APPENDIX I

PATTERN RECOGNITION IN AN ELECTRONIC READER

BY

PHILIPP M. PAHL

AND

DAVID L. JOHNSON

P. M. Pahl

D. L. Johnson

It is no longer news to read of an electronic or electromechanical device which has been developed to duplicate a specific human function and to perform that function far more efficiently and accurately than human capabilities permit. Even where perception of environment is involved, such devices have shown marked superiority over humans in measuring quantities and distinguishing such qualities as color, light, pressure, and sound. One function of extreme interest and significance is the ability to recognize general shapes and forms. The electronic reader is a device to distinguish among a variety of forms: a machine able to identify and describe the various language units it perceives. These language units are simply a collection of different forms—the many squiggles that are the letters, words, and other symbols comprising all written language. The electronic reader and its associated problems are the basis of this discussion.

In order to make a comprehensive study of this problem, the nature of the reader input and output must first be examined, i.e., the written language itself and the description which the device should make of what it has perceived. It is also of value to review what is known of the original reader, the human system. Many clues to the internal operation of any reading device may be found in analysis of the system operation of the eye, nerve, and brain.

Application of an Electronic Reader

Where might automatic high-speed reading devices be profitably used? Operations requiring the transformation of large amounts of printed matter to some coded equivalent, or involving the repeated examination of printed material to determine some particular bits of information, could use such a reader to advantage. An efficient automatic reader would make it possible to automatically sort and route a great deal of our mail, resulting in faster service and lower costs. Automatic reading devices would be doubly effective in the handling of many business problems, especially those connected with keeping records and processing numerical information where the transformation of routine alphabetical and numerical material to some machine code is involved. The reader would reduce the costs of these operations by speed and efficiency increases; it would also free operating personnel from the boring routine aspects of the job to concentrate on difficulties the reader might not be able to solve. In the fields of library research, weather data examination, and the development of reading aids for the blind, automatic reading devices would be extremely important.

A research project which has been of considerable interest at the University of Washington is the study of mechanical translation of languages from Russian to English. Before the mechanical translation of languages can compete with human translation, either on an economic or time basis, some means must be devised for the automatic introduction of the text language into the translation system. The development of an electronic reader is vitally important to the success of the over-all translation system.

The Reader Input

In order to appreciate the problems any automatic reading device must solve, some of the details of written language must be examined. The various language units which may be used for analysis and description of a written language include letters, syllables, words, clauses, phrases, sentences, and paragraphs. In most written languages, the number of letters is relatively small. English and other languages using the Roman alphabet generally use 26 letters, Russian has 32, German, 31, and Arabic, 28. In each alphabet some letters are quite similar in form, while others are totally different. Letters of different alphabets may be identical in form, while the sounds they represent are quite dissimilar. It is also known that certain letters are used more frequently than others. Samuel Morse took advantage of letter probability when developing his original telegraphic code. The relatively small number of letters in many alphabets indicates that letters may be advantageously used as basic identification units by an electronic reader. When letters are combined with each other, the recognition becomes more complex. Considering the English alphabet, the possible number of two-letter combinations is 26^2, or 676; the possible three-letter combinations is 26^3, or 17,576. Fortunately, not all of these combinations are allowable, and others are used only infrequently. All languages seem to possess a preference for certain combinations which are predominately used. An analysis of English shows that about 70 syllables make up one-half of ordinary speech, but that 1,370 syllables are required to characterize 90 per cent of the language. If 93 per cent is assumed to represent essentially all of the English language, then the number of syllables is approximately 50 times the number of letters. The relationship between syllables may further reduce the recognition problem. Studies of telephone speech have indicated a preference for the consonant-vowel-consonent type of syllable, which occurred in one-third of the syllables spoken in the study. This type and the consonant-vowel and the vowel-consonant types made up three quarters of all the syllables examined. Assuming, then, that syllables could be determined and identified in written material, the use of certain syllables or letter groups would provide quick recognition of a large part of any written language.

Another language unit, the word, is more easily recognizable in print than are syllables. For almost the past 400 years it has been common practice to leave rather uniform spaces between printed and written words. The highly developed languages all contain large numbers of words, many of which are seldom used. Whereas the Oxford English Dictionary contains about 500,000 word entries, it has been estimated that good written fiction could be composed using only about 10,000 of these word types.

The problem of word recognition, rendered very complex by the number of words involved, does have simplifying aspects. First, there are certain preferred sequences of letters or syllables which make up the various words. Thus "nidificate" can be seen to be a very likely candidate for an English word, while "gelifkwaardig" has only a minute chance of belonging to the group. As is true with syllables, it has been observed that certain words of any language are used with more frequency than others. These words have a tendency to be quite short and are frequently monosyllables. Because of these simplifications, it may be useful to consider the use of words as basic identification units for an electronic reader, even though the number of words involved may be 1,000 times the number of letters.

The remaining language units, such as clauses, phrases, and sentences may all involve combinations of two or more words. Because of the large number of words to choose from, the totality of possible combinations is staggering. Combinations of words of any form would probably be completely impractical for automatic reader identification units.

In addition to the operating restrictions imposed by the choice of some specific language identification unit, a number of other considerations would cause difficulty for any automatic reader. All languages are written with a variety of punctuation marks which are normally quite significant to the exact meaning of the author's ideas. Thus it is certain that provision must be made for the recognition of punctuation marks. Various peculiarities of the printing process contribute to producing a non-uniform input. Examples are: the imperfect letter or the letter set just above or below the others on the same line, type faces of various sizes and styles, the practice of hyphenating words, the practice of squeezing or expanding the type to avoid hyphenating at the end of a line, and even misspelled words. Other imperfections such as smears and small spots in the paper might also cause confusion. Obviously the large variations in individual handwriting styles would be a formidable obstacle to overcome.

The Reader Output

In the previous section where possible applications of an automatic reader were considered, it was implied that the device would be of value as one of the parts in a larger system. That is, the reading device would sense the printed text, organize its responses into a unique description, then deliver this description to the next component in the system which would use what the reader had perceived. The principal requirement for the reader output is compatibility with the following component in the system.

It is necessary that the receiving component be supplied with information in a form which it can readily interpret and efficiently process. To meet this requirement, the reading device may have to express its output in a language unit different from the one the device itself uses for identification. This involves a transformation of what has been perceived to some different form; i.e., a translation process within itself. The reading device may recognize only the individual letters of the text, yet be required to produce an output in word form to yield optimum operation of the receiving component. In addition to the form of the reader output, the rate of output of the device must be compatible with the information-using component. The ultimate design of successful electronic reading devices may well depend upon knowledge of human visual processes. Human ability to see, classify, remember, and identify a great variety of shapes and forms is responsible for the widespread use of written materials as a form of communication. For this reason, certain of the aspects of human sight, in particular those related to the recognition of form, must be considered. Research in the field of human recognition has by no means been completed, but many conclusions have been reached which may be of value.

The phenomenon of recognizing standard patterns may be illustrated by the following example. A literate person is able to recognize a given letter of the alphabet under a great variety of conditions. The letter may be large or small, close to the person or relatively distant, situated against a number of different backgrounds, normally oriented or upside down, written by hand or formed by one of several styles of type, and may even be in a plane almost parallel to the reader's line of sight. Despite these diversities, recognition of the given letter is still possible. Recognition is achieved so long as the letter retains enough of its invariant qualities to make it distinguishable from the remainder of the alphabet. Furthermore, establishment of the letter's identity assumes a decision that the letter was not one of the others in its group.

One reason for human success in pattern recognition is that our varied visual inputs become highly standardized after being received. This is accomplished by the complex action of various eye-muscle feedbacks and other mechanisms which are less well understood. There are several ways in which the human system attempts to simplify recognition by standardization of visual inputs. First, objects which attract our attention are viewed so that their images fall on the fovea, the central portion of the retina where perception of form is enhanced. The images are also focused on the retina to make them as sharp as possible. Next, when the object is one which is most familiar to us in a particular orientation, a certain mechanism tends to produce this orientation, regardless of the object's actual position. Finally, a process of accommodation causes the eye to receive its most intense impressions at the boundaries, i.e., discontinuities of color or illumination. The result is that our visual images are reduced more or less to outlines. This last process may be compared with the "stopped-spot" principle of television scanning. In this system, the scanning spot passes rapidly over areas of no contrast but halts temporarily at boundaries, the positions of which are coded for transmission. The advantage of the system is that signal redundancy is greatly reduced.

The standardization of visual information may be thought of as one step of a data reduction process. It has been estimated that the information capacity of the human eye is roughly 4×10^6 bits per second; but that the perceptual intake rate of the mind is on the order of a few hundred bits per second. The mind can therefore use only a minute fraction of the information available at the eyes. It can be concluded that we retain only the most significant portions of our visual images for perceptual purposes and discard the remaining quantity of relatively useless information. After the input standardization, the structure and operation of the optic nerves and visual cortex apparently accomplish the remainder of the data reduction.

How do these visual inputs, reduced by the human system to standardized images, relate to the problem of pattern recognition? While not conclusively proved, it is fairly certain that recognition is a comparative process taking place in the mind. The comparison is between some image presently seen and the multitude of previously seen images stored in memory. Storage of images occurs as one of the three established human memory processes. One of these processes may be termed a picture memory which in many ways may resemble a series of snapshots of our environment, including our familiar patterns. Each snapshot may be made up of several hundred spots. The number of spots has not been conclusively measured and probably varies from picture to picture. Each spot has a binary bit or yes-no characteristic. It is estimated that the pictures may be stored at a maximum rate of about 10 per second and a person may well store as many as 10^{10} of these pictures in his lifetime. The stored snapshots are available for recall in the comparative process of recognition. Just how this comparison is made is not definitely known. Evidence obtained thus far indicates that the comparison is some sort of statistical process which involves picking the most likely picture, considered as a population of binary bits, with which to associate our perceived data. The minimum time required for recognition to take place is approximately 0.1 second.

Reading is a highly complex type of pattern recognition worth special consideration. In reading, it is known that the eyes are stationary about 90 per cent of the time, moving only with rapid jerks or saccades. Various experiments suggest that perception occurs almost entirely during the stationary periods. The eyes see a sharp image only at the fixation points. Sufficient clues as to general contours of other letters or words in the immediate vicinity of the fixation point are simultaneously received to enable recognition of several letters, a word, or more than one word. Recognition, then, takes place not only at the fixation point, but over a relatively large area. A significant point is that the eye receives information about what lies ahead of the fixation point as well as what follows it. This result, coupled with the reader's intuitive knowledge of statistical and syntactical properties of the language, seems to provide a feedback which determines where the next fixation point will occur. The distance covered by the saccades, and hence reading speed, is highly flexible, depending on the rate of perception of the text information.

Finally, there are recognition problems which the human cannot resolve on the basis of one particular object. The object must be viewed or associated with other related patterns before absolute identification is possible. If one were asked to identify the symbol "p" by itself, one could not determine whether it was a "p" or a "d" until its orientation with respect to other letters was known.

Reading Devices Already Developed

During the past forty-five years, numerous attempts have been made to construct devices which would effectively recognize printed material. The objectives of such devices may be classified generally into three categories: to furnish the blind with reading aids, to expedite the feeding of alphanumeric data to business machines, and to study theories concerning human visual processes. Study of these devices shows several common characteristics. The following statements apply generally to all of the previously developed reading devices, especially those of recent design:

1. All use individual characters, letters or numerals as identification units.
2. Characters are vertically scanned by a single light source or some "flying spot" equivalent.
3. There are numerous passes of the light source over each character, resulting in a minute point-by-point examination of each character.
4. Recognition is restricted to a small group of characters which each device has a predetermined ability to recognize.
5. Compared with the human visual system, the devices are only moderately effective in recognition of a variety of type size, style, and quality.
6. Few have any provision for automatically shifting from one line of print to the next.

Recognition Logic

Consideration of the operating characteristics just enumerated provided the direction for the experimental portion of this project. It was believed that reading devices of the type just described might realize much better performance, particularly increased speed, if either or both of the following techniques were used: horizontal scanning of an entire line of print, especially if scanning could be accomplished at a speed comparable to a single vertical scan of a single character; and parallel handling of the scan information, assuming that more than one scanning source be effectively generated and used.

It was decided to investigate the possibility of developing and mechanizing a recognition logic based on multiple horizontal scanning and parallel handling of the resultant information. Speed of operation was not an objective of the device to be constructed. It was also decided to examine the effectiveness of a coarse scan of each language unit to see whether accurate recognition was possible when a minimum of horizontal scans was used.

The first step in the development of a recognition logic was the choice of a language unit to be used for identification purposes; i.e., selection among such possibilities as letters, syllables, words, etc. In the first stages of investigation it was hoped that a relatively straightforward means might be devised to use words as reader-identification units. The problems in mechanization of such a device, however, were so involved as to require extremely extensive and expensive equipment. To permit a comparison between the efficiency of vertical and horizontal scans, the use of individual alphabetic letters as the basic identification unit was adopted—specifically the use of small printed Russian letters.

Following the choice of the 32 Russian characters as basic identification units, it was necessary to determine the properties of each letter which distinguished it from the others and permitted unique identification. These properties might be termed the invariant characteristics of each letter. The method used to determine the characteristics to establish a given category was placing the letters upon a fine grid. To approximate the information that might be available to a narrow horizontal scanning spot, a slotted card was placed over the letters, leaving only a horizontal display of one section of the print. Five equally spaced horizontal scans were assumed for the study. The top scan intersected the print of only two letters; the bottom intersected only six letters. Based on this observation, the first two categories of letter characteristics were chosen. Any letter which extended above its fellows was assigned to Category 1; those which extended below were assigned to Category 2.

The next result of the analysis was the observation that a scanning spot would intersect the print of each letter a certain number of times and that the number of intersections depended on the vertical position of the particular scan with respect to each letter. The number of scan intersections per letter is usually one, two, or three. These scan intersects with the print may be converted to electrical pulses which can be counted and used as identifying characteristics. On this basis, eight additional categories were established which are entirely dependent upon the intersection-counting process, i.e., the number of scan intersections along the top, center, and bottom of a given letter. It should be noted that the length of the intersection will not affect the counting. A scan through the center of an "H" and an "X" would each produce one pulse, and although the pulses are of different length, they are both counted as one. The scan intersect count is not required for the top and bottom scans since they always have a one count when referred to the Russian alphabet.

Subsequent letter analysis revealed the usefulness of what might be considered as letter extremities. For the purposes of this work, an extremity was considered to exist whenever one or two portions of a letter extended appreciably to the right or left of the main body of the letter. Further, existence of an extremity was based on the upper and/or lower portion of the letter extending beyond the center portion. No extremity was assumed to exist when the reverse was true. The extremity indication would consider only the three central letter scans. Because multiple horizontal scans must be used some of which would intersect at different times from others, it was decided to use this time difference in the reading device to determine letter extremities. The letters shown in Fig. 1 serve to illustrate the extremity definition. Picture the three central scanning spots moving simultaneously in a vertical row along the paths indicated. Scans 2 and 4 will give both right and left extremities on the first letter. The next letter has only a left extremity on Scan 2, while the following two letters have no extremities existing. With the establishment of extremities as identifying characteristics, four more categories were added, making a total of 14, and the analysis was essentially complete. Table I lists the fourteen categories.

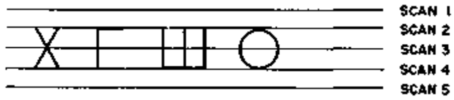

FIGURE 1

TABLE I

Category	Characteristic	Scan Number	No. Letters Included
1	upward extending	1	3
2	downward extending	5	6
3	one pulse	2	17
4	two or three pulses	2	14
5	one pulse	3	16
6	two pulses	3	10
7	three pulses	3	6
8	one pulse	4	20
9	two pulses	4	9
10	three pulses	4	2
11	one extremity left	2,3,4	4
12	two extremities left	2,3,4	4
13	one extremity right	2,3,4	3
14	two extremities right	2,3,4	5

Note that Category 4 includes both those letters with two and three intersection pulses on Scan 2. These letters can be uniquely determined without establishing a three-pulse category for Scan 2.

An examination of the categories when related to the individual letters shows that four letters all belong to the same categories and therefore cannot be uniquely determined. Several methods could be applied to solve this problem. One would be to sense the difference between large and small dark areas so that a short pulse would not be considered the same as a long one. Another would be an increase in the number of scans from a total of five to seven. For the purposes of this research, five scans were considered sufficient to demonstrate the developed recognition logic and no attempt was made to resolve the ambiguity of the four letters.

A Russian Character Reader

The next step following development of the recognition logic was the design and construction of an electronic device capable of using the logic to identify Russian letters. For design purposes, the proposed device was divided into three functional sections: scanning, pulse forming, and logical circuitry. It was considered that the logical circuitry section in which actual identification of each character would take place presented the greatest degree of difficulty. Accordingly, design of the reader was initiated in this area. Over-all simplicity was a design objective.

The recognition logic developed in the previous section requires accomplishment of three tasks for recognition to take place. In order to identify a given character the logical circuits must decide whether a portion of the letter has been seen by Scans 1 and/or 5, count the number of returns from Scans 2, 3, and 4, and also determine whether the letter has extremities and count and identify the extremity returns. To minimize information storage, the operations should occur nearly simultaneously so that each letter may be identified by the time the five scanning spots pass over it. This operation implies that additional logical circuitry would be required to sense when the scanning spots had completely passed each letter and to use this information to reset the totality of the circuitry. The use of relays in all logical circuitry was assumed, both because of ease of procurement and the correspondence of relay action with Boolean logic.

The counting function requires a capability to count to one, two, or three and produce a distinct output corresponding to the resultant count, then be reset to the zero-count condition regardless of where the counter has stopped. A separate counter is required for each of Scans 2, 3, and 4. Suitable relay counting circuitry was designed to accomplish the requirement.

The design of circuitry to determine extremities was accomplished by using a time-dependent form of Boolean algebra. It should be recalled that a letter with an extremity is defined as having at least the upper or lower portion extending well to the right or left compared with its center. Considering the left side of a letter, an extremity exists if returns from Scans 2 or 4 can be sensed by the reader before there is any return from Scan 3. The scanning spots are assumed to move from left to right. In concise terms, Category 11 is established by the condition, $2(t_1)\sim 3\ (t)$, or $4(t_1)\sim 3\ (t)$, where t_1 is the time when returns are sensed from Scans 2 or 4 and t is any time from the reset, t_0, up to but not including t_1. Categories 12, 13, and 14 may be established by the same form of Boolean notation. Relay operations in the extremity circuits are basically the same as in the counting circuitry with additional "hold" and "dropout" relays to provide the time of dependence.

During the development of the counting and extremity-determining circuitry, it was noted that certain letters might not be identified correctly unless provisions were made to obtain maximum information from their return pulses. Differentiation of the scan pulses permitted identification to be used for counting and identifying left- and right-side extremities.

With the circuitry discussed thus far, it is possible for a reading device to determine the categories to which a scanned letter belongs. In short, this is accomplished by energizing a combination of three more relays out of a possible 14. The energized relays, in turn, supply voltage to their respective category output leads. The next requirement is for circuitry which can sense and indicate which letter has been scanned, given that some combination of leads is at some non-zero potential. Each unique combination of active terminals may be identified with the scanned letter by at least two methods. One is by use of a logical "tree" structure. Another method might be by means of a logical diode matrix. Because of materiel difficulties, it was decided to eliminate this portion of the reader and make use of visual recognition attained by matching patterns of active lights. This method throws the burden of decoding the simplified logic essentially upon the human receiver.

The final requirement of the logical portion of the reader was the ability to determine when one letter had been completely scanned and then to reset the other logical circuitry to its original state. In this way, the reader, after scanning the first letter, would automatically reset to process information from subsequent ones. It was hoped originally that, when each of the five scanning spots no longer touched any portion of a letter, scanning could be considered finished and the reset action initiated, but there are several letters which frustrate such a plan. Consider the letter "X." When the scanning spots are in the position shown in Fig. 1, none of the scans are intersecting the marks of the letter. Yet it is clear that if reset were to occur at this instant, the letter would be incorrectly identified. The use of a single vertical scan used only for reset operation would seem to be the optimum solution of this problem. In the reading device itself, manual reset was used.

The construction of the pulse-forming circuits required extensive experimentation. A type-930 gas phototube was found adequate when used with triode amplifiers. The scanning mechanism consisted of two spotlights mounted behind a flat plate with five holes. The plate was mounted so that the five holes were in a vertical row. At one side of the plate were mounted the five phototubes; each of the holes in the plate was aligned with one of the semicyclindrical phototube cathodes. A short grooved track was mounted on the same chassis with the plate, and the phototubes and a piece of plate glass with source letters marked with a grease pencil was passed down the track. The interruption of the light to the phototube caused by the letter passing through the beam activated the logical circuitry and permitted identification of the letter. Practical readers must scan the printed page with minute beams of light which will be reflected from the page in different intensities for mark or no-mark. The photo-sensitive transducer receives the reflected light and this information is delivered to the logical circuitry for identification.

Discussion of Results

As previously stated, the group of characters to be identified was the Russian alphabet, specifically the 32 small printed letters. The characters tested were modeled after those shown in Mueller's *Russian-English Dictionary*. Of the 32 letters in the Russian alphabet, the reader placed 30 in the categories predicted. Categorization other than that predicted was caused by more sensitivity of the reader than had been expected. The device gave these final results:

Letters correctly identified . . . 30 of 32, or 94%
Letters uniquely determined . . . 26 of 32, or 81%

The results obtained indicate that a recognition logic based on horizontal scanning is highly effective in identifying lower-case printed Russian letters. The results could have been degraded or enhanced, however, by several factors worth considering. First is the shape of the letters themselves. A consistent and exact reproduction of the shapes from which the logic was derived would give essentially 100 per cent identification. For different types of letter formation (type forms) it might be necessary to adjust the relative scan positions.

The horizontal scan recognition logic was based to a certain extent on human visual principles. Instead of using a point-by-point scan of each letter, an attempt was made to develop a recognition technique based on sensing only the most obvious characteristics of each letter. This approach was shown to be feasible by the results obtained. In reference to the particular recognition technique, however, six of the 32 Russian letters were not uniquely identified. This implies that another specific recognition coding method might be examined for complete definition. Such a study is currently in progress in the Electrical Engineering Department under Engineering Experiment Station sponsorship.

An important feature of the horizontal scan is that the reader can generate scan information and process it on a parallel basis. Because of this feature a single pass of the scan spots over the letter was sufficient for identification. This fact permits recognition at roughly five times the rate possible with serial scanning. Parallel or simultaneous techniques can certainly be applied to practical electronic reading devices.