Excerpt: Simplifying the ABC's

Jay Doblin with Inder Agrawal, Marianna Porter
and Robert Peterson

How efficient is the roman alphabet? Not very in an age when man-to-man and man-to-machine exchanges are so vital to our communication processes. The efficiency of the alphabet is discussed in terms of information theory, and a new system of letterform design—an extension of the familiar seven-stroke electronic numerals—is proposed.

Surprisingly, in spite of all the advances in technology, no concentrated effort has been made to fit the roman letterforms to present communications media. The roman letterforms served their purpose when the space, time, or costs for storage or transmission were under no great strain and most communication was man to man. But today man-machine and machine-machine communications are an integral part of our daily lives. In this environment these archaic letterforms have severe limitations. While graphic designers produce endless style variations of roman letterforms and designers struggle to accommodate these letterforms for keyboards, displays, and readouts, technology already has evolved as a solution that could be expanded into a more efficient alphabet....

Two-state Codes in Communication

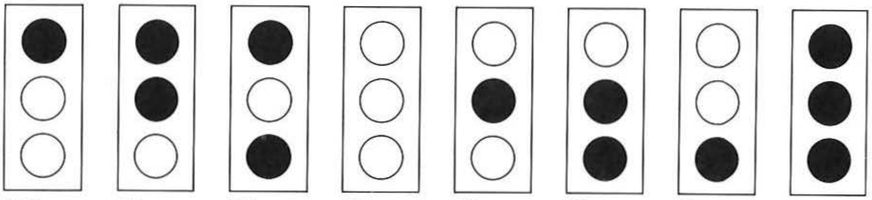
The problem with the present letterform lies mainly in the inherent lack of economy and efficiency in converting the complex hand drawn letterforms into a two-state code which is essential for transmission. The terms "economy" and "efficiency," lead to the conclusion that a measure of information is possible. Measurement of the meaning is difficult. Information that is valuable to one person in a certain circumstance

Visible Language, IX 1 (Winter 1975), 73-86.
Visible Language, c/o The Cleveland Museum of Art, Cleveland, Ohio USA 44106.
Jay Doblin's address: 233 East Ontario Street, Chicago, IL 60611.

could be totally useless in another circumstance or to another person. For the purposes of designing improved letterforms, consideration of the meaning of the letter "A" is irrelevant. What is important is to determine how much information must be transmitted to distinguish the message "A" from other possible messages and, furthermore, how much information is required to visually communicate the form of "A." Both these aspects of information can be readily measured.

In information theory the smallest unit of information is termed a "bit," an abbreviation for the term binary digit. One bit is the amount of information needed to make a decision between two possible alternatives. The amount of information conveyed by the indicator lamp of an oven, for instance, is one bit; the "on" or "off" status of the lamp tells us about one of the two possible states of the oven. This one bit signal can be expressed as two to the first power (2^1), two possible states of one signal. Two bits of information are required to decide among four (2^2) alternatives, three bits to decide among eight (2^3) alternatives; four bits to decide among sixteen (2^4) alternatives; and so on. Every time a bit is added the number of alternatives is doubled.

To demonstrate the communication capacity of a three-bit code, an ordinary three-lamp traffic signal with red, yellow, and green lights is now used to communicate only three messages—stop, caution, go. But the three lights could communicate eight (2^3) different messages. For example, if the red light were on, the signal could mean stop, as it does now. If red and yellow were on together, it could signal the message, "Proceed cautiously at ten miles per hour." If red and green were on, the signal could communicate, "Proceed cautiously at twenty miles per hour." When all three lights were off, the signal could mean, "Drive carefully at normal city speed limit of thirty miles per hour." If yellow alone were on, the signal could communicate the message, "Go forty miles per hour." If green and yellow were on, the message would be, "Go fifty miles per hour." Green alone could be, "Go sixty miles per hour." If all three lights were on it would signal, "Go as fast as you can." This stop-light system is not being recommended, but is cited



The standard three-lamp traffic signal, although in practice used to communicate only three messages, demonstrates how, with various combinations of the colored lights, a three-bit code could be used to communicate eight different messages.

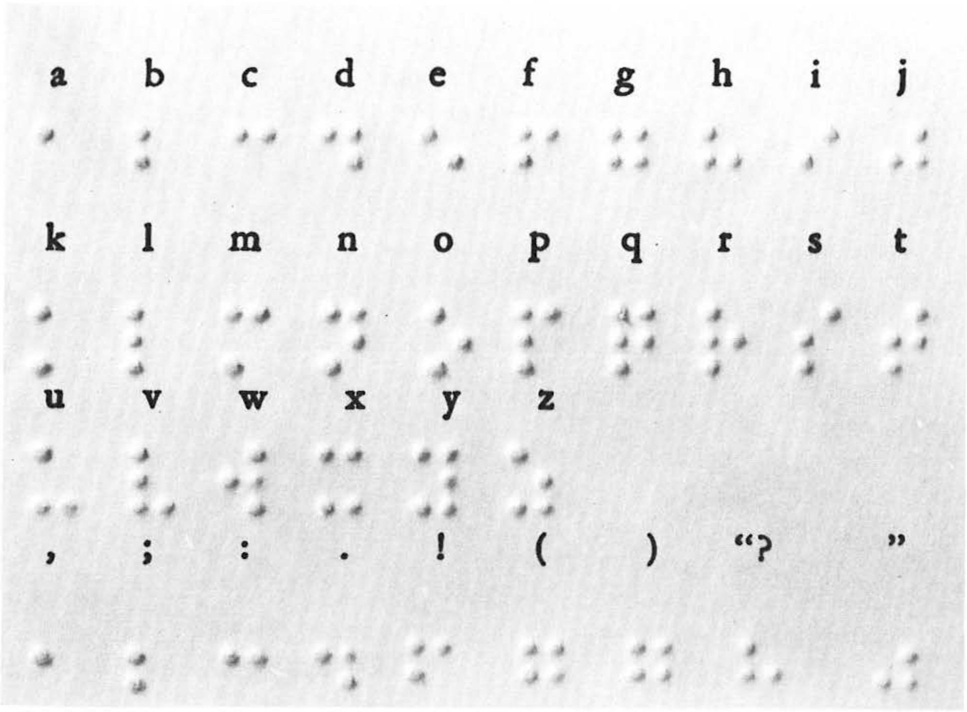
to demonstrate the communication capacity of a three-bit system.

The alphabet consists of twenty-six letters. When we see an "A" the message conveyed is one of twenty-six possible messages. This would be true irrespective of whether "A" is upper- or lower-case, in any size, or in whatever typeface. Since it takes four bits of information to decide among sixteen (2^4) possible alternatives and five bits of information to decide among thirty-two (2^5), the information contained in the message "A" lies between four and five bits (4.71 to be exact).

Six bits (2^6) permits encoding sixty-four different messages which is sufficient to contain all the twenty-six letters, ten numerals, and all the punctuation marks.

Braille, a Six-bit Language

Early recognition of the communication potential of six efficiently used bits occurred in 1829 when Louis Braille published his alphanumeric system that has become the universal written code for the blind. As early as the seventeenth century embossed letterforms for the blind had been introduced, but they were cumbersome to handle, difficult to learn, and could not be easily written. Braille characters use combinations of six embossed dots in a matrix three high, two wide. Of the sixty-four possible figures, twenty-six are used for letters, ten for numerals, and the rest for punctuation, some small common words, and a few common letter sounds. Other combinations have been designated for musical or scientific notation. Their value is that the



The lack of information economy in roman letter-forms which require thirty-five bits to communicate is demonstrated by the Braille alphanumeric system which requires only six bits to communicate sixty-four figures.

low bit count makes them easy to perceive tactilely, and easy to produce manually.

It is now possible to see the inherent lack of information economy in the roman letterforms. While it takes only six bits to transmit alphanumeric messages, it takes at least thirty-five bits to communicate the roman letterform. It was shown by Shannon in his famous theory of communication that all the available information in a continuous signal can be collected by sampling it at finite intervals. The length of the interval, of course, depends on the nature of the signal. The halftone process used in printing and the scanning for television broadcast are good examples of this. Similarly, the form of a letter can be communicated by a series of discrete dots. The minimum number of dots required would then be an indication of the information required to communicate the form. It takes a 5 x 7-dot matrix or thirty-five bits of information to render roman letterforms. This is the format used in most electric signs, scoreboards, CRT displays, and in "hard copy" data terminals.

This thirty-five bit (2^{35}) matrix has the capacity to produce 34,358,738,368 different characters—millions more than needed. The reason for this astronomical capacity to produce characters is caused by the complexity of the archaic roman letterform, and even the thirty-five bit letterform is marginal in both design and readability.

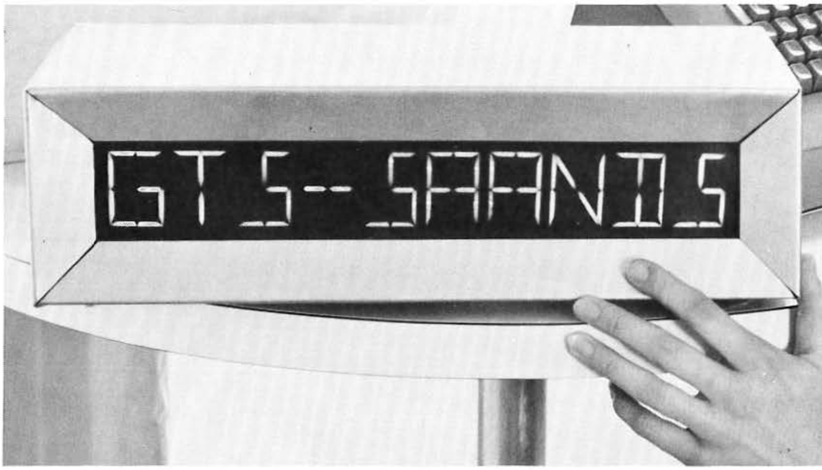
To get around this problem, manufacturers of signs and other readout devices use coding and decoding devices that convert the thirty-five bit code (as read by an optical character reader, for instance), into a six or seven bit code for transmission. For a display, a seven-bit code received in a communication channel must be decoded to a thirty-five bit code for display or printout. What is needed is a six- or seven-bit letterform that can be read directly.

BASIC *Letterform*

To answer this requirement the BASIC letterform has been devised by Design Planning Group. Derived from Braille, it uses seven strokes to connect the six dots. Seven strokes double Braille's six-bit message capacity to 128 (2^7) different messages. The alphabet is an extension of the familiar seven-stroke set of numerals which is widely used on scales, scoreboards, calculators, digital instruments, and electronic cash registers.

The BASIC electronic alphabet is based on the standard number stick figure. This matrix can be narrowed and spacing reduced so that more characters can be packed into the same line length. For print communication BASIC typeface can be thickened, reproportioned and respaced so that the visual effect is more balanced and elegant. Letter spacing for printing could be computerized so that typographic output would be easy to read and would present an even greyness of color.

A variety of technical methods are already used to display the electronic seven-stroke numerals. But for this article the interesting aspect is not these technological mechanisms that produce the letterforms, but the ease with which all of us have



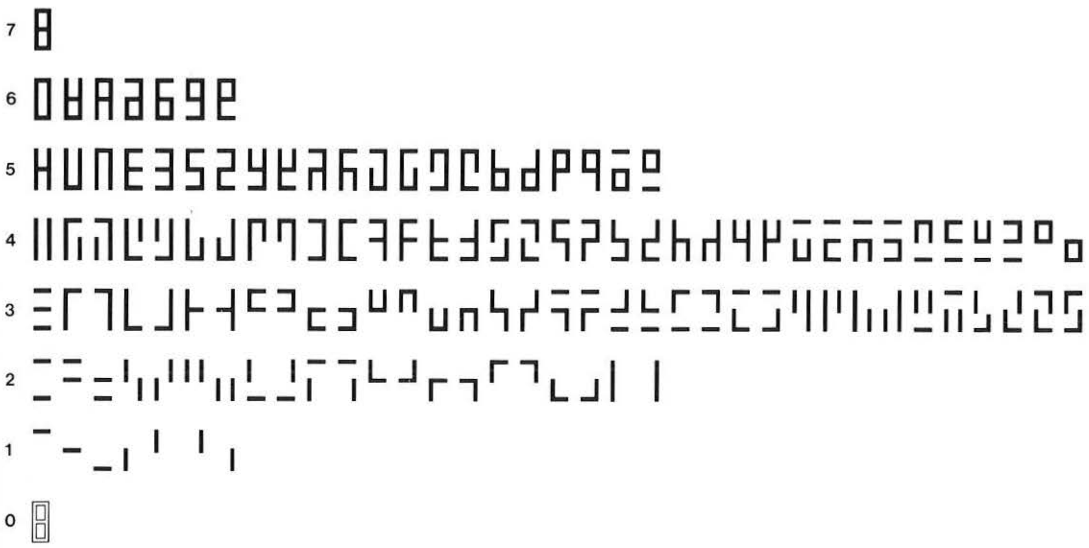
A host of equipment is now in use employing the seven-stroke numerals. There is no reason why the same equipment cannot be used to communicate letters as well as numerals. However, no known alphabet has been developed for this purpose.

adapted to reading these seven-stroke numerals that use no curves or diagonals.

Before any of the letters were chosen from the 128 BASIC characters, the numerals were assigned because of their widespread use. Because of this, BASIC compromises a few of the letters. S and Z are sacrificed because preference is given to 5 and 2; since there is no diagonal or vertical center stroke, M, W, and V are revised. X is the worst problem since there are no diagonals, so the H form is used; h uses its lower-case form. The rest of the letters are self-explanatory since they resemble their roman counterparts. There are plenty of characters left for punctuation, formulae, music, contractions, etc. We have found that learning the BASIC is quick and painless.

The Legibility of BASIC

But there is more to designing an alphabet than achieving a low bit count. In every communication system the sender and receiver have an agreed upon set of possible messages or message elements called "message space." In an alphanumeric code the message space consists of the letters, numerals, and whatever punctuation marks or additional characters are mutually under-



Of the 128 characters formed by the seven-stroke code, forty-five are used for the BASIC letters, numerals, and operators. This leaves quite a large selection of characters for other uses—scientific, musical, and other forms of notation.

One of the best examples of a simplified alphabet, by Christopher Evans and Timothy Epps, is based on a twelve-stroke square grid capable of producing over 4,000 different characters. Neither curves nor diagonals are used so the rendering of some letters, like X, must be compromised. Because of its simplicity, the Evans-Epps letterform lends itself to optical character recognition.

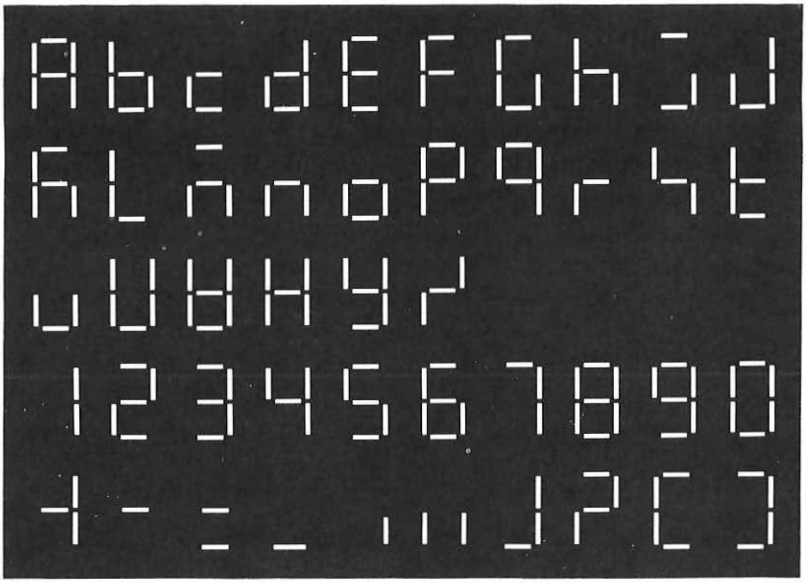
stood by the communicating parties. Thus, physicists can have their peculiar symbols, as can mathematicians or biologists. In communicating, then, familiar characters are recognized which are then assembled into meaning. Faced with an unfamiliar form a person is either taken aback or will interpret it as something familiar. In theory, it ought to be easy to read a seven-bit code consisting of seven sequential dots (like the code used on paper tape). But the difficulty of recognizing 128 different dot forms is impossible and so realization comes that the nature and quantity of information needed for recognition is different than that needed for transmission.

Pattern recognition is a complex process about which little is known. It is generally believed, though, that recognition is the result of our ability to discriminate relevant features of a form. These features, or invariants as they are called, contain the minimum information required for the pattern to be identified. Thus, a letter in its various forms, handwritten in various styles or printed in various typefaces, is recognized by virtue of some information-bearing elements—the strong triangular form of "A" or the circular "O" or the two lines crossing to form an "X." Our alphabet now uses combinations of all three basic visual forms: the square, triangle, and circle, and therefore may tend to deliver the greatest letterform discrimination. It is possible that the eye, with its powerful discrimination capacity, perceives differences better when forms are more varied, even at the expense of a high bit count?

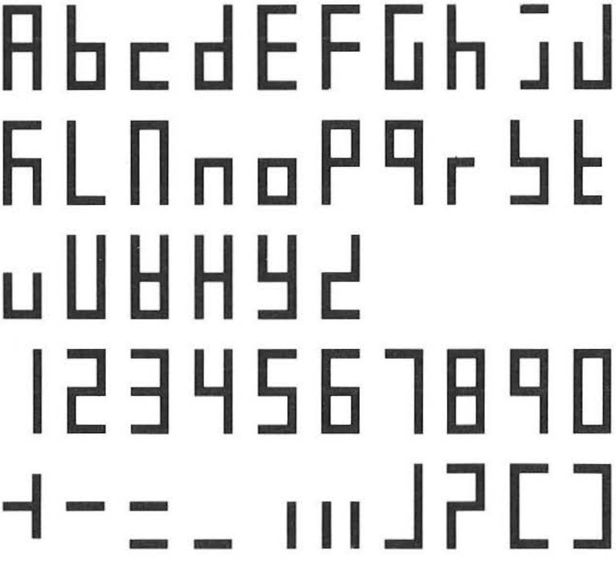
Under adverse visual conditions BASIC might fail more quickly than standard letterforms because there is no redundancy in the characters. Most roman letterforms can be cut in half and still read. BASIC couldn't stand this. The only way to determine these unknowns is through adequate testing in the visual laboratory....

The Aesthetics and Efficiency of BASIC

At first, BASIC may look harsh and angular. Gone are the elegant thicks and thins, there are no juicy O's, no serifs to hold the lines of letters together. But the lower bit count simplicity of BASIC is in line with the long-term trend toward simplicity in



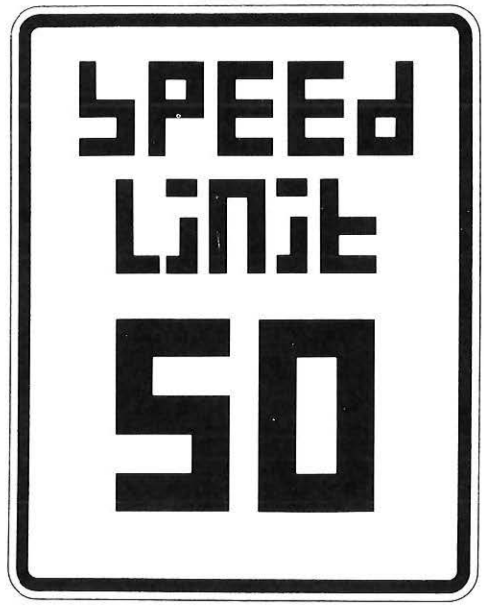
The twenty-six letters, ten numerals, and other operators of BASIC are illustrated using the seven-stroke electronic display matrix. Most of the characters resemble the roman predecessors, but I, M, S, V, W, and Z require some relearning. Of the operators, + does not have a crossbar and ? loses its dot.



the old alphabet, A to Z
was changed to fit technology
but when it was read
it was frequently said
these new letters are all greek to me

Legibility of BASIC can be judged by the facility with which this sentence is comprehended.

An existing road sign is compared to one using BASIC.



design, a result of the shift from hand to machine production. Another reason is a need for greater visual order in an increasingly complex environment. It should be noted that the more visual training the eye gets, the more it tolerates, prefers, or demands greater simplicity of form. Hand-carved furniture became progressively simpler until the Bauhaus, and then it was propelled to pure geometric form. This trend is exemplified by the products of Braun and the architecture of Mies. Obviously, there is a pernicious symbolic drag of the old handcrafter "royal" forms on mass-produced products, but the long trend of design is to cleaner form. This trend to simplicity is true of letterforms as well. Written symbols started as very complex ideograms which over centuries became simpler phonetic marks which were then "geometricized." Complex German script became simple with its sans serif, single-thickness strokes and circular O's. BASIC carries the visual economy of letterform toward a logical conclusion.

A most important attribute of BASIC is its efficiency. The increasing demand on time, space, and cost for storage or transmission of records and messages has led to the common use of micrographic, magnetic, and electronic media. The quantity of the information stored or transmitted is limited by the inherent constraints of the media. For micrographic recording, for instance, the image is reduced until the resolution limit is reached. This limit can be a factor of the resolution limits of cameras, of enlargers, the fineness of the "grain" of photo-sensitive materials, etc. Within these constraints, an elementary typeface can stand greater reduction than can a complex typeface. A letterform with a lower bit count could increase communications capacity In addition, a more efficient letterform designed for the man-machine interface could lead to inexpensive optical readers that would enable huge quantities of printed material to be digested by computers at a low cost. Excessive coding and decoding of information could be avoided since display or printing of an efficient form would require no more information than that needed for transmitting the message. Because BASIC has the potential to require less equipment to produce it, a compact typewriter for it has been projected.

The BASIC *Typewriter*
Considering the technical advances that have made possible miniaturized dictation and calculating equipment, there is no reason why such technology should not be applied to a typewriter. The introduction of electric-powered typewriters lowered the typist's effort, but they are still a long way from being refined equipment. As they are, typewriters are heavy, bulky, noisy, fragile, complicated, and expensive devices. The virtues of BASIC can be combined with current advances in producing typed copy (such as Texas Instruments' Silent 700 electronic data terminals) where the character printer, instead of being a cluster of individual mechanically impacted type characters, is a monolithic 5 x 7-dot matrix silicon heat element. Typing is done on thermographic paper which forms characters at the rate of thirty per second. Printing is entirely electronic, there is no mechanical typing impact and no noise.

The physical shape of the projected BASIC typewriter is flat, and its light weight permits it to be carried in a briefcase. It can operate on a line cord or rechargeable batteries. Extensible legs at the rear permit it to be set at any angle so as to slope the keyboard to suit the operator. Its thinness allows it to be used at standard desk height without the "drop area" necessary for most typewriters. Because there is no carriage slam or mechanical hammering, the unit does not "walk" on the desk as do standard typewriters.

The keyboard, a design proposed by the Max Planck Institute, divides the keys into right and left hand areas for touch typists. The space between the keyboards straightens the arms and reduces fatigue. The main reason for staggering the rows of keys on conventional typewriters is to permit the mechanical type arms to pass by each other. But since this unit has no mechanical connections from the keys to the typing head, the key rows need not be staggered. There is little perceptible motion or sound when the keys are touched which at first gives an eerie feeling, but which is quickly accommodated.

Various modules can be added to the BASIC typewriter. An acoustic coupler can hook up the typewriter to feed or be driven by a computer. A cassette tape drive can store information and

ABCDEFGHIJ
KLMNOPQRST
UVWXYZ
"#$%&'+><!
1234567890

Texas Instruments' Silent 700 data terminal uses a dot matrix on a monolithic solid state printhead to provide silent electronic printing. The five by seven matrix is composed of thirty-five heating elements which form the characters on heat sensitive paper. Other devices for dot matrix printing use paper that is sensitive to electric current or electrostatic ink jets.

The projected BASIC typewriter has, among its advantages, a keyboard proposed by the Max Planck Institute which divides the keys into right and left hand areas.

in turn drive the typewriter. A paper tape drive can be used. A typical numerical keyboard can be added to convert the typewriter to an outprinting desk calculator. The unit can also be used as terminals for point-to-point communications networks.

To sum up, BASIC appears to meet the communications requirements of today: BASIC, after a minimum of learning, is at least as readable as existing letterforms; BASIC typographic characters have a good balance of white to black to allow greater readability and superior micrographic reduction; BASIC requires the minimum information for communicating its forms; BASIC is machine sensible as well as man sensible; BASIC permits display using current technology—light emitting diodes, liquid cryst.

Excerpted with kind permission from an article in the January/February 1974 issue of *Industrial Design* magazine.