U. S. Naval Weapons Laboratory
Dahlgren, Virginia

CALLIGRAPHY FOR COMPUTERS

by

A. V. HERSHEY
Computation and Analysis Laboratory

NWL REPORT NO. 2101

Task Assignment
NO. R360FR103/2101/R0110101

1 Aug 1967

TABLE OF CONTENTS

ABSTRACT

Consideration is given to the possibility of providing a computer and a cathode ray printer with an unlimited repertory of characters. Digitalizations are presented for mathematic, cartographic, and calligraphic characters. The repertory is available to any computer through FORTRAN IV programming. The latest cathode ray printers are almost adequate for the preparation of mathematical reports. Some progress has been made toward development of a mnemonic code for the recording of a mathematical text on tape.

FOREWORD

The work of this report represents an advance in the application of computers. Programming and computation were charged to the Foundational Research Program of the Naval Weapons Laboratory, Project No. R360FR103/2101/R0110101. Character displays were programmed for the NORC cathode ray printer by W. H. Langdon, and for the STRETCH cathode ray printer by Mrs. E. J. Hershey. The photomicrogram of Figure 1 was prepared by J. P. Rucker. Dot plots were prepared on an S-C 4010 printer at the Naval Weapons Laboratory and vector plots were prepared on an S-C 4020 printer at the Naval Ship Research and Development Center. The manuscript was completed by 1 Aug 1967. The Japanese Lexicon was checked by Educational Services of Washington, D. C.

APPROVED FOR RELEASE:

/s/ BERNARD SMITH
Technical Director

INTRODUCTION

Although computers are used primarily for arithmetic, there are other ways in which computers can be used for the saving of labor. The use of computers and cathode ray printers for typesetting[1,2] is receiving much attention at the present time. Publishers are interested in the possibility of reducing the cost of printing and scientists are interested in the possibility of improving the versatility of printing.

The objective of the present investigation is to explore the feasibility of utilizing the computers and cathode ray printers at the Naval Weapons Laboratory for the preparation of mathematical reports. In this connection a large repertory of digitalized characters has been prepared. The repertory was intended to correspond in scope to the repertories of the American Institute of Physics[3] and the American Mathematical Society[4]. The virtuosity of the cathode ray printer has been explored further with a number of calligraphic digitalizations.

Although a number of printer systems[2] currently are under development, it is assumed in the present report that the Linotron equipment of the Mergenthaler Linotype Company and the Charactron equipment of the Stromberg-Carlson Corporation may serve as examples to illustrate representative qualities, speeds, and versatilities. The repertory in the present report is intended to fill a need for a system which does not sacrifice too much quality or speed, but is unlimited in versatility.

A digitalization of characters was undertaken originally at the Naval Weapons Laboratory[5] for use on dot plotters. An improved version of the original digitalization is presented herewith as Appendix A. With the exception of a few of the characters, no attempt was made to vary line thickness.

A digitalization of characters has been prepared recently at the Bell Telephone Laboratories[6] for use on vector plotters. Line thickening was achieved through the use of multiple lines one raster unit apart. The style of character has been limited so far to Roman and Greek lower case and upper case. The remarkable success of the line thickening has been a stimulus to an extension of the same technique to exotic graphics.

The digitalizations at the Naval Weapons Laboratory and at the Bell Telephone Laboratories complement each other insofar as they do not overlap from the standpoint of style or height of character.

A digitalization of characters is currently under preparation at the Naval Weapons Laboratory for use on vector plotters. Details of the current digitalizations are presented herewith as Appendix B. The scope of the digitalizations is indicated by the following table.

CHARACTER DIGITALIZATIONS

I SIMPLEX

Roman, Greek, Script, Numeric, FORTRAN, Electronic, Cartographic.

II DUPLEX

Roman, Greek, Italic, Futura, Script, Russian, Numeric, Mathematic, Astronomic, Musical.

III TRIPLEX

English Gothic, Italian Gothic, German Gothic.

IV JAPANESE

Hiragana, Katakana, Kanji.

Some of the alphabets in the table have been given new names because they are not identical with existing alphabets. The word simplex has been selected to describe those alphabets which are composed of lines of uniform thickness and have no serifs or flourishes. The simplex style of character is known otherwise as gothic*, sans serif, grotesk, light face, or block letter. The word complex may be applied to those alphabets which are composed of lines of variable thickness and do have serifs or flourishes. The complex style of character includes those which are known otherwise as standard, modern, boldface, or black letter. The words uniplex, duplex, multiplex may be used to express the number of lines which are used in parallel to obtain a variation in line thickness.

Three sizes of characters are provided by the repertory in Appendix B. Characters 9 raster units in height are available for FORTRAN or cartographic applications. Characters 13 raster units in height are available for indexical lines of print. Characters 21 raster units in height are available for principal lines of print.

PRINTING SYSTEMS

Character Generation

In cathode ray printing systems, characters are displayed on the face of a cathode ray tube and are photographed by a camera. Two distinct methods are used for the creation of a character on the face of the cathode ray tube. In one method, a character is created by a beam of electrons which is shaped by its passage through an aperture in a matrix. In the other method, a character is created from the strokes of an electron beam with a constant sweep rate.

---

*Only in America is the term gothic applied to this style of character.

The space occupied by a character and the time required to create the character are constant for shaped characters but depend upon the size and complexity for stroked characters. In order to compare the methods of creating characters, weighted averages of space and time are required. Weighted averages may be derived through summation of the product of space or time for each character by the frequency of occurrence of the character as utilized in cryptology[7].

Shaped characters and stroked characters both may be created with the Charactron printers.

<u>Charactron Printers</u>

The cathode ray printers at the Naval Weapons Laboratory consist of an S-C 4010 printer[11] on line to the Naval Ordnance Research Computer, and an S-C 4010 printer[12] off line to the STRETCH computer. These are dot plotters and have no vector plotting capability beyond axis generation. The shaped characters occupy 8 raster units of width and require 58 microseconds of time. The matrix contains only 64 characters. Stroked characters can be plotted with the aid of vector simulation subroutines, or the characters can be created out of dots as in Appendix A. A representative weighted average of width for dot plots is 17 raster units and a representative number of dots per character is 22. The plotting of each dot requires 85 microseconds of time.

In the S-C 4020 printer[13] a vector plotting capability is added to the dot plotting capability of the S-C 4010 printer. Stroked characters can be created out of vectors as in Appendix B. A representative weighted average of width for vector plots is 18 raster units and a representative number of vectors per character is 19. The time to plot each vector depends upon the time to decode the plot instruction and the time to sweep the vector. A representative decoding time is 85 microseconds and a representative sweep rate is $\frac{1}{2}$ raster unit per microsecond. The size of the raster is 1024×1024.

In the S-C 4060 printer[14, 15] the speed and repertory have been increased. Four sizes of shaped characters are provided, and the shaped characters require 11 microseconds of time for creation. The matrix contains 115 characters and includes both lower case and upper case. Four sizes of plotting dot are provided. A representative decoding time is 15 microseconds and a representative sweep rate is 2 raster units per microsecond. The size of the raster is 3072×4096 and the size of the raster unit is the same on both axes. The longer dimension of the raster is in the longitudinal direction on the camera film. The fineness of the raster cannot be utilized fully for stroked characters because of limitations on the fineness of resolution. The smallest plotting dot is three raster units in diameter according to measurements on a specimen of hard copy.

Linotron Printers

In the Linotron printer the characters are stored as photographic images on four glass plates. Any selected character is scanned photoelectrically in a succession of horizontal sweeps across the character block. The photoelectric signal is displayed on a cathode ray tube. The selection, enlargement, and deflection of each character all are performed electrically. The time to create a character depends upon the size of character. For 6, 8, 10 point sizes of character the printing speed is quoted[16] at 1000, 800, 620 characters per second, respectively. The characters are of graphic arts quality on an $8 \times 10\frac{1}{2}$ inch page size. The repertory includes 1020 characters of which a few are mathematical. However, the present scope of the Linotron project does not extend to chemical and mathematical composition.*

Relative Speeds

Insofar as the data in the above considerations are representative

---

*The existing repertory does not include the integral sign or the partial differential symbol.

of actual performance, the data in the following table are representative of printing speeds.

Printing System (stroke vs shape)	Printing Speed (characters/second)
Dot plot on S-C 4010	530
Vector plot on S-C 4020	550
Vector plot on S-C 4060	2200
Print on Linotron	620
Print on S-C 4010	17400
Print on S-C 4020	17400
Print on S-C 4060	90000

The above estimates do not include the time on a general purpose computer which would be required for the preparation of input to the cathode ray printers.

RESOLUTION

Model

In order to gain some insight into possible factors in the resolution of a cathode ray printer, an analysis will be made on a specific model in which the raster on the cathode ray screen covers an area 10 cm × 10 cm square and contains 1024 × 1024 raster units. It will be assumed that hard copy from the cathode ray printer covers an area 6" × 6" and is viewed by a reader's eye at the conventional distance of 10".

Acuity

A limiting factor is the acuity of the eye. Any resolution in excess of the amount which can be perceived would be wasted. The acuity of the eye varies among individuals, and the acuity varies with the type of perception. Insofar as the perception of separation between lines is a gauge of acuity, the angle of resolution[10] is 30" of arc or a quarter of a raster unit.

Diffraction

An interesting factor is the diffraction of electrons or light in the printer system. The diffraction pattern of a circular aperture consists of alternating bright and dark rings around the geometric center. The angle θ which is subtended by the diameter of the first dark ring is given by the equation

$$\theta = 2.44 \frac{\lambda}{d} \qquad (1)$$

where λ is the wave length and d is the diameter of the circular aperture. The wave length for electrons is given by the equation*

$$\lambda = \sqrt{\frac{150}{V}} \times 10^{-8} \text{ cm} \qquad (2)$$

where V is the voltage through which the electrons have been accelerated before diffraction.

The paths of the electrons which enter an aperture of the matrix have some dispersion of direction because of the finite aperture of

---

*This equation is given in the Encyclopaedia Britannica[8] but not in the Handbook of the American Institute of Physics[9]!

the electron gun, and the dispersion is increased further by diffraction at the aperture. Regardless of the dispersion, all electrons which emanate from a given point in the aperture would be brought to a focus at a common point on the screen if the focusing were perfect.

The effect of diffraction applies to the aperture of the focusing system. It is assumed that the electrons are at 3300 volts when they are diffracted at an aperture of 1 cm diameter and at a distance of 50 cm from the cathode ray screen. The diameter of the first dark ring is computed to be less than 3×10^{-5} raster units and the effect of electron diffraction is negligible.

It is assumed that the cathode ray screen is coated with RCA phosphor No. 11 which has a peak intensity of emission at a wave length of 4600 Å. It is assumed that the camera is operated at a lens aperture of $f/5.6$. The diameter of the dark ring of optical diffraction is calculated then to be 0.064 raster units.

Grain Size

It is assumed that the diameter of the grains of the phosphor is 5 microns. The grain diameter then corresponds to one twentieth of a raster unit. That the grain size is small also on the film in the camera is indicated by Figure 1. This photomicrogram is a 650 × magnification of a dot which has been recorded on film in the NORC cathode ray printer.

Aberration

One factor which affects resolution is the effect of aberration on the focusing of the electron beam. A diffuse character of the plotting dot can be discerned in Figure 1. The diffuseness may be greater still in a cathode ray printer which is not maintained in perfect adjustment. The diffuseness has the beneficial effect in a dot plotter of making it possible for a series of closely spaced dots to merge into a continuous line. The diffuseness has the deleterious effect in a vector plotter of bridging small gaps or of filling small openings in the characters. Due allowance must be made in the design of the characters to avoid these unacceptable effects. A gap in a line may be smaller than the opening within a circle without undue bridging or filling.

Dot Size

From densitometer readings it has been determined that the effective diameter of the plotting dot is 2.9 raster units for the S-C 4010 printer. A diameter of 2.3 raster units has been reported[6] for the S-C 4020 printer. That the diameter could be as small as one raster unit for the same printer is implied by measurements on the hard copy sample from the S-C 4060 printer. It is evident that the cathode ray printers do not achieve the ultimate in resolving power.

The diameter of the plotting dot in a vector plotter should be a minimum in order to give a maximum control of line thickness. The diameter must be no less than one raster unit in order that solid areas may be swept out. The fineness of strokes which can be printed on current cathode ray printers is limited by dot size and not by raster size.

Raster Size

A line of text in a mathematical document should be long enough so that the mathematical equations which are inserted in the text only rarely need to be broken with part on one line and part on another line. With the model herein adopted for analysis, the length of a line of text is 6". If this were typewritten in elite style at 12 characters per inch there would be 72 characters per line of text.

If the line of text were printed with stroked characters at 18 raster units per character, then 1296 raster units would be required per line of text. This is not too many characters per line. Although the number of characters per line is less than 72 for the texts of the American Institute of Physics[3] or the American Mathematical Society[4], it may be more than 72 for the texts of the Cambridge University Press[17].

Requirements

It seems apparent that the S-C 4010 and the S-C 4020 cathode ray printers do not have small enough plotting dots and large enough rasters to meet the requirements for the printing of mathematical texts. The S-C 4060 cathode ray printer could meet the requirements if the plotting dot were truly 2 raster units in diameter and the starting and stopping of vectors were controlled to within a raster unit.

CHARACTER DESIGN

Design Criteria

There would be no problem in copying any existing character if the cathode ray printer did not have a finite plotting dot and a finite raster size. The problem of design arises from the need to make a compromise between the three factors of smallness, smoothness, and legibility. It is desirable to make the characters as small as possible so that as many characters can be printed on a line of print as possible. It is desirable to make the edges of curved lines smooth so that characters may have a professional appearance. It is essential that there be no loss of legibility because of bridging or filling of small gaps. The finest detail in any character of an alphabet sets a limit on the smallness of character for the whole alphabet.

The problem of digitalization is to locate successive points in a relatively coarse grid such that vectors can be drawn between the points with optimum results. The absolute position of the successive vectors is not so important as the relative orientation of the successive vectors. With an application of ingenuity it often is possible to achieve a pleasing effect with the polygonalization of curved lines. The limitation on digitalization which is imposed by the finiteness of the grid constitutes an artistic challenge. It is not obvious a priori that all of the characters of interest can be digitalized.

Character Size

A satisfactory polygonalization of a small circle is not possible for a circle of any arbitrary size. The number of sides of the polygon is related to the size of the polygon. The smallest sizes are an octagon of 4 or 6 raster units diameter and a dodecagon of 8 raster units diameter. The next two sizes are hexadecagons with 10 or 14 raster units diameter. The choice of diameter is related to the fact that the polygon appears round only if it has the same radius at 45° inclinations as it has at 0° or 90° inclinations. The products of $\sqrt{2}$ and the smallest integers are approximately integral only if the integers are 5 or 7.

From a mathematical standpoint, an ellipse would be polygonalized by a polygon which is tangent to the ellipse at the point of contact between ellipse and polygon. The ellipse may be found by simultaneous solution of the equation

$$\frac{x^2}{a^2} + \frac{(y - b)^2}{b^2} = 1 \tag{3}$$

for the ellipse, and the equation

$$\frac{dy}{dx} = - \frac{b^2}{a^2} \frac{x}{(y - b)} \tag{4}$$

for the slope of its tangent. In these equations a and b are principal radii of the ellipse. Solution leads to the equation

$$\frac{y}{x} = \frac{-1 + \sqrt{1 + \frac{a^2}{b^2}\left(\frac{dy}{dx}\right)^2}}{\frac{a^2}{b^2}\,\frac{dy}{dx}} \tag{5}$$

Along a side of the polygon, x and y are related linearly, and the slope dy/dx is constant. The point of tangency between ellipse and polygon may be found by the solution of two simultaneous linear equations in x and y. A number of solutions have been obtained, but only the solutions in the following table are within reasonable bounds.

Side of Polygon	Height of Ellipse
$y = \frac{1}{4}(x - 2)$	$2a = 22.0$ for $\frac{b}{a} = \frac{2}{3}$
$y = \frac{1}{3}(x - \frac{3}{2})$	$2a = 18.5$ for $a = b$
$y = \frac{1}{2}(x - 1)$	$2b = 18.5$ for $\frac{a}{b} = \frac{2}{3}$

The height for polygonalization is not well defined but seems to range from 18 to 22 raster units.

Professional printers measure the size of type in points such that one inch equals 72 points. The point size of type is the normal distance from the base line of one line of type to the base line of the next line of type. The design of character within a character block depends upon the amount of white space which is to be provided between lines of type. Printers often increase the white space to more than normal with additional leading between lines of type. The normal distance from one line to the next is one em, which is subdivided further into printers units such that one em equals 18 units.

A natural correlation between mechanical printing and cathode ray plotting would be achieved if a printer's unit were equated to an integer number of raster units. Insofar as a representative height of character is 12 printer's units, a representative height of character would be 12 or 24 raster units.

In the printing of mathematical texts the principal line of type is printed in 10-point type while the indexical lines of type are printed in 6-point type. The sizes of character in raster units should be compatible with two kinds of line of type.

In the Roman alphabet some lower case letters are two-thirds as high as the upper case letters. The height of the upper case letters should be a multiple of three. Many lower case letters are round, while several upper case letters are oval. The Arabic numerals have round parts. The various round characters should be coordinated with small circles. In the Italic alphabet there are slant lines of various lengths. The projection of each slant line on the horizontal axis is a small integer. For a given slope of line the height of line can have only a few values. Typical slopes for actual Italics are 1 to 3 or 4.

The above considerations have led to a choice of 14 raster units as the basic width and 21 raster units as the basic height of the upper case letters of principal lines of type, and a choice of 10 raster units as the basic width and 13 raster units as the basic height of the upper case letters of indexical lines of type.

Character Space

Calligraphers[25] advocate the use of the style of Roman lettering on the Trajan column. This style may be appropriate for architecture but the letters vary greatly in width. Inasmuch as the lettering in the present alphabets is intended to be used interchangeably in words of a text or as symbols in a graph, the letters have been designed to appear uniform in width.

Calligraphers[21, 22] agree that the white spaces within letters and between letters should have a uniform distribution along a line of print. This is not really possible in the presence of the letter pairs AA or VV, but these letter pairs are rare. The spacing which should be allotted to each letter varies with the environment in which the letter is situated, and it even has been proposed that the width of the letter itself should vary with its environment. In the present alphabets each character block is allotted its own width, but the width can be changed to any other value as may be desired under program control in the computer.

Character Style

The digitalizations of simplex alphabets are adaptations of the alphabets on Le Roy lettering sets. The digitalizations of complex Roman, Greek, Italic, Russian alphabets are adaptations of the alphabets to be observed in newspapers, text books, and dictionaries[18, 19].

Script and Gothic Alphabets

Originally there was only one style of Roman lettering, but the need for a rapid cursive handwriting resulted in a rounding of angularity with the formation of the uncial style of lettering. Now there are two sets of characters for each style of lettering. The majuscules are used for initials and are known otherwise as capitals or upper case letters. The minuscules are used for text, and are known otherwise as small letters or lower case letters. Further evolution of the minuscules resulted in Script for writing and Gothic for printing.

Characters from these alphabets are borrowed occasionally by mathematicians to represent special quantities.

Digitalization of the script alphabet has been adapted from a Headliner Typemaster of the Varityper Corporation. The first Gothic alphabet has been adapted from a Le Roy lettering set for Old English and is called English Gothic. The second Gothic alphabet represents a large family of alphabets for which there does not seem to be a consistent nomenclature. Some writers refer to it as Gothic uncial while others call it Lombardic Gothic. It seems to have been developed in Lombardy while the best examples[23, 24] seem to come from Spain. The present version is an adaptation of a font of the American Type Founders Company[20]. It is being named Italian Gothic because of its Lombardic origin. The third Gothic alphabet is an adaptation of Fraktur[25] and is named German Gothic.

Musical Symbols

The digitalization of musical symbols depends upon the spacing between the lines of the staff. A whole note can be centered over a line only if its height is an even number of raster units. The note can be centered between lines if the spacing between lines is even. A whole note can straddle a line without undue filling and numerals 13 raster units high can be used for measure signs if the spacing between lines is selected to be 10 raster units.

Japanese Characters

The ultimate challenge to calligraphy for computers is the imitation of brush strokes in Chinese and Japanese characters. An investigation has been made to determine the feasibility of digitalization of the Japanese characters. The results are given in Appendix C. The results even have been used for the preparation of an abstract of a Naval Weapons Laboratory report in Japanese as well as in French and German.

Originally the Japanese had no way to write the Japanese language[31]. Chinese characters were introduced into Japan along with Confucianism and Buddhism. The structure of a majority of Chinese characters consists of two parts. One part defines the meaning while the other part defines the pronunciation. The two parts often are so selected as to express a logical or poetic meaning for the character.

The Chinese characters are used as stems of many words. Two or more Chinese characters often are grouped together to form compound words. The Chinese characters are called *kanji* by the Japanese. A character dictionary lists 5500 Chinese characters of common occurrence in the modern literature. There are many more in the classical literature. Many of the *kanji* have been simplified, and in November 1946 the Japanese Ministry of Education selected 1850 *kanji* to be used in newspapers and official documents. These are called *Tōyō Kanji* or current characters. They constitute much too restricted a list for technical writing, and even the abstract which is referred to above is not confined to the list.

Parts of certain Chinese characters have been abstracted by the Japanese to form two phonetic syllabaries. The phonetic characters are called *kana* by the Japanese. The *hiragana* syllabary is used as the inflection of words and the *katakana* syllabary is used for foreign words or telegrams. There are 48 basic characters in each phonetic syllabary. Some of these may be modified by diacritical marks or *nigori* to make 25 additional characters. The number of phonemes is 73 for each syllabary.

Each Chinese character has one or more pronunciations of Chinese origin which are called *on*. The Chinese characters for common things also have a Japanese pronunciation which is called *kun*. When Chinese characters are used individually or with a Japanese inflection they are given the *kun* pronunciation. When they are joined together in a compound word they are given the *on* pronunciation. There are only 326 *on* pronunciations to be distributed among 5500 characters. Each *on* pronunciation applies therefore to many characters. Ambiguity is avoided insofar as each *on* occurs only within the context for which it has a unique interpretation. The pronunciations can be transliterated into the Roman alphabet in accordance with the Hepburn system. The Romanization is called *rōmaji* by the Japanese. Certain vowel sounds are suppressed while others are lengthened in certain pairs of *kana* which are transliterated into distinct phonemes. There are 114 phonemes in the *rōmaji*.

The structure of each Chinese character consists of one or more parts. One part of every character is called a radical. There are 214 radicals. Many of the radicals are themselves complete characters, while other radicals no longer are used except as parts of characters. To find a character in a character dictionary the first step is to recognize the radical in the character. The radicals are listed serially in the order of increasing number of strokes in the index of the dictionary. All characters with the same radical are listed together in the order of increasing number of strokes in the body of the dictionary. The problem of finding a character thus is reduced to the scanning of a relatively small number of pages in the dictionary.

Character Selection

In view of the large number of characters in a character dictionary, severe limitations had to be imposed on the selection of characters for digitalization. The scope of selection of characters was limited to three sets of characters. The first set includes those radicals which are members also of the *Tōyō Kanji* list. The second set includes those characters which are taught to the Japanese children in the first grade. The third set is a selection of characters of scientific interest. A character which was found to be a component of two or more compound characters was certain to be included. If one character of a pair of antonyms was accepted, the other character was included also, or if one character of a set of characters was accepted, other characters in the set were included. It was impossible to cover more than a small part of any one subject, and the list of characters is illustrative rather than comprehensive, but it should be well balanced as far as it goes.

The choice of characters was checked by a closed circuit through the dictionaries [26-35]. Starting with an English to *kanji* dictionary, the *kanji* for a selected English word was found, then continuing with the character dictionary, the *rōmaji* of the given *kanji* was found, and ending with a *rōmaji* to English dictionary, the *kanji* and English for the given *rōmaji* were found. Thus the final English word could be checked against the initial English word.

In the character dictionaries each character is followed first by the *on* pronunciation, second by the *kun* pronunciation, with English translations wherever possible, and finally by a table of compounds wherein the character appears. Although many of the individual characters no longer are used alone and appear only as components of compounds, they still are given archaic English translations, which would unbalance an abridged list of morphemes. Furthermore, certain grammatical morphemes do not occur in the character dictionaries because they have only phonetic renderings. It appears that the best way to illustrate the use of digitalized characters is by a dictionary listing analogous to Sanseido's[33]. Each entry in the listing is punched on a separate punch card in the order *rōmaji*-*kanji*-*kana*-English. The deck of cards may be sorted, abridged, or augmented easily. Its present status is illustrated in Appendix D.

Each character in Nelson's dictionary[32] is assigned its own number, whereas the characters in other dictionaries are located by page number. Inasmuch as the numbering in Nelson's dictionary provides a natural and definite identification, it has been adopted for the numbering of digitalized characters. It is easy to recover the character by its number from the dictionary.

The style of character which seems most promising for digitalization is represented by the simplified square characters in Nelson's dictionary [32]. These contain hairline horizontal strokes, tapered inclined strokes, and heavy line vertical strokes. Before the characters can be digitalized a decision must be made as to the conversion factor to be used for length from inches to raster units.

Character Conversion

The simplest character of all is No. 0001 (*ichi* = one). It consists of a horizontal line with a triangular spot at the right end. The thickness of the line is 0.010 in. and the length of the line is 0.270 in. The triangle has a base line of 0.060 in. and an altitude of 0.040 in. The vertex of the triangle is 0.010 in. to the left of the center of its base line.

Character No. 0768 (*jū* = ten) differs from character No. 0001 by the addition of a vertical stroke. The horizontal stroke is reduced to a thickness of 0.005 in. and a length of 0.260 in. The triangle has a base line of 0.055 in. and an altitude of 0.034 in. The vertical stroke has a thickness of 0.032 in. and a height of 0.258 in.

Character No. 2170 (*ki* = tree) differs from character No. 0768 by the addition of a pair of diagonal and curved strokes which extend downward to the left and to the right from the center. The horizontal stroke has a length of 0.254 in. and the vertical stroke has a height of 0.263 in. This character occurs as the radical of an especially large number of other characters. When it is used as a radical it is compressed horizontally. In character No. 2379 (*ki* = opportunity) the horizontal stroke has a length of only 0.093 in. The triangular spot has a base line of 0.030 in. and an altitude of 0.020 in.

Thus the thickness and size of components vary in ranges which depend upon the range of fineness of detail. In order to reproduce the above ranges of line thickness and triangle size the conversion may be determined to be 0.011 inches per raster unit. This provides two widths of vertical stroke and three sizes of triangle provided the plotting dot is not more than one raster unit in diameter, and due allowance is made for the thickness of line.

A critical determination of the conversion of length is provided by those characters where there is a set of equally spaced parallel strokes. The space between strokes must conform to an integral number of raster units. Any change of space between strokes then is magnified to a large change in the space allowance for the set. Measurements of spacing have been made upon sixty characters. From the measured distance which spans each set of equally spaced strokes it is possible to compute a distance per raster unit for every possible number of raster units per space. When these distances are plotted together for comparison it becomes apparent that there is a tendency for certain distances per raster unit to persist from character to character. There is some persistence around 0.011 inches per raster unit while there is a stronger persistence around 0.0055 inches per raster unit. The second value would allow the horizontal strokes to have just the right thickness for a full representation of detail but the characters would be twice as large.

Critical examples of characters with many equally spaced strokes are given in the table on the next page.

Character Number	Inches per Raster Unit	Inches per Raster Unit	Translation
0272	0.0115	0.0057	*koto* = fact
2141	0.0098	0.0059	*ryō* = quantity
2160	0.0108	0.0054	*kumoru* = cloud up
3113	0.0117	0.0053	*sara* = dish
3127	0.0103	0.0055	*me* = eye
4608	0.0112	0.0056	*kuruma* = vehicle
4883	0.0108	0.0054	*hagane* = steel

This table illustrates the degree of correlation between values for the conversion factor.

Although all characters are centered within the same square block, the overall size of many characters is not well defined because pointed strokes radiate outward in all directions from the interior. The size is really well defined only for those characters which are enclosed in a square radical. Examples with square enclosures are illustrated in the following table.

Character Number	Width	Height	Stroke Count	Translation
0868	0.165	0.155	3	*kuchi* = mouth
2994	0.178	0.188	5	*ta* = rice field
1028	0.190	0.202	6	*mawaru* = go around
1037	0.202	0.220	8	*kuni* = country
1045	0.208	0.233	12	*ken* = circle

The dimensions in the table are center to center between horizontal strokes or between vertical strokes in the external enclosure. The dimensions increase with complexity to a maximum of 21 raster units when the conversion factor is assumed to be 0.011. This is compatible with the standard size of Roman alphabet.

The digitalizations in the present investigation are limited to characters with a nominal height of 21 raster units. With some omission of detail in tight spaces and some overflow in complicated cases this size is believed to be adequate for all characters in Nelson's dictionary except No. 5444. Inasmuch as this character represents dragons in motion, it is of doubtful utility. The remaining characters either have been simplified or can be digitalized without too much distortion provided the minimum spacing between lines can be as small as two raster units. Even character No. 5444 can be digitalized when the nominal height of character is 42 raster units.

DOT DATA

Smooth straight lines can be generated with a dot plotter only in limited directions where the discrete increments ΔX, ΔY from one dot to the next have simple integral values. Primary directions are generated when the lines are defined by the increments

$$(\Delta X, \Delta Y) = (2, 0)$$

$$(\Delta X, \Delta Y) = (2, 1)$$

$$(\Delta X, \Delta Y) = (1, 1)$$

or by any permutation of magnitude or reversal of sign among these increments. Secondary directions are generated when the lines are defined by alternation between the following pairs of increments

$$(\Delta X,\ \Delta Y) = (2,\ 0),\ (2,\ 1)$$

$$(\Delta X,\ \Delta Y) = (1,\ 0),\ (2,\ 1)$$

$$(\Delta X,\ \Delta Y) = (1,\ 1),\ (2,\ 1)$$

or by permutations or reversals among these. Jogs in the lines become perceptible when more elaborate patterns are used. The linear characters A, K, M, N, V, W, X, Y, Z contain a variety of inclined lines and limitations on the possible inclinations determine the shapes of the characters. The Roman style of character is available to a dot plotter, but the inclinations for an Italic style of character would be too exaggerated.

Dot plotting on NORC is accomplished by either of two character plotting routines. Block No. 0130 gives a mathematical repertory while Block No. 0160 gives a cartographic repertory. These NORC subroutines have been converted recently to FORTRAN IV by the Control Data Corporation.

The digital data for each character are packed in the data array of each subroutine. The data consist of decimal digit pairs. The first digit pair gives the half width of the character. The second digit pair gives the X-displacement and the third digit pair gives the Y-displacement to the first dot. The subsequent digit pairs give displacements to successive dots. In each of these digit pairs the first digit is the X-displacement and the second digit is the Y-displacement. Negative displacements are expressed by 9's complements. Whenever the first digit is 5, the previous displacement is repeated a number of times equal to the second digit. If the digit pair is 00, the next four digits are interpreted in the same way as the second and third digit pairs, except that displacements are relative to the last plotted dot. The digit pair 50 signifies the end of character.

The decimal format for NORC data is not suitable for STRETCH programming. Inasmuch as the NORC word is 16 decimal digits long and the STRETCH word is 64 binary bits long, there can be a one to one correspondence between the BCD datum word for NORC and the binary datum word for STRETCH. One decimal digit with 9's complements in NORC is mapped into three integer bits and one sign bit in STRETCH. An array of coordinates for dot plotting is recovered from memory by interrogation of a pair of STRAP subroutines.

Replacement of FORTRAN programming by STRAP programming in the character plotting routines has achieved a 7-fold reduction in machine time.

VECTOR DATA

Smooth straight lines are no problem for a vector plotter, but curved lines are approximated by polygons. Small polygons are constructed from short vectors whose components ΔX, ΔY have the following integral values

$(\Delta X, \Delta Y) = (1, 0)$ $(\Delta X, \Delta Y) = (1, 1)$

$(\Delta X, \Delta Y) = (2, 0)$ $(\Delta X, \Delta Y) = (2, 1)$ $(\Delta X, \Delta Y) = (2, 2)$

$(\Delta X, \Delta Y) = (3, 0)$ $(\Delta X, \Delta Y) = (3, 1)$ $(\Delta X, \Delta Y) = (3, 2)$

$(\Delta X, \Delta Y) = (4, 0)$ $(\Delta X, \Delta Y) = (4, 1)$

$(\Delta X, \Delta Y) = (5, 0)$

or have any permutation of magnitude or reversal of sign among these values.

In the composition of a character, the ordering and the direction of vectors are immaterial for any cathode ray printer which is correctly adjusted. In order to minimize chaos in the sequence of vectors, the vertical strokes are recorded first and the horizontal strokes are recorded last. Directions are consistently from left to right and from top to bottom. This conforms more or less to the stroke sequence for hand drawn letters. A different sequence might improve the efficiency of a mechanical plotter by a reduction of the amount of motion in a pen up status.

The traditional origin of coordinates for digitalization would be on the base line of the character and at the left edge of the character block. The origin of coordinates for the alphabets at the Bell Telephone Laboratories is situated in the upper left corner of the character block. The origin of coordinates for the characters at the Naval Weapons Laboratory is situated centrally in the interior of the character. This simplifies the centering of isolated characters in cartographic applications and provides a common center line for mixtures* of fonts. Otherwise the origin is arbitrary and the data may be referred to any other origin by a relatively simple subroutine.

The digital data for each character are recorded in a separate block on tape. Each block consists of 16 decimal digit words. Each word is divided into four fields of four digits each. The first word is a beginning-of-block word and the last word is an end-of-block word. Each field of digital data is divided into two digit pairs. The first digit pair of the first field gives the left edge of the character block. The second digit pair of the first field gives the right edge of the character block. Each of the remaining fields give coordinates of a point. The first digit pair gives the X-coordinate and the second digit pair gives the Y-coordinate of the point.

---

*Examples of mixtures include large parentheses around built up fractions or Roman symbols in a Japanese text.

Negative coordinates are expressed by 9's complements. A vector is plotted between each sucessive pair of points. A field of 5000 signifies the end of a string of connected vectors. When this field is sensed, plotting is terminated at the last point and is resumed at the next point. A field of 5050 signifies the end of the character.

The raw data are not suitable for efficient machine computation. They must be reformated in binary mode in such a way as to minimize the memory which is required to store them and to minimize the programming which is required to synthesize printer instructions from them. Although the synthesis of printer instructions could be done in FORTRAN, it is doubtful if this would be as efficient as a synthesis of printer instructions in machine language. STRAP routines are under development for conversion and extraction of data on STRETCH.

REPORT PREPARATION

The usual method for preparing reports at the Naval Weapons Laboratory consists in the typing of a manuscript with an ordinary typewriter which is fitted with Typits. The report herewith was prepared on a Varityper. Six decisions must be made before a character can be struck. These are concerned with horizontal position, vertical position, character style, character size, keyboard bank, and typewriter key. The many errors which occur are painted over or are cut out and replaced laboriously with corrective patches. The alternative would be the typing of the report on a paper or magnetic tape, which could be rewritten and corrected as many times as necessary. Once a correct tape has been achieved, all further conversion and printing becomes automatic. Writing on tape has the disadvantage that the typist must be trained to use function codes. All coding should be mnemonic or phonetic as far as possible without undue complication.

DISCUSSION

The effective utilization of a large repertory depends upon the development of an adequate mnemonic code which a typist can be trained to use. Experimental codes have been described by Barnett[36]. Certainly the alphameric characters will serve as input to Roman alphabets. There is available a convenient transliteration of Greek into Roman for mathematical applications. This transliteration is more nearly isomorphic than isophonic. The phonetic transliterations of Greek, Russian, and Japanese should serve for linguistic applications.

The primary criterion for a choice between character designs is based on what looks best. Attempts to apply mathematical rules have not been entirely adequate. The ultimate criterion certainly is subjective and is an aspect of gestalt psychology. The end of a line seems to have less importance geometrically than it has psychologically. The apparent interaction between a character and the environment in which it is situated may be an application of the adjacency principle of Gogel[37].

CONCLUSION

It can be concluded that the preparation of mathematical reports is almost within the reach of the latest cathode ray printer equipment.

REFERENCES

1. *Symposium on Electronic Composition in Printing.* National Bureau of Standards, Gaithersburg, Maryland, June 15-16, 1967

2. *Automatic Typographic-Quality Typesetting Techniques. A State of the Art Review.* M. E. Stevens and J. L. Little, National Bureau of Standards Monograph 99 (7 April 1967)

3. *Style Manual.* American Institute of Physics, New York (1959)

4. *A Manual for Authors of Mathematical Papers.* Bulletin of the American Mathematical Society, Vol. 68, p. 429 (1962)

5. *Subroutines for the NORC CRT Printer.* A. V. Hershey, Naval Weapons Laboratory Report No. 1686 (9 March 1960)

6. *Three Fonts of Computer Drawn Letters.* M. V. Mathews, Carol Lochbaum, Judith A. Moss, Bell Telephone Laboratories, Murray Hill, N. J. (Private Communication, 1967)

7. *Cryptology.* Encyclopaedia Britannica, Vol. 6, p. 848 (1965)

8. *Electron Diffraction.* Encyclopaedia Britannica, Vol. 8, p. 245 (1965)

9. *American Institute of Physics Handbook.* D. E. Gray, Editor (McGraw Hill Co., New York, 1963)

10. *The Physiology of the Eye.* H. Davson (Little, Brown and Co., Boston, 1963) p. 103

11. *NORC High-Speed CRT Printer Programming Manual.* F. M. Urvinitka, NPG Report No. 1620 (October 1958)

12. *CRT Printer Programming Manual.* W. G. Ward, NWL Technical Memorandum No. K-12/62 (July 1962)

13. *Programmers' Reference Manual, S-C 4020 Computer Recorder.* Stromberg-Carlson Corporation Document No. 9500056 (August 1965)

14. *S-C 4060 Stored Program Recording System. Description and Specifications.* Stromberg-Carlson Corporation Document No. 9500209 (October 1966)

15. *S-C 4060 Stored Program Recording System. Software Description and Specifications.* Stromberg-Carlson Corporation Document No. 9500236 (April 1967)

16. *Electronic Composing System.* (U. S. Government Printing Office, Washington, D. C. 1966)

17. *A Course of Modern Analysis.* E. T. Whittaker and G. N. Watson (Cambridge University Press, London, 1952)

18. *Webster's New International Dictionary. Second Edition.* (G. and C. Merriam Company, Springfield, Mass., 1959) p. 75, p. 2750, p. 3001.

19. *Specimens of Type Faces.* (U. S. Government Printing Office, Washington, D. C.)

20. *Specimen Book and Catalog.* (American Type Founders Company, Jersey City, N. J. 1923) p. 785

21. *The Design of Lettering.* E. Weiss (Pencil Points Press, Inc., New York, 1932)

22. *The Alphabet and Elements of Lettering.* F. W. Goudy (University of California Press, 1942)

23. *Alphabets, Ancient and Modern.* J. B. Russell (Padell Book Co., New York, 1945)

24. *Lettering from A to Z.* C. P. Hornung (Wm. Penn Publishing Corporation, New York, 1954)

25. *Treasury of Alphabets and Lettering.* J. Tschichold (Reinhold Publishing Corporation, New York, 1966)

26. *Japanese Grammar Self-Taught.* H. J. Weintz (E. Marlborough & Co., London, 1907)

27. *New Critical Japanese-English Dictionary.* (Tokyo Government Bookstore Publication, 1921)

28. *Japanese Conversation Grammar.* O. Vaccari and E. E. Vaccari (Vaccari's Language Institute, Tokyo, 1965)

29. *English-Japanese Conversation-Dictionary.* O. Vaccari and E. E. Vaccari (Vaccari's Language Institute, Tokyo, 1963)

30. *Pictorial Chinese-Japanese Characters.* O. Vaccari and E. E. Vaccari (Vaccari's Language Institute, Tokyo, 1964)

31. *A. B. C. Japanese-English Dictionary, Parts I and II.* O. Vaccari and E. E. Vaccari (Vaccari's Language Institute, Tokyo, 1952)

32. *The Modern Reader's Japanese-English Character Dictionary.* A. N. Nelson (Charles E. Tuttle Company, Rutland, Vermont, 1962)

33. *Sanseido's New Concise Japanese-English Dictionary.* (Japan Publications Trading Company, Rutland, Vermont, 1959)

34. *Kenkyusha's New English-Japanese Dictionary.* (Kenkyusha Limited, Tokyo, 1960)

35. *Encyclopaedic Dictionary of Physics, Volume 9. Multilingual Glossary.* (MacMillan Company, New York, 1964)

36. *Computer Typesetting.* M. P. Barnett (M. I. T. Press, Cambridge, Massachusetts, 1965)

37. *The Visual Perception of Size and Distance.* W. C. Gogel, Vision Research, Vol. 3, p. 101 (1963)