

PATENTS ACT 1977

Agents Ref: P3669GB-JDB/JAC/eq

Description of Invention

5

Title: A method of and apparatus for encoding a character.

10

THIS INVENTION relates to a method of and apparatus for encoding a character.

15

20

25

Chinese characters have a uniqueness of their own and their forms have been arrived at through a long historical development. Chinese characters are essentially square in pattern and are formed by a number of basic strokes. To the present time, essentially three methods are available for locating a Chinese character in a dictionary: (1) according to the Pin Yin system which utilizes a Romanized pronunciation of the character to identify the same; (2) utilizing a Boushu within the character as a guide, a Bushou being a part of the character which is common to a large number of other characters, that is a root or basic portion of the character; and (3) using the actual configuration of the character as a guide by assigning a cypher or code symbol to each of the four corners of the character, the so called four-corner indexing method.

30

35

However, none of the above-mentioned methods is particularly suitable to use with computers and similar machines such as word processors. Thus, the Pin Yin system is not easily adaptable for use with computers and similar systems because of the innumerable homonyms which the system produces. Furthermore, the standard romanised pronunciation of characters used in the Pin Yin system can be difficult to grasp, particularly as far as the

provincial populations in China are concerned because of the numerous dialects which prevail in China. Further, if a Bushou identified within the character is used as a guide, it is then necessary, after having located the Bushou, to count the number of strokes in the character to identify the particular character in the numerous list of characters which incorporate that particular Bushou and it is often the case that a large number of characters have the same number of strokes. Furthermore, in the case of many characters, it is not easy to ascertain the correct Bushou under which the character may be found. The prevailing four-corner index classification also has the disadvantage that the same four symbol digit code is allocated to a large number of characters/patterns and, if a fifth digit is added, there are still a large number of characters which are allocated identical codes under this system so that further identification is needed. The four digit telegraphic code being used by the Chinese post office is connected with the identification of a Bushou but the system has no logical basis and is not easy to memorise so that it is only suitable for output from a machine and not for input thereto. Other systems of coding which have been proposed which allocate code symbols, in particular numbers, to characters and/or patterns require that a certain sequence of writing the strokes of the character be followed or that a certain sequence of selection of the strokes be used which results in many complications and consequently such systems cannot be assimilated or learnt in a very short time period.

It is an objections of the present invention to overcome or at least mitigate the above-mentioned disadvantages.

According to one aspect of the present invention, there is provided a method of displaying a Chinese character, which method comprises encoding the character by a Bushou within the character, allocating to the character a ^{identifying} a

13

first code dependent on the identified Bushou and allocating to the character a second code dependent on the remainder of the character ignoring the identified Bushou, and inputting the resultant code into an apparatus for displaying the character corresponding to the code.

In a second aspect, the present invention provides an apparatus for encoding a Chinese character, which apparatus comprises means for identifying a character as Chinese, means for allocating to the character a first code representing the Bushou and means for allocating to the character a second code representing the rest of the character.

A third aspect of the present invention provides apparatus for ~~producing~~^{displaying} a character comprising means for inputting to the apparatus codes obtained using apparatus in accordance with the second aspect, means for identifying the code input by the inputting means as being associated with a particular character, and means for supplying the identified character to display means.

For a better understanding of the present invention, and to show how the same may be put into effect, reference will now be made, by way of example, to the accompanying drawings, in which:

FIGURE 1 illustrates chematically the Finger signs for each of the different stroke patterns used in encoding Chinese characters using a method in accordance with the invention;

FIGURE 2 illustrates schematically one arrangement for a keyboard for enabling codes produced using a method in accordance with the invention to be input to, for example, a computer, a microprocessor or possibly a word processor; and

FIGURE 3 shows in schematic block diagram form apparatus embodying the invention.

Various examples of methods of encoding characters in accordance with the invention will now be described, by way of example, with reference to Figures 1 and 2 and to Tables 1 to 9, which Tables, for the convenience of the reader, have been grouped together at the end of this description.

In a method in accordance with the invention for encoding Chinese characters, ten main stroke patterns are identified and each main stroke pattern is allocated a unique numeral code. Table 1 shows the numeral codes allocated to each of the ten main stroke patterns together with the subordinate patterns associated with the main stroke pattern. Figure 1 shows a table of finger signs giving the finger sign for each of the main and subordinate stroke patterns identified for Chinese characters.

In order to encode a Chinese character the character is divided into two parts, a main root or "Bushou" portion and a remaining portion or remainder. A two-digit numeral code is allocated to the Bushou whilst a four-digit numeral code is allocated to the remainder

as will be described below.

5 The main root or Bushou of a chinese character must
occupy at least two corners of the character, Chinese
character being generally of 'square' or rectangular
shape. In this regard, it may be helpful to imagine the
character to be encoded as being placed in an imaginary
10 rectangular frame. In order to locate the Bushou, the
'sides' of the character are considered in turn starting
from the top side and then proceeding in an anticlockwise
direction. Once the Bushou has been identified one of
the codes shown in Table 1 is allocated to each stroke
pattern of the Bushou, to provide a two -digit code in
which, generally, the first or left hand digit is the
15 code symbol for the stroke pattern of the left hand
side/or top side of the Bushou and the second or right
hand digit is the code for the stroke pattern of the
right hand side or bottom side of the character as shown.

20 Each of the Chinese characters has, as mentioned
above, only one Bushou. Thus, for example if the Chinese
character  is considered, then the Bushou should
be determined to be  (the part of the character
which occupies the top left and right hand corners of the
imaginary rectangle). As indicated above, the Bushou
must occupy at least two corners of the character so that
25 as far as the character  is
concerned,  should be taken as the Bushou and in
the character  ,  should be taken as the
30 Bushou. Tables 2a to 2h show the 2-digit codes allocated
to each of the Bushous recognised in this system. Where
one stroke or stroke pattern occupies two corners, then
the second or right hand digit of the code is taken to be
zero. Thus, for example, the two - digit code for the
35 Bushou  mentioned above is 80 and the two-digit code
for the Bushou  mentioned above is 00.

Where the Bushou occupies three corners, it is treated as if two corners are occupied. Thus, the Bushou  is given the code 33 although three corners are occupied. Where the Bushou occupies all four corners i.e. the Bushou comprises a closed frame. The Bushou will be allocated the appropriate code for that frame; for example, if the frame is a square the Bushou which will be allocated one of the codes 60 to 68 (see Tables 2 e and 2f). Where the Bushou occupies all four corners, then the remainder of the character is the miniature character inside the Bushou.

Ten main Bushous are identified in Tables 2a to 2h by the two digit-codes 00, 10, 20.... to 90. Each main Bushou is associated with a number of other or subsidiary Bushous which, in addition to the stroke pattern forming the main part of the Bushou, include other strokes or stroke patterns. The second digit of the code for the subsidiary Bushous is determined as described above so that, for example, the Bushou  is allocated the two digit code 08 the top part  being a 'head' represented by the code or cypher 0 and the bottom part  being a (Chinese) eight represented by the code or cypher 8).

As shown in Tables 2a to 2h, no Bushou is associated with any one of the two digit codes 05, 15, 36, 43, 46, 52, 53, 54, 55, 56, 57, 65, 69, 70, 76, 83, 86, 89, 91, 92, 93, 94. These codes may therefore be used for other characters as will be described below or could be allocated to new Bushous which may be developed.

After the two - digit code has been allocated to the Bushou, the Bushou is then ignored or "axed" and the remainder of the character is considered. As mentioned above, the system allocates a four-digit code to the remainder of the character. Thus, again imagining an

imaginary rectangle around the remainder of the character, the code for each 'corner' of the remainder of the character is determined in accordance with the codes given in Table 1 proceeding in the following order top left hand corner, top right hand corner, bottom left hand corner, and then bottom right hand corner, to form a four-digit code.

When one stroke or stroke pattern occupies two corners of the remainder of the character, the second of the two corners (in the order mentioned above) is allocated the code 0. For instance, in the character 法 the Bushou is 𠂔 and the two digit code for the Bushou is '31'. The code of the remainder 去 is '4073'. Thus, as the upper pattern 十 occupies two corners, the second corner must be allocated the code '0', '7' and '3' being the cyphers of the two lower corners of the pattern 𠂔. Therefore the code for the complete character is '314073'. When the remainder of the character occupies four corners or comprises only one stroke or stroke pattern, the code for that remainder will include three '0's. For instance, the code for 和 is '296000' and the code of 太 is '483000'.

If the Bushou occupies all four corners of the character, the remainder is taken from the miniature character inside this four-corner frame. For instance, the code for the character 𠂔 is '661010'. When the remainder occupies four corners, the third and fourth digits of the code are used to represent the frame and the fifth and sixth digits are determined by the lower two corners of the miniature character inside the frame. For instance, the character 𠂔 is allocated the code '497790'. When the Bushou occupies three corners, it is treated as if it occupies two corners. For instance, the code of the character 进 is '335500'. Thus, although

the Bushou 乚 occupies three corners, it is still given the code '33' and treated as having occupied only two corners. Therefore the last four digits are to be taken from the two 'thrust' symbols or stroke patterns 5 丩 which have the code '5' the fifth and sixth digits will necessarily be zero.

For certain characters, there are specific ways of coding, apart from the regular code. For instance, the character 在, should be coded '568050' according to the system but can mistakenly be coded '805000'. Both codes are found in character lists of the system, but for easy memorizing, a new code '000365' has been created to mean '365 days' and for any duplicated characters in a Chinese text, a simple code has been invented. For instance in 人人, 往往, the second 人 or 往 or any other character to be duplicated or repeated can be automatically obtained by just pressing '99' or '0' on the keyboard shown in Figure 2.

A six-digit numeral may be applicable to two or more than two Chinese characters. Accordingly the following procedure is adopted to avoid duplication.

- 25
- i) The most commonly used character is given preference and is
- 30
- 35

allocated the original code.

ii) The next most frequently used character is given a larger code number by +1 (adding 1) to the last digit of the code to produce a new code, and so on.

iii) When the simplified form of a character and the regular script character could be determined to have the same code number, the simplified characters are always given a number larger by +1 than the regular script characters

iv) If accidentally the +1 step was missed out while writing a message and the completed document has already been sent out, it will not be difficult for the receiver to find out the mistake because the system is arranged such that characters which have sequential codes are in no way related in meaning. When the user becomes familiar with the +1 step by repeated use the user will automatically go directly to the correct code number. For instance, the character 程 is arranged to have the code '296010' which code can also be derived for the character 香 ; however by having repeatedly used the character 香 the user will go directly to that correct code for that character, namely '296011' instead of arriving at the correct code by first keying in '296010' and then adding one to the last digit.

It is possible that a user of the system may select an incorrect but conventionally used Bushou. For instance, in this system 人 should be taken as the Bushou for the character 拿 and the character should be given the code '801050,' but, because of tradition, 手 may be taken as the Bushou, producing the wrong code '258060'. Accordingly, both code numbers have been included thus saving time in finding the correct code number and consequently accelerating the speed of input of the codes to a machine such as a computer or wordprocessor using the keyboard shown in Figure 2.

Using a method embodying the invention the number of Bushous has been reduced from the more than 240 generally seen in average Chinese

dictionaries to 160.

Table 3 illustrates the one hundred and ten Bushous required in the Xiandai Hanyu Cidian System which are not required in the present system. Amongst those not required are twelve Bushous which are really modern Bushous created for the simplified form of Chinese characters. Elimination of Bushous is chiefly effected by ignoring any Bushous which control less than six characters.

The present method has created six new Bushous not previously recognised,

namely 六, 二, 1, 11, ' , ' .

Furthermore in addition to the stroke patterns conventionally considered to occupy four corners of a character

namely 口, 17, 15, 13, two new such stroke patterns are added,

namely 1, 1. Table 4 illustrates examples of characters including the newly-recognised Bushous and stroke patterns. With the stroke patterns classified as the four-corner stroke patterns it is not necessary to know the sequence in which the strokes were written. Rather, a user of the system has just to determine the code based on the pattern of the character. The system is specifically useful and very manageable for any foreigners whose native language is not Chinese.

The codes for more than 9,000 characters have already been determined using the method without having to have recourse to duplicating any code number and still providing the user with the convenience of creating new characters or encoding new numbers for his private use using the free two-digit codes in the Tables 2a to 2h of Bushou codes and determining the remainder of the code in the manner described above. Although the new codes are

determined by the user, they can still be logical and easy to remember. For instance, the rare surname 佟 can be encoded by the code 912723.

5 The method can also be used for groups of characters. Thus, the first two digits of the six-digit code are determined from the stroke pattern(s) of the upper two corners of the first character in the group whilst the second two digits are determined by the stroke
10 pattern(s) of the lower two corners of the last character in the group or phrase. The last two digits of the six digit code are predetermined as a serial number for groups or phrase. Thus, for example, 000001 is the code for 广州, 001401 is the code
15 for 意见一致, the serial number for groups or phrases thus being 01.

 Characters such as punctuation marks, numbers, (in particular arabic numerals), and mathematical symbols
20 are necessarily a special case and the codes given above cannot normally be applied.

 As will be noted from Tables 2a to 2h, no Bushou is allocated to the two-digit 'Bushou' code 15.
25 Accordingly, this two-digit code is reserved for characters such as punctuation marks, numerals and mathematical symbols. Table 5a shows various punctuation marks allocated respective codes in the ranges 155000 to 155099 and 155900 to 155999. It should be noted that the
30 code for any form of opening quotation mark invariably terminates with the digit 1 while the code for any form of closing quotation marks invariably terminates with a 9. A '5' is usually the terminating digit of a code for an intermediate symbol while '0' is usually the last digit
35 of commonly used punctuation marks.

 Numerals are treated into two different ways. If

the numeral is a Chinese numeral, then a single digit code preceded by five zeros is allocated to each numeral so that, for example, the code allocated to 九 is '000009'. For 'mathematical' or 'arabic' numerals the Bushou code 15 is used. Thus, as shown in Table 5b, codes 150000 to 150009 are allocated to the arabic numerals 0 to 9. Similarly the codes 151111 to 151234 and 159001 to 159401 are allocated to various mathematical symbols. Thus, as shown in Table 5b, the mathematical symbol "+" is represented by the code 150071, the Chinese pronunciation of '71' being similar to that of the character 加 while the symbol "x" is represented by the code 150015, the number '15' being homonymous with 'x' in Chinese. The symbol "÷" is denoted by the code 150045, '45', being equated in Chinese pronunciation to the pronunciation of the character 除. The symbols "<" and ">" are represented by the codes "150049" and "150051", respectively, 49 being less than 50 and 51 being greater than 50.

Similarly the symbol "≠" is represented by the code 150089 while the symbol "=" is represented by the code 150099.

25

A forward (or right hand) arrow is denoted by the code 150201, a downward arrow by the code 150301 a backward (or left hand) arrow by the code 150209 and an upward arrow by the code 150309. The code 150101 represents @ or a unit, the code 150109 represents & or a total, the code 150144 represents a $\sqrt{\quad}$ (144 being a square), the code 151234 represents a 1234, (1234 being consecutive numbers), the code 151111 represents the mathematical symbol '.' for "because" the code 152222 represents the mathematical symbol '.' for "therefore", the code 159009 represents the operation of squaring, and the code 159027 represents the operation of cubing a

35

number (9 and 27 being a square and a cube of three respectively).

5 Various of the other two digit codes listed in
Tables 2a to 2h and not allocated to Bushous are also
used to signify special forms of characters. Thus, the
'Bushou' two-digit code 56 is allocated to regular
scripts and variants of Chinese characters which also do
not include a Bushou while the two-digit code 57 is used
10 for simplified forms of Chinese characters which cannot
be distinguished by a Bushou, and for rarely seen
characters the two-digit code 54 is used. Where a
character is a Bushou as well as being a character in its
own right, the two-digit code 55 is used. This manner of
15 coding has a specific intended meaning. Thus, being
right in the centre between '1' and '9', '5' is utilized
in the first position of the two-digit code to indicate
special kinds of characters for the code in the second
place: '4' pronounced (si) in Chinese is a homonym
20 of 死 to mean 'dead or that the character is not
much used'; '7' (qi) rhymes with 簡 (jian) in
certain dialects to mean 'simple' and '6' is a homonym of
俗 (lu) meaning 'common'.

25 The use of the coding method is not restricted to
Chinese characters and may be used with other types of
characters. Thus, the two-digit or Bushou code '53' is
adopted to represent Japanese Kanji, '53' as pronounced
in Japanese being the first two sounds in the Japanese
30 phrase 御参考 meaning 'for
your reference'.

The method is also applicable to writing Japanese
Kanas and the codes allocated are shown in Table 6. The
35 two-digit code 65 is allocated for Katakana.

Using the Japanese syllabary as the base the codes

for Kanas of the first 7 row must take '1' as the last digit and the last digit of the codes for those of the fifth row must be '5'. The third digit is invariably '0', while the fourth digit is '0' for Seion, '1' for Dakuon and '2' for Handakuon. The fifth digit indicates 1 column. Therefore for the first 7 column the fifth digit is '0' and for the tenth 7 column, the fifth digit '9'. digit. For Hiragana the two-digit code or indicator is '69' instead of '65' (see Table 6a).

The coding system is also suitable for encoding letters of other languages including English, German, French, Spanish, Russian and Greek; however some modification of the code is necessary. Thus, to encode such characters, the numeral '3' is used to indicate the pinnacle of a triangular pattern, '8' to represent a round corner, '6' to represent an incomplete round corner and '0' to indicate an unoccupied or void corner. Other types of strokes or stroke patterns follow the codes given in Tables 1. Table 7a illustrates the modifications for European alphabets. As shown in Table 7a specialised letters in Russian, German etc and capital letters are denoted by adding a specified number to the last digit of the six letter code for the basic English language letter. Thus, English language capital letters which are the same in shape as the corresponding lower case letters are obtained by adding 1 to the code for the corresponding lower case letter and so on as illustrated in Table 7a. Table 7b illustrates the codes for the English, alphabet plus the modifications for the French, German and Spanish alphabets. The numbers in brackets illustrate the number added to the basic code. Thus, the code for 'â' is 051881 (the code for a) plus 1 equals 051882. Tables 8 and 9 illustrate the six letter codes for the Greek and Russian alphabets, respectively.

The coding system may also be applicable to Korean Hangul characters.

5 The codes obtained using a method embodying the invention can be used for confidential communications, since for any character a six-digit code or number is used. The six-digit codes may be scrambled in accordance with a prearranged procedure agreed by a sender and a receiver of a message. Thus, for example, the positions
10 of one to three of the digits in a derived code can be exchanged.

15 The present method uses only digits to formulate the codes and recourse to letters of, for example, the English alphabet or other signs is not necessary. Further, although the system uses a six-digit code, it is not necessary to use (or enter if the code is being input to a computer or word processor) the first zero. Thus, in actual use, it may frequently only be necessary to use
20 3 or 4 digits. Also, since six-digit codes from 000001 to 999999 are available, many codes are unused and are therefore free for the user to allocate special symbols or other characters.

25 Figure 2 illustrates schematically a keyboard 1 for use in entering codes determined in accordance with the present invention into a machine such as a word processor or computer (not shown), which word processor or computer includes a character generator for generating the
30 character corresponding to the code input via the keyboard. The characters so generated may be displayed on a VDU and/or printed-out using a printer in accordance with normal practice. As shown in Figure 2 the keyboard 1 includes a 0-9 numeric key pad 2 for entering the
35 codes. As mentioned above, the user need only use the numeric keypad and need not be concerned with letters or other characters. As illustrated in Figure 2, the keys of the keyboard each have a number of functions

determined by operation of one of eight function keys 3. Thus, the keys of the numeric keypad 2 and are other key 4 can be used to input directly the letters of the English alphabet by depressing the appropriate function key. Selection between the first and second halves of the English alphabet is determined by use of the function key 3.

The function key 3" at the upper left of the keyboard may be used to facilitate entry of the numeric codes so that it is, for example, not necessary to press the six keys of the numeric keyboard 2 required to enter, ie, the code '150001' for '1' or the code '150009' for '9' but merely to press the key 3". Therefore the average speed of input of codes using the keyboard 2 is not, for every character, determined by the requirement to press six keys because, quite frequently, it is sufficient to press only three or four keys to enter a six-digit code. If the keyboard is provided with a key for entering three successive '0's by one key stroke, then there are many cases where the code for a character can be entered using two less keystrokes (for example the code for the character 和).

Keying in numerals is not at all difficult, the distances between all of the numeric keys being small. Therefore one can operate the keyboard with one hand, leaving the other hand free to do something else, for instance, to hold the draft or a drink, so reducing the amount of wrist movement which cannot but mean an increase in input speed. Furthermore, when a user is familiar with the numeric keys, the user will automatically place his finger on the key the moment one determines the numeral which is no less than a conditioned reflex. No second thinking by the brain will be required and as a bonus it will again necessarily enhance the input speed.

The codes of more than 9,000 characters have already been compiled by the Applicant using the method without having to duplicate any code and while still providing the user with the convenience of creating new characters or encoding new numbers for his own private use (which will of course be prearranged with the receiver of the code) i.e. to utilize the unused two-digit codes or numerals or Bushou codes (Table 2). Although the new codes are determined by the user, they can still be logical and easy to remember. For instance, one can code the rare surname 佟 by assigning '912723' to it.

Figure 3 illustrates in block diagrammatic form one example of apparatus for carrying out the method described above.

As shown in Figure 3, the apparatus comprises a central processing unit (CPU) 5 in the form of a computer having internal memory storage for containing a program for carrying out the method. The CPU5 is connected to a first external memory 6 storing look-up tables containing the information given in Tables 1 to 9 below and a second external memory 7 for storing codes determined by the CPU 5. The CPU 5 is also connected to one or more output devices, for example a visual display unit 8 and a printer 9. The CPU5 may additionally be arranged to supply encoded data to a remote location, for example by means of a modem connection 10 to another CPU.

Scanning means 11, for example a video camera or a solid state

camera incorporating a matrix of photosensitive devices, for example charge-coupled devices, is provided for scanning a character to provide signals representing the character to the CPU so that the CPU can then identify the character to be encoded. Alternatively, a digitizing tablet (not shown) may be connected to the CPU 5 so that a person wishing to encode a message merely draws or traces each character on the digitizing tablet to input signals representing the character to the CPU 5.

Once a character has been input to the internal memory of the CPU, the CPU first determines the group to which the character belongs, for example the group of Chinese characters, the group of punctuation marks, a Roman alphabet etc to determine the first code. The various group, except the group of Chinese characters, are, of course, shown in Tables 5a, 5b, 6, 6a, 7b, 8 and 9 below. If the character belongs to one of the groups shown in tables 5a to 9 below, then the appropriate second code can be determined by using the look-up tables stored in the memory 6. If, however, the character is a Chinese character, then, because of the large number of Chinese characters, it is not practical to store all the Chinese characters which may be used in memory. Moreover, the amount of time which would be needed to locate the correct character in memory to determine the code associated with that character would be prohibitively long. Accordingly, the Bushous shown in Tables 2a to 2h and the code rules given in Table 1 are stored in the memory 6. Of course, as far as the European alphabets are concerned, the code rules shown in Table 1 as modified by Table 7 below may also be stored in memory either to enable a particular character within such a European alphabet to be determined or as a safety check to ensure that the correct code has been allocated when the character is identified in memory.

When the CPU 5 identifies the scanned character as being a Chinese character, the CPU first considers the four corners of the characters starting at the top and working in an anti-clockwise direction round the character to identify at least part of a Bushou within the character. The CPU 5 then compares the first - identified stroke pattern occupying two corners of the character with each of the ten basic Bushou patterns indicated by the single digit codes 0 to 9 in tables 2a to 2h and stored in memory to determine to which group of Bushous the Bushou within the character belongs. Next, having identified the group of Bushous, the CPU 5

compares the character with each of the Bushous stored in the memory 6 in that group to determine which Bushou in that group is closest to the Bushou within the character. Having identified the Bushou for the character, the CPU 5 then ignores or "axes" the Bushou possibly be means of a software mask and considers the remainder of the character. The CPU is programmed to carry out the rules laid down in Table 1 below so that the stroke (or stroke patterns) in each corner of the remainder of the character is (are) compared with the stroke patterns of Table 1 stored in the memory 2 to determine the code to be allocated to the stroke pattern in each corner of the remainder of the character. The corners of the remainder of the character are, of course, considered in the order given above, that is top left hand corner, top right hand corner, bottom left hand corner and then bottom right hand corner. The CPU 5 thus determines the four digit code for the remainder of the characters. The comparison performed by the CPU in allocating the first and second codes may for example a bit by bit comparison or any other suitable known form of computer comparison technique may be used.

The complete code comprising in sequence the first and second codes determined for the character can then be stored in the second memory and possibly also displayed on the VDU 8 and/or printed out by the printer 9. As mentioned above, the code may also be output via a suitable communications link such as a Modem connection 10 to a remote terminal or other CPU. Of course, the output from the CPU 5 may be supplied directly to a word processor provided with its own visual display unit and/or printer, the word processor having, of course, a character generator for regenerating the characters in response to input thereto of the codes by the CPU 5 or

manually by a user.

As described, a method and/or apparatus embodying the invention, as far as Chinese characters are concerned, adopts a process of "axing" or ignoring an identified Bushou to determine part of the code for a character. The method is such that each Chinese character is associated with only one particular Bushou and a corresponding code is assigned to that Bushou whereas the remainder of the character, after "axing" or ignoring the Bushou, is assigned another code. The combination of the first and second code makes up the complete code. As described above, a special

5

10

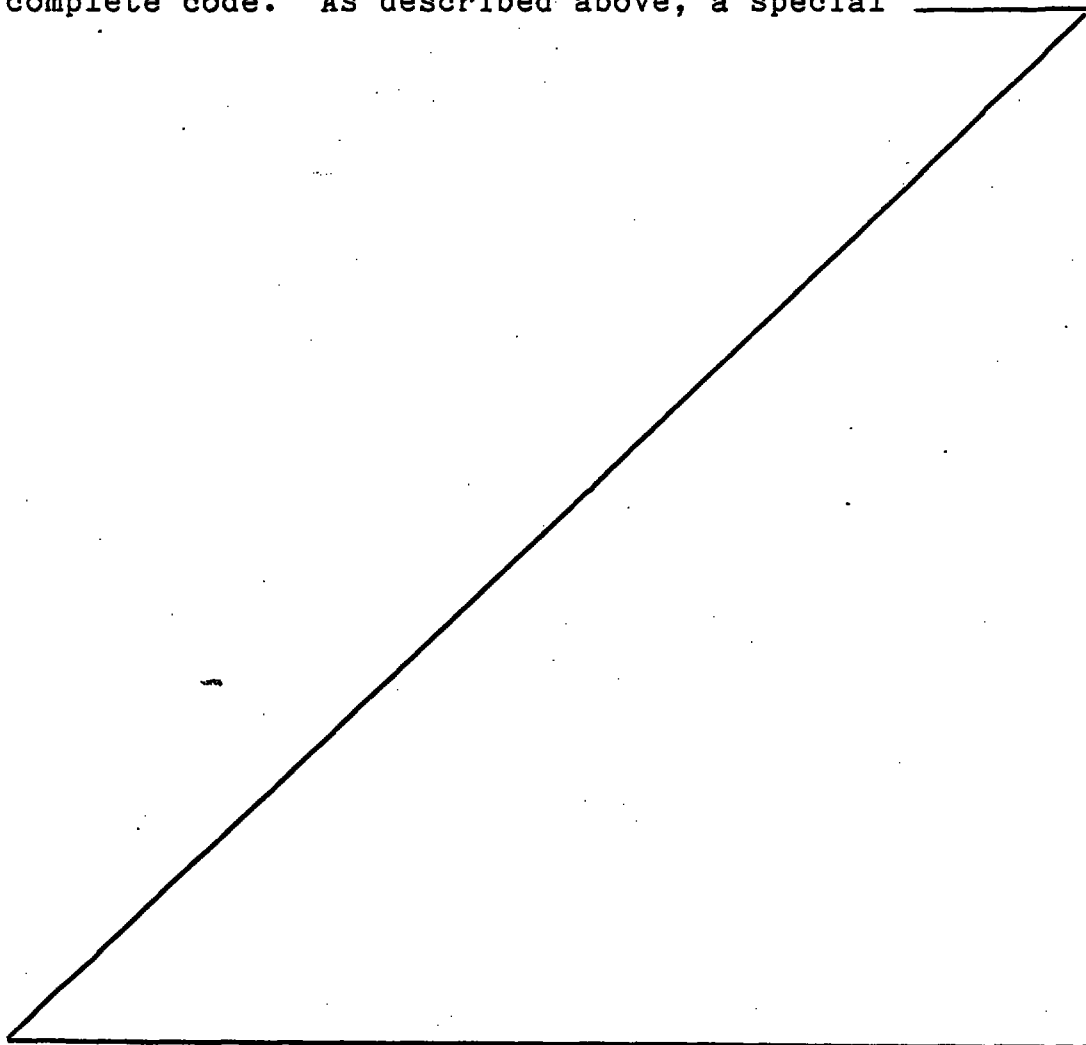
15

20

25

30

35



keyboard may be provide consisting of single cypher keys from 0 to 9 and a small number of additional keys to enable all the necessary functions to be performed.

- 5 It is not necessary when using a method and/or apparatus described above to know the sequence of strokes used in constructing a character. Rather, the code to be allocated to a character can be determined immediately by identifying the pattern of the character. Also, as described above, a simple numeric input keyboard is provided so that, when a person is
- 10 familiar with the coding system, the codes can be input to a computer or wordprocessor using the keyboard so that both Chinese nationals and foreigners can write Chinese characters without any difficulty on a specially designed word processor incorporating such a keyboard.

15

20

25

30

35

TABLE 1

The correspondence of Cipher and Stroke/Pattern

Cipher	Name	Main Pattern	Subordinate Pattern
9	(Chinese) Small	小	小 个 十 〃 〃
8	(Chinese) Eight	八	人 八 八 八 八 八 八 〃
7	Corner	「 ㄟ 」	ㄟ ㄟ ㄟ ㄟ ㄟ ㄟ ㄟ
6	Square	□ □	□ □ □
5	Thrust	丰	丰 丰 丰 丰 丰 丰 丰 丰
4	Cross	十 x	十 十 十 x 十 十
3	Dot or right slide	· \	· ~
2	Vertical or left slide	/	/
1	Horizontal	- (-)	/ 丿 丶
0	Head	丷 (丷)	either of the unoccupied or void lower corners

TABLE 2a

First two-digit code Bushou (not as a character) Wherefrom

	0	一
00	一 *	一
01	广 *	广
01	鹿	鹿
01	立	立
02	广	广
02	方	方
03	亦	亦
04	文, 文'	文, 文
04	辛	辛
05	(blank)	

First two-digit code Bushou (not as a character) Wherefrom

06	言	言
07	亡	亡
07	衣, 衣'	衣, 衣
08	六, 六'	六, 六
09	麻	麻
	1	一
10	一	一
10	弓	弓
10	雨	雨
11	二 *	二
11	瓦	瓦

24

TABLE 2b

First two-digit code	Bushou (not as a character)	Wherefrom
12	工	工 (+1)
12	夕	夕
12	羽, 羽'	羽, 羽
13	王	王 (+2)
14	耳, 耳'	耳, 耳
14	子, 子'	子, 子
15	(blank)	
16	石	石
16	西, 西'	西
17	酉	酉 (+1)
17	刀	刀
18	頁, 頁'	頁, 頁

First two-digit code	Bushou (not as a character)	Wherefrom
19	示	示
2	ノ	ノ
20	**	可
20	片	片
20	隹	隹
21	乚*	乚
21	彡**	彡
21	鳥†, 鳥'	鳥
21	魚†, 魚'	魚
22	亻**	亻
22	月*	月
22	虍*	虍
22	豸	豸

TABLE 2c

First two-digit code	Bushou (not as a character)	Wherefrom	First two-digit code	Bushou (not as a character)	Wherefrom
23	亅*	亅 ⁽⁺¹⁾	27	夕*	夕
23	厶	厶	27	夕	夕
23	鳥, 鳥	鳥	27	夕*	夕
23	魚, 魚	魚	27	毛	毛
			27	角	角
24	彳	彳 ⁽⁺²⁾			
24	舟	舟	28	齒, 齒*	齒, 齒
24	身	身	28	欠	欠
25	手	手	29	禾, 禾*	禾, 禾
25	牛, 牛*	牛, 牛	29	糸*	糸
25	鬼	鬼	29	爪, 爪*	爪, 爪
26	白, 白	白	3		
26	舌	舌	30	一*	一
			30	彡*	彡
27	山, 山	山			
27	月*	月	31	彡*	彡

Ø Note that in this Bushou the numbering is from right to left.

TABLE 2d

First two-digit code	Bushou (not as a character)	Wherefrom	First two-digit code	Bushou (not as a character)	Wherefrom
32	ㄣ	ㄣ	4	×	+
32	戶	戶	40	×	×
33	ㄣ, ㄣ	ㄣ, ㄣ	40	+	+
33	心	心	41	土, 土	土, 土
33	...	...	41	士	士
34	寸	寸	42	力	力
35	戈	戈	42	巾, 巾	巾, 巾
36	(blank)		42	彡	彡
37	ㄣ	ㄣ	43	(blank)	
37	門	門	44	女, 女	女, 女
37	一	一	44	++, ++, ++	++, ++, ++
38	穴	穴	45	革	革
39	ㄣ	ㄣ	45	车, 车	车, 车

Ø Note that in this Bushou the numbering is from bottom to top.

Ø Note that in this Bushou the numbering is from right to left.

TABLE 2e

First two-digit code	Bushou (not as a character)	Wherefrom
46	(blank)	
47	𠂇*	𠂇
48	大	大
48	走	走
49	木, 𣎵*	木, 𣎵

5 丰 丰

50	扌*	扌
50	車	車
51	虫	虫
51	主*	主
52	(blank)	

First two-digit code	Bushou (not as a character)	Wherefrom
53	Japanese Kanji	
54	Rare characters	
55	Bushou	
56	⊙ Unclassified Regular Script	
57	⊙ Unclassified Simplified Form	
58	𠂇*	𠂇
59	来, 𠂇*	来, 𠂇
6	□ □	
60	□ □	
61	日 日	(+1)

⊙ Both include a few characters which clearly possess a Bushou but are squeezed out from the congested areas.

TABLE 2f

First two-digit code	Bushou (not as a character)	Wherefrom
61	見 ⁺	見
61	見	見
62	目	目 (+2)
62	目 [*]	目 (+2)
63	黑	黑
64	田	田
65	(blank)	
66	口 [*]	口
67	日	日 (+1)
68	貝	貝
68	足	足
69	(blank)	

First two-digit code	Bushou (not as a character)	Wherefrom
7	厂	厂
70	(blank)	
71	厂	厂
71	皿	皿
71	马 ⁺ , 马 ⁺	马
71	见 ⁺ , 见	见
72	尸	尸
72	尸 [*]	尸
72	户 ⁺	户
72	月	月
73	尸 [*]	尸 (+1)
73	馬, 馬	馬
73	馬 [*]	馬
74	又, 又 ⁺	又, 又
74	爻	爻

TABLE 2E

First two-digit code	Bushou (not as a character)	Wherefrom
75	戊	戊
76	(blank)	
77	門	門
77	骨, 骨	骨
77	白, 𠂔	白, 𠂔
77	鼠	鼠
77	冂*, 冂*	冂, 冂
77	几*	几
78	風, 风*	風, 风 (+1)
78	鬥	鬥 (+1)
78	贝*, 貝	贝
79	匚*	匚 (+2)
8	八	
80	八, 八	八, 八

First two-digit code	Bushou (not as a character)	Wherefrom
80	人, 人*	人, 入
80	一*	一
81	气	气
81	金	金
82	丷*	丷 (+2)
83	(blank)	
84	父	父
84	女*	女
85	羊, 羊*, 𦍋*	羊, 羊, 𦍋
86	(blank)	
87	𠂔*	𠂔

TABLE 2h

First two-digit code	Bushou (not as a character)	Wherefrom	First two-digit code	Bushou (not as a character)	Wherefrom
87	缶, 缶	缶	92	(blank)	
87	食, 食†	食, 食	93	(blank)	
88	矢	矢	94	(blank)	
88	矢*	矢	95	类*	类
89	(blank)		96	尚, 尚†	尚, 尚
9		小	97	然*	然
90	小, 小†, 小†	小, 小, 小	98	火, 火†	火, 火
90	小*	小	99	米, 米†	米, 米
90	水, 水†	水, 水			
91	(blank)				

- Notes:
- 1) Bushous marked with * are Non-characters.
 - 2) Bushous marked with † have in their respective simplified or aberrant forms the same two-digit code.
 - 3) Bushous marked with ‡ are simplified or aberrant form of other Character Bushous or Non-character Bushous also in the list.

41

TABLE 3

Bushous Abolished

卵流車非鼃鼃隶韭是香音首隔高芦黃帶鼎黍鼓鼻禽

光肉自色齊齊艮辰麥麥赤東豆豕县鹵鹵里谷采龜龜

尺玉甘龍龍业申由瓜玄民疋皮夬矛戔亞亞臣東而至

三土巳己屮少么《天韦韋夬廿廿支比无牙支氏斗聿

丨ノ、一乙卜厂匕儿几クマ又干兀九弋夕丸斗ヨ



TABLE 4

' 六 '	衰	086073
	产	087120
	產	087121
	彦	087122
	商	087760
' 二 '	二	110000
	三	111000
	亏	112002
	云	112073
	正	112100
' 1 '	至	112340
	修	232720
	條	232790
' E1 '	條	232780
	與	774080
	興	775080
' 〃 '	並	821010
	前	821022
	普	821060
	关	821080
	曾	826060

TABLE 5a

Codes for 155000 - 155099
Punctuation Marks 155900 - 155999

,	comma	155000	(	parenthesis (opening)	155061
.	full stop	155001	)	parenthesis (closing)	155069
•	Chinese full stop	155008			
!	exclamation mark	155009	[	square bracket (opening)	155071
?	question mark	155010	]	square bracket (closing)	155079
:	colon	155011			
;	semicolon	155019	«	book name quote (opening)	155081
			»	book name quote (closing)	155089
'	single quotation mark (opening)	155021			
'	single quotation mark (closing)	155029	-	hyphen	155091
			—	dash	155092
“	double quotation mark (opening)	155031	/	slant	155095
”	double quotation mark (closing)	155039			
			†	dagger	155098
「	single Chinese quotation mark (opening)	155041	*	asterisk	155099
」	single Chinese quotation mark (closing)	155049			
			o	circle	155900
、	enumerations divider	155045	•	dot	155990
『	double Chinese quotation mark (opening)	155051			
』	double Chinese quotation mark (closing)	155059			

TABLE 5b

How to code 150000 - 150109
Numerical Symbols 151111 - 151234
159001 - 159401

0	150000	½	150050
1	150001	¼	150025
2	150002	¾	150075
3	150003	{	150081
4	150004	}	150091
5	150005	%	150100
6	150006	@	150101
7	150007	&	150109
8	150008	√	150144
9	150009	#	151234
+	150071	∴	151111
-	150041	∴	152222
x	150015	²	159009
÷	150045	³	159027
<	150049	'	159001
>	150051	"	159010
≠	150089	°	159100
=	150099	¥	159200
~	150199	£	159300
→	150201	\$	159400
←	150209	¢	159401
↓	150301		
↑	150309		

TABLE 6

Japanese Katakana

650001	ア	650022	シ	650042	ニ	650062	ミ
		650122	シ			650063	ム
650002	イ			650043	ヌ		
		650023	ス			650064	メ
650003	ウ	650123	ス	650044	ネ		
						650065	モ
650004	エ	650024	セ	650045	ノ		
		650124	セ			650071	ヤ
650005	オ			650051	ハ		
		650025	ソ	650151	ハ	650073	ユ
		650125	ソ	650251	ハ		
650011	カ					650075	ヨ
650111	カ	650031	タ	650052	ヒ		
		650131	タ	650152	ヒ	650081	ラ
650012	キ			650252	ヒ		
650112	キ	650032	チ			650082	リ
		650132	チ	650053	フ		
650013	ク			650153	フ	650083	ル
650113	ク	650033	ツ	650253	フ		
		650133	ツ			650084	レ
650014	ケ			650054	ヘ		
650114	ケ	650034	テ	650154	ヘ	650085	ロ
		650134	テ	650254	ヘ		
650015	コ					650091	ワ
650115	コ	650035	ト	650055	ホ		
		650135	ト	650155	ホ	650095	ヲ
				650255	ホ		
650021	サ					659999	ン
650121	サ	650041	ナ	650061	マ	651000	ー

TABLE 6a

(69 indicates Hiragana and the rest is same as Katakana whose indicator is '65')

690001 あ	690022 し	690042 に	690062 み
	690122 し		
690002 い	690023 す	690043 ぬ	690063 む
	690123 す		
690003 う	690024 せ	690044 ね	690064 め
	690124 せ		
690004 え	690025 そ	690045 の	690065 も
	690125 そ		
690005 お	690031 た	690051 は	690071 や
		690151 は	
690011 か	690131 た	690251 は	690073 ゆ
690111 か			
690012 き	690032 ち	690052 ひ	690075 よ
690112 き	690132 ち	690152 ひ	690081 ら
		690252 ひ	
690013 く	690033 つ	690053 ふ	690082 り
690113 く	690133 つ	690153 ふ	690083 る
		690253 ふ	
690014 け	690034 て	690054 へ	690084 れ
690114 け	690134 て	690154 へ	690085 ろ
		690254 へ	
690015 こ	690035 と	690055 ほ	690091 わ
690115 こ	690135 と	690155 ほ	690097 を
		690255 ほ	
690021 さ	690041 な	690061 ま	699999 ん
690121 さ			

97

TABLE 7a

1.	Pinnacle of a triangular pattern	= 3
2.	Round Corners	= 8
3.	Incomplete round corner	= 6
4.	Void Corner	= 0
5.	Part of the Capital Letters of the English language	= Small Letters +1 (in English are ZWOC SXV)
6.	English Capital Letters used as Small Letters in Russian	= English Capital Letters -1 (in Russian are BHKMT)
7.	English Small Letters used as Russian Capital Letter	= English Small Letter +1 (in Russian y)
8.	English Small Letters with additional signs in German	= English Small Letters +1 (ä, ü) (ö +2)
9.	English Capital Letters with additional signs in German	= English Capital Letters +1 (Ä, Ü) (Ö +3)
10.	English Small Letters with additional signs in French	= English Small Letters +10 (ç é ê è ë)
11.	English Small Letters with additional signs in Spanish	= English Small Letters +2 (ñ)

	s	058118	L	052071	é	058897(+11)
English	t	054061	M	053322	ê	058898(+12)
a	u	052280	N	053223	ë	058899(+13)
b	v	053230	O	058001(+1)		
c	w	053233	P	057821(+1)		
d	x	054000	Q	058003	German	
e	y	053200	R	057823	ä	051882(+1)
f	z	051771	S	058119(+1)	ö	058002(+2)
g			T	051020	ü	052281(+1)
h	A	053023	U	052288		
i	B	057878	V	053231(+1)	Ä	053024(+1)
j	C	058687(+1)	W	053234(+1)	Ö	058004(+3)
k	D	057870	X	054001(+1)	Ü	052289(+1)
l	E	057171	Y	053220		
m	F	057120	Z	051772(+1)		
n	G	058688			Spanish	
o	H	052200			ñ	052604(+2)
p	I	052000	French			
q	J	050206	Ë	058696(+10)		
r	K	052203	è	058896(+10)		

TABLE 8

How to code Greek Letters

α 058281	ρ 058020	Α 053023	Ρ 057821 (+1)
β 058828	σ 056180	Β 057878	Σ 057777
γ 056200	τ 058668	Γ 057621 (+1)	Τ 051020
δ 056080	υ 057720	Δ 053033	Υ 056620
ε 056161	φ 052260	Ε 057171	Φ 055001 (+1)
ζ 056168	χ 055000	Ζ 051772 (+1)	Χ 054001 (+1)
η 056802	ψ 054000	Η 052200	Ψ 059041 (+1)
θ 058080	ω 059040	Θ 058081 (+1)	Ω 056611
ι 052060	α 056688	Ι 052000	
κ 052204 (+1)		Κ 052203	
λ 053021 (+1)		Λ 053025 (+2)	
μ 052201		Μ 053322	
ν 052230		Ν 053223	
ξ 052162		Ξ 051010	
ο 058000		Ο 058001 (+1)	
π 051022		Π 057723 (+1)	

TABLE 9

How to code Russian letters

а 056886	р 057820	А 053023	Р 057821 (+1)
б 056280	с 058686	Б 057178	С 058687 (+1)
в 057877 (-1)	т 051019 (-1)	В 057878	Т 051020
г 057620	у 053200	Г 057621 (+1)	У 053201 (+1)
д 057773	ф 055000	Д 057774 (+1)	Ф 055001 (+1)
е 058886	х 054000	Е 057171	Х 054001 (+1)
ж 053224 (+1)	ц 052273	Ж 053225 (+2)	Ц 052274 (+1)
з 056868	ч 052205 (+5)	З 056869 (+1)	Ч 052206 (+6)
и 052332	ш 052277	И 052333 (+1)	Ш 052278 (+1)
я 056032	щ 052275 (+2)	Я 052334 (+2)	Щ 052276 (+3)
к 052202 (-1)	ы 052270	К 052203	Ы 052271 (+1)
л 057724 (+2)	ь 052078	Л 057725 (+3)	Ь 052079 (+1)
м 053321 (-1)	э 056860	М 053322	Э 056861 (+1)
н 052199 (-1)	ю 052801 (+1)	Н 052200	Ю 052802 (+2)
о 058000	я 058722	О 058001 (+1)	Я 058723 (+1)
п 057722		П 057723 (+1)	



51

CLAIMS.

1. A method of displaying a Chinese character, which method comprises encoding the character by identifying a Bushou within the character, allocating to the character a first code dependent on the identified Bushou and allocating to the ~~second~~ character a second code dependent on the remainder of the character ignoring the identified Bushou, and inputting the resultant code into an apparatus for displaying the character corresponding to the code.
2. A method according to Claim 1, wherein the first code allocated to the identified Bushou is determined by the configuration of the identified Bushou.
3. A method according to Claim 2, wherein the step of identifying the Bushou comprises considering the four corners of the character and identifying a stroke, a stroke pattern or patterns occupying two or more corners of the Bushou.
4. A method according to Claim 3, further comprising considering the four corners in an anti-clockwise order.
5. A method according to Claim 3 or 4, wherein the four corners are considered starting at the top of the character.
6. A method according to Claim 3, 4 or 5, wherein the step of allocating a first code to the Bushou comprises identifying a first stroke or stroke pattern occupying at least two corners of the Bushou dependent on the configuration of the stroke or stroke pattern.
7. A method according to Claim 6, wherein the step of allocating the first code includes determining the configuration of a second stroke or stroke pattern occupying two other corners of the character.
8. A method according to Claim 6, wherein a designated first code is allocated to the Bushou when the Bushou occupies all four corners.
9. A method according to any one of Claims 2 to 8, wherein the step of allocating the first code comprises identifying



a Bushou within the character as being one of the Bushous given in Table 2 and allocating the code given in Table 2 for that Bushou.

10. A method according to any one of Claims 2 to 9, wherein the step of allocating the first code comprises using the codes identified in Table 1 for strokes and stroke patterns.

11. A method according to any one of Claims 1 to 10 wherein the step of allocating the second code comprises considering each corner of the remainder of the character and allocating a code symbol to each corner of the character dependent on the stroke pattern, if any, occupying that corner.

12. A method according to Claim 11, wherein the step of allocating the second code comprises using Table 1 to determine the code symbol or numeral for each corner of the remainder of the character.

13. A method according to any one of Claims 1 to 12, wherein, if two or more characters are allocated the same first and second codes, the allocated first and second codes are used for the more common character and the second code for the less common character is modified in a predetermined receiver.

14. A method according to Claim 1, wherein the step of allocating the first code comprises allocating first and second code symbols to first and second, respectively, upper corners of the Bushou, and the step of allocating the second code comprises allocating first and second code symbols to first and second, respectively, lower corners of the remainder of the character, the second code also containing a symbol or symbols indicating that the first and second codes in combination represent a character.

15. A method according to Claim 13, wherein the step of modifying the second code comprises changing a last symbol of the second code.



16. A method according to any preceding claim, wherein the first code is a two-digit code and the second code is a four-digit code, the first code being placed to the left of the second code.

17. A method of displaying a character substantially as hereinbefore described with reference to Tables 1 to 9 and Figures 1 to 3 of the accompanying drawings .

18. Apparatus for encoding a Chinese character, which apparatus comprises means for identifying a character as Chinese, means for allocating to the character a first code representing the Bushou and means for allocating to the character a second code representing the rest of the character.

19. Apparatus according to Claim 18, wherein the means for identifying a character comprising scanning means.

20. Apparatus according to Claim 19, wherein the scanning means comprises a video camera.

21. Apparatus according to Claim 18, wherein the means for identifying a character comprises a digitizing tablet for identifying a character drawn or traced thereon by a user.

22. Apparatus according to any one of Claims 18 to 21, wherein the means for allocating the first code comprises means for comparing the identified character with data stored in a memory of the apparatus to determine the Bushou.

23. Apparatus according to Claim 22, wherein the memory stores data representing Bushous identified in Chinese characters and the means for allocating the first code is arranged to compare the identified character with the data stored in memory representing the Bushous to identify any particular Bushou within the character, when the character has initially been identified as being a Chinese character.



-39-

24. Apparatus according to any one of Claims 18 to 23, wherein the means for allocating the second code comprises means for considering the four corners of the remainder of the character in turn and, for each corner, comparing any stroke or stroke pattern occupying the corner with predetermined stroke patterns stored in memory to determine the second code, each predetermined stroke pattern stored in memory being associated with a particular code symbol.

25. Apparatus according to any one of Claims 18 to 24, wherein the means for allocating the first and second codes comprise a microprocessor or computer.

26. Apparatus for encoding a character, substantially as hereinbefore as described with reference to, and as illustrated in, the accompanying drawings.

27. Apparatus for displaying a character, comprising means for inputting to the apparatus codes obtained using apparatus in accordance with any one of claims 18 to 26, means for identifying the code input by the inputting means as being associated with a particular character, and means for supplying the identified character to display means.

28. Apparatus according to Claim 27, wherein the inputting means comprises a keyboard.

29. Apparatus according to Claim 28, wherein the keyboard is substantially as shown in Figure 2 of the accompanying drawings.



-40-

30. Apparatus according to any one of Claims 27 to 29, wherein the display means comprises a visual display unit and/or printer.
