O.C.R.—benefits and pitfalls

By R. M. Paine*

This paper follows up an earlier paper, which was concerned with the preparations required for O.C.R., by discussing the problems of the actual installation of O.C.R. and the results achieved in production. Certain of the pre-installation plans had to be changed to achieve good throughput and accuracy. Various pointers are given to the desirable features of any O.C.R. equipment. It is emphasised that a working system is far more than the electronics to read characters and marks.

(Received May 1968)

It has long been commonplace to state that one of the main problems with computers is obtaining the input data to be fed into the computer system. In an earlier paper (Paine, 1966) I quoted Sir Edward Playfair's statement that 'In business applications the difficulty lies in the so far unsolved problem of the interface between the human brain and hand, and the machine. No one can regard rows of girls reading figures and punching holes as an elegant solution of the problem in the computer age; but so long as computers have to deal with data produced by ordinary people, the problem must be faced. Progress is being made but all too slowly.'

Three years have passed, and Eastern Electricity have taken delivery of a Farrington Optical Scanner and Document Reader Model 3010 (see **Fig. 1**), yet I feel inclined to repeat the last sentence 'Progress is being made, but all too slowly'—to relate not mainly to Eastern Electricity's experience, but to the O.C.R. world generally. At the moment the initials O.C.R. seem to stand for either 'Overall Caution Required' or 'Only Courage Repays'.

Background

The theory is, of course, that a scanner will save the use of many operators, and will get input to the maw of the computer quicker in elapsed time, and more accurately, and that although the capital cost of a scanner is large, it will bring economies because of the replaced punches and verifiers and the salaries of the punch operators. In Eastern Electricity we believe this theory to be true, but the benefits of O.C.R. cannot be obtained without careful planning involving the costly time of programmers, operations managers, clerical supervisors and maintenance engineers, and without a great deal of experimentation on paper quality, document handling, printer ribbon control, methods of batch control and staff training.

Ordering the Farrington scanner

The model 3010 Farrington Document Reader with output on to magnetic tape cost £54,625, before devaluation, and that is quite a lot of mini-skirted punch girls to replace.

We ordered the scanner in August 1966 but it did not reach Ipswich until 31 August, 1967. All this time our billing operation was continuing to grow as we transferred more and more consumers to the centralised computer system. At the moment we have 2 million consumer accounts on the computer, and it was mainly in the consumer account application we wanted to use O.C.R.

Turnround documents

Our O.C.R. application was for turnround documents, and there were two types:

(*a*) A bill would be printed on the computer and sent to the consumer. When he paid, the stub would be detached and sent back to the computer centre. The stub containing consumer reference number, type of payment, and amount, would be read by the scanner onto magnetic tape, and the information used to update the consumer's main file to show that he had paid (see **Fig. 2**).

(*b*) A meter reading slip, including the consumer reference number, would be printed by the computer printer and sent to the meter reader. He would mark on it in ordinary lead pencil the reading of the consumer's meter, and a code describing the type of reading, and return the slip to the computer centre. The slip would be scanned, and the information about the consumer's reference number, his meter indicator number, the reading, and the reading code would be transferred to magnetic tape. This tape would be

Fig. 1. The Farrington optical scanner

* *Eastern Electricity, Russell Road, Ipswich*

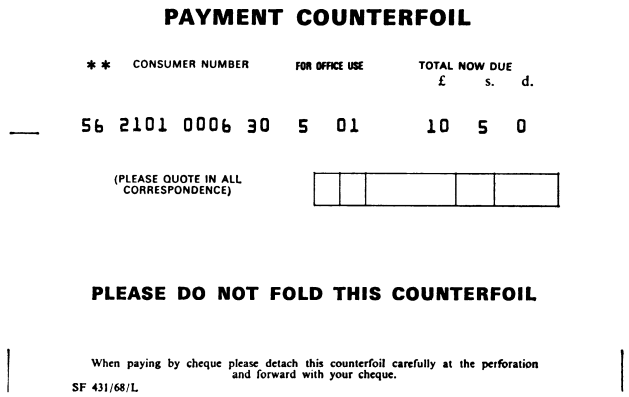

Fig. 2. The payment stub

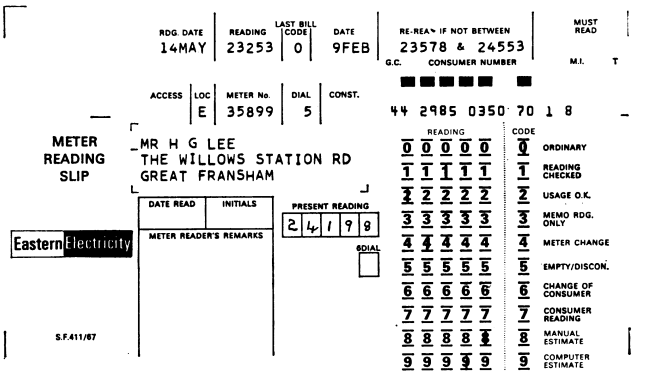

Fig. 3. The meter reading slip

processed against the consumer's main file, and in conjunction with other information on the file about the consumer, used to produce his bill (see **Fig. 3**).

In theory this system was fairly automatic with few difficulties. We are anticipating that when all our 2·2 million consumers are transferred, we will deal daily with 36,000 cash stubs from the bills, and 36,000 meter slips. This is not as heavy as Giro's future load at Bootle, but still a very large load.

Linking the scanner to the computer

Our Honeywell 1800 and 200 installation uses ¾ in. tape, and the Farrington scanner produces ½ in. tape and we had to consider how to link the two systems. We could perhaps have tied the scanner on-line to the 200, but the manufacturers at that time did not seem interested in the development required. We were against putting a slow speed device on-line to the 1800, even if it had been possible at the time, since we thought lack of core store, and the tying up of one of our 1800 magnetic tape units, operating at 133 KC a second, to a slow scanner would only cause us operating problems and loss of valuable 1800 time. So we eventually decided to instal a ½ in. 20 KC tape deck and associated equipment on our 200 computer, at a round cost of £15,000, which brought our scanner investment to about £70,000. This, of course, brought an additional advantage that we could now read industry-compatible ½ in. tapes on our 200, which might be useful in dealing with the Banks and Giro, and in exchanging information among the nationalised industries. It also means that we can read ½ in. tape from magnetic tape encoders such as the Honeywell KEYTAPE or the MOHAWK recorder.

The conversion process is to read documents on the scanner onto its ½ in. tape unit (actually made by Ampex), transfer the ½ in. tape to the Honeywell ½ in. tape unit on the 200, convert the information, by means of a 200 program, to a ¾ in. tape unit on the 200, then take the ¾ in. tape and process it on the 1800. This was only a slightly longer procedure than the previous theory of on-line working to the 200.

O.C.R. font on documents

We decided to use the Farrington self-check font on the scan line on our two documents, but elsewhere on these documents, and on our other printing, we chose to use normal Honeywell font. This was in case there was any difficulty in consumers or staff reading the Farrington self-check font. (The S. E. Gas Board decided, however, to use self-check font on all positions on their scanning documents.) I personally do not think there is any difficulty in reading it. Our decision meant we needed a special print barrel for the Honeywell 222/4 printers, since we needed normal font for numerals and alpha and a few special symbols, plus self-check font for numerals—but we managed to fit them into the 64-character set by sacrificing some special symbols such as 'question mark' and 'not equal to'. The two special barrels added about £5,000 to our costs, bringing the scanner investment up to £75,000. We had estimated in our planning that we could save about 40 to 45 punch operators and that the scanner would read our 72,000 documents in a shift of 8 hours. Taking the 25% investment grant into consideration, we would more than break even, and in fact save our punch card section from bursting the seams of the punch room. We might still require 115 to 120 girls to cope with our other work, and scanner rejects, and the punch room could just about accommodate these, but it could not accommodate 155 to 160 girls. This was one of the most important facts driving us on to use a scanner.

Delivery and commissioning

Eventually we had our scanner at the end of August 1967. But installation and engineering commissioning problems, such as electrical noise and earthing, delayed the start of acceptance tests to the middle of October. In these tests we soon found pitfalls, some due to the scanner, some due to document handling in our showrooms and some due to the quality of print on the computer printers. Actual production work did not really commence until January 1968.

Determining throughput speed

One of the first shocks we had was the actual speed obtained of documents read per minute. The stated reading speed of the documents by Farrington was 312 a minute. Relying on our past knowledge that punch card handling reduces rated speeds significantly, we took about two-thirds of this speed and used a rate of 200 documents a minute in our cost calculations. But in October and November 1967 we were only obtaining 20 documents a minute for cash, and 45 a minute for

meter readings. The reasons behind this are, I think, helpful to anyone considering a justification for O.C.R. First of all the transport mechanism was a fairly simple affair and unfortunately very sensitive—the documents stand vertically on their horizontal bottom edge (feeding foot) in the input hopper, and are pressed forward by the weight of other documents behind them and by a spring loaded metal block (see **Fig. 4**). Documents are separated by a combination of an overhead air-jet and a vacuum belt. When documents are of varying heights in a batch, the air-jet is less effective in separating the documents and causes feeding difficulties which require operator intervention, and thus a stoppage, and hence a slowing down in throughput speed.

Only our cash stubs were of varying heights. This was because a lot of our bills were paid in the shops, and the cashier's receipting machines were sometimes guillotining the stubs too short, and sometimes too large—so they varied greatly in a batch even from the same shop, and even more so between shops. We had to reinforce our previous instructions to cashiers on the importance of cutting on the right position, and the receipting machines (of several different types and vintage) were marked as to where a particular position on the bill should be placed before guillotining, and a line was included on the next printing of the bills to indicate clearly the particular position on the bill. In the long run, we intend to standardise as far as possible our receipting machines and where necessary, because of the needs of decimalisation of the currency, to replace very old machines with new ones, but with 150 shops this is a large number of machines to replace.

Three pockets

The optical characters are read while the document is in the stationary position, and then the document is moved on its bottom edge past the mark sensing position and thence into one of three pockets representing:

(*a*) invalid characters
(*b*) invalid marks in the matrix area and
(*c*) valid documents.

In the event of a transport 'wreck' or mis-feed of some kind, there is a very useful display of lights which inform the operator of the intended destination, so that the document can be taken out of the hopper or incorrect pocket and placed in the correct output pocket. But this has to be done manually, and the machine stops while it is being done, and then takes quite a few cycles to get up to the correct vacuum pressure to read the next document—and 'wrecks' and 'mis-feeds' were happening far too often. So again the throughput rate was reduced. A tolerant, fast, document transport system is one of the first things to look for when choosing scanners, and is very important in determining the throughput speed.

Double feeds

At the beginning of the transport feed, there is a mechanical sensing device to check that the thickness of the document is correct in order to trap double feeds—which mean the first document has been scanned but the second document has not been scanned and has been merely moved through the machine with the first one. This caused us problems initially because when a double feed was signalled, the machine stopped and the operator might take time to decide which document had not been read. If it was not in fact a real double feed but merely the corner of the document folded over and sensed as two thicknesses, then the operator might be feeding the same document twice, with consequent trouble on the computer run and for our clerical section trying to sort out what had happened. Also in the beginning, the payment stubs which consumers sent with cheques direct to the Central Accounting Office were date stamped in the cashiers section by perforating the date right through the batch. This was a good audit precaution, but unfortunately it made all the stubs stick together and we could not get them through the scanner feed—we either had stoppages for double feed or for not feeding at all. We soon ceased the perforation method—but I hasten to add that we still took very good audit precautions.

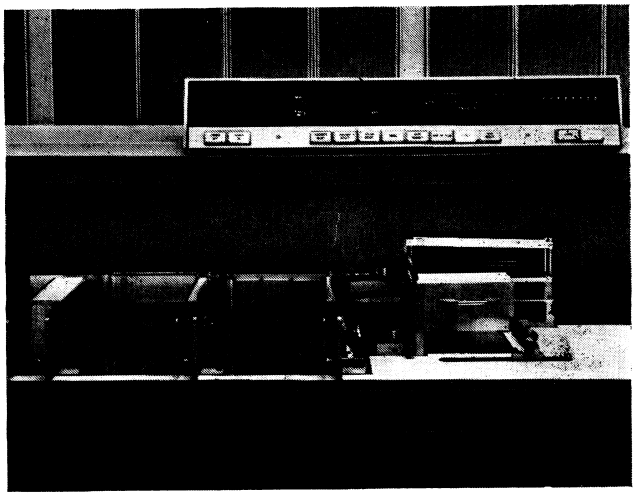

Fig. 4. Transport mechanism of the Farrington scanner

The double feed mechanism had a tendency to slip after a time and become too wide; we have had the situation where 10,000 meter slips have been read but 10 or 20 have passed through undetected and without bills being produced for them. It is far too costly to check clerically every day that a bill has been produced for each meter slip produced, and then to find the missing meter slip. On cash transactions we do have batch totals which tell us quickly if a cash item has not been read—though it is still very time consuming to find the offending document clerically and to remedy the error. But on meter slips we do not have batch totals, since this would have involved introducing a new operation which was not required previously when punching the readings. Close control was kept on the double feed mechanism to avoid such problems. If, nevertheless, a meter reading is not processed three days after the reading date, a bill is automatically estimated by means of the computer program and despatched. We have now replaced the mechanical double feed sensor with an optical sensor checking on the opacity of the documents, and this has considerably improved the position.

All rejects are sent for punching. On the whole we prefer punching and scanning the same document twice, and sorting this out by program, to missing a document—that is neither scanning it or punching it. For this

reason, double feeds on cash stubs are now diverted by the plug-board program to the middle stacker. This is normally used for matrix rejects, but the matrix is not used on cash documents. This gives the operator a chance to examine the documents, unencumbered by other documents already in the other two stackers, and re-feed if necessary. It would be advantageous if there was a fourth stacker to allow this to be done for meter slips. It is therefore useful in selecting equipment to consider the number of pockets an O.C.R. machine has, and how they could be used in your system. Operator experience has now enabled us to deal with a lot of these problems, but it is a tedious business operating a scanner and human errors can creep in, so it is better if the machine can make all the necessary decisions.

Condition of returned documents

One of the causes of double feeds and mis-feeding is the poor condition of the documents when they are returned to us by consumers; the problem is not so bad with meter slips handled by our own staff. Some cash stubs are creased, folded, torn and stained, and our cashiers or other gremlins do not help by adding staples, paper clips, pins and sticky tape. Meter slips are very troublesome only when the weather is extremely wet, when the slips curl or stick together and refuse to go through the machine at speed. We have introduced a new design of meter binders to try to overcome this dampness problem, and these have had a fair measure of success.

The quality of paper for O.C.R.

This poor condition of the documents (and it is by no means universal) has led us to consider the quality and weight of our paper. Before introducing the scanner we moved from 18 lb. to 21 lb. paper, which Farrington said would be very suitable for the machine. We have now moved, however, to 24 lb. paper for the following reasons.

(*a*) Greater resistance to bad handling of the cash stub by consumers.
(*b*) The extra stiffness will improve the transport feeding of the scanner.
(*c*) The extra weight and stiffness will assist the 'knocking-up' of the documents that is required prior to feeding.
(*d*) The extra opacity will eliminate any risks of reading through the document and picking up shadows from the next document.
(*e*) Receipting machines can cut more accurately on good quality paper and will not 'snatch' at the paper, causing a slanting cut.

This has certainly assisted the throughput and efficiency of the equipment without requiring a large increase in the cost of the stationery.

Current throughput rates

Our throughput rate has now reached 120 documents a minute for cash, and 150 for meter slips. We estimate that we need a rate of about 150 documents a minute to complete all the work in an 8-hour shift, though the scanner costs would probably break even operating at around 80 documents a minute for more than one shift. We do not have much more than one shift to read the full load of documents, since we plan to start our billing production runs on the computer at 3.30 p.m., and the scanner work commences at 6.30 a.m. With a speed of 150 documents a minute we will save about £15,000 a year over the cost of doing the work by means of punch operators.

If we did not have the scanner it would be impossible to punch all our cash receipts and meter slips between 9 a.m. and 3.30 p.m., and then assemble all the cards ready for feeding to the computer. Thus the scanner gives us the opportunity of a much quicker start to our billing operations and allows us to deliver the bill to our customers perhaps a day earlier—and this reduces the interest on our working capital, though this has not been taken into account in our estimates of annual savings.

Operator requirements

To maintain a high speed we have had to man the scanner with two operators—one to knock-up the documents, feed them into the hopper, operate the buttons and interpret the display on the lights, and one to remove the documents from the pockets and place them in their correct trays for the next action—i.e. rejects for punching, valids to be returned to clerical departments. It might even be preferable to use three operators—with the extra one carrying out meticulous knocking-up, making sure the documents are all facing the same way, removing any foreign objects, allowing air between documents, etc. Operating experience seems slow to be acquired, but nimble fingered girls are much better than clumsy men, provided they know the significance of the decisions they have to take when there is a stoppage.

Check digits and batch headers

The electronic side of the scanner has given us relatively little trouble. In designing a scanner system very much more has to be planned than reading a special font. In the past I feel too much attention has been given to the electronics and the 'gee-whizzery' and not enough attention as to how the scanner will really fit into clerical, consumer and computer procedures.

There are, however, one or two points on the electronics or optics side, apart from the need to make sure that very good maintenance and spare part facilities can be supplied by the manufacturer, and that the manufacturer has sound knowledge both of the use of O.C.R. systems and installing the machines. We had initially designed the cash stub with two check digits—one for the reference number and one for the amount and payment code. Even so, we have found a few cases, before the scanner was correctly tuned, of a reading error not being detected by the check digit. On meter slips we initially had a check digit only on the reference number, but have now added another check digit to control the meter indicator number and the last digit of the reference number, since occasionally in these two unchecked digits a '0' was being read as an '8', and vice-versa, and not being rejected.

We have also inserted check digits on our batch headers, after a certain amount of form re-design, to make sure false information is not carried through to the magnetic tape; there is nothing worse than an

incorrect batch header since it tends to make all the following data wrong as well. In scanning you must use check digits for all the data to be read—even a fairly simple method like the Luhns check (Farrington, 1966). It is also very important that if a batch header is rejected, the scanner stops; otherwise data will be put onto tape without the important header information. This may mean that two batches are incorrectly merged together which, in the case of cash stubs, can cause batch differences which take a large amount of clerical effort to track down and reconcile. One possible way of avoiding this is to feed in a batch at a time; but this slows down the feeding rate significantly. We thought our initial system design enabled us to stop the scanner if a header was rejected, but we found in practice that we had to bring in some additional features of form design to cope with the possibility a header document being fed off-centre, and, therefore, not being recognised as a header but rejected as unreadable without stopping the scanner.

It is worthwhile paying a great deal of attention to the design and function of batch headers, and also trying to make it as easy as possible for the clerical section to find batch differences. It must not be forgotten that the task of the clerical section is made more difficult because a batch is always split in two—the scanned documents, and the reject documents which have been punched. If possible one must avoid dividing it into further sub-batches, because this would make their task even more difficult. This was one of the main reasons why we did not try to re-read rejects on the scanner.

Matrix marking

The standard of matrix marks by the meter readers has been fairly good, and shows the benefits of thorough training and of bringing the staff concerned into the computer centre to see how their actions affect the operations of the scanner. There were some difficulties—mainly attributable to fitting our meter reading system to the scanner or vice-versa rather than to inherent defects of the scanner. Most of our meters have five dials, but some have four dials. The matrix has five columns for the recording of five dials, so we plugged the scanner program board to ignore a blank in the fifth column on the assumption that it would only occur in the case of a four-dial meter. This we found could be dangerous since it could also occur for a five-dial meter where the meter reader had forgotten to record the fifth dial, or where there is only a faint pencil mark in the fifth column which was not detected by the scanner. Thus false information was put on to the tape. We have now programmed the computer to print the meter slips for four-dial meters with a mark in the '0' position of the fifth column, and thus we are able to use the valuable hardware feature of a blank column, double mark detection system to reject all slips with a blank in the fifth column. Moral—it is dangerous to make assumptions!

Occasionally the scanner has been so good at reading marks that it has passed on, as valid data, blemishes in the paper or indentations or smudges made by the meter reader. Ghost data rejected by the computer is sometimes very difficult to check against a seemingly correct meter slip.

Timing marks

The black timing marks, pre-printed by the suppliers of the stationery, are very important, and have caused us some troubles. If the black ink is patchy, leaving voids (not very discernible to the naked eye) and perhaps dividing a timing mark into two, the scanner may act as if there were two timing marks. This means that the scanner repeats the mark in the column underneath twice, counts to its prescribed number of columns, switches off its sensing, and ignores the mark in the fifth column. This, as may be imagined, is very difficult to check, since the meter reading on the slip is correct and does not resemble the reading on the tape. You can have your timing marks any colour you like, as long as it is *black*–dense, consistent, black! At one time we had to drop scanning meter slips for about two weeks because of a bad batch of pre-printed forms. It pays to insist on very tight specifications to stationery suppliers and carry out tests on all deliveries. To make sure that the slips are being read correctly, and to know this early enough in the day to take action if there are difficulties, we print out the first few readings of each batch for clerical checking.

Print quality

This leads to consideration of the quality of printing by the computer. Here I would offer a word of warning: do not, if at all possible, have different manufacturers for the scanner and the printer. It is extremely difficult to tie them down, in the contract, as to what their respective specifications and obligations are, and to decide who is responsible if the scanner cannot read the printing one hundred per cent of the time. Of course, outside the contract, and in the computer room, both sides are very co-operative. We thought our standard of printing was good until we tried to persuade a scanner to read it with minimum rejects. The straightness of the line of print is not such a problem as we anticipated that it might be. The reader seems fairly tolerant provided that we are consistent in printing the correct distance from the foot of the form. More critical is the formation of a computer character without any loss of image at the top or the bottom—particularly the top. The Honeywell engineers were quite good in overcoming the problem of the missing parts of characters—this defect may seem obvious but again it was not very apparent to the naked eye. The density of the character is equally important and it pays to be critical with your standard of printing. But you cannot set up a printer correctly once and for all; it needs daily retuning. Also with 3 line-printers, we found that not all the printers gave the same quality of print or were as consistent in keeping to that quality. So we tend to use one particular printer for O.C.R. work.

O.C.R. ribbons

The correct carbon ribbons are very important in ensuring good quality print, and so is the cleaning of the printer several times a day. We experimented and found that for good scanner results we had to use the special O.C.R. ribbons—not the normal ribbons—although the O.C.R. ones are much dearer. We now purchase only O.C.R. ribbons, and put a new ribbon on the O.C.R. printer each day immediately before the bills

are printed. This ribbon is used initially for O.C.R. work and later that day for ordinary printing work. The next day a new O.C.R. ribbon is put on and the old ribbon transferred to a printer used for non-O.C.R. work for the remainder of its life—which means it is used for about 10 hours on O.C.R. work and 22 hours on non-O.C.R. work. Operators must watch the print quality very closely since ribbons in a production batch may vary, and it may be necessary to change the ribbon before 7 hours.

Control procedures

One problem when operating the scanner is that you do not know what data has actually gone onto the magnetic tape. If you do not find out that there are errors in the data on tape until you start the billing run in the evening, and all the punch operators have gone home, you are in a very difficult position. So we have instituted more control procedures and checks than we had originally planned; these are mainly concerned with printing out selections of the data that is on the tape, such as the batch numbers. There is a very heavy burden on the clerical staff if bad results are obtained, and the overall system must be designed with the needs of the clerical staff and punch operators in mind.

Rejects

We read about 155,000 cash stubs a week and about 160,000 meter slips, and these volumes are steadily increasing. At the moment, our reject rate on cash is about 5%, and on meter slips 10%. An analysis of the cash rejects shows that about half of the rejects are caused by part payments which involve the cashier crossing out the original amount, and by poor printing. A large load has been lifted off the punch room, and the clerical sections say that scanning produces about the same number or fewer cash batch differences than punching used to do—in other words it is probably slightly more accurate. We are not happy, however, with the reject percentages, and the occasional crises brought about by some malfunctioning of the scanner. Our contract stipulates that Farrington will guarantee a reject rate of 0·5% wholly attributable to the malfunctioning of the scanner—but it is difficult to prove what is wholly attributable to the scanner, since paper and print quality play such a large part. Farrington also guarantee that the error rate (wholly attributable to the scanner) shall not exceed one character in ten million characters; but it is, of course, very difficult to know what errors you have—especially undetected errors!

Conclusion

Eastern Electricity are definitely committed to Optical Character Recognition, and we are obtaining sound and valuable benefits from it. We are hoping to extend its application, and at the moment we are planning to use O.C.R. for reconciling cheques issued with the bank statement of cheques cashed, using the computer to avoid a tedious clerical job. This involves printing an O.C.R. line below the M.I.C.R. line of the cheque. In the future, we hope to make use of O.C.R. in the stores application and for the statistics of sales of appliances in our shops. We are also exploring other ways of improving our data preparation system, such as the possible use of keyboards recording directly on to magnetic tape, such as Honeywell Keytape or Mohawk equipment. There are a lot of improvements that could be made in O.C.R., such as faster and more tolerant transport systems, without going to the extreme of reading handwriting. It is a slow process perfecting one's installation, but I think it is rewarding in the end. You can do nothing, however, without skilled manpower and a lot of hard work, and the success of O.C.R. in Eastern Electricity could not have been brought about without them.

Acknowledgement

I would like to thank the Chief Accountant of Eastern Electricity (Mr. F. S. Grindrod) for permission to publish this paper.

References

PAINE, R. M. (1966). Preparation for optical character recognition, *Computer J.*, Vol. 9, p. 221.
Farrington Electronics Inc. (1966). *The optical character reader operations manual*, Section 5.