful automatic acoustic recognition are known, and are discussed at length in a paper published last year[10], it was decided not to elaborate this part of the circuit any further, but, for reasons already discussed, to devote more attention to the use of linguistic information in the operation of the recognizer.

4. The Use of Linguistic Information

The linguistic knowledge built into the existing recognizer is of a very simple kind. It consists of information about the so-called digram frequencies of phonemes or, in other words, about the relative probability of occurrence of the phonemes as a function of the preceding sound. In practice, this means that the recognizer has two separate memories. First it has to remember the identity of the previous phoneme recognized; this we shall call the "phoneme memory." It needs a second memory, to be called the "store of linguistic knowledge," to remember the digram frequency of every phoneme following every other phoneme. Whenever a phoneme recognition is carried out, the machine must automatically recall what the previous phoneme was and make available to the computer a set of figures representing the probability of occurrence of every phoneme following this particular preceding phoneme.

4.1. *The phoneme memory*

The requirements of the "phoneme memory" can be best explained by considering what should happen when a phoneme sequence *a, b. c* is being dealt with. As soon as *a* has been recognized the memory must register the identity of *a*. When *a* ceases and *b* is about to be recognized, the memory must produce a signal indicating that the previous phoneme was *a* so that the set of numbers representing the digram frequencies of all the phonemes following the phoneme *a* can be produced from the "store of linguistic knowledge." This signal must be maintained throughout the duration of *b* and must be replaced by another signal, that signifying that the previous phoneme was *b*, as soon as *b* ceases, and so on. At this point is is necessary to digress for a moment to consider where in fact is the beginning and end of a phoneme, or, in other words, to discuss the so-called "gating" problem. Theoretically, this is a very complex question, but as far as the operation of the recognizer is concerned the phoneme boundaries are taken to be at the instants when the maximum detector changes over, that is, when a different maximum detector relay becomes energized; these are then the instants when the output of the "linguistic knowledge" has to be changed.

Returning now to the phoneme memory, this consists of a number of identical units and each of these memory units is connected to a different maximum detector relay. The coils of these relays are in the anode circuits of the maximum detector triodes, and the contacts in the phoneme memory units. Each set of maximum detector relay and memory unit is associated with one of the phonemes that the recognizer can distinguish. The circuit of one of these memory units is shown in Fig. 6. Relay A is

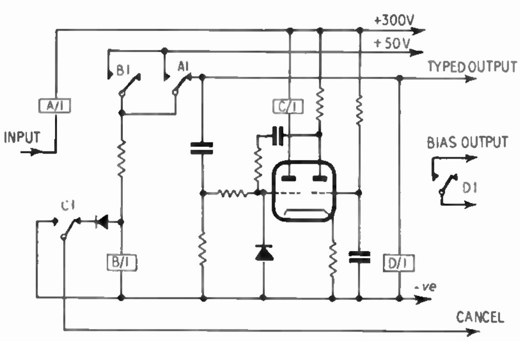


Fig. 6. A phoneme memory unit.

the maximum detector relay, and when it operates, it energizes self-holding relay B. This will register the identity of the phoneme. When this phoneme ceases, and the A relay releases—and of course the A relay of another memory unit energizes—two things happen: the flip-flop circuit is triggered and relay D is energized. The flip-flop keeps relay C energized for a limited period of time. The contacts of relay C short the line marked "cancel" to earth. The "cancel" lines of all the memory units are connected together so that when the C relay contacts operate they will de-energize the B relays of all memory units, except the one in which the C relay is energized, and thus cancel the memory of any previous phoneme except the immediately preceding one. The diode is included in the "cancel" circuit to prevent the simultaneous operation of all B relays via the common "cancel" line. The contacts of relay D will operate the store of linguistic knowledge. Relay D remains energized as long as relay B is energized. When the next A relay drops out, it will operate its own C relay which in turn will now short out the B relay of the first memory unit, which returns to normal and is ready for another operation. In other words, the A relay activates the memory unit, the B relay retains the memory. When the phoneme ceases the C relay cancels all previous memories but maintains that of its own unit and the D relay operates the store of linguistic knowledge. Another function of the C relay is that whilst it is energized and is earthing the "cancel" line, no other B relay can energize even if its A relay does operate. This prevents spurious operation during the memory change-over period.

4.2. *The store of linguistic knowledge*

This store contains information about all possible digram sequential probabilities relevant to the speech material to be tackled by the recognizer. That is for n phonemes n^2 digram frequencies have to be remembered. Whenever the phoneme memory indicates that one particular phoneme has ended and the next one is to be recognized, that set of n digram frequencies which is relevant in the light of the identity of the previous phoneme has to be selected from the store and sent to the computer. The information sent to the computer is in the form of a set of voltages with values proportional to the digram frequencies. The information is stored by the position of the sliders of a matrix of potentiometers as shown in Fig. 7. The potentiometers in one column refer to the digram frequencies of one particular sound being followed by the other sounds, and the potentiometers in a row refer to the probability of one particular sound following other sounds. Thus in the column headed /i/ the slider of the first potentiometer is set proportional to the digram frequency of /i/ being followed by /i/, the slider of the potentiometer underneath this one is set proportional to the digram frequency of /i/ being followed by /m/, and the one underneath, of /i/ being followed by /s/. In the next column, the sliders are set proportional to the values of digram frequencies of /m/

being followed by /i/, /m/, /s/ respectively, and so on. The function of the diodes is simply to prevent the potentiometers connected together from shunting each other.

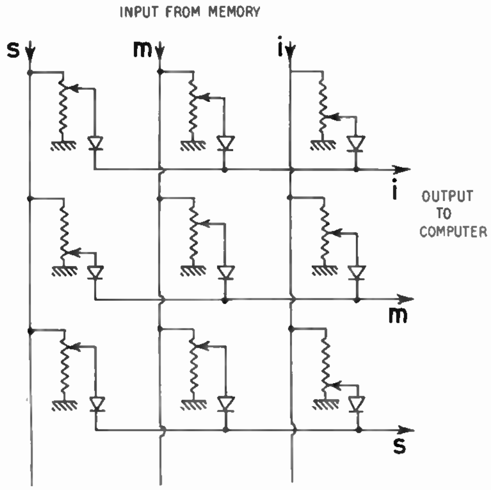


Fig. 7. A typical section of the "store of linguistic knowledge" circuit.

In the recognition process, when a phoneme, say /s/, ceases and the next sound is being recognized, a fixed d.c. voltage is applied to the /s/ column of potentiometers and they in turn supply the computer with a set of voltages proportional to the digram frequencies of the various phonemes following /s/.

4.3. *The computer*

The computer makes the final choice of the phoneme to be recognized. It receives information from the acoustic recognizer and from the store of linguistic knowledge about which phonemes are most likely to occur at a particular moment and selects the one that is most likely in the light of information from both sources. The circuit arrangement is shown in Fig. 8. The computation is carried out by a number of multiplying circuits, one for each phoneme. The outputs of the multipliers in the acoustic detector are not sent directly to the maximum detector as stated previously. Instead, these outputs are multiplied once more by the statistical voltages derived from the store of linguistic knowledge. The voltage showing how probable it is that any one of the phonemes, say /m/, will follow the preceding phoneme is multiplied with the output of the /m/ multiplier in the acoustic recognizer whose output indicates how likely /m/ is from the acoustic point of view.

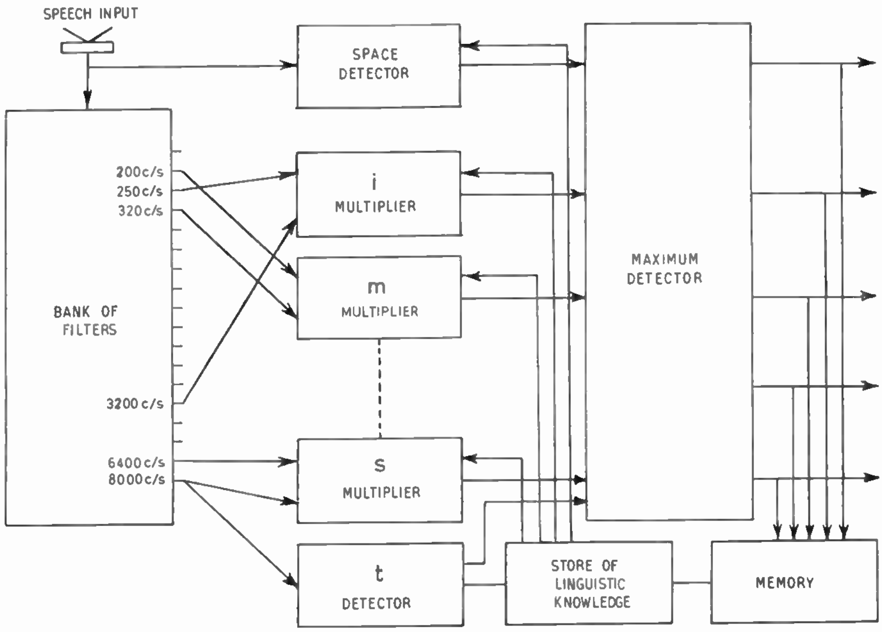


Fig. 8. Schematic diagram showing the arrangement by which statistical information is combined with acoustic information in the mechanical speech recognizer.

This second multiplication is carried out by a circuit similar to that used in the acoustic recognizer. The outputs of this second set of multiplying circuits are then sent to the maximum detector. In the case of the consonants /t/ and /k/, where acoustic recognition is achieved without multiplication, the appropriate probability voltages are connected to the /t/ and /k/ circuits respectively and are then transmitted to the maximum detector directly whenever the /t/ and /k/ detectors operate. The same applies to the space detector. Space has its own relay in the maximum detector and its own unit in the phoneme memory, in the same way as the phonemes. This is because when recognizing sounds in initial position the digram frequencies relevant to the position following space are taken into consideration.

5. The Typewriter Memory

The decision of the maximum detector operates an electric typewriter. The maximum rate at which the recognizer will normally deliver its output is too high for the typewriter, but the mean rate is lower because of the pauses between the words spoken into the recognizer. The decisions of the recognizer are therefore not sent to the typewriter directly: they are stored in a special memory and then sent to the typewriter at a constant rate that is lower than the typewriter's peak rate of operation but higher than the mean rate of producing phonemes in speech. In this way the typewriter never falls far behind the recognizer.

The typewriter storage device consists basically of two uniselectors each with six levels of 25 contacts, both connected to the same bank of 150 capacitors arranged in six rows. The first uniselector steps once every time the recognizer produces an output and remembers the phoneme recognized in the form of a 6-digit binary code by putting a pattern of charges on six capacitors. The second uniselector moves at a constant rate and "reads" the binary code off the capacitors. The binary code is converted by means of a relay-tree into a form suitable for operating the typewriter.

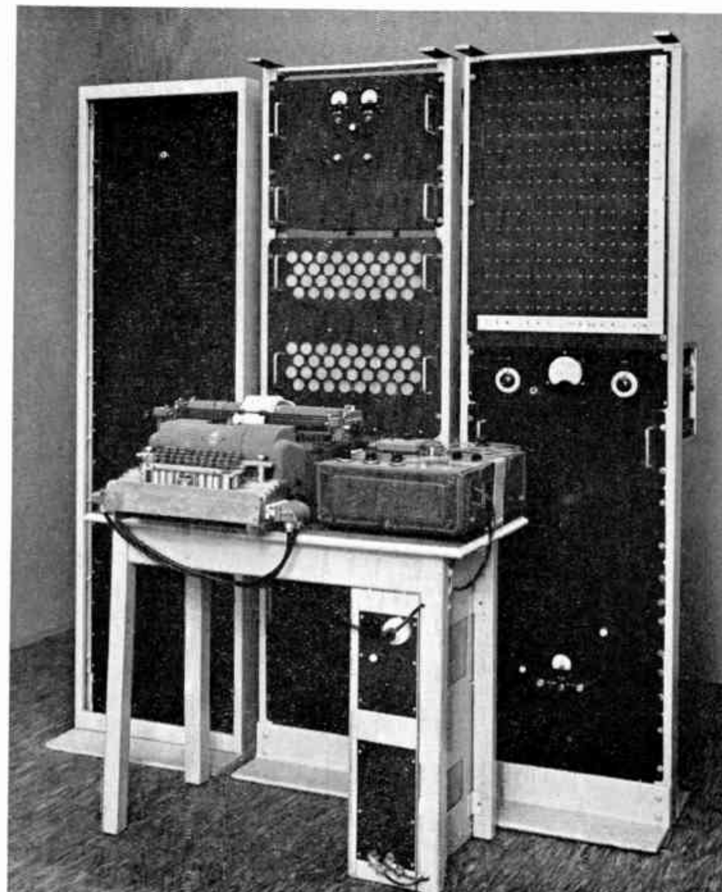


Fig. 9. The automatic speech typewriter. (By courtesy of the National Physical Laboratory).

6. The Electric Typewriter

The keys of the electric typewriter are operated by solenoids mounted above the keyboard. When a solenoid is energized, its plunger will push downwards, depress a key and operate the typewriter in the normal way. Each solenoid carries an additional plunger with the ordinary typewriter key tops mounted on them. These form a second keyboard above the level of the solenoids and, when one of these keys is depressed, the additional plunger presses on the solenoid plunger and this in turn depresses the real key. In this way manual operation of the typewriter is also possible.

Figure 9 shows a photograph of the equipment. The tape recorder and the typewriter are on the central table. The rack on the left carries the filters, the rack on the right the various memories, with the potentiometer matrix of the "store of linguistic knowledge" at the top. The multiplying circuits are mounted on the central rack.

7. Speech Material and Results

The recognizer in its present form deals only with the thirteen phonemes, four vowels and nine consonants, listed in Table 1. This table shows side by side the phonetic symbol for each sound, a key word containing the sound and, in the third column, the symbol typed by the typewriter. The typewriter, of course, cannot spell the words recognized, but tries to give a phonetic transcription. Some of the symbols used on the typewriter are different from those generally accepted in phonetics, so as to make them easier to read by phonetically untrained people and the choice was also influenced by what was easily obtainable from the typemakers.

Table 1

List of Phonemes used with the Recognizer

i	heat	i	ʃ	she	š
u	hoot	u	f	fee	f
a	heart	a	z	zoo	z
e	hurt	ë	m	me	m
t	tea	t	n	knee	n
k	cool	k	l	leap	l
s	see	s			

The speech material used with the machine consists of a list of words, each word pronounced in isolation. The words chosen were all those words to be found in an English dictionary[11] which contain only the 13 phonemes in the repertory of the machine and which have no consonant clusters. Table 2 shows some examples of the output of the machine. The column on the left gives the words read into the recognizer, and the middle column the corresponding output typed automatically. The last column shows what the output should look like if the recognizer makes no mistakes.

There are several ways of assessing the success of recognition. One is simply to compare the words read in with the words typed out. Another, perhaps more valid, method is to ask someone to read the typed output and to com-

Table 2

Specimen of Recognizer Output

Original Word	Recognized Word	Correct Recognition
Accoutre	ëtatë	ëkutë
Anneal	ënil	ënil
Are	a	a
Arm	an	an
Armour	ana	anë
Art	at	at
Calf	katë	kaf
Car	ka	ka
Carcass	kakës	kakës
Cartel	katël	katël
Carter	katël	katë
Cart	ka	kat
Cartoon	katun	katun
Catarrh	këta	këta

pare the words read into the recognizer with the interpretation of the reader. Both methods have been tried and are more fully described in a previous paper[12]. Table 3 shows the results giving both sound and word articulation scores, and gives in separate columns the results obtained when the acoustic recognizer only was used and when both acoustic and linguistic information were used in the machine. Comparison of the two columns shows the effect of using linguistic information.

These results were obtained using only one speaker. When a second and third speaker were used without making adjustments to the machine, the score dropped from about 70 per cent. to about 45 per cent.

8. Plans for Future Work and Practical Applications

The results in Table 3 show the effect of using digram frequencies only. Many other experiments—using different forms of linguistic information in various ways—have to be carried out to gain a better understanding of the effect and of the usefulness of linguistic information in automatic speech recognition. Some of the more important of these experiments are: to see the effect of increasing the time span of the linguistic information by using trigram or n-gram statistics; to see the effect of severely restricting the number of words in the vocabulary of the machine and so increasing the effectiveness of language statistics; to see the effect of using words as well as phonemes as the units of recognition. Another important question to be studied is how far the reader of the speech typewriter output can correct the mistakes of the machine by using his own knowledge of linguistic rules. These problems are discussed elsewhere[12, 13].

A considerable amount of additional equipment is required for carrying out these experiments. For example, if the linguistic knowledge to be used in the recognizer is to be increased from digrams to trigrams, the store of linguistic knowledge has to contain over 2,000 items instead of about 200 as at present, and the phoneme memory also has to be extended. If, instead of using only 13 phonemes as at present, the recognizer is made to deal with all the 40 phonemes of English, 64,000 values of trigram frequency would have to be remembered. As the construction of all this equipment would be very costly and time-consuming, the possibility of using instead commercially available electronic computers is being considered. Many of the operations carried out by various parts of the recognizer just described, such as multiplication, maximum detection, pattern matching, storage of information, facilities for carrying out actions as a function of the nature of preceding decisions, are available in the ordinary electronic computer. Once a satisfactory computer programme is found, it should be possible to design an analogue circuit performing the same function. Such questions as the best way to programme these operations, how much to programme and how much to do outside the computer, how much it will cost, how quickly a programme can be changed, and many others, will have to be settled before the first computer programme for automatic speech recognition is designed, and they are being investigated at the moment.

As far as the practical application of an automatic speech recognizer is concerned, it is quite clear that there is no immediate hope of using it literally as a speech typewriter and such a use was, in fact, never contemplated. The recognizer was terminated by a typewriter

Table 3

Recognition Scores obtained with Automatic Speech Recognizer

	Acoustic recognizer only		Recognizer using linguistic information	
	Sound articulation %	Word articulation %	Sound articulation %	Word articulation %
Input/output	60	24	72	44
Reader	41	28	57	43

because this seemed a convenient way to recording its output. Serious consideration is however being given to the question of using an automatic speech recognizer in connection with the voice control of various operations and in bandwidth compression telephone transmission.

There are many operations at present controlled by hand which could be controlled more conveniently by the spoken word. Examples are automatic telephone dialling, the operation of some input devices for digital computers, certain actions for controlling machinery in factories, and so on. The number of commands used in these situations is quite small and this very restricted dictionary is a great help in improving the efficiency of an automatic speech recognizer. Therefore it may well be that automatic speech recognizers will find practical application in this field first of all.

As for bandwidth compression telephony, it is well-known that the transmission of ordinary speech requires a bandwidth of about 4,000 c/s with a 30 db signal/noise ratio, which represents an information rate of some 40,000 bits per second. Some very promising bandwidth economy systems are being developed at the moment which only transmit information about certain features of the speech sound wave, but they still need a channel capacity of about 1,000 bits per second. If an automatic speech recognizer could be used in telephone transmission, then instead of transmitting the sound wave, only information about the phoneme sequence need be sent along the line and this information could then be used to operate a speech generator at the receiving end to reconstitute the information in its acoustic form. As there are 40 phonemes in English, produced at the rate of, say, 20 per second, it would need a channel capacity of about 100 bits/second, or a bandwidth of 10 c/s at 30 db signal/noise ratio to transmit this information. Such a transmission system, even if it worked perfectly, would not be equivalent to a normal telephone system, because such features as stress, intonation, individual voice quality, would be absent. Although for these reasons such a system might never be usable in a public telephone system, there are many communication links where these restrictions would be no great disadvantage.

9. Acknowledgments

The author wishes to express his thanks to Mr. J. E. West, of the Phonetics Laboratory, University College, London, for the inventiveness and ingenuity he has contributed to the design of much of the circuitry described in this paper. The same applies to Mr. H. J. Crabbe, also an assistant in the Phonetics Laboratory.

The work described in this paper was supported in part under Contract No. AF61(514)-1176 with the Air Research and Development Command, United States Air Force.

10. References

1. C. E. Shannon, "Prediction and entropy of printed English," *Bell System Tech. J.*, **30**, p. 50, 1950.
2. D. B. Fry, "The Experimental Study of Speech," in "Studies in Communication," pp. 147-167 (Secker & Warburg, London, 1955).
3. D. B. Fry, "Perception and Recognition in Speech," in "For Roman Jakobson," pp. 169-173 (Mouton, The Hague, 1956).
4. A. M. Liberman, P. Delattre and F. S. Cooper, "The role of selected stimulus variables in the perception of unvoiced stop consonants," *Amer. J. Psychol.*, **65**, p. 497, 1952.
5. A. M. Liberman, P. Delattre, F. S. Cooper and L. Gerstman, "The role of consonant-vowel transitions in the perception of stop and nasal consonants," *Psychological Monographs*, **68**, no. 8, 1954.
6. J. D. O'Connor, L. J. Gerstman, A. M. Liberman, P. D. Delattre and F. S. Cooper, "Acoustic cues for the perception of initial /w, j, r, l/ in English," *Word*, **13**, p. 24, 1957.
7. A. M. Liberman, "Some results of research on speech perception," *J. Acoust. Soc. Amer.*, **29**, pp. 117-123, 1957.
8. P. Denes, "Effect of duration on the perception of voicing," *J. Acoust. Soc. Amer.*, **27**, pp. 761-4, 1955.
9. R. K. Potter, G. A. Kopp and H. C. Green, "Visible Speech" (Van Nostrand, New York, 1957).
10. D. B. Fry and P. Denes, "The solution of some fundamental problems in mechanical speech recognition," *Language and Speech*, **1**, pp. 35-58, 1958.
11. Daniel Jones, "An English Pronouncing Dictionary," 11th ed. (J. M. Dent, London, 1956).
12. D. B. Fry and P. Denes, "On presenting the output of a mechanical speech recognizer," *J. Acoust. Soc. Amer.*, **29**, pp. 364-7, 1957.
13. D. B. Fry and P. Denes, "An analogue of the speech recognition process" (*Proceedings of the Symposium on the Mechanization of Thought Processes, 1958*; in the press).

DISCUSSION

Dr. A. D. Booth (Chairman): There seems to me to be two assumptions implicit in this work, the first that the phoneme is in fact the best recognition unit which can be conceived, and the second that digram frequencies should be applied irrespective of the reliability of the recognition in any current phoneme.

The first of these points, of course, is one of personal taste and we have found in my Laboratory that a method of recognition based upon axis crossings is more satisfactory for computer application. This preference is, of course, completely dictated by the application which we ourselves have in mind.

The second point seems to me a more serious one because I believe that a straight-forward application of the digram frequency principle would lead one to make far more mistakes than would a more selective application. The sort of thing which we have considered in the context of recognition for spoken word standard English is that if any particular segment of a word has several nearly equi-probable interpretations, then, and only then, shall cognizance be taken of a previous phoneme or phonemes.

Personally, I would be extremely doubtful as to whether a spoken word typewriter would come into general use, not because I think that it is technically unfeasible, even at the level of producing standard English with all its vagaries, but rather because it is extremely difficult to produce a perfect message in the written sense by means of directly spoken words—this, I think, will be borne out by anybody who has attempted to use a dictating machine.

Finally, I would suggest that a far more satisfactory output for the spoken word typewriter of Professor Fry and Mr. Denes would be a high-speed paper-tape punch. The record so produced could then be fed directly to an electric typewriter. Equally, and probably more important, this could be fed directly to a digital computer.

Communicated: Since the lecture was given, I have had the opportunity to devote a little thought to the problem of digram and trigram frequencies and especially to that of programming on a digital computer, and the opinion which I expressed at the meeting that programming was a trivial operation is now reinforced. In particular I would say that the programme for obtaining digram and trigram frequency analyses can also be tackled most expeditiously by means of the punched record directly produced by the recognizer which I have previously mentioned.

Dr. R. Fatehchand: I think Professor Fry and Mr. Denes have certainly shown how formidable is the design of mechanical speech recognizers. And yet it is one of those problems which at first sight appear quite simple, if one considers analysis-synthesis telephony and the synthesis of speech sounds.

The Vocoder system is well known—if the spectrum channels which contain voltage maxima are transmitted and all other channels suppressed, a further considerable reduction in band-width is possible.† High sentence intelligibility can be attained with a channel capacity of less than 100 bits/sec. If speech is synthesized artificially, an even smaller channel capacity is needed. Lawrence‡, for example, synthesizes good quality speech sounds from only four parameters. Three of these parameters specify the resonant frequencies of the vocal cavity analogue, and the fourth specifies whether the synthetic sounds should be voiced or fricative in character. The necessary information for the control of these parameters requires a channel capacity of about 50 bits/sec.

Mr. Denes estimated that a channel capacity of about 100 bits/sec would be required for a phoneme coding system, and since the analysis-synthesis and synthesis systems only require a channel capacity of this order, it would seem that the significant parameters which specify the phonemes could be extracted quite simply. However, Professor Fry has implicitly shown the fallacy of this view—that speech from which

† E. Peterson and F. S. Cooper, "Peakpicker—a bandwidth compression device," *J. Acoust. Soc. Amer.*, **29**, p. 776, 1957.

‡ W. Lawrence, "The synthesis of speech from signals which have a low information rate." In "Communication Theory," p. 460 (Butterworths, London, 1953).

frequency spectrum information, say, has been removed, is correctly recognized because the human brain has a detailed knowledge of language structure, and that in fact a successful mechanical recognizer, if its store of language structure is not to be prohibitive, must make a detailed analysis of the frequency spectrum, duration, etc., of the speech sounds.

It seems that digram frequencies for phonemes depend, among other things, on the number of sounds in a syllable.† How would this be coped with in a phoneme recognizer which made use of digram frequencies? Secondly, has there been any particular difficulty in recognizing consonants which follow vowels? I ask this because of work which we have been doing at Birkbeck College. We found in CVC words for example, that if one attempted to hear the end consonant in isolation, by gating off the rest of the word, the vowel sound in many cases could always be heard, no matter how much of the consonant was removed.

F. P. Thomson (Associate Member): The success of the speech-operated typewriter depends largely upon the speaker. I would say that the authors have blamed their machine for inaccuracies that are the fault of the speaker. Even if a standardized sound for a phonetic symbol is established there is little likelihood of getting a speaker (far less speakers), to repeat that sound accurately each time he says a word containing that phonetic sound. My experience of teaching English by phonetics and with tape and wire recorders in Sweden is that you can teach a really healthy person to say a phonetic sound in isolation providing they have a good hearing response frequency, but it is practically impossible to expect a person to repeat constantly and with the same tonal and timing values a series of phonetic sounds which make even a simple word as one of two syllables. I said *a really healthy person* because the slightest cold or catarrh can change voice clarity and you cannot accurately teach a person to say what they cannot hear; that is one of the main reasons why foreign languages should be taught to young children by native speakers; children's ears have such a wide frequency response that they can hear all the frequencies of native speech, which can be heard by few adults who, learning a foreign language after adolescence, are usually the victims of their own ears' lowered frequency response. People with a "good ear for music" often retain into adulthood a good ear frequency response.

So I would like to ask the authors how their machine can be made to react faithfully to words with glottal stops, also to words with a very high percentage of sibillant sounds such as the Swedish word for *sea*, which, correctly pronounced, is so full of hissing and whistling sounds from the throat that it is difficult to transcribe phonetically; many Scandinavian words have this characteristic.

R. A. Waldron (Associate Member): There is one faculty that is used by a human listener in recognizing speech that was not brought out by Professor Fry in his paper, and which finds no analogue in the machine described by Mr. Denes. The human listener not only observes sounds and compares them with certain standard sounds which he carries in his memory, but also compares the sounds made by a speaker with each other.

The phoneme is an abstract concept which may be represented in a variety of ways—written or printed marks on paper; series of dots and dashes in the form of electric currents, flashes of light, buzzes of vibrators, or the waving of flags; vocal sounds, and artificial sounds similar in quality to vocal sounds. The vocal sounds made by various speakers for a given phoneme show a considerable variation, and the sounds made by one speaker for a given phoneme may overlap those made by another speaker for a different phoneme. For example, an American's "t" is often nearer to an Englishman's "d" than to his "t." It is the relationships between phonemes that constitute a language, not the actual sounds used to represent the phonemes.

The art of understanding speech is therefore to identify the phonemes that the speaker uses, and while, as Professor Fry has said, we do this to a large extent by comparing the sounds produced with standards that we carry in our memories, this is not the whole story. It

† J. W. Black, "The information of sounds and phonetic digrams of one and two-syllable words," *J. Speech Hearing Disorders*, **19**, p. 397, 1954.

frequently happens that a speaker is understood even though the sounds he produces are widely different from the hearer's standards—so different that a direct comparison may be impossible. Under these circumstances the hearer takes a little time—a few seconds, minutes, or months, depending on how outlandish the speaker's accent may be—before he begins to understand. During this interval, his brain is analysing the sounds received and formulating a new set of standards, applicable only to that particular speaker. The process of establishing this set of standards involves comparison of the various sounds with each other, bearing in mind possible combinations of phonemes, and hence a decision as to which sound represents which phoneme. Having discovered the relationship of speech sounds to phonemes, the process is still not finished, for the new relationships will be forgotten if not repeatedly reinforced, so that in fact the process of comparison of a speaker's speech sounds with each other is always going on in the listener's mind whenever he is listening to speech. This applies no matter who the speaker may be, although the process is relatively less important, the closer the speaker's speech habits to those of the listener.

The mental processes involved are similar to those used in listening to music, when chords and melodies are understood by the relative pitches of the notes used; very few listeners are able to judge the absolute pitch of a note, and this is immaterial to musical appreciation.

Mr. Denes's machine is less well able to understand the speech of persons other than the one who provided its memory store. This can be attributed to departures of their speech sounds from those stored as patterns for the various phonemes. At first sight, it might appear that the way to improve performance would be to incorporate a large number of memory stores, each with a different set of sounds to represent the phonemes. There would still be the difficulty of incorporating a means for the machine to decide which memory store was appropriate for each speaker. If the machine could be provided with a faculty similar to the human one of learning each speaker's sound system and, by comparison and analysis, relating his sounds to the various phonemes of the language, this might, in the long run, prove a simpler method of enabling the machine to understand a large number of speakers with different speech habits.

G. R. G. Plant: The subject of the papers is of particular interest to me because I have undertaken various experiments on behalf of deaf children during the past few years, and some of these have related to the visual analysis of speech sounds. In a prototype speech analyser we have avoided the necessity for a large number of passband filters by employing a sweep oscillator working in conjunction with a single crystal filter, whereby the whole of the speech spectrum is swept repetitively and without the use of mechanically moving parts. The information derived is then presented across the screen of a cathode ray tube and provides, if somewhat crudely, different visual patterns for the different speech sounds. I wonder whether Professor Fry experimented with any similar form of panoramic sweep analyser before deciding upon the particular multi-filter acoustic analyser which is employed in the mechanical speech recognizer.

Dr. T. B. Tomlinson (Associate Member): The use of linguistic information has been shown to be of exceptional importance and in the equipment described in this paper the recognition of a particular phoneme is conditioned to a marked degree by the identity of the preceding phoneme stored in the memory. It occurs to me that perhaps excessive "weight" is being given to this aspect of speech recognition: clearly it is advantageous to make use of linguistic information in this manner *provided that the previous phoneme has been recognized correctly*. As used in the equipment illustrated by Fig. 8, the selection of an incorrect phoneme will prejudice the chances of correct recognition of the next one and so on. There is present a form of "positive feedback" for errors which could be a distinct disadvantage.

I would suggest that an improved system might be as follows:—

The multiplication products of the acoustic detector are fed into a first selection stage where only those outputs above a certain marginal level are allowed to pass on to subsequent selection circuitry. In this modified selection stage the "bias" or marginal level might be automatically set by the magnitude of the largest signal in such a way that only two. or possibly three, of the most probable phonemes are selected. The signals corresponding to these two or three phonemes are then applied to a second multiplier stage the other input being derived from the linguistic knowledge store. The outputs of these second multiplier stages are connected to a maximum detector stage, as in the equipment described by the authors.

In this manner, the use of redundant information is not applied until the choice has been restricted to several of the most probable phonemes and the possibility of a continued series of errors is likely to be reduced.

We have been given a very clear outline of the complexity of the problems of speech recognition and we have been shown that the equipment necessary to recognize a restricted number of phonemes is already quite elaborate. Would the authors care to give us some idea how much more complex the equipment might have to be in a practical situation, operating from a microphone and against a background of industrial noise? For instance, the noise of a speech typewriter, as used in the equipment described, could give rise to acoustic feedback and the only output might be a continuous repetition of the typewriter's most potent and favourite phoneme.

Dr. W. Saraga : In the case where the output device is a typewriter, it is necessary in order to recognize speech to supplement the acoustic analysis by a statistical linguistic analysis. I understand that the reason for this necessity is that speech sounds are not uniquely defined by their acoustic characteristics and that a human listener uses linguistic criteria in addition to acoustic criteria for speech recognition. On this basis it seems to me that in the case of a speech synthesis as output device the statistical analyser should be unnecessary as a human listener capable of performing the statistical analysis is available as the ultimate output device. Would the authors please state whether this conclusion is correct or not?

Dr. G. L. Hamburger (Member): Dr. Saraga's point recalled a children's party game where a circle is formed and one child thinks of a very simple sentence and whispers it into his neighbour's ear. He in turn whispers what he thinks he understood into his next neighbour's ear. This process continues until it has gone full circle. The last participant then pronounces the message which invariably bears no resemblance at all to the original. Surely we can draw an analogy between this innocent game and Dr. Saraga's suggestion, namely, that if speech synthesis as described this evening is used for the transmission of intelligence without the beneficial supplement of a statistical linguistic analysis it would seem that results may be most hazardous.

Dr. C. W. Miller (Member): As Mr. Denes explained how his equipment made use of a store of linguistic information and described the success attained, I am sure that we were all preparing to ask "How much difference does it make if this feedback is open-circuited?" Mr. Denes anticipated this question and in doing so he gave what is perhaps the most important result of the experiment. The absolute value of success in recognition of sounds with the equipment is obviously limited in many ways, but the improvement attained by using the linguistic information in a most elementary form is so obviously significant and is a delightful demonstration of principle. One immediately thinks of a converse experiment, i.e. how would a human equipment behave if deprived of linguistic information? Obviously the shorthand writer finds difficulty in dealing with material of unfamiliar context. Perhaps Professor Fry could comment on the recognition of sounds in an unfamiliar language or indeed in a series of nonsense sounds.

E. J. Wender (Member): Professor Fry gave us a clear idea on how individual speech mannerism would effect acoustic speech recognition and showed that the listener has to make use of his knowledge of the constraints operating in his language to understand and recognize the speech.

It follows. therefore, that under such circumstances, mechanical recognition becomes extremely difficult and the presence of a variety of errors must be expected.

If we consider two similar sounding phonemes or words in which only the first and perhaps the second letters or syllables are different, the percentage of correct acoustic reception of the articulations is not very high in the absence of further clues. In such cases in mechanical recognition, the wave-form analysis would only show deviations in the wave-form of the wave-front which, being more often than not distorted by speech mannerism, give rise to failure in acoustic and mechanical speech recognition.

I formed the impression, therefore, that the complex system of the mechanical speech recognizer would gain accuracy if the speech wave motion analysis, i.e. the resolution of the compound wave-form of the phoneme. could be taken a step further by emphasizing the wave-front forms of the phonemes or units.

Of the several ways that suggest themselves, I would choose a method of separating acoustically and/or mechanically, the voice modulations from the frequency pattern of each phoneme or unit. This would have the advantage that by way of a single operation, one could obtain a clearer picture of the emphasized wave-fronts and the separate wave-forms of modulations and pauses, etc.

Thus, would it not be just possible that in this manner, the verdict of the various recognizer units would be more often in unison, particularly when the memory section, as was described, has no clue to offer.

AUTHORS' REPLY

We agree with Dr. Tomlinson as to the importance of sequential errors in the machine and Dr. Hamburger's example shows how far they can affect recognition in a human listener. We have discussed this in a previous paper.[10] Both Dr. Tomlinson and Dr. Booth discuss the question of the relative weight to be assigned to acoustic and linguistic information. At present they are given equal weight. While there is no theoretical reason for altering this, future experiments may show the need for a change.

The main difficulty of operating the speech recognizer from a microphone input in a noisy background is that the recognizer cannot separate speech-like sounds from noise and would therefore deal with the two kinds of input in the same way.

Dr. Fatehchand gives estimates of channel capacity in various compression systems which differ somewhat from those given in the second paper. We estimate that a formant-tracking compression system with a Lawrence synthesizer would need a capacity of about 1500 bits/sec—250 bits/sec for each of the six channels—and hence a phoneme coding system would still provide a considerable saving.

We have experienced no greater difficulty in recognizing final consonants than initial ones. The gating difficulty Dr. Fatehchand refers to is due to the fact that it is not always easy to find a point in the acoustic sequence that corresponds to a phonemic division.

We quite agree with Dr. Saraga and have stated elsewhere[12] that errors by the machine can be corrected by presenting the output in a form familiar to the recipient—as is the case when the output is reconstituted as speech. Even in these conditions a certain level of accuracy in the output would be required and we doubt whether this can be achieved without the use of linguistic information in the machine. Experiments on the lines suggested by Dr. Miller have been done in the past and results can be found in Harvey Fletcher's book.† They confirm that the less contextual information the listener receives the less successfully can he recognize speech.

Mr. Waldron underlines some important points made in the first paper. The whole phonemic system is a system of relations and hence cannot be realized on the acoustic plane in absolute values. It is dangerous to talk about a set of standards used by the listener; we have tried to make it clear that he associates a whole range of acoustic values with a given phonemic category.

We have already suggested a self-adjusting system to correct for individual variations from speaker to speaker.[10]

† Harvey Fletcher, "Speech and Hearing in Communications," (Van Nostrand, New York, 1953).

NEW BRITISH STANDARDS

The following is a selection of the new and revised British Standards and Codes of Practice of interest to members which have been issued during recent months. Copies may be obtained from Sales Dept., British Standards House, 2 Park Street, London, W.1.

B.S. 308A. **Students' edition of Engineering Drawing Practice.** *Price 4s. 6d. (3s. to students).*

This specially abridged, low-priced version of B.S. 308: Drawing Office Practice, is intended for students who are in the early stages of an engineering education. Each page of the abridged version carries a cross-reference to the equivalent page of the parent edition.

B.S. 448. **Electronic-valve bases, caps and holders.** *Price 2s. each section.*

Four new sections, and one new replacement section have been issued. Their numbers and titles are: B5E, Base (Issue 2); B7G/F, Base, valve outline; B8D, Base, base pin position gauge valve outline; B8D/F, Base, valve outline; B9A/F, Base, valve outline.

B.S. 2134: Part 1: 1959. **Fixed electrolytic capacitors (aluminium electrodes) for use in telecommunication and allied electronic equipment. (Part 1: General requirements and use).** *Price 7s. 6d.*

This is one of the series of specifications dealing with components intended primarily for use in telecommunication and allied electronic equipment; Part 1 deals with *general requirements and tests* and Part 2—to be published later—will specify sizes, ratings, etc. of a standard range of electrolytic capacitors. Use of this standard requires reference to B.S. 2011, "British Standard climatic and durability tests for components," which fully describes a range of tests to which components may be subjected in accordance with the requirements of the relevant standards for the individual components.

B.S. 2948: 1958. **Slides and opaques for television.** *Price 3s.*

Relates to the general dimensions of two sizes of slides and two sizes of opaques for television. It defines those areas of the slides and opaques which are transmitted and the areas which are available for the essential pictorial and written matter in order that it may be reproduced on the screen of the average domestic receiver.

B.S. 2951: 1958. **Glossary of terms used in high vacuum technology.** *Price 7s. 6d.*

Gives definitions of terms accepted in high vacuum technology. The terms are arranged in sections dealing respectively with general terms, vacuum systems, pumps and pump components, manometers and gauges, leak detection and vacuum applications; the last named section includes terms used in well known vacuum processes including electronic-valve and lamp manufacture. Comprehensive notes on pressure units, critical backing pressure, and sensitivity definitions in leak detection are given in an appendix. A list of the symbols occurring in the Glossary is included; there is also a detailed index.

B.S. 2962: 1958. **Picture areas of motion picture films for television.** *Price 4s.*

Defines those areas of 35 mm. and 16 mm. films which are telerecorded and transmitted, and the areas which are available for the essential pictorial and written matter in order that it may be reproduced on the screen of the average domestic receiver.

B.S. 3045: 1958. **The relation between the sone scale of loudness and the phon scale of loudness level.** *Price 3s.*

In this new publication the sone scale is specified by its relation to the phon scale, the relation used being an approximation based on assessments of the available experimental evidence. It does not necessarily represent the degree of accuracy required for research purposes, nor should it be deemed to serve any function other than the conversion of loudness levels to loudnesses in sones, and vice versa.

The Foreword to the standard explains "The loudness level of any sound may be expressed uniquely in accordance with the procedure defining the phon scale (B.S. 661, Glossary of acoustical terms). Owing, however, to the arbitrary nature of the phon scale, loudness values expressed in phons have to be interpreted by the user on the basis of previously heard sounds of known loudness level.

"For some purposes, especially to convey to the non-specialist the relative loudness of different sounds, it is preferable to express loudness values in the units of an alternative scale. This scale, in which the loudness is expressed in sones, is so constructed that the apparent ratio of the loudnesses of two sounds as judged by normal observers, is given directly by the ratio of the sone values of the sounds."

B.S. 3081: 1959. **Basic dimensions for printed wiring.** *Price 3s.*

The basic dimensions and requirements specified in this new publication apply to the printed wiring used in telecommunication and allied electronic equipment. The 6-page standard does *not* deal with printed circuits or with strip lines for microwave techniques.

An appendix gives recommendations on those dimensional aspects of printed wiring for which, in view of the comparatively early stage of development, it is considered unwise to specify rigid requirements at the present time. Recommendations relating to components for use with printed wiring are also given. The content of the standard will be extended as experience is gained in the production and use of printed wiring.

# of current interest . . .

Tercentenary of the Royal Society

On 28 November, 1660 "a group of originally minded men meeting together in the City of London" decided to form themselves into a scientific society. King Charles II being informed of this constituted himself the formal Founder, and the new society, now the premier scientific society in the world, came to be known as The Royal Society of London, or more generally "The Royal Society."

The Society announces that it proposes to celebrate the tercentenary of its foundation from 18-26 July, 1960, and that a committee with Sir Cyril Hinshelwood, President of the Royal Society, as Chairman, has been set up to organise the celebrations. Leading scientists from all nations are expected to attend the events which are to include special visits to the City of London, and the Universities of Oxford and Cambridge, both of which have important links with the Society's development over three centuries.

R.E.C.M.F. Council and Officers

At the annual general meeting of the Radio and Electronic Component Manufacturers' Federation, Mr. E. M. Lee, B.Sc. (Member), was elected President. Mr. Arthur F. Bulgin, M.B.E. (Member) was re-elected a vice-president and Mr. C. M. Benham, B.Sc. (Member), was elected treasurer.

A New Higher Technological Award

The National Council for Technological Awards has announced that in order to encourage qualified men and women to undertake further study beyond the level of the Diploma in Technology and carry out original investigations, they have decided to create an award higher than the Diploma in Technology. Their intention is that this award shall be a mark of outstanding distinction granted to a student who has proved his ability by completing a substantial programme of work demanding the application of his knowledge to the solution of a problem of value to industry.

The Council will establish a college to be known as The College of Technologists, and the new award will take the form of Membership of this College (M.C.T.). To qualify for this new award a student must undertake a programme of work to be carried out jointly in industry and at a technical college. This programme may be concerned with any technological aspect of industrial activity, such as research, development, design, production or market investigation.

The Institution of Production Engineers

Subject to the appropriate resolution being passed at an Extraordinary General Meeting of the Institution of Production Engineers on April 30th, a petition is being submitted on behalf of that Institution for Incorporation by Royal Charter.

The Institution of Production Engineers was founded in 1921, incorporated in June 1931, and has a membership of 10,850 (December 1958).

Chinese Scientific Literature

The Lending Library Unit of D.S.I.R. has started to collect Chinese scientific literature and about 150 periodicals are now on regular order. An article in the Unit's "Bulletin" suggests that we cannot ignore the scientific output of China, which may be growing rapidly. Mention of this vital problem of significant technical information in languages other than English was made in an editorial in the December 1957 *Brit.I.R.E. Journal*.

The Lending Library has a large collection of Russian scientific literature and organises a scheme for translations in collaboration with the National Science Foundation in the United States. This may possibly be extended in the future to include scientific literature from China. Meanwhile. the contents of these Chinese publications must be assessed, and a scientist with a knowledge of Chinese is being recruited by L.L.U. to select and promote use of Chinese scientific and technological literature.

The Measurement of the Time Delay of Ultrasonic Delay Lines†

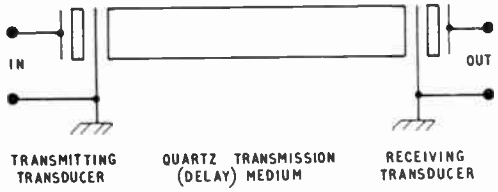
by

J. SEARS, ASSOCIATE MEMBER‡

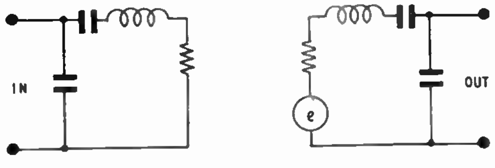
Summary: A method of determining the delay of wide band ultrasonic delay lines to an accuracy of better than 0·1 microseconds is described. The experimental technique and estimation of errors is given. Alternative methods are discussed.

1. Introduction

Two types of electro-mechanical sonic delay lines are in general use, those using a guided wave in which the wavelength is large compared with the transmission aperture[2, 3], and those in which the aperture is large compared with wavelength[1, 3], so that the information is beamed through the transmission medium in a manner analogous to radiation from an aerial. The "guided" lines are commonly used for the lower frequencies and employ wire as the transmission medium while the "beamed" type used for higher frequency carrier applications usually have quartz as the transmission medium. The latter type is illustrated schematically in Fig. 1.



(a) Mechanical arrangement.



(b) Electrical equivalent.

Fig. 1. Quartz delay line.

The transducers (usually quartz elements) exploit the piezo-electric effect to transform the electrical input to mechanical energy at the input to the delay medium, and to convert it back to an electrical signal at the output.

Figure 2 similarly illustrates a wire delay. The wire (at least, that part enclosed by the

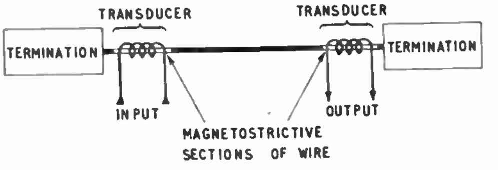


Fig. 2. Schematic arrangement of the wire delay line.

transducer coil) is of magneto-strictive material so that an input consisting of an electric current to the transmitting transducer coil is converted into mechanical energy in the wire and reconverted at the receiving transducer.

This note is concerned with a particular case of the measurement of the delay of the "beamed" type of line; the technique, however, should be extensible to other types.

The lines in question used a carrier frequency of 15 Mc/s. with approximately 6 Mc/s bandwidth and nominal delays from 10 to 650 microseconds. The error in measurement was to be small compared with 0·1 microseconds.

2. Possible Methods

A number of methods were considered and are briefly outlined as follows.

2.1. *Calibrated shift on c.r.o. display (with time markers)*

The shift is first calibrated from the markers and then the markers into, and out of, the delay are displayed. The time delay is read directly.

† Manuscript first received 23rd November, 1958 and in final form on 3rd February 1959. (Paper No. 497.)

‡ Formerly with the Ministry of Supply, Royal Aircraft Establishment, Farnborough; now with the U.K. Atomic Energy Authority, A.E.R.E., Harwell.

U.D.C. No. 534.22:621.374.32

The accuracy depends on the definition of the markers and the accuracy of the p.r.f. and is better than 0·1 per cent. for delays greater than 250 microseconds.

2.2. *Beat method*

A c.w. oscillation is fed into the delay line and the output is added to the input; as the frequency is varied the amplitude of the sum varies. By counting maxima and noting the corresponding frequency change the delay can be computed, i.e.

$$\tau = \frac{N}{\Delta f}$$

where N is the number of maxima
and Δf is the corresponding frequency change.

2.4. *Decaying pulse train*

A pulse is generated in synchronism with the leading edge of the gate of a counter. The pulse is introduced into a delay-amplifier ring adjusted to produce a decaying pulse train. The pulse frequency is sampled during the counter gate period.

The method has not been considered in detail by the writer but it appears to be operationally simple. The main limitation and source of error is in the progressive deterioration of pulse shape due to bandwidth limitations.

2.5. *Pulse modulated carrier*

This method was chosen as providing the best compromise between accuracy and ease of

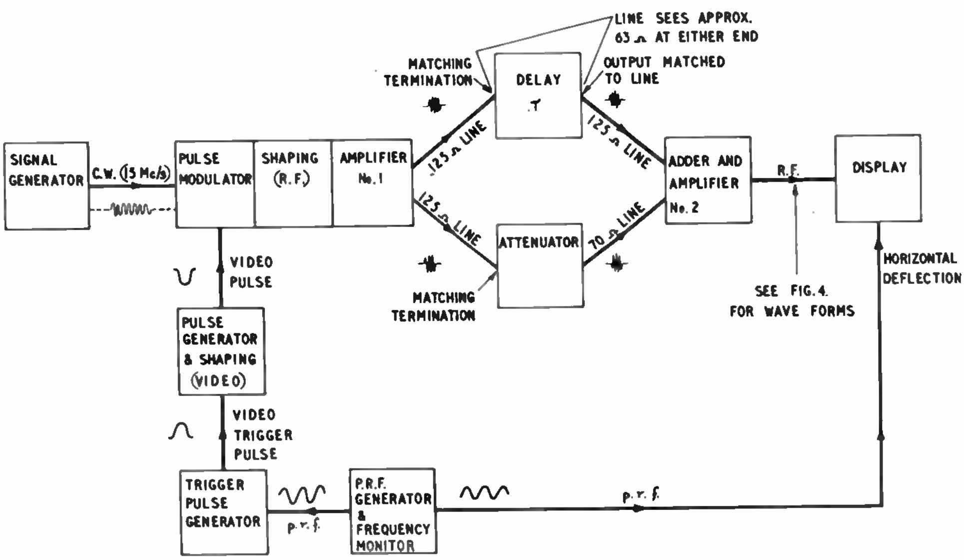


Fig. 3. Block diagram of equipment.

The method is capable of a good degree of accuracy provided the multipath transmission is small and is best suited to the longer delay wide band lines where the integer uncertainty in N (as a fraction of N) corresponding to a defined value of Δf can be less than the acceptable error.

2.3. *Counter method*

A counter may be switched on by an input pulse to the delay line and switched off by the output pulse. The method is rapid and simple. The accuracy is limited by the counting rate and inaccuracies due to waveform distortion; delay in associated amplifiers, etc. is also present.

operation in the particular circumstances of the work.

The method of measurement used (see Fig. 3) was to feed a pulse modulated carrier to the delay under test and then to add together the input and output pulses, the input pulse being attenuated before addition so that the two pulses were of comparable amplitude at the summation point. The resulting waveform was then a train of pulse pairs with a time separation between input and output pairs equal to the delay time and a pair repetition period equal to the reciprocal modulation pulse recurrence frequency. It is apparent that when the reciprocal pulse recurrence frequency is equal to the line delay time, input and output pulses will be coincident, and if the carrier phasing is correct this coincidence will be indicated by a null in the waveform amplitude. Now the value of the pulse recurrence frequency can easily be determined to a high degree of accuracy with the aid of a digital frequency meter; the time delay of the line can thus be determined.

2.6. *Use of a sinusoidally modulated carrier*

This method, similar in some respects to 2.5., was also considered. If the pulse modulation in the method described is replaced by sinusoidal modulation, minima of output will again be related to the modulating frequency and the delay can be found from the expression

$$\tau = \frac{2n-1}{f}$$

where f is the modulating frequency and n is the order of minimum as f is increased.

Accuracy increases with modulating frequency but care is necessary to avoid ambiguities. The effects of phase distortion and multipath transmission can be serious.

3. Experimental Procedure

Before analysing the method it is necessary to appreciate in some detail the experimental procedure.

3.1. *Apparatus*

The apparatus was set up as shown in Fig. 3. The signal routing is indicated by the arrows and the nature of the signal shown pictorially at appropriate points. A 15 Mc/s signal from the signal generator is pulse modulated and passes via amplifier No. 1 to the delay line and attenuator. The outputs of these are summed at the input to amplifier No. 2 and the resulting signal displayed on an oscilloscope.

The p.r.f. generator produces a variable-frequency sine wave which provides horizontal deflection for the display and drives the pulse trigger generator. This latter has an adjustable delay provision which is of value in centering the pulses on the display. The pulse generator delivers a pulse of closely defined shape to the pulse modulator, the final shape of the r.f. pulse being determined by a stage of shaping (bandwidth restriction) between the modulator and amplifier No. 1. Pulse shaping follows from the need to define bandshapes accurately so that the mathematical assessment of the measurement errors is reasonably easy. The matter is dealt with further in Section 4 and in the Appendix.

3.2. *Measurement of Delay*

The experimental procedure consists of the following steps.

(i) The p.r.f. is set at about $1/\tau$, where τ is the nominal value of the delay; the display then appears as shown in Fig. 4 (a).

(ii) The attenuator is adjusted for approximate equality of pulse height.

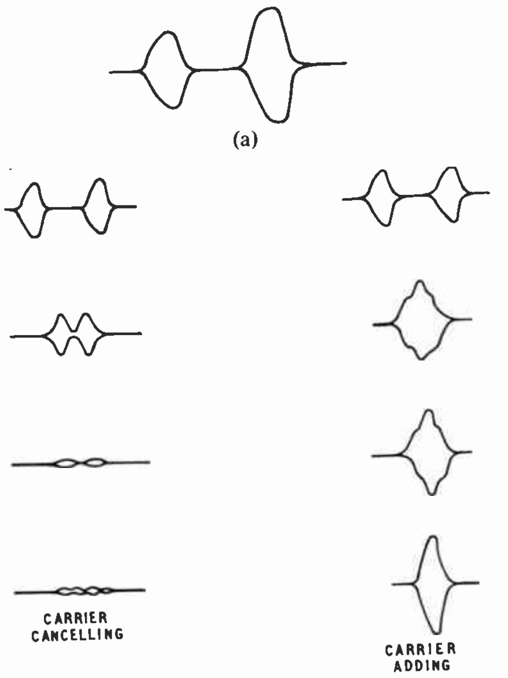


Fig. 4. (a) and (b) Display wave forms (r.f. envelopes).

(iii) The signal generator output is adjusted until the pulses are of an amplitude just short of saturation.

(iv) The p.r.f. is adjusted until the delayed and undelayed pulses "coincide." This is apparent when, as the pulses move over one another, they interact instead of merely "co-existing."

The appearance of the display as the pulses approach coincidence is shown in Fig. 4 (b).

(v) Carrier frequency f_0, recurrence frequency f_p, and attenuator setting are now adjusted sucessively for minimum output. In the desired condition the carriers of the delayed and undelayed pulses cancel, and the delay is then given by

$$\tau = 1/f_p.$$

4. Estimated Accuracy

For any one measurement, it will be seen that the measured value of delay τ is given by:

$$\tau = \frac{n}{f_p \pm \Delta f_p \pm \Delta' f_p} + \Delta\tau \pm \Delta'\tau$$

where τ is the indicated delay
f_p is the true p.r.f.
Δf_p is the systematic error in measuring f_p
$\Delta' f_p$ is the random error in f_p
$\Delta\tau$ is the systematic error in τ
$\Delta'\tau$ is the random error in obtaining pulse coincidence
n is an integer (unity in the experiment described).

These errors will now be considered in detail.

4.1. *Systematic errors*

Systematic error in f_p is due to departure from its nominal value of the counter sampling gate width due, for example, to error in crystal oscillator frequency.

The accuracy is better than 0·005 per cent. (manufacturer's figure) so that $\Delta f_p < 5 \times 10^{-5} f_p$.

Systematic error in τ is due to:—

(i) Asymmetry of delay in the connecting leads.
(ii) Delay in the attenuator.
(iii) Errors caused by circuit bandwidth limitations.

Only (iii) was considered to be of any significance.

Now it is shown by Cherry[4] that the delay of a filter is equal to the phase/frequency slope. For the quartz delay medium this is sensibly linear over the pass-band (set by the transducers) but for the transducers it is not linear, thus these latter may be considered to add a true delay to the quartz delay plus distortion to the delay/frequency characteristic due to the non-linearity of their phase/frequency characteristic.

The effect of this distortion on the present method of measurement is discussed in the Appendix and is found to be of the order of − 0·007 microseconds for a delay line bandwidth of 6 Mc/s. The delay of the resonant circuits is also determined and is found to be 0·069 microseconds total. This is by definition included in the delay to be measured but will vary inversely with the transducer bandwidth.

4.2. *Random errors*

Random errors arise from:—

(i) Counting uncertainty in the frequency monitor. ($\Delta' f_p$).
(ii) Coincidence error ($\Delta'\tau$) i.e. random errors in the determination of the coincidence minimum.

The counter counts to the nearest integer and so give rise to a frequency error equiprobable in the range $\pm 0{\cdot}5$ c/s in the counting period. This corresponds to $\pm 0{\cdot}5$ c/s for a 1 second count and $\pm 0{\cdot}05$ c/s for a 10 second count.

The coincidence error, which is due primarily to limitations (enhanced by thermal noise and frequency stability limitations) in the operator's powers of discrimination and skill, was for the conditions of the experiment 0·006 microseconds r.m.s., established from some hundred observations.

Table 1

Delay	Limit of Constant Error (a)	R.M.S. Random Error (b)	Overall Error to 99% confidence (c)	Counter Gate Width (seconds)
10	0·0005	0·006	0·0165	1
100	0·005	0·0067	0·025	1
100	0·005	0·006	0·021	10
500	0·025	0·0094	0·053	10
750	0·0375	0·0173	0·081	10

Assuming a rectangular probability distribution for the counting uncertainty and a normal distribution for the coincidence error, the combined error distribution is their convolution.[5] This distribution was used to establish the limits of error given in Sect. 4.3.

4.3. *Overall Accuracy*

Columns (a) and (b) in Table 1 opposite show the constant (systematic) and random errors for the measurement of different delays in the range initially specified. The figures were determined from the equation for τ and from the values of the different forms of error considered in Sections 4.1 and 4.2. The overall errors are shown in column (c) to better than 99 per cent. confidence. All error values in the table are in microseconds.

5. Details of Circuitry

Certain items, the p.r.f. generator and frequency monitor, trigger generator and attenuator, were commercial products whose function relative to the subject has been dealt with; other items were designed for the particular application and a brief treatment of their functioning is pertinent.

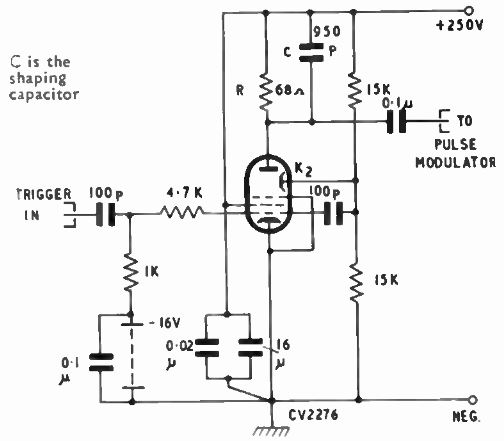


Fig. 5. Pulse generator and video shaper.

5.1 *The pulse generator*

The circuit is shown in Fig. 5, the valve being a secondary emission pentode.[6] In the absence of a trigger voltage it is cut off. When a positive-going trigger pulse is applied the valve conducts and the dynode potential rises. Positive feedback to the grid via the 100pF capacitor causes a very rapid current build up into the anode load. The action reverse when the grid potential falls below zero.

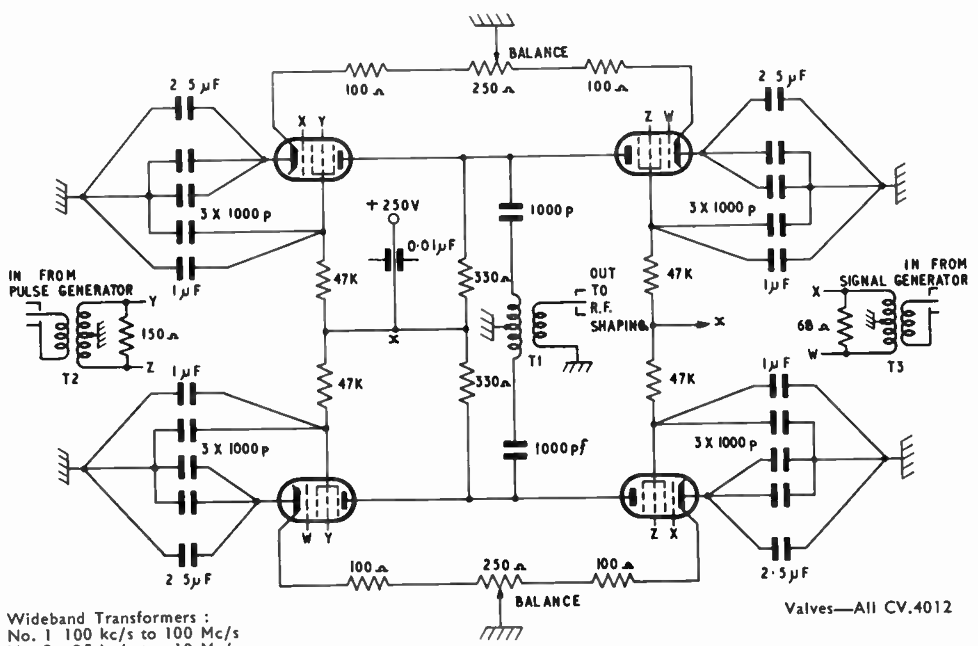


Fig. 6. Pulse modulation.

The anode load consists of the valve stray capacitances to which are added a comparatively large capacitance C1 and a parallel resistance, R. The resultant approximates closely to the ideal of a rectangular pulse of current injected into a parallel CR circuit, the voltage change across the capacitor then being substantially exponential with time.

shaping already introduced is desirable to remove any remaining traces of video breakthrough. It is placed in the first stage of Amplifier No. 1 (Figs. 3 and 7) by passing the r.f. pulse through a single LCR circuit of appropriate bandwidth. The gain/frequency characteristic of the remainder of the amplifier is flat within one decibel from 5 to 24 Mc/s.

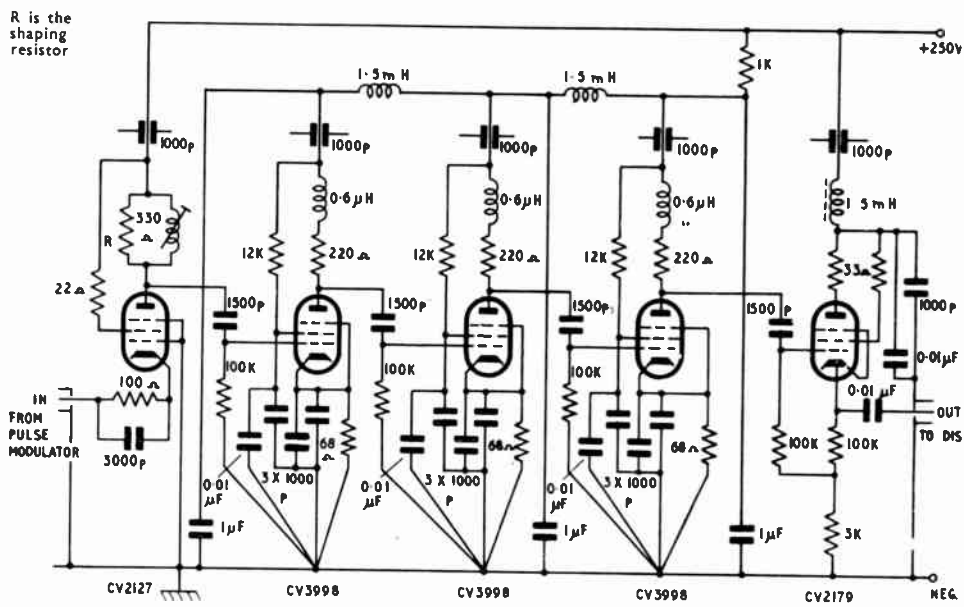


Fig. 7. R.f. shaper and amplifier No. 1.

The arrangement further provides bandwidth limitation to reduce video harmonic components at carrier frequency which confuse the appearance of the display. This effect will be referred to as video breakthrough.

5.2. *The modulator* (Fig. 6)

A balanced modulator is required to reduce video breakthrough which results in spurious minima. In addition, prevention of carrier break-through (i.e. transmission of the carrier frequency between pulses) was desirable for other applications. A double balanced arrangement was therefore used resulting in a video and carrier breakthrough more than 30 db down on the required output, and a substantially linear modulation characteristic.

5.3. *R.F. shaper*

A further stage of shaping as well as the video

5.4. *Adder and amplifier* (Fig. 8)

As described, the output and the attenuated input of the delay under test are fed to a summation point and are then amplified before being displayed on the oscilloscope. The summation circuit and amplifier are combined in one unit as shown in the figure.

The amplifier employs the feed-back principle[7, 8] to provide a pass-band flat to better than 0·5 db at 15 ± 3 Mc/s.

6. Conclusion

The method achieved the required degree of accuracy, and in practice, particularly where numbers of nominally similar lines were tested, was quite rapid.

7. Acknowledgment

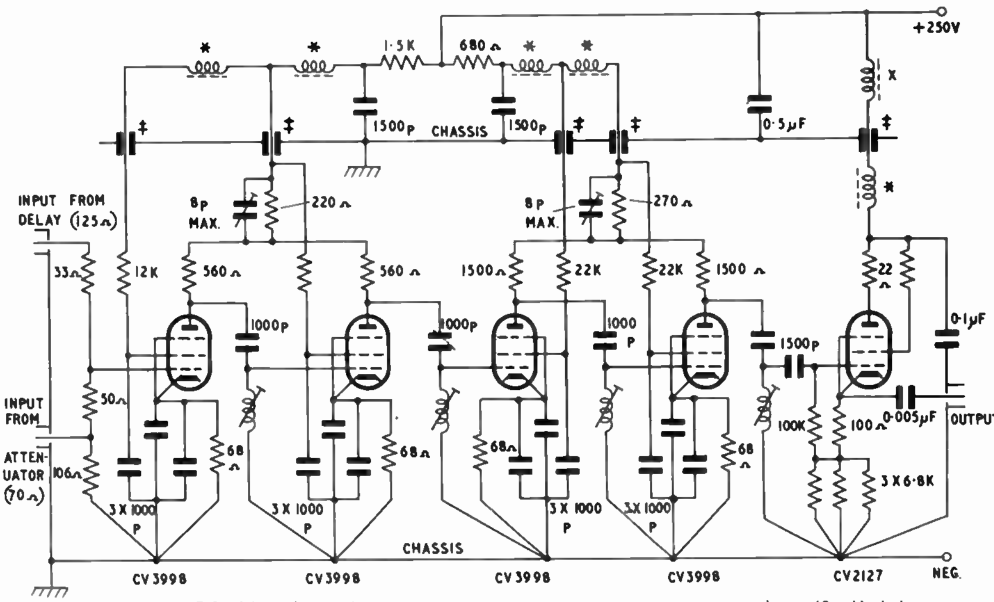
Thanks are due to Mr. N. R. Bailey of the Royal Aircraft Establishment for his general

advice and encouragement.

This paper is published by permission of the Controller, H.M. Stationery Office.

8. References

1. David L. Arenberg, "Ultrasonic delay lines," *1954 Convention Record of the I.R.E.*
2. G. G. Scarrot and R. Naylor, "Wire-type acoustic delay lines for digital storage," *Proc. Instn Elect. Engrs*, **103B**, Supp. No. 3, pp. 497-508, 1956.
3. S. Morleigh, "A survey of delay lines for digital pattern storage," *Electronic Engineering*, **30**, pp. 380-7, June 1958.
4. E. C. Cherry, "Pulses and Transients in Communications Circuits," para. 4.38. (Chapman and Hall, London, 1949.)
5. P. M. Woodward, "Probability and Information Theory with Applications to Radar." (Pergamon Press, London, 1955.)
6. F. H. Wells, "Pulse circuits for the millimicrosecond range," *J. Brit.I.R.E.*, **11**, pp. 491-503, November 1957.
7. H. S. Jewitt, "Wideband i.f. amplifiers," *Wireless World*, **60**, pp. 86-90, February 1954.
8. H. S. Jewitt, "Component tolerance effects in feedback i.f. amplifiers," *Electronic Engineering*, **28**, pp. 165-7, April 1956.
9. E. C. Cherry, *ibid.*, para. 4.41.



* are 10,000 pF feed branch capacitors ‡ are 40 μH chokes

Fig. 8. Summing circuit and amplifier No. 2.

9. Appendix: Consideration of the effect of the Transducers on the Measurement

Assuming the transducers to behave like simple symmetrical resonant (LCR) circuits they add a real component to the ultrasonic delay[4] plus an experimental error due to phase distortion. For a linear phase characteristic it is shown[4] that the undistorted delay is equal to phase/frequency slope, which in this case is taken as the slope at the band centre.

Now, the phase shift in one transducer is

$$\phi = \tan^{-1} Q\left[\frac{\omega_0}{\omega} - \frac{\omega}{\omega_0}\right]$$

where ω_0 is the band centre frequency
Q is circuit amplification factor
ω is frequency (radian/sec.).

Differentiating with respect to ω, the slope at the band centre ($\omega = \omega_0$) is found and hence the undistorted delay is shown to be $1/\pi B$ seconds where B is the half-power bandwidth of the transducer in cycles.

The experimental error, occasioned by distortion of the delayed pulse due to non-linearity of the transducer phase characteristic is determined as follows.

If the shape of the input pulse is known, the output pulse shape can be determined by application of DuHamel's integral[9] and (using graphical methods) the relative timing of delayed and undelayed pulses can be found for their sum to be a minimum. Comparing this relative timing with the undistorted transducer delay the measurement error due to distortion is found.

For simplicity in the present case, the shaping circuits were made similar in character to the transducers and it was assumed that the LCR circuits behaved like their zero-frequency carrier equivalents[9]. Then the instantaneous amplitude of the pulse envelope at the output of any of the equivalent RC sections is:—

$$e_0(k) = E\left(1 - e^{-k}\sum_{1}^{n}\frac{k^{n-1}}{(n-1)!}\right)$$

where $k = t/T$

$T = R.C$ (or $1/\pi B$ for the equivalent LCR circuit)

E = amplitude of the driving voltage step

n = number of sections

t = time from the origin of E.

From this expression the curves of Fig. 9 were obtained and by manipulation of appropriate pairs (two and four in this case), the corresponding minimum and origin difference (relative timing) was obtained (Fig. 10). It will be seen that the difference between the position of the minimum and the transducer delay ($2/\pi B$) is about 0·007 microseconds which is small compared with the required accuracy.

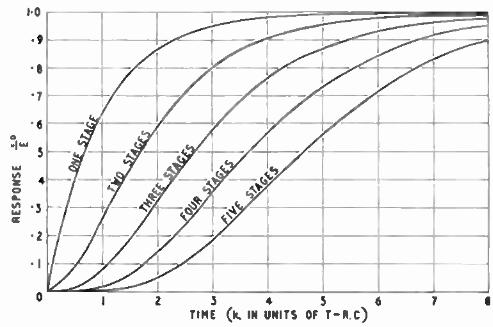


Fig. 9. The response (e_0) to a step input (E) of a series of similar RC low pass filters.

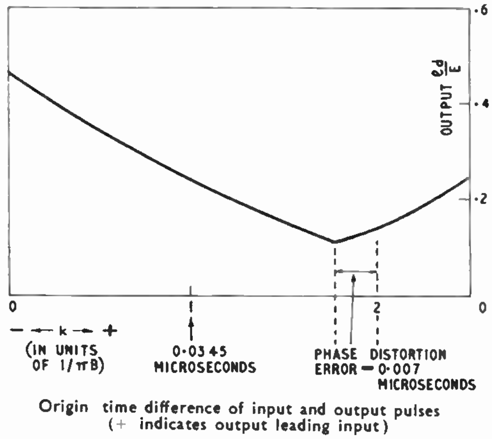


Fig. 10. The output of the summing circuit as a function of the relative pulse origin timing, with two stages of shaping.

RECOMMENDATIONS FOR MEDICAL ELECTRONIC INSTRUMENTATION

One of the terms of reference of the Medical Electronics Committee is to make recommendations on standards and specifications for medical electronic instruments where the medical requirement has been established and agreed.

In all these cases it will be necessary, since the report is for engineers, to give the medical background briefly, supported by as much additional material in the way of references and bibliography as is available. This will enable the engineer who is particularly interested in this subject to extend his knowledge through the existing literature.

It is proposed that a number of different instruments will be covered, the first of which will be constant voltage electro-diagnostic stimulators. The Committee is indebted to the Editor of *The British Medical Journal* for permission to incorporate the electrical specification of a constant voltage electro-diagnostic stimulator which formed part of the report of a Medical Research Council sub-committee on nerve injuries.

1. Constant Voltage Electro-Diagnostic Stimulators†

A report prepared for the Medical Electronics Group Committee by R. Brennand (Associate Member).

1. Introduction

When a person sustains an injury resulting in muscle paralysis, early diagnosis and prognosis is of considerable importance. For example, if recovery is not proceeding normally, then appropriate surgical treatment can be undertaken as soon as possible. One method of assessing muscle recovery is by the determination of its response to electrical stimulation.

A variety of methods of employing electrical stimulation for this purpose has been discussed in the medical literature, for example, by Ritchie[1] and Wynn-Parry[2], while the design of electronic stimulators for clinical use has been described by Grey Walter and Ritchie[3] and Sneath and Mayer[4].

However, finality of specification or design has by no means been achieved to date and it is thought that the present note may be of assistance to engineers and others contemplating work in this field.

1.1. *Scope*

This communication is mainly concerned with constant-voltage stimulators for routine clinical use, this type of instrument having been favoured by the Medical Research Council sub-committee.[9] For the present purpose a constant-voltage stimulator may be defined as one having an output impedance which is low when compared with that of the patient. The specialized requirements of stimulators for physiological research (see, for instance, references 5 and 6) will not be considered here.

2. Principles of Electro-diagnostic Stimulation

Modern electro-diagnostic stimulation is based on the differential response of nerve and muscle fibres to percutaneous electrical stimulation. In a normal healthy muscle the response to an electric stimulus depends on the continuity of the associated nerve fibres. A normally innervated muscle will respond readily to short duration (< 1 millisecond) electrical stimuli.

When a muscle loses its nerve supply due to injury or disease it is said to be denervated. In this case, on stimulation, it is found that the muscle readily responds only to relatively long duration impulses (> 3 milliseconds). This is because the muscle fibres have now to be stimulated directly, and muscle fibre is inherently much less excitable than nerve.

2.1. *Intensity duration curves*

If rectangular d.c. pulses of varying amplitudes and durations are applied to a muscle so

† Report No. 17. Approved for publication by the Council on 14th January 1959.

U.D.C. No. 616-073:621.374

as to produce a standard response, a curve relating pulse amplitude and pulse width may be drawn as in Fig. 1. The technique is briefly as follows. (For a detailed explanation of the procedure see reference 2, page 235). After making suitable preparation to ensure repeatability, pulses of 300 milliseconds duration at a frequency of 30 pulses per minute are applied to the muscle under test via two lint covered brass electrodes 1 cm in diameter. The lowest pulse amplitude required to produce the smallest observable contraction of the muscle is determined. This value of pulse amplitude is known

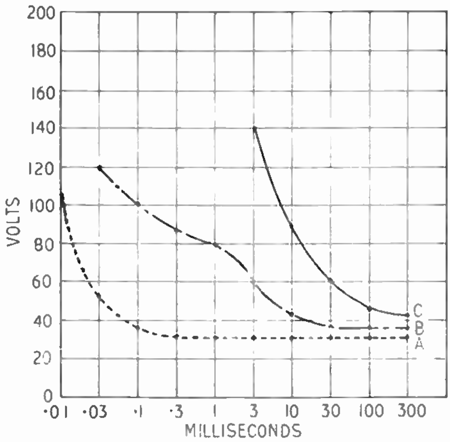


Fig. 1. Typical constant-voltage I.D. curves.

as the rheobase. The stimulus duration is then successively reduced, and the amplitude required to produce the standard contraction at each pulse width is recorded. A curve may then be drawn with pulse amplitude as ordinate and pulse width as abscissa: it is known either as an intensity-duration (I.D.) or a strength-duration (S.D.) curve. (Fig. 1, curve A).

It may be of interest to note that I.D. curves may be plotted on other excitable tissues, e.g. Bjorn[13] has described the stimulation of teeth. In this case the subjective reaction of the patient is employed as the standard response.

Another way in which the response of the muscle to electrical stimulation may be expressed is in the determination of chronaxie. Chronaxie is defined as the stimulus duration required for the smallest observable muscular contraction with a stimulus amplitude of twice the rheobase.

Depending on the type of stimulator employed (constant voltage or constant current) the pulse amplitude may be expressed in either volts or milliamperes. (See Sect. 2.2).

The stimulus amplitude for smallest observable contraction in normal muscle is the same over a wide range of pulse widths and rises only when very short duration stimuli are used. (Fig. 1, curve A). This curve represents a measure of the excitability of nerve and indicates that the stimulus has excited the intramuscular nerve fibres.

When denervation is present the muscle fibres alone are stimulated, and since these are very much less excitable than nerve, the I.D. curve rises steeply from the rheobase and no response is obtained at the shorter pulse widths. (Fig. 1, curve C).

During reinnervation (the process by which a regenerating nerve re-establishes contact with the muscle fibres) the I.D. curve shows progressive change from the muscle to the nerve type of response. Thus the curve is typically discontinuous, showing the combined response of both types of excitable tissues (curve B). A slight discontinuity is the earliest sign of reinnervation which may precede the first usually observed signs of recovery by as much as one and a half months[2]. Recovery can be followed by the progressive return to the "normal" type of curve found by the serial plotting of the I.D. curve.

Darcus[14] has discussed in detail the function and anatomy of the neuro-muscular system, and includes a short account covering all aspects of nerve degeneration and regeneration.

2.2. *Types of stimulator employed in electro-diagnosis*

As stated in the previous section, the pulse amplitude may be expressed in milliamperes or volts, according to whether the apparatus used is of the constant current or constant voltage type. In the case of the former, the output impedance is typically 100,000 ohms or greater, and for the latter, 500 ohms or less. It has been suggested that constant current techniques are more accurate since the patient current is independent of skin and electrode resistance. However, the constant current stimulus is not so well tolerated by the patient[1, 7], and it is doubtful if any greater accuracy results[7, 9] since it is the shape of the curve and not the absolute value which is of significance.

A recent report[9] by a sub-committee of the Medical Research Council's Nerve Injuries Committee has decided in favour of the constant-voltage stimulator for routine purposes and the electrical specification recommended by the Committee is reproduced in Section 3.

2.3. *Patient voltage–current relationships*

If the patient impedance were purely resistive, the two types of stimulator would produce the same result. In practice the patient impedance is complex, consisting of both capacitance and resistance, so that the voltage and current waveforms are not comparable.

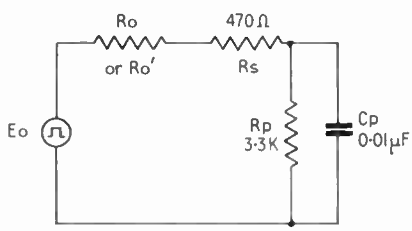


Fig. 2. Simplified patient equivalent circuit for electro-diagnosis.

A somewhat over-simplified equivalent circuit approximating to electro-diagnostic conditions is shown in Fig. 2, while Fig. 3 gives equivalent circuits during the pulse period. The following notation is employed:—

- E_o open circuit output voltage of the stimulator
- R_o stimulator output impedance at the selected value of E_o
- R_o' stimulator output impedance during the trailing edge of the pulse
- R_s series resistance of patient equivalent circuit
- R_p shunt resistance of patient equivalent circuit
- C_p patient equivalent capacitance
- E_p instantaneous patient voltage
- I_p instantaneous patient current
- $I_{p1} = I_p$ during pulse leading edge
- $I_{p2} = I_p$ during pulse top
- $I_{p3} = I_p$ during pulse trailing edge

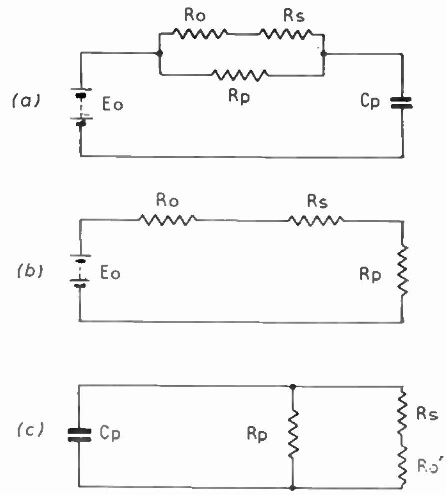


Fig. 3. Equivalent circuits for (a) Pulse leading edge; (b) Pulse plateau; (c) Pulse trailing edge.

2.3.1. Patient current

With an idealized square pulse input the initial overshoot of the current waveform (see Fig. 4) would have a peak amplitude of

$$I_{p1} = \frac{E_o}{R_o + R_s} \qquad \text{.........(1)}$$

At any value of I_p this overshoot will have a duration of

$$t_1 = C_p R' \log \frac{I_{p1} - I_{p2}}{I_p - I_{p2}} \qquad \text{.........(2)}$$

$$\text{where } R' = \frac{(R_o + R_s) R_p}{R_o + R_s + R_p}$$

and I_{p2} is defined by eqn. (3).

During the pulse the current will be

$$I_{p2} = \frac{E_o}{R_o + R_s + R_p} \qquad \text{.........(3)}$$

When the pulse finishes C_p will discharge producing a peak current of

$$I_{p3} = -\frac{E_o R_p}{R_o + R_s + R_p (R_o' + R_s)} \qquad \text{.........(4)}$$

For any particular value of I_p this undershoot will have a duration of

$$t_2 = C_p R'' \log \frac{I_{p3}}{I_p} \qquad \text{.........(5)}$$

$$\text{where } R'' = \frac{R_p (R_o' + R_s)}{R_p + R_o' + R_s}$$

It should be noted that, in general, $R_o' > R_o$ (see Sect. 2.4).

A typical patient current waveform for the case of the small electrodes quoted in Sect. 2.1. is shown in Fig. 4, where the effect of increasing R_o' to 10,000 ohms may also be seen. The

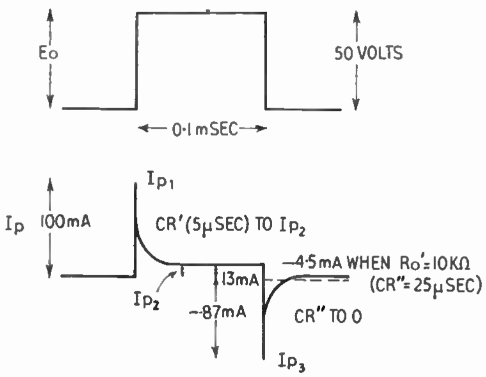


Fig. 4. Typical output voltage and patient current waveforms for constant-voltage diagnostic stimulator.

stimulator employed had no-load rise and decay times of approximately 0·5 microseconds, while R_o and R_o' were 10 ohms and 25 ohms respectively. The peak output current capability was 3 amperes.

2.3.2. Patient voltage

If R_o and R_o' are low compared with R_s, the patient voltage E_p approaches the open circuit stimulator voltage E_o. However with the majority of stimulators now in use this is not the case. Figure 5 shows typical waveforms when $R_o = 500$ ohms, $E_o = 50$ volts and R_o' is 500 ohms or 10,000 ohms.

The patient electrode voltage first rises immediately to an amplitude

$$E_{p1} = I_{p1} . R_s \qquad \text{.........(6)}$$

and then continues to rise exponentially with a time constant of C_pR' to a maximum value of

$$E_{p2} = I_{p2} (R_s + R_p) \qquad \text{.........(7)}$$

The trailing edge of the patient voltage waveform will at first fall immediately by

$$E_{p3} = (I_{p2} - I_{p3}) R_s \qquad \text{.........(8)}$$

followed by an exponential decay of time constant C_pR''.

The constant-voltage stimulator thus produces a patient current waveform which may be regarded as differentiation of the stimulating pulse. The constant-current stimulator on the other hand produces a patient voltage waveform which is integration of the output current pulse.

The practical significance of this waveform distortion is that the I.D. curves and the chronaxie index obtained with the two types of stimulator are not comparable.

The chronaxie index for example is about ten times longer with constant current than with constant voltage stimulation.

2.4. *Stimulator output impedance*

One result of the complex patient load is the large variation in patient current occurring during the pulse period.

Certain types of stimulator present different values of output impedance during the various phases of the pulse period. As is well known, a cathode follower providing a positive pulse output into a capacitive load may be cut off on the negative-going trailing edge of the input waveform[11, 12]. Since for efficiency, the cathode resistor is normally made high compared with $1/g_m$ the output pulse will have a long tail.

The patient current pulse from a stimulator employing such an output circuit would have a large amplitude short duration overshoot on the leading edge with a relatively long duration low amplitude, negative-going undershoot corresponding to the trailing edge of the applied

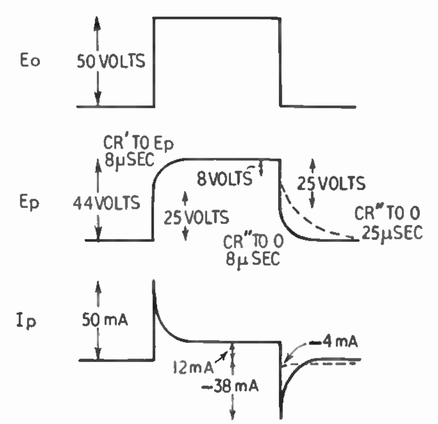


Fig. 5. Stimulator output voltage (E_o) patient voltage (E_p) and patient current (I_p) waveforms for constant voltage stimulator when $R_o = R_o' = 500\Omega$. (Waveform for $R_o = 500\Omega$, $R_o' = 10\text{k}\Omega$ shown dotted).

voltage pulse. Stevens[8] has discussed the effects of varying stimulator output impedance on the I.D. curve, and gives typical waveforms for different electrode sizes.

The effects of unduly high R_o and R_o' are most noticeable at the shorter pulse widths. Some typical results of changes in R_o and R_o' when using the small electrodes recommended by Wynn-Parry[2] are shown in the I.D. curves of Fig. 6. The diagnostic significance of the intensity duration curve should not be materially affected by the stimulator output impedance

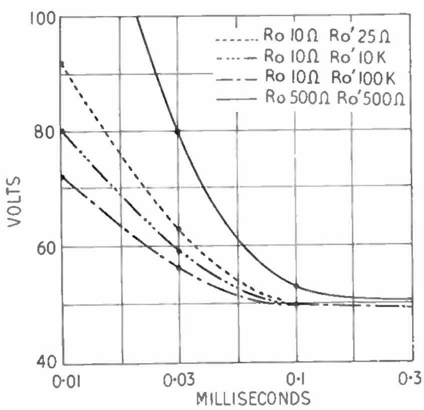


Fig. 6. The effect of variation in R_o and R_o' on I.D. curve.

changes shown in Fig. 6, although these would assume considerable importance if, say, the response to short duration stimulation was being studied by itself.

It is considered that a distinction should be made between the "pulse" and "quiescent" values of stimulator output impedance and the relevant values should be quoted when specifying the parameters of a diagnostic stimulator. In general it may be said that, provided R_o is less than 500 ohms and R_o' less than 10,000 ohms, reliable results will be obtained with the technique described in reference 2.

3. Performance Requirements

The following electrical specification is reprinted by permission of the Editor of *The British Medical Journal* and is taken from the report of a sub-committee of the Medical Research Council's Nerve Injuries Committee. The specification is that of a stimulator which was found by the committee to be satisfactory in performance.

(a) *Pulse Duration.* 10 pulse widths shall be available as follows: 300, 100, 30, 10, 3, 1, 0·3, 0·1, 0·03, 0·01 milliseconds. Pulse width accuracy shall be ±5 per cent. or 1 microsecond, whichever is the greater.

(b) *Pulse repetition frequencies* shall be 0·5, 5, and 50 pulses per second.

(c) *Pulse amplitude* shall be 200 volts maximum for all pulses. An amplitude control in conjunction with a calibrating system shall be employed to set the output to the maximum value of 200 volts.

(d) *Output Control.* The delivered output shall not differ from the indicated output by more than 10 per cent. for all pulses and repetition frequencies, including switching transients. An output dial, directly calibrated in volts, shall be employed in order to provide a greater effective scale length than the size of meter commonly employed.

(e) *Wave Form.* This shall be essentially flat-topped for all pulses. Amplitude droop shall not exceed 10 per cent. on the 300 milliseconds and 5 per cent. on the remaining pulses. Rise and decay times shall be negligible compared with pulse width. In the case of the 0·01 millisecond pulse, permissible rise and decay times are 0·001 millisecond. Leading and trailing edges shall have no more than a single overshoot which shall not exceed 10 per cent. of amplitude.

(f) *Output impedance* shall not exceed 500 ohms at any point on the scale.

(g) *Audible Monitoring.* There shall be an optional audible monitoring signal synchronized with the pulses.

(h) *Safety Precautions.* One of the patient terminals shall be connected to earth and the case. Dangerous potentials shall be excluded from the patient's circuit in the event of valve or component failure.

Construction Requirements

Minimum constructional requirements shall be those set out in Inter-Service Specification DEF 5000.

A stimulator designed to meet the M.R.C. specification has been described by Brennand[10].

4. Other Requirements

In addition to the performance requirements specified by the M.R.C. sub-committee, the following are considered necessary.

4.1. *Output impedance*

If the output impedance changes appreciably during the pulse cycle, the values of R_o and R_o' should be stated.

4.2. *Stability*

Stability of performance with respect to time and valve and mains voltage variations is essential, since the apparatus must function for long periods with a minimum of attention. The effect of mains supply voltage changes on the pulse width, pulse amplitude and p.r.f. accuracy should be stated.

4.3. *Ease of operation*

Wynn-Parry[2] allows approximately four minutes for obtaining a strength duration curve on abnormal muscle and less than one minute for a normal muscle S.D. curve. Thus the stimulator should be designed for rapid and easy operation by non-technical personnel.

For example, the pulse amplitude control (see 3.3) should cover the range 0·200 volts without range switching. Range switching merely increases the difficulty of plotting the I.D. curve and confers no real advantage.

4.4. *Ease of maintenance*

It is important that ease of maintenance is not sacrificed in attempts to produce apparatus of small physical size.

4.5. *Low initial cost*

If the stimulator is to be acceptable the initial cost of the instrument should be as low as possible.

5. Design Considerations

Obviously there are many possible arrangements of circuit elements which will meet the specification, and the designer must seek a balance between performance and economy. Typically a circuit would consist of (i) p.r.f. generator, (ii) pulse width generator, (iii) output circuit. It is not proposed to discuss (i) and (ii) here, since many references to this type of circuitry are to be found in the literature[15, 16, 17]. Landee *et al.*[18] give equations and worked examples covering the design of a wide range of multivibrators, phantastrons, etc.

5.1. *Output Circuit*

The stimulator output circuit must be able to handle the large variation of load current which occurs during the pulse. A suitable design would provide as a minimum, the current demanded by the equivalent circuit of Fig. 2 for the output voltage specified in Sect. 3(c). Suitable circuits include cathode followers, over-driven anode loaded amplifiers and pulse amplifiers employing overall negative feedback[23]. The output circuit should be arranged so that both output terminals are at earth potential in the interval between pulses. There is little doubt that adequate performance for routine clinical use can be obtained from a single valve cathode follower. However, when rise and decay times of one microsecond or less are required together with low values of R_o and R_o' the simple cathode follower (with or without additional amplification) is inadequate.

Several writers have described circuits for reducing the decay time of the simple cathode follower. These methods are based on the use of a second valve for discharging the load circuit capacitance. In the arrangement described by West[19] and Goddard[20] the second valve is turned on when the anode voltage of the cathode follower rises on anode current cut-off. Deming[21] gives a circuit for producing 100-volt pulses with 0·1 microsecond rise time but gives no details of the load requirements. Schoen[22] has recently described transistor circuits for producing low voltage (6-8 volt) pulses at peak load currents of 3 to 5 amperes for computer applications. These output voltages are, of course, too low for percutaneous stimulation but the circuitry may be adaptable for higher output voltage. For the work described in this report the present writer employed a cathode follower —amplifier arrangement in which R_o was reduced by overall negative feedback. A low minimum value of R_o' was obtained by means of a clamping circuit driven by the pulse generator and both R_o and R_o' were independently variable over a wide range.

6. References

1. A. E. Ritchie, "Thermionic valve stimulators," *Brit. J. Phys. Med.*, **11**, p. 101, 1948.
2. C. B. Wynn-Parry, "Electrical methods in diagnosis and prognosis of peripheral nerve injuries and poliomyelitis," *Brain*, **76**, p. 229, 1953.
3. W. Grey Walter, and A. E. Ritchie, "Electronic stimulators," *Electronic Engng*, **17**, p. 585, 1945.
4. C. B. Sneath, and E. G. Mayer, "Electro-medical stimulators," *Wireless World*, **55**, p. 129, 1949.
5. V. H. Attree, "An electronic stimulator for biological research," *J. Sci. Instrum.*, **27**, p. 43, 1950.
6. W. J. Perkins, "Electronic stimulators for physiological use," *Electronic Engng*, **27**, p. 434, 1955.
7. C. B. Wynn-Parry, in "Electrodiagnosis, Electromyography" ed. S. Licht. (Newhaven, Conn,, 1956.)
8. W. G. S. Stevens, "Output impedance of constant voltage diagnostic stimulators," *Brit. J. Phys. Med.*, **19**, p. 217, 1956.
9. M. R. C. Sub-Committee Report, "Electrodiagnostic stimulators," *Brit. Med. J.* **II**, p. 714, 1958.
10. R. Brennand, "Electronic Stimulator Type 5." Technical Repair & Development Section, No. 248 Maintenance Unit, R.A.F., Chessington, Report No. 24, 1949.
11. B. Y. Mills, "Transient response of cathode followers in video circuits," *Proc. Inst. Radio Engrs*, **37**, p. 631, 1949.
12. A. J. Shimmins, "Cathode follower performance," *Wireless Engineer*, **27**, p. 289, 1950.
13. H. Bjorn, "Electrical excitation of teeth and its application to dentistry," *Swedish Dental J.*, 1946.
14. H. D. Darcus, "Functional anatomy of the locomotor system." (In "Physical Medicine and Rehabilitation" ed. B. Kiernander, Blackwell Scientific Publications, Oxford, 1953.)
15. For example, in "Principles of Radar," Chap. 2.53 (McGraw-Hill, New York, 1946).
16. F. C. Williams, "Ranging circuits, linear time base generators and associated circuits," *J. Instn Elect. Engrs*, **93**, Part III A, No. 7, 1946.
17. For example, in "Waveforms" ed. B. Chance (McGraw-Hill, New York, 1949).
18. R. W. Landee and others. In "Electronic Designer's Handbook." (McGraw-Hill, New York, 1957.)
19. J. C. West "Textbook of Servo-mechanisms," p. 153. (English Universities Press, London, 1953.)
20. M. J. Goddard, "Direct coupled cathode-ray oscilloscopes," *J. Brit. I.R.E.*, **15**, p. 194, 1955.
21. C. R. Deming, "Amplifier for fast rise—fall pulses," *Electronics*, **28**, p. 170, September 1955.
22. S. Schoen, "Transistors provide computer clock signals," *Electronics*, **32**, p. 70, 27th February 1959.
23. J. G. Thomason, "Multi-valve cathode follower circuits," *Wireless World*, **63**, pp. 310-313 and pp. 373-377, 1957.

INTERNATIONAL CONFERENCE ON MEDICAL ELECTRONICS

The Second International Conference on Medical Electronics will be held at the new Unesco Building in Paris from June 24th to 27th.

The first conference was held in Paris last year with the intention of bringing together medical workers and engineers and to explore the possibilities of such co-operation. The Institution was officially represented by Mr. W. J. Perkins of the National Institute for Medical Research, who was subsequently appointed a member of the Executive Committee of the International Conference.

Following this Conference a Medical Electronics Group was formed within the Institution, and Mr. Perkins was elected its first chairman. He will again represent the Brit.I.R.E. at the forthcoming Conference.

Four rooms in the Unesco Building have been allocated for the 1959 conference. Simultaneous translation in French and English will be available in the two larger rooms which have a seating capacity for 300 and 150 persons. The two smaller rooms, each with a seating capacity of 40 persons, will be available for round table discussions and meetings. In addition there will be scientific demonstrations and a display of manufacturers' equipment in the Exhibition Hall.

Papers are being allotted fifteen minutes for presentation, with five minutes for discussion. Any members wishing to submit a paper may send their contribution to the Institution. Similarly, manufacturers wishing to demonstrate equipment can obtain further information from the Institution.

The membership fee for the Conference is 9 dollars which, in addition to participation at the meetings, includes advance abstracts of papers and other conference publications. W. J. P.

News from the Sections . . .

SOUTH MIDLANDS SECTION

A paper dealing with future developments in electronics was read in February by Mr. G. W. A. Dummer, M.B.E. (Associate Member) of the Royal Radar Establishment, on the subject of "Micro-Miniaturization."

Mr. Dummer first gave a brief review of the reduction in size of components from "normal" to "miniature" and then to the present "subminiature." He described the influence of the low voltage operation of transistors on component design. The limit of packaging of subminiature components on printed wiring boards had now been reached and yet more miniaturization was still required for military purposes, examples being army infantry radio and airborne radar equipment.

Experimental work now being carried on thin film components such as resistive films, dielectric films and magnetic films produced by evaporation techniques was described. The problems of mounting of these microminiature components on ceramic bases was also dealt with.

The superposition of these component films to form solid circuits was discussed, and a new micro-modular system of construction proposed in the U.S.A. was described. Mr. Dummer said that the future might lie in still further developments in solid state electronics such as molecular spin amplifiers and masers. R.D.

NORTH WESTERN SECTION

At a well attended meeting of the Section in March, Mr. R. E. Blythe read a paper on "Closed Circuit Television Equipment." After a brief review of the development and manufacturing effort in the field of television he explained the rapid growth and progress made in establishing it as an important tool in industrial and scientific instrumentation.

Mr. Blythe then discussed the fundamental principles of a system suitable for industrial usage. After stating the design requirements, he described the various camera pick-up tubes which had been evolved during the development of television. The function of the Vidicon camera pick-up tube and how it operated was described in some detail. He pointed out that in closed circuit television any one of the three television transmission standards could be adopted. The 625-line, now an international standard, was generally adopted to suit the export market but Mr. Blythe stressed that the equipment could be converted to serve any other system by minor changes in circuitry.

After describing the applications of closed circuit television under three main categories, Monitoring, Surveillance and Communication, a number of slides were shown embracing the equipment used and how, if necessary for safety, the camera could be remote-controlled. The illustrations also included applications in the steel and electrical industries, hospitals and nuclear power. Finally, Mr. Blythe gave a practical demonstration using a camera, camera control unit and a 14-in. monitor. F.J.G.P.

WEST MIDLANDS SECTION

At the February meeting members heard about some of the problems encountered in nuclear engineering when Dr. L. W. J. Newman read a paper on "Some Aspects of the Control of Nuclear Reactors." He first gave a brief elementary introduction to the basic concepts of atomic energy in order to establish the essential features of a typical reactor system. Possible methods of controlling the fission chain reaction were briefly examined, and the requirements for a control system were then stated.

Dr. Newman then discussed the ways in which the requirements had been met in a particular case, namely the control of the heavy water moderated and cooled reactor of the DIDO type. He dealt at some length with the special precautions which have to be taken in the design, choice of materials, and manufacture of components for the system.

THE 1959-1960 SESSION

The local Section Committees are now drawing up plans for next Session's programme of Meetings and Visits. Members who are prepared to read papers are invited to get in touch with their local Honorary Secretary.

An Automatic Standing-Wave Indicator for the 3-cm Waveband†

by

ELIZABETH LAVERICK, B.SC., PH.D.‡ and J. WELSH, B.SC., A.R.C.S.‡

Summary: The equipment is based on the rotary type of standing-wave indicator, the mechanical rotation of the detector being replaced by a ferrite polarization-rotating section and a fixed detector. The v.s.w.r. information is extracted from the output signal and fed to a meter calibrated directly in v.s.w.r. and/or reflection coefficient. Advantages over other similar equipment are that there are no rapidly moving parts liable to wear, it does not rely on similarity of crystal characteristics, and the response at any one frequency is effectively instantaneous. The indicator can be used as a second grade microwave bench, for testing microwave systems and as production testing equipment for microwave components.

List of Symbols

- H_z Longitudinal magnetic field component of H_{01}-mode in rectangular waveguide.
- E_y Vertical electric field component of H_{01}-mode in rectangular guide.
- H_x Transverse magnetic field component of H_{01}-mode in rectangular waveguide.
- λ'_g Design guide-wavelength in rectangular waveguide.
- λ_g Guide wavelength in rectangular waveguide.
- λ_c Cut-off wavelength for H_{01}-mode in rectangular waveguide.
- λ'_c Cut-off wavelength for H_{11}-mode in circular waveguide.
- ε Ellipticity or ratio of minor to major axes of polarization pattern.
- D Diameter of circular waveguide.
- a Broad dimension of rectangular waveguide.

1. Introduction

There are several equipments described in the literature for the automatic measurement of voltage standing wave ratio of microwave components and systems[1-13] some of which are highly accurate and, of necessity therefore, large and complex. The simpler instruments are not on the whole entirely satisfactory as regards accuracy and/or serviceability, and it is felt that there is still a requirement for an instrument which will measure quickly and accurately the v.s.w.r. of any component or system over a range of frequencies. Such an instrument should be reliable, serviceable, easy to set up and operate, and not too costly.

The equipment to be described was developed with these requirements in mind. It has certain advantages over other equipments in that it has no rapidly moving parts which would be prone to wear, does not rely on similarity of crystal characteristics, does not involve the use of a waveguide switch, and the response at any one frequency is effectively instantaneous. It is based on the rotary type of standing wave indicator, the mechanical rotation of the detector being replaced by a ferrite polarization-rotating section and a fixed detector. The v.s.w.r. information is extracted from the output signal and fed to a meter calibrated directly in v.s.w.r. and/or reflection coefficient.

2. Principle of Operation

2.1. *The Rotary Standing-Wave Indicator* (Fig. 1)

The rotary standing-wave indicator[14], around which this equipment is built, is a four-arm junction consisting of a rectangular waveguide coupled by means of a suitable coupling element in its broad face to a section of circular waveguide perpendicular to it. The coupling element is designed to couple equally from the

† Manuscript received 16th January, 1959. (Paper No. 498.)

‡ Elliott Brothers (London) Limited, Radar Research Laboratory, Elstree Way, Borehamwood, Herts.

U.D.C. No. 621.372.812.029.64:621.317.42

longitudinal and transverse components of the magnetic field, H_z and H_x, associated with the wave travelling in the main guide. Since H_z and H_x are in space and phase quadrature, the two

Fig. 1. The rotary standing-wave indicator.

H_{11} modes excited in the circular guide combine to form a circularly-polarized wave whose direction of rotation depends on the direction of propagation of the energizing wave. Hence, if the main waveguide is terminated by a load giving rise to a reflected wave, a circularly-polarized wave of the opposite hand of rotation is coupled into the circular guide. The amplitude and phase of this wave is related to the complex reflection coefficient of the load. The two circularly-polarized waves in the circular guide combine to give, in general, an elliptically-polarized wave whose polarization pattern[15] can be measured, using, for example, a rotatable plane-polarized detector. The ratio of the major and minor axes gives a direct measure of the v.s.w.r. of the load, and the angular position of these axes is directly related to the phase of the standing wave.

In its simplest form, the coupling element is a small circular hole placed at distance x_0 from the waveguide wall. With a matched termination

$$\left|\frac{H_z}{H_x}\right| = \frac{\lambda'_g}{\lambda_c} \cot \frac{\pi x_0}{a} = 1$$

at the design guide wavelength, λ'_g, i.e. for a matched termination to couple a circularly polarized wave, x_0 must be chosen such that

$$x_0 = \frac{a}{\pi} \tan^{-1} \frac{\lambda'_g}{\lambda_c} \qquad \text{.........(1)}$$

At any other guide wavelength λ_g

$$\left|\frac{H_z}{H_x}\right| = \frac{\lambda_g}{\lambda_c} \cot \frac{\pi x_0}{a}$$

The coupled wave is elliptically polarized and the ellipticity

$$\varepsilon = \left|\frac{H_z}{H_x}\right| \propto \frac{\lambda_g}{\lambda'_g}$$

That is to say, with a single-hole coupling element the apparent v.s.w.r. of the instrument varies directly with λ_g and the instrument is narrow band. For example, an instrument designed at $\lambda_0 = 3{\cdot}2$ cm would have a v.s.w.r. $\geqslant 0{\cdot}98$ over a 2 per cent. frequency band only.

A multiple hole coupling element has been designed which covers a somewhat broader band[16]. This has of necessity a low coupling. (about 30 db). In order to cover as wide a band as possible with an accuracy of 1 per cent. v.s.w.r. it was decided in this case to use a single hole coupling for which the coupling can be about 20 db, and equalize the coupling to the transverse and longitudinal magnetic fields at each frequency by adjusting the position of the circular guide with respect to the coupling hole.

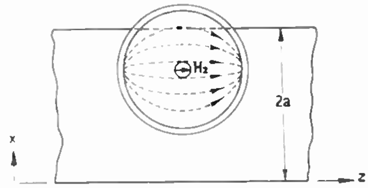


Fig. 2. Coupling to the H_z-component

Consider the coupling from one component only, say the H_z component (Figure 2). If the circular guide is moved in the z-direction, then it can be shown from the field equations for the H_{11}-mode in circular guide that the coupling will vary according to the law

$$J_1'\left(\frac{2\pi r}{\lambda'_c}\right)$$

where r = distance between the centre of the circular guide and the coupling hole. That is to say, the coupling is a maximum at $r = 0$, when the circular guide is concentric with the coupling hole, and zero at $r = D/2$ where D is the circular guide diameter. Moving the circular guide in the x-direction, the coupling to H_z varies according to the law

$$\frac{1}{r}\, J_1\left(\frac{2\pi r}{\lambda'_c}\right)$$

The variation with r is seen to be much less in this case. Obviously for the H_x component the situation is reversed—the coupling from H_x varies from maximum to zero as the circular guide is moved in the x-direction. It follows therefore that by moving the position of the circular guide in the z- and x-direction the ratio of the coupling from H_x and H_z can be varied over the entire range, 0 to ∞ or, more specifically can be maintained equal over a band of frequencies. Figure 3 shows the locus of the

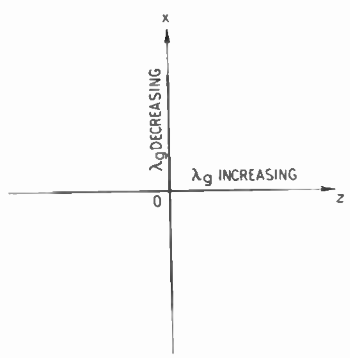
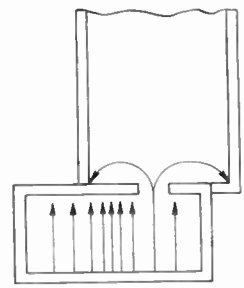


Fig. 3. Locus of circular guide adjusted for circular polarization at each wavelength.

circular guide adjusted for circular polarization at each wavelength. The origin, 0, represents the design wavelength, λ'_g, given by eqn. (1) at which the circular guide and the coupling hole are concentric. If λ_g increases, the longitudinal magnetic component becomes greater than the transverse, and to compensate for this the circular guide is moved along the z-axis. For $\lambda_g < \lambda'_g$ the circular guide is moved along the x-axis. Since the variation in coupling is symmetrical about the centre, there should be two possible positions of the guide giving circular polarization at each wavelength for a matched load termination. If the matched load is then replaced by the component under investigation the resulting polarization pattern will give the v.s.w.r. and phase of the component.

In practice, the coupling to the electric vector, E_y, may also play a part, giving rise to small errors when the circular guide is offset (Fig. 4).

Fig. 4. Coupling to the electric vector E_y.



2.2. *The V.S.W.R. Indicator System*

In order to measure the v.s.w.r. the elliptically-polarized wave in the rotary standing-wave indicator is passed into a ferrite polarization rotator, which rotates the axes of the ellipse with respect to a plane-polarized crystal detector. In fact the axes are caused to oscillate through ± 90 deg about their mean position at a frequency of 100 c/s. Thus an amplitude-modulated signal is produced at the detector, whose maximum and minimum correspond to the major and minor axes of the ellipse. Since the microwave source itself is modulated at 15 kc/s the output from the crystal detector is a 15 kc/s signal amplitude modulated at 100 c/s. This is amplified by means of a tuned automatic-gain-control amplifier, the automatic gain-control bias being derived from the largest peak of the modulated 15 kc/s signal. In this way the peak is held constant for all values of ellipticity (or v.s.w.r.). The signal is then detected and filtered, yielding the 100 c/s modulation, whose peak to peak amplitude is a direct measure of the v.s.w.r. to be measured. This is displayed on a valve-voltmeter calibrated in v.s.w.r. The output is also fed to a "pass-limits" circuit, in which two electronic relays, operating a red and green light respectively are preset to trigger at certain values of v.s.w.r.

3. Description of the Equipment

The equipment is contained in a two-tier rack (Fig. 5). A block diagram of the equipment is shown in Fig. 6. The bottom tier contains the

Fig. 5. The automatic standing-wave indicator.

klystron and calibrated cavity, the ferrite isolator, the waveguide and rotary standing-wave indicator, the stabilized power supply for the amplifier and the pass-limits circuit. The upper tier contains the amplifier, the polarization rotator and associated equipment, the crystal detector, the klystron power supply, modulator and triangular waveform generator.

3.1. *Microwave System*

The microwave source is a square-wave modulated reflex klystron, type R.5222, used in an external cavity resonator which can be tuned over a frequency range. This is followed by a transverse field ferrite isolator having a backward loss of 20 db over the waveband, which reduces multiple reflections in the system and which prevents "pulling" of the klystron frequency. The signal then passes through the rotary standing-wave indicator to the load under investigation. The output from the rotary standing-wave indicator is rotated by the ferrite polarization rotator before passing into a plane-polarized detector.

3.1.1. The rotary standing-wave indicator

This instrument was designed to operate over a frequency band 8,550 Mc/s to 10,000 Mc/s. The rectangular guide used was specially selected accurately-drawn brass waveguide. The coupling hole, diameter 0·405 in., was placed 0·205 in. from the side wall corresponding to

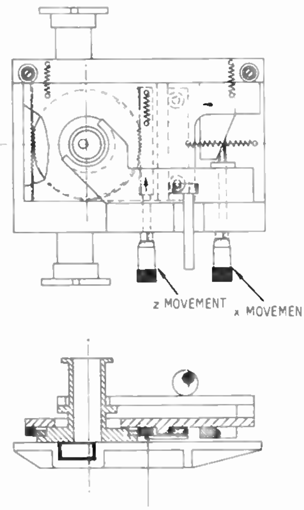


Fig. 7. Cross-section of the rotary standing-wave indicator.

$|H_x| = |H_z|$ at 10,000 Mc/s. The coupling factor was about –20 db. The circular guide diameter 0·84 in. was chosen to eliminate higher order modes, in particular, the E_{01}-mode which tends to be excited. The circular guide is adjusted with respect to the coupling hole by means of two micrometers. The instrument is shown in cross-section in Fig. 7.

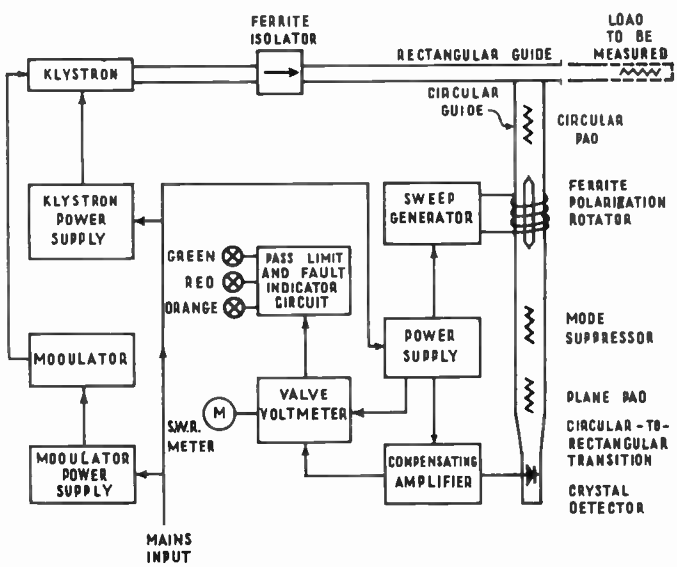


Fig. 6. Block diagram of the automatic standing-wave indicator.

Tests were carried out on the instrument using a precision matched-load termination and the setting-up positions over the frequency band were determined. Typical contours giving the locus of the circular guide for various apparent v.s.w.r. are shown in Fig. 8. The centre of each

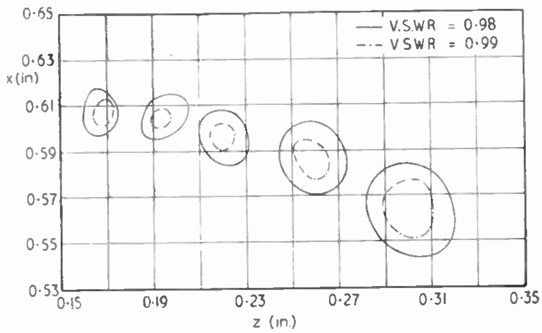


Fig. 8. Locus of circular guide for various apparent v.s.w.r.'s over the wave-band.

set of contours gives the setting-up position for v.s.w.r. $\geqslant$ 0·99. To measure the polarization pattern from which the v.s.w.r. was obtained a circular-guide rotating joint was used in conjunction with a plane polarized detector. The latter consists of a taper from circular to rectangular waveguide containing a mode suppressor to suppress one polarization followed by a low-level crystal detector by means of which the orthogonally polarized component is measured.

3.1.2. The polarization rotator

This device is based on the Faraday rotation effect in ferrites. A rod of Ferramic R1 is supported by a foamed p.v.c. holder in the circular guide, and a longitudinal magnetic field is

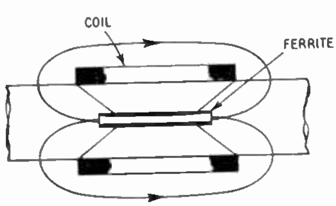


Fig. 9. The polarization rotator.

applied to it by passing a current through a coil wound round the waveguide (Fig. 9). As the current and hence the magnetic field is increased, the plane of polarization of the incident microwave signal is rotated through an angle θ by the ferrite, until the ferrite reaches saturation, ($H = H_{sat}$). Figure 10 shows a typical rotation curve. If the magnetic field is changed through a cycle from $+H_{sat}$ through zero to $-H_{sat}$ and back again, the rotation curve follows a hysteresis loop (Fig. 11).

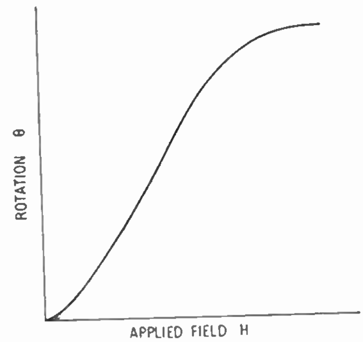


Fig. 10. Variation of rotation with applied magnetic field.

In the polarization rotator the ferrite specimen is a Ferramic R1 pencil of diameter 0·21 in. and overall length about 5 in. A 100-c/s triangular current is applied to the coil, the resulting field (about 15 Oersted) being such that

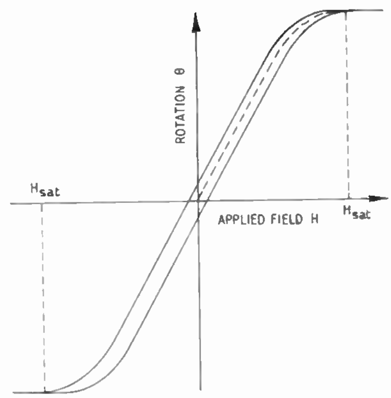


Fig. 11. Hysteresis loop (rotation v. applied field).

a rotation of at least ± 90 deg is achieved over the band. Only the linear part of the rotation characteristic is used. The ferrite is matched at each end by a "quarter-wave" section of ferrite of small diameter. Since the permeability of the ferrite varies with the applied field, it is not possible to achieve a matched condition which will apply throughout the cycle of operation of the ferrite. In fact the phase and amplitude of the wave reflected by the ferrite varies considerably with the instantaneous magnetic field and with the microwave frequency. The amount of rotation also varies with the microwave frequency, and with this particular ferrite, the rotation increases from 90 deg to 220 deg over the frequency band, with a given field. To overcome this it is so arranged that the current through the coil is automatically adjusted to give the correct field when the instrument is set up at the required frequency. When a magnetic field is applied to the ferrite polarization rotator, unwanted ellipticity is produced arising from two sources—multiple reflections between the ends of the ferrite and preferential absorption of one hand of polarization by the ferrite. This ellipticity has been measured at zero rotation and at 90 deg rotation and values of 47 db and 43 db were obtained respectively.

3.2. *The Associated Electronic Circuits*

3.2.1. Compensating amplifier

The output from the microwave crystal detector, consisting of the 15 kc/s square wave, amplitude modulated at about 100 c/s is passed into the "compensating amplifier." This consists of a low-noise triode-connected valve, with fixed bias followed by four variable-mu RC-coupled pentodes, the last of which has a tuned circuit in its anode circuit which feeds the detector and the peak comparator. The detector "demodulates" the 15-kc/s signal, yielding the 100 c/s component. The peak comparator accepts only the peaks of the modulated signal which exceed a stabilized reference voltage, and these peaks, each consisting of a short burst of 15 kc/s waves, are amplified and rectified and after passing through a long time constant circuit (about 1 second) are used to derive the automatic gain control bias.

The degree of gain control achieved is such that a change in amplitude of the input signal by a factor of 10 causes a change in output signal of about 1 per cent. The long-term stability of output for a constant input signal is about 1 part in 1,000. The performance is more than adequate for this particular equipment, but is probably rather below the maximum attainable from this type of circuit.

3.2.2. Valve voltmeter

The 100 c/s wave whose peak-to-peak amplitude is directly related to the v.s.w.r. to be measured, is rectified by a voltage doubling circuit, in order to measure the true peak-to-peak voltage as opposed to the mean-to-peak. The rectified signal is fed to the input of a stable valve voltmeter. This is of the cathode follower type, employing a double triode, with a 6 in. dia. 1 mA f.s.d. meter calibrated directly in v.s.w.r. connected between the cathodes of the triodes. The linearity of meter deflection with input voltage is better than 1 per cent. of the indicated deflection, and the dependence of the meter reading on the input wave shape was found to be negligible when equal amplitude sine- and square-waves were compared.

3.2.3. Pass limits and fault indicator circuit

The steady potential on one side of the meter is suitably "backed-off" and amplified in a single stage d.c. amplifier, the output of which feeds a pair of single-stage electronic relays, based on the Schmitt trigger circuit. These can be preset to trigger at any indicated v.s.w.r. by means of controls on the front panel. The anode currents of the trigger circuits operate small Post Office type relays which light red and green lights. In this way the v.s.w.r. scale is divided into three categories—for example, good, acceptable and fail—depending on whether the green, green and red, or red lights are energized.

The automatic gain control line of the compensating amplifier feeds another similar d.c. amplifier-relay circuit which energizes an orange light when the a.g.c. bias is such that the compensating amplifier is operating within the range of good compensation. This light acts as a fault indicator, and fails to light if for any reason the signal level at the second detector is insufficient for reliable measurement.

3.2.4. The sweep generator

This is a straightforward 3-valve circuit consisting of a 100 c/s triangular-wave generator and power amplifier supplying a maximum of about 5 watts of power to the solenoid of the ferrite polarization rotator. Provision is made for altering the amplitude of the triangular wave fed to the gyrator in a non-linear manner by means of a linear potentiometer ganged to the klystron reflector control. By this means the Faraday rotation was limited to the required value at all frequencies. (Section 3.1.).

There are three main reasons for requiring the rotation to be approximately constant and only just in excess of 90 deg:

(a) the preferential absorption of one hand of circular polarization by the ferrite increases with the angle of rotation, resulting in the introduction of unwanted ellipticity into the microwave signal.
(b) an increased rotation causes higher frequency side-bands to the 15 kc/s wave which will suffer distortion in the selective compensating amplifier. (The amplifier is made as selective as possible consistent with no attenuation of the sidebands, in order to maintain a good signal to noise ratio).
(c) The mismatch of the ferrite rod becomes progressively worse with increasing magnetic field.

3.2.5. Power supplies

Klystron Power Supply: This supplies—

(i) stabilized −350V to the klystron cathode, and has an output impedance of less than 100 ohm.
(ii) stabilized −500V (approximately) to the klystron reflector.
(iii) stabilized −200V to the d.c. amplifiers in the trigger circuits where it is used to derive various "backing-off" potentials.
(iv) stabilized −200V derived from (i) above to feed a 15 kc/s square wave generator which "chops" the −350 volt cathode potential and thereby amplitude modulates the klystron.

Amplifier Power Supply: This is a high stability unit with an output impedance of less than 1 ohm supplying +350V to the compensating amplifier, trigger circuits and relays, valve voltmeter, and sweep generator.

4. Some Applications of the Equipment

The equipment can be used as a second-grade microwave bench, for testing microwave systems, and as a production testing equipment for microwave components.

In the development of microwave components it is often more efficient to use a second-grade measuring bench for initial measurement and setting-up purposes, this being followed at a later stage by more precise measurement and analysis using higher-grade equipment. It is important that the early stages of the work should be covered rapidly and thoroughly and yet should not if possible involve the development engineer himself full-time. By the use of the automatic standing-wave indicator these objectives are readily achievable. The v.s.w.r. of the component under test is observed over the waveband, in a few minutes. The component can then be matched using, say, a sliding post, at one frequency, the v.sw.r. being under continuous observation during the matching process. The results over the waveband are then quickly checked. If a v.s.w.r. greater than 0·98 is required the higher grade measuring bench can then be brought into use.

If it is required that a complete system or part of a system be tested this equipment has an added advantage. With many systems it is necessary to connect the test gear by means of a flexible waveguide link or by a series of waveguide bends, etc. These have an appreciable v.s.w.r. and will normally contribute to the apparent v.s.w.r. of the system under test. With this type of rotary standing-wave indicator however it is possible to correct for these unwanted effects without seriously impairing the accuracy of measurement. A matched load termination is connected in place of the system under test and the rotary standing-wave indicator is adjusted to balance out the reflections from the flexible link, etc. so that the apparent v.s.w.r. is again unity. If the matched load is then replaced by the system under test the meter then registers the v.s.w.r. of the system alone. In the same way, measurements can be made at any point within a system, provided a matched load can be connected at that point.

With careful operation too it is possible to test a system or component continuously with frequency and pick out any sharp resonances, etc., which may not emerge if normal methods of measurement were used. This, of course, would be further facilitated by the use of the "broad-band" rotary standing-wave indicator provided the bandwidth requirements are compatible.

The third application of this instrument is for the production testing of components. If the number of components is small, the v.s.w.r. can be quickly read from the meter for each component in turn, if necessary, at several frequencies. For large batches of components the "pass-limits" system can be used, in conjunction with a quick-release flange to connect the components under test, and the components can be quickly graded into categories, as previously mentioned. (Sect. 3.2.3.).

5. Errors in the System

There are four main sources of error in the equipment. These arise from the rotary standing-wave indicator, the polarization rotator, the circuitry and the meter calibration.

5.1. *Errors from the Rotary Standing-wave Indicator*

Theory indicates that the accuracy of the rotary standing-wave indicator decreases with v.s.w.r. (v.s.w.r. measured < 1). Tests were carried out to assess the order of magnitude of the error in the rotary standing-wave indicator for various terminations. For this purpose wedges of graphite-loaded araldite were made up to slide in a piece of precision waveguide, the front surface of the wedge being shaped to set up the required reflection. The wedges were tapered and of such a length that apart from the initial reflection all the energy was absorbed by the load. The reflection coefficient was measured for various positions of the load within the precision waveguide. Typical results at one end of the waveband for two loads with v.s.w.r. of 0·46 and 0·145 respectively are shown in Table 1. These errors arise largely as the result of undesired coupling from E_y and from interaction effects between the load under investigation and the coupling element itself. For comparison the results obtained using a first-grade slotted-line standing-wave indicator are also indicated.

Table 1

($\lambda_0 = 3{\cdot}5$ cm)

Rotary S.W.I.		Slotted Line S.W.I.	
Mean v.s.w.r.	Variation in v.s.w.r.	Mean v.s.w.r.	Variation in v.s.w.r.
0·466	±0·0065	0·458	±0·0075
0·145	±0·0125	0·146	±0·0035

5.2. *Errors arising from the Polarization Rotator*

Multiple reflections in the circular arm of the rotary standing wave indicator can normally give rise to errors only as a result of cross-coupling between the two planes of polarization, arising for example, from ellipticity of the circular waveguide. These errors can easily be made negligible within the rotary standing-wave indicator itself. The introduction of the ferrite polarization rotator, however, complicates matters considerably. The reflection and transmission coefficients of the ferrite may well be different for different incident planes of polarization and will also vary in phase and amplitude as the current through the coil changes. These effects will result in errors in the measured v.s.w.r. Reflected waves passing through the ferrite more than once will give rise to components in the output rotated through 3θ, 5θ, etc., which will also contribute an error signal. To reduce these errors to an acceptable value it is necessary to reduce multiple reflections in this part of the system to a minimum.

The coupling element as seen from the polarization rotator appears very much like a short-circuit. The conventional crystal detector has a reflection coefficient of, say, $|\rho| = 0{\cdot}4$. The reflection coefficient of the ferrite was found to vary from 0·005 to 0·035 and its phase by as much as 90 deg as the field is swept. These effects are reduced by means of suitable well-matched attenuating pads in the system in front of the crystal detector and between the coupling hole and the ferrite.

The attenuating pad between the coupling element and ferrite is rather special in that it is in circular guide and must present the same characteristics to all incident planes of polarization. In order to obtain the necessary symmetry a pad was developed consisting of crossed resistive vanes. The vanes were tapered to minimize reflections and trimmed to give equal attenuation. Their phase lengths were then equalized by means of small polystyrene stubs in the appropriate plane. The attenuation achieved was 18 db. The ellipticity introduced by the pad was at least 46 db down, for all planes of polarization, and the v.s.w.r. greater than 0·97. It can be shown that the ellipticity arising from the ferrite (see section 3.1) introduces an error of less than 0·005 in v.s.w.r.

Under these conditions it was found that the errors introduced in this part of the instrument were in the worst case $\pm 0{\cdot}01$ in v.s.w.r.

5.3. *Errors arising from the Circuitry*

5.3.1. Crystal law

It can be shown that owing to the non-ohmic law associated with the semiconducting materials used in most crystal rectifiers, the current flowing in an external low impedance circuit, or the potential across an external high impedance circuit is proportional to the power of the low intensity microwave input, i.e. the rectified d.c. output is proportional to the square of the amplitude of the microwave input.

Departures from the square law arise from several causes, including the fall in reverse resistance with increasing inverse voltage, and the change in ratio of the low or high external circuit resistance to the crystal forward or reverse resistances with change in input.

Where a silicon crystal feeds an external high impedance circuit it is usually true that for low signal inputs (below 100 microwatts) the crystal law stays constant although it may differ from the ideal square law by a few per cent. For input signals of between 100 microwatts and 10 milliwatts the index of the law usually lies between 1·9 and 2·1 and above 10 mW the power of the law progressively falls, eventually becoming approximately linear. Although the effect of d.c. bias on the crystal detector was found to improve the sensitivity, in some cases by considerably more than the theoretical 6 db, the effect on the crystal law is negligible over the range of small input powers[17].

In the equipment described in which it is necessary to know the crystal law with some accuracy, the only suitable method found was to select crystals whose law was square to within about 2 per cent. over the required range of input powers. About 80 per cent of new silicon crystals fell within this category. The effect of a small departure from the ideal square law on the measured v.s.w.r. is very small in the range $1 > S > 0{\cdot}5$.

5.3.2. Non-linearity in the amplifier

Although the amplifier is of the automatic gain control type, whose output is substantially independent of input the automatic gain control action has a long time-constant and does not interfere with the approximate linear amplification of the much shorter period modulation envelope. The non-linearity which occurs is due mainly to the inherently curved characteristics of the variable-mu valves. By keeping the signal levels on the last variable-mu valves down to a suitable level, the error in the output signal level due to this non-linearity was kept to below 2 per cent. of that level. This source of error has no effect on the ends and maximum effect towards the centre of the v.s.w.r. scale, where it causes a maximum decrease in indicated v.s.w.r. of 0·01 at a v.s.w.r. of 0·7.

5.3.3. Gain control limitations in the amplifier

Considerable precautions are taken as described in Section 3.2 to ensure that the gain of the amplifier is automatically adjusted so that the largest peak of the modulation envelope is maintained at constant amplitude. However, due to the changing interval between maxima and their changing shape and amplitude, with microwave frequency and with load phase and amplitude, some change in the amplitude of the maximum occurs, but this is less than 1·5 per cent. for variations of all the above parameters encountered in practice. This error affects the meter reading by a similar magnitude at f.s.d. but causes only very small errors for $1 > S > {\cdot}7$.

5.3.4. Errors arising from the meter calibration

The meter is calibrated directly in v.s.w.r. and because of the square-law of the crystal detector at low levels the meter deflection

$$\theta \propto 1 - S^2$$

where S = v.s.w.r. measured ($1 > S > 0$). It is clear therefore that the v.s.w.r. scale becomes somewhat cramped for small values of S, and the accuracy with which the meter can be read decreases with S. In general, the accuracy required also decreases with S, and this source of error is not important.

5.5. *Overall Accuracy*

The overall accuracy of the instrument was tested over the waveband using the graphite-loaded araldite wedges described in Section 5.1. For v.s.w.r. $\geqslant 0{\cdot}85$ the error was $\pm 0{\cdot}015$. For $0{\cdot}85 \leqslant$ v.s.w.r. $\leqslant 0{\cdot}50$ the error was $\pm 0{\cdot}02$. For very low values of v.s.w.r. the accuracy is limited by the accuracy with which the meter can be read.

6. Conclusions

The equipment as described above has proved its usefulness as a production test equipment, as a second-grade measuring bench and for testing microwave systems and parts of systems. For certain applications its usefulness would be enhanced by continuously sweeping the klystron so that v.s.w.r. can be observed at every point in the frequency band. For this purpose it would be necessary to incorporate a broadband coupling element in the Rotary Standing-Wave Indicator as mentioned in Section 2.1. Such elements have been developed and would yield the same order of accuracy as the present instrument over a slightly reduced waveband (about 9 per cent.).

The rotary standing wave indicator itself yields both amplitude and phase information, but in the instrument described above the phase information is not used. Although it would be possible to extract the phase information and use it to give, say, a polar plot of reflection coefficient over the waveband, the complexity of the instrument would be increased accordingly.

7. References

1. H. Ritzl and H. Groll, "Ring-shaped waveguide measuring section with automatic display," *Funktechnische Zeitung*, **5**, pp. 281-286, May 1955.
2. N. L. Pappas, "Measuring microwave impedance with the reflectometer," *Tele-tech and Electronic Industries*, **15**, May 1956, pp. 72-3 and 128-9.
3. H. L. Bachman, "Automatic plotter for waveguide impedance," *Electronics*, **29**, March 1956, pp. 184-187.
4. A. F. Harvey, "Instruments for use in the Microwave band," *Proc. Instn Elect. Engrs*, **98**, Part II, pp. 781-9, December 1951.
5. J. A. C. Kinnear, "An automatic swept-frequency impedance meter," *British Communications & Electronics*, **5**, pp. 359-361, May 1958.
6. E. A. N. Whitehead, "A microwave swept-frequency impedance meter," *Elliott Journal*, September 1951, p. 57.
7. J. A. C. Kinnear, Unpublished work by Elliott Brothers (London), Ltd.
8. "The Microwave Megamatch." (Kay Electronic Co., Pine Brook, New Jersey.)
9. Jorgen P. Vinding, "The Z scope, an automatic impedance plotter," *Convention Record of the I.R.E.*, 1956, Part 5 (Microwaves and Instrumentation), p. 178.
10. S. B. Cohn, "Impedance measurement by means of a broadband circular-polarization coupler," *Proc. Inst. Radio Engrs*, **42**, pp. 1554-58, October 1954.
11. E. M. Purcell, "Measurement of Standing Waves," p. 473. "Principles of Microwave Circuits," Ed. C. G. Montgomery, R. H. Dicke and E. M. Purcell, Radiation Lab. Series, Vol. 11. (McGraw-Hill, New York, 1958.)
12. W. J. Duffin, "Three-probe method of impedance measurement," *Wireless Engineer*, **29**, pp. 317-320, December 1952.
13. R. S. Cole and W. N. Honeyman, "Two automatic impedance plotters, *Electronic Engineering*, **30**, pp. 442-446, July 1958.
14. Brit. Pat. Spec. 592,224. 11th September 1947.
15. J. D. Kraus, "Antennas," Chap. 15, pp. 465-470. (McGraw-Hill Book Co., New York, 1955.)
16. S. B. Cohn, loc. cit., p. 1555.
17. J. Welsh, "Effect of a d.c. bias on crystal sensitivity." Unpublished work by Elliott Brothers (London), Ltd.

Radio Engineering Overseas . . .

The following abstracts are taken from Europe and Commonwealth journals received in the Library of the Institution. Members who wish to borrow any of these journals should apply to the Librarian, stating full bibliographical details, i.e. title, author, journal and date, of the paper required. All papers are in the language of the country of origin of the Journal unless otherwise stated. The Institution regrets that translations cannot be supplied.

NETWORK ANALYSIS

The electrical network laboratory of the University of Ljubljana has recently built an analogue computer for solving problems of passive linear networks. The results are displayed on a measuring instrument or on a cathode-ray tube. On its screen the attenuation or phase angle curves or the absolute value curve of the transfer function for a given circuit can be directly observed. The resistance potential network represents by analogy the complex frequency plane on which one can see the position of roots (poles or zeros) of the transfer function which determines the read out attenuation or phase angle. Thus it is feasible to obtain for a required attenuation or phase angle the corresponding transfer function (i.e. its poles and zeros) which truly represents that attenuation or phase angle.

"Application of resistance potential networks in passive linear network engineering." J. Virant. *Elektrotehniski Vestnik*, **26**, No. 9-10, pp. 281-284, 1958.

MICROWAVE POWER MEASUREMENT

The measurement of high peak powers in the microwave range with conventional liquid calorimeters is difficult because of the possibility of electric breakdown. A broadband calorimeter has been developed at the Zurich Technical University in which the dissipative structure is not inserted in the main guide but coupled to the main guide by a grid of wires in the side wall. The power is measured by a substitution method and the total error is less than 2 per cent.

"A broadband microwave calorimeter for high powers." T. Jaeger and M. V. Schneider. *Archiv der Elektrischen Ubertragung*, **13**, pp. 21-25, January 1959.

CIRCUIT TECHNIQUES

A note from the Institute of Radio Physics and Electronics, University of Calcutta, describes a discriminator, employing transmission lines, designed to have linear response over a wide range of frequency deviation as is required in f.m. wideband work. It is shown that linearity of response can be obtained without difficulty up to a fractional frequency deviation as high as 50 per cent.

"A new wideband discriminator." N. B. Chakraborti. *Indian Journal of Physics*, **32**, pp. 537-546, December 1958.

DIRECTIONAL COUPLERS

Under certain simplifying assumptions, a perturbation calculation can yield statements concerning long-slot directional couplers of a general cross-sectional shape. A perfect exchange of energy between two waveguides coupled by a narrow longitudinal slot can thus occur only if the cross-sectional dimensions meet the following condition: If the slot is visualized closed, the mode to be coupled must have a somewhat lower phase velocity whose magnetic longitudinal field strength H_z is greater at the slot position for a given transmitted power. The slot width required for attaining a predetermined transfer coefficient increases with decreasing guide wavelength, decreasing H_z (at the place of the slot in each case), and decreasing coupling length. The measuring results available for the case of H_{10} (rectangular guide) to H_{01} (circular guide) coupling agree well with the values found by analysis. Finally the well known analogy to intercoupled pendulums is pointed out, making a generalization for unequal masses.

"A long slot directional coupler for H-waves." H. Pascher. *Archiv der Elektrischen Ubertragung*, **13**, pp. 76-82, February 1959.

OSCILLATOR CRYSTALS

The requirements for crystals used in filter networks are considered by a U.S. Army research worker in a recent German paper. An essential requirement for filter crystals is that the excited resonance frequency be sufficiently isolated from other responses. For low-frequency quartz plates, this requirement is normally fulfilled except in the case of a few dimensional ratios. Thickness modes, in particular the thickness-shear type, usually have a variety of modes in the vicinity of the main response rather than a single mode. For high frequency quartz filter crystals of the AT-type in the range 7 to 30 Mc/s a triangular plate shape has been found to form boundaries resulting in a satisfactory reduction of unwanted modes.

"Quartz filters for the frequency range 7-30 Mc/s." R. Bechmann. *Archiv der Elektrischen Ubertragung*, **13**, pp. 90-93, February 1959.

TRANSISTOR TECHNOLOGY

A new type of transistor is described in a recent French paper. Its field of use is intermediate between that of a drawn transistor (low frequency) and that of a base diffused transistor (high frequency). It is produced by two successive diffusions, one with indium the other with arsenic. The physical principles and the calculations and the curves for determining the parameters are described. It is claimed that its advantages include technical simplicity in production

"Double diffusion germanium npn transistors." R. Deschamps. *L'onde Electrique*, pp. 74-87, February 1959.

COLOUR TELEVISION

Color transmission by the N.T.S.C. system, combining the luminance signal and the carrier-chrominance signal to a common signal, introduces the possibility of "crosstalk interference" between luminance information and chrominance information. Double modulation of the chrominance sub-carrier can further give rise to "crosstalk" within the chrominance information itself, hence between saturation signal and hue signal.

The paper discusses such crosstalk interference as makes its appearance already in the case of a linear transmission channel, and illustrates by reference to experimental results the crosstalk effects appearing in an equivalent video-frequency transfer quadripole with strong linear and non-linear distortion and with a combination of both types of distortion.

An analysis is subsequently given in particular of those circuits, as are critical in studio amplifiers for transfer and distribution of the N.T.S.C. signal, and it is investigated how far circuits used at present will come up to the increased demands placed with transmission of the N.T.S.C. signal.

"Transmission failure in the N.T.S.C. system." H. Schonfelder. *Archiv der Elektrischen Ubertragung*, **12**, pp. 497-509, November 1958.

TELEVISION CAMERA TUBES

The supericonoscope with potential equalization, "Rieselikonoskop" (Sprinkling Iconoscope), is employed as a camera tube in most German and some other European studios (Austria, France, Luxemburg, Netherlands, and Sweden). A paper from the Fernseh Company deals with the distortion the picture signal suffers from the electronic mechanism in the tube. The interference signals which appear are investigated quantitatively for the supericonoscope IS 9/35. The measurements lead to certain setting rules for the sprinkling current and the intensity of the scanning beam and to precise rules on admissible tolerances.

"Signal distortion in the supericonoscope." W. Dillenburger. *Archiv der Elektrischen Ubertragung*, **13**, pp. 63-75, February 1959.

ELECTRONIC COMPUTERS

A large scale digital computer using parametrons as logical elements was completed in March 1957 in the Electrical Communication Laboratory of the Nippon Telegraph and Telephone Public Corporation. A series of papers has been published describing the general features of the system, logical design and engineering details. The computer is of a fixed point system in a parallel operation, with forty binary digits for a number word and with a pair of single address orders for an order word. Major components are about 5400 parametrons which constitute logical elements of the arithmetic and control units, and 519 vacuum tubes mainly for the excitation power source, the storage selection circuit and the neon indicators.

The parametron circuit is composed of a pair of the ferrite toroidal cores on which a wire of nine turns is wound, connected to a capacitor of 5000 pF and a coupling resistor of 300 ohms.

Stores consisting of magnetic cores with 256 word capacity and magnetic tapes are employed. A photo-reader with the maximum speed of 200 characters per second is used for the input; the output uses a paper punch with a speed of 12 characters per second and a high speed printer of a type belt construction with the maximum speed of 1500 characters per second and a cathode ray tube display. Six-hole paper tape with its sixth hole as an odd parity check, is used.

"The parametron digital computer MUSASINO-1." I. "System and logical design." S. Muroga and K. Takashima. II. "Engineering description." K. Takashima, S. Muroga and K. Nishida. III. "Magnetic core memory." S. Yamada, T. Bessho and T. Koshiba. *The Journal of the Institute of Electrical Communication Engineers of Japan*, **41**, No. 11, pp. 1132-1155, November 1958.

MAGNETIC CORE STORAGE

A recent Dutch paper describes a memory system which contains 1024 words of 44 bits and employs cores of ferroxcube 6D3, a material having a rectangular hysteresis loop. After a general discussion of the structure of the memory some features are dealt with in detail. These include a circuit to prevent blocking of the output amplifiers (this shortens the cycling time to 3 microsec, about $\frac{1}{3}$ of the normal value); an output amplifier in which advantage is taken of the fact that a transistor can act for some microseconds as a temporary storage element; and a method of producing current-pulses of stable shape and height, based on the saturation of a ferrite core in a choke.

"A fast method of reading magnetic-core memories." H. J. Heijn and N. C. de Troye. *Philips Technical Review*, **20**, (No. 7), pp. 193-207, January 1959.