SPECIFICATION

Ideographic coding

The invention relates to ideographic coding suitable for providing representations, in a machine-processable and reproducable form of ideographic characters or words of written Chinese, Chinese-related, or any other ideographic languages.

Languages where written characters have a unique single character-to-word correspondence are obviously much more difficult for machine processing than languages using a small, finite number of characters, such as letters of an alphabet, from which words are represented by sequential groups of the characters. Effectively, the "alphabet" is coextensive with the vocabulary.

The Chinese, Japanese and Korean languages typify these problems. For example, the standard telegraphic code for Chinese characters is at present limited to 10,000 words and identifies these by a numbered character list. Encoding and decoding is clearly a time-consuming process. Also, a Chinese typewriter usually has a flat bed of word-character printing slugs in a 70 × 35 matrix. These are individually located by sight using a slug selector, which, apart from obvious reading difficulties, is limited to only some 2,450 characters and has a normal speed of only some 10 characters per minute. Others, for example some Japanese typewriters or printing machines, have used an abbreviated alphabet of relatively simple ideographs which have to be combined to represent a normal single-character word, but in an unnatural way.

The essential identity of such word-characters is, of course, the relationships of its component strokes. It is an interesting fact that, despite often startling dialect differences of pronunciation, the written characters are generally immediately recognisable in much the same way as Arabic numerals are immediately recognised irrespective of individual language variations as to their word designations.

It is an object of this invention to further facilitate machine processing of ideographic characters of Chinese and related languages, especially given the position of Chinese as the most widely used of primary languages.

One prior attempt to provide encoding of such languages is based upon the old Chinese typewriters, effectively by forming the bed on a rotatable cylinder equipped with a cursor and operative on naturally formed rather than reversed characters. Encoding is then by digital drum-rotational and cursor-position signals. This is claimed to be substantially faster than the flat-bed typewriter but is nonetheless rather inflexible and dependent for speed on the operator's knowledge of the actual layout of the characters on the drum, speed of drum rotation (bearing in mind that the operator has to see and recognise the markings on the moving drum), and limitation on vocabulary size to fit onto a physically managable drum.

Direct reading of hand-written, or otherwise produced, original written information, presents similar if not greater difficulties than for Western alphabet style script. Thus problems arise not only from different type faces or writing styles and sizes, but particularly from the scanning resolution required, for example ideographic characters could have as many as 30 or more component strokes. Some success has been achieved on a "during writing" basis and has shown that the making of only a few, say six, strokes may suffice in most cases for character identification. However, this is not of any immediate general applicability except perhaps, that the work done by Susumo Kuno has established that a 16 × 16 point matrix is adequate to represent at least all of the Chinese characters in the Chinese telegraphic code. That may be employed herein if desired instead of some other size, for example 20 × 20 points that should accommodate virtually the entire Chinese vocabulary of some 50,000 characters.

According to this invention there is provided ideographic character selection apparatus comprising a first means having a plurality of individually-operable signal generating devices each corresponding to a different one of a predetermined plurality of character identifier components and each operative to produce a unique signal; second means responsive to a combination of said unique signals resulting from several device operations of the first means for causing display of a corresponding unique set of possible characters; and third means operable for selection of any particular one character from the displayed set.

The phrase "character identifier components" is intended to mean identifiers or criteria that are parts from which ideographic characters are built-up or by which ideographic characters can be successively represented, such as phonetics, component strokes, etc.

We have developed a highly advantageous scheme of encoding based upon any convenient character classification system or analysis that provides a finite and keyboard-accommodatable number of character component criteria that will ensure identification, if not of a single character itself, then of a manageable group or file of possible actual characters, usually for simultaneous display. From such a display subsequent selection of the desired one character of such a group or file can be readily achieved herein. Clearly such a proposal is applicable to, for example phonetics, Western A-toZ alphabet or Japanese or Korean restricted alphabets, or a specifically developed analysis basis, and is inherently capable of utilising the stroke-type identification of Susumo Kuno either on a first *n* stroke basis or simply on overall, or even part-content, of specific types of stroke.

The last mentioned stroke identification basis could, of course, both be available on the same keyboard or touch-pad type entry unit (which the term keyboard type unit is intended to encompass), perhaps with a

mode switch, key or pad device thereon. The part or overall contents approach might even allow entry of written material by an operator ignorant of or poorly acquainted with the language concerned.

It is important that the character component criteria be of a sufficiently basic nature to result in a total number thereof that is readily handled by an operator, i.e. incorporated into a keyboard of a manageable size and number of keys. These character components will necessitate plural selection of criteria to identify the desired manageable groups or files of actual characters. Allowing for selection of pluralities of such criteria for group or file identification and display, facilitates operation of embodiments of the invention, i.e. multiple key operations prior to an automatic decoding and addressing step that locates the group or file concerned.

The direct use of larger character parts or building blocks such as the Chinese radicals of which there are some 200 or more tends to be impractical as they are too numerous to be accommodated on a workable keyboard. It is noted, however, that the preferred input character criteria hereof will usually allow for the identification of such larger character parts or building blocks within the total number of actual input component criteria required for each character identification, and may assist data base organisation. One aim of the invention, of course, is to avoid two distinct selection and display stages, e.g. one to provide a selection from possible character parts and a subsequent one to provide for selection from possible actual characters including those parts, as that would be slow and possibly require greater operator skill due to its complexity. However, it would be possible to arrange that the first few input character component criteria did identify a character part such as a radical and that might be displayed automatically as a check while further input component criteria were selected.

As viewed within the overall framework of the problem of machine handling of ideographic characters, whether the machine be an electronic computer or other system, or of another type, such as mechanical, there are three quite distinct components, namely input to the system; representation for internal system handling of the characters; and output from the system. Of these components, representation is readily dealt with, for example by numbers directly related to the characters on a one-to-one basis, and satisfactory output can be obtained using, for example, a dot-matrix display unit or a printer. The principal impact of the invention is probably best viewed as being on the input component, typically using a wholly keyboard information presenter, or a device of a "voice recognition type" analogous to a vocoder, for character group or file identification purposes and a visual display for selection from within a character group, either as a conventional electronic visual display unit accessing a computer type store or as a photographic film or hologram reader and display. However, this invention may also be used to advantage in the representation and output components if the confidence factor on identification is high enough, as the amount of information needed for group identity, thus saving transmission, gating and storage hardware or utilisation thereof, and the identification within the groups may be done at the output stage.

The last-mentioned facility may well be of most significance in a typewriter or printer where output quickly follows input, but could also be of value in other circumstances, say where fast "course" input was required or became necessary, e.g. due to high temporary demand for input facilities in a multipurpose or multi-user system, and subsequent "fine" character identification could be left until later, or where the information either only requires checking on output of the "course" input or is known to be unlikely to produce ambiguities not resolvable easily on output.

More specific description of manners of practical implementation of the invention will now be given, by way of example, with reference to the accompanying drawings, in which:

Figure 1 is a block diagram;

Figure 2 shows one layout of input device, specifically of a keyboard or analogous type of entry device;

Figure 3 shows a vocoder input arrangement;

Figures 4 and *5* show keyboard layouts for Chinese and Japanese characters, respectively;

Figures 6 and *7* show a keyboard layout for English language letters;

Figures 8, 9 and *10* show successive visual displays during one file selection; and

Figure 11 is a further block diagram of an embodiment of the invention.

In Figures 1 and 2, a keyboard or touch-pad operated entry device is indicated generally at 10 and has character component entry keys or pads 12, digit entry keys or pads 14 for arabic numerals and control keys or pads or switches 16 for purposes to be described. Shown as a separate block 20, but normally located within the entry device 10 is a coding circuitry operable according to key or pad 12 in order to give a corresponding coded binary digit word of a sufficient number of digit positions to give a unique such word for each key or pad 12. For this purpose the circuitry 20 may comprise a parallel-to-binary converter having individual key or pad input lines 22 energised on operation of the keys or pads and binary word output lines 24, one per digit thereof.

Each of the possible values represented by outputs 24 is shown as identifying and giving access to a different part of a store 30, if necessary via store access drive or addressing means 32 as may be required, for example, for a store of the so-called floppy disc, microfilm or microfiche, hologram, cartridge disc, or tape types, or a core or semiconductor store, magnetic bubble memory, charge-coupled devices, etc. This store 30 comprises the dictionary of the language concerned and contains at least one separate entry for every word-character of the vocabulary of that language as required for the particular purpose of the overall system. These entries may be actual representations of the word-character if of the microfilm or halogen type, but, as shown, otherwise need be no more than unique identifiers for some other store 30A of the physical forms of the characters themselves, say as a dot matrix library, microfilm, etc. form.

The store entries representing characters are indicated at 36 and are organised into files 38 each of characters that are the same, similar or confusable within the resolution capability of the keys or pads 12 after the permitted number *n* of operations thereof in the input of a particular desired character. More will be said about the number *n* later but, at present, regarding Figure 1, it is noted that *n*-1 further store addressing or accessing stages are shown at 40 with output lines 42 that identify further portions of the store parts 28 until a single file 38 is reached.

These files 38 need not, and normally will not, be of the same size in terms of their number of entries so we show a store readout device 44 which may be a comparator, counter or shift register or any other suitable device, which will respond to file start tags in the store when a read out is required to provide all of the entries of a particular file regardless of its length.

The use of such file tags could allow a certain flexibility of operation in that, if further tags were normally associated with multifile store portions within a said part 28, the encountering of the file tag before the normally permitted *n* input key or tab operations could be automatically indicated to show that further input operations were not required. Except in the case of characters having fewer than *n* key or tab components, that would, however, appear to sacrifice some of the intrinsic discrimination potential of the system, but there might be some files that were incapable of rational further discriminations once a certain level of input operations was reached.

There is, of course, another, and perhaps preferred manner of store organisation and system operation, which is to defer any application of store accessing until all of the input operations up to *n* are complete for the character concerned and the character entries can then be accessed via store access control 32.

At all events, and however done, once a file is located by the input operations and read-out, Figure 1 shows a visual display unit 50 for presenting the characters thereof for final selection by the operator either on a separate set of keys or tabs or, and as shown, via the numeral entry key or tabs 14 or some other control device 16. Once that has been done via lines 52 and circuitry 54 the output of the latter can be used as an "in file" identifier tag, say to control the counter or shift register in order directly to obtain the character from some other similar remote store and/or simply to specify that the relevant displayed character should be selected for display, processing, transmission, printing or storing either as a whole or in number coded form.

Reverting to the normal maximum input operation number *n*, it is preferred that such be so related to the characters by their analysis into component characteristics as to allow, at least for the vast majority of the vocabulary, adequate identification, within that number, of a file containing no more individual characters than can be simultaneously displayed and quickly identified by the operator. The aim is for *n* not to exceed 6 and for files normally to contain less than, say, 15 or 20 characters each, but possibly as many as 72 if the file is extended to include rarely used words. Perhaps most conveniently, the display screen will show file characters in one or more line and in discrete areas labelled 1 to whatever maximum is required and thus allow selection by the numeral keys or tabs 14.

For example, these objectives can be met using phonetics for the character components to be entered, say the standard Chinese language phonetics or the Katakana or Hiragana Japanese alphabets, or the Korean alphabet.

The markings on the keyboard or input device 10 are arranged to suit the particular language and the skill of the operator.

For example, Chinese language keyboard can be marked (apart from numerals 0 1 2 -- 9) in either

(i) Chinese phonetics symbols (ㄅㄆㄇㄈ ㄚㄛㄜㄞ) or
(ii) English alphabets (A B C --- Z) or
(iii) some combination of (i) and (ii) or
(iv) any other phonetics markings familiar to the operator.

For Japanese, the markings can be (apart from numerals):
(i) Katakana alphabets, (ㄎ ㄎ ㄟ) or
(ii) Hiragana alphabets, (ㄎ ㄎ ㄟ) or
(iii) English alphabets, (A B C -- Z) or
(iv) some combination of the above three alphabets, or
(v) any other markings familiar to the operator.

For Korean language, the markings can be (apart from numerals);
(i) Korean alphabets (ㅂ ㅍ ㅃ) or
(ii) English alphabets (A B C -- Z) or
(iii) some combination of the two, or
(iv) any other markings familiar to the operator.

In the general case of an ideographic language, the markings (apart from numerals) can be:
(i) a phonetics marking of that particular language, or
(ii) English alphabets (A B C -- Z), or
(iii) some other markings familiar to the operator, or
(iv) any combination of the above.

If a voice recognition device is used as input device 10, then it is designed to recognise the phonetics symbols as listed above, appropriate to the language and the operator. The main components of the voice recognition device are shown in Figure 3, where 71 represents a voice-to-electrical signal generator, 72 represents comparator search and accessing circuitry for a memory 73 of recognition pattern information, and 74 represents a coded output.

Clearly, the system is also applicable to any other identification system that may be specially designed or in widespread use or known to the operators of a particular system.

One matter which has not been mentioned this far is that of dealing with the possibility that a particular input sequence may, perhaps because of lack of familiarity or complete understanding by the operator, produce a coding that does not correspond exactly to a stored file. Then, comparison circuitry may be used to select those actual file headings that are closest, say with only one characteristic component difference and/or via a table of likeliest variants, followed by successive display of those files.

Control circuitry for the various system components is indicated only generally at 60 and may take any convenient logic and function circuit form.

A more detailed system implementation will be described later with reference to Figure 11.

Specific examples of keyboards are now considered. There are some 56 Chinese phonetics symbols now in common use and a keyboard containing 56 appropriately marked keys will make up a keyboard suitable for use as input device 10, see Figure 4. Figure 5 shows the equivalent for Japanese phonetics. (The details of the control function keys will be more closely specified in the next example). These Chinese and Japanese phonetics symbols systems are capable of specifying the entire vocabulary of the corresponding dictionary.

A voice recognition device designed to recognise the 56 symbols read out by the operator can also be used in place of the keyboard.

The next example is an alternative approach utilising a keyboard marked in English alphabets and both such a keyboard as an input device and operation of a particular word processing system will be described.

The keyboard may be marked as shown in Figures 6 and 7.

In order to input to the system the character 布, meaning "cloth" and pronounced in Mandarin as "BU", the operator will use "B" and "U" on the keyboard (and optionally an "end of file name" control key). This is decoded by device 20 and a search for a file with file name BU is initiated.

The file BU may be positioned in store between the files BO and CA, with all the files being arranged in alphabetical order. The contents of the file BU will be brought out of store 30 (or 30A) and displayed together with identifying numerals, preferably on the bottom one or two lines of the display unit 50. The display device 50 may well be capable of displaying up to 10 lines each of 10 characters, each formed by a matrix of up to 24 × 24 dots, see Figure 8 for the whole BU file including the desired character at location 5 and other confusable or more complex characters having or including the Mandarin pronunciation BU.

The display format can be across the screen from left to right as in Figure 8 or can be from top to bottom if so desired.

The operator will inspect the bottom one or two lines of file characters and type in the numeral 5 to indicate his choice of the selected character. Preferably, the character, 布, displayed under the numeral 5 is then moved into the upper area of the display and the one or two bottom "file lines" erased.

The display will now look like Figure 9 if the entered character was the first on the display. Note that a cursor marker is also displayed to show the position on the screen where the next character should go.

The operator now repeats the above operations of keying in the pronounciation of the next character criteria and choosing the correct character from the displayed file, for display at the position indicated by the cursor, the cursor then advanced by one character position.

When the non-file display area screen is filled up with text characters, a "scrolling" operation is useful to shift the text lines upwards losing the top line and leaving empty the line above the area used for file display. The top line will, of course, be stored in memory for subsequent output or other processing.

When an entire script has been inputted, it is convenient if the operator can recall any part of the whole of the text and edit it using the control function keys, which may thus include: delete character, insert character, delete a line of characters, insert a line of characters, cursor movement control (up, down, left, right etc.), scroll, merge two lines, create new file, close file, recall text file, delete file, etc, as marked on them.

An "alphanumeric" key is provided because, even for texts written in an ideographic language, there is often need to use English letters and Arabic numerals, e.g. in scientific and technical texts. Technical terms are often left in English and specifications left in Arabic numerals. As an example, it may be desired to insert the transistor type BU 109. The operator will thus want to put in the English letters B and U. In order to distinguish this from the character attribute (phonetics) entry BU, the "alphanumeric" key is used. To input the English letters B and U, the operator will first press the key marked "alphanumeric" and then proceed to enter in B and U. These letters will then appear on the screen and, on completion, the operator may press the "alphanumeric" key again or otherwise release it, to revert back to the normal encoding input mode.

Such a system has many advantages, for example:

1. The speed of response is very fast. In a system utilising floppy discs as dictionary memory, the contents of a file can be displayed for inspection within half a second of the last file-name-defining key operation. A practical memory of 0.25M Byte capacity floppy discs has a 0.39 second respond time. A solid state memory would have a response time of micro-seconds.
2. The dictionary can be expanded at will by adding files into the memory.

3. The file structure is such that, with English alphabet keyboard and a maximum of 6 character criteria key operations, a dictionary of more than 10,000 characters is catered for using a total of some 415 files with file names starting from A, AI, AN, ... to ZUN, ZUO etc. The file names and/or contents will vary depending on the language and the dialect pronounciation used, for example Chinese represented by the dialects Mandarin or Cantonese, or some other language, e.g. Korean.

4. The keyboard marking is simple. Using the English alphabet, a total of some 58 keys (alphabet letters, numerals, punctuation and other markings plus control function keys) is sufficient to identify any one of 10,000 Chinese characters and allows a simple and practical word processing system. The decoded output 24 from device 10 need be only 7 binary lines wide.

Considering Japanese, the keyboard can be marked in Katagana or Hiragana alphabets, or some combination of the two, in addition to numeral, punctuation and control function keys.

Some 49 alphabet keys will be sufficient for substantially the entire Japanese dictionary with control keys to suit the particular requirements. However, the presence of one particular control key, the "Kanji" key, is discussed in detail.

This is because Japanese text consists of two groups of characters, namely Kanji (Chinese characters), and Hiragana or Katagana alphabets. For example, the Kanji 漢 can be represented by Hiragana as かん, or カン in Katagana. The same two alphabets かん, however, can also appear in groupings in non-Kanji form, say みかん meaning "orange".

Clearly, there is a need to distinguish the two because the latter group みかん will stay as it is throughout whereas the input かん representing the Kanji is to be converted into the charater 漢.

Therefore, a "Kanji" control key is provided to distinguish between the alphabet inputs which are used to pick up a Kanji, and the straight forward alphabet inputs which will eventually be processed and outputted in the same form. In order to input the Kanji 漢, the operator first operates, the "Kanji" key, then either mechanically or otherwise keeping it operated, keys in the character criteria かん (or カン). On release of the Kanji key, a search for the file with file name かん is initiated.

Obviously, the Kanji key may be os designed that the operator can depress it once before and once after the two input key strokes and hence delimit the character criteria grouping concerned from the rest of the text. Indeed the Kanji key may take the form of a foot switch so that the operator can, without using his fingers either depress it before and after the alphabet grouping or keep it depressed through the group's input.

The search for the file かん will be as described previously and its contents (all Kanji including かん) will be displayed on the bottom two lines of the screen, the dictionary lines, see Figure 10. The operator, on inspecting the display, now keys in the numeral 2 to indicate his character choice and the correct Kanji character is displayed in the upper area of the screen as in the previous example.

If, however, the operator wishes to input みかん to mean "orange", he will key the three character criteria in without operating the Kanji key, or switch, and these will be displaced on the screen immediately without any processing.

Suitable editing and word processing operations will be similar to that already described and will not be repeated.

Again, a voice recognition system designed to recognise the alphabets read out by the operator can also be used instead of the keyboard. However, to avoid the need of recognising all Kanjis, some equivalent means of identifying a Kanji grouping from alphabet text must be present, a switch or key or any other device performing the function already explained above will suffice.

Figure 11 shows in more detailed block form, units for carrying out the functions mentioned in the above examples. To avoid cluttering the drawing unnecessarily, contents of the various blocks will be mentioned in a manner that is believed to be readily implemented by a person of reasonable skill and knowledge in the art of digital data processing.

The input device, keyboard 10, together with its control keys 16 (and a Kanji key 11 in case of Japanese) may be of the Hall effect or gold contact types available from, say Honeywell. Key top markings appropriate to various languages are described previously and will not be repeated here. Device 20 can be a normal integrated circuit type encoding device giving representative binary signals for command decode unit 50'.

The signals on lines 24 can be in accordance with standard 8-bit ASC II codes if the keyboard is based on that shown in Figure 7.

The command decode unit is made up of buffers and logic elements as required and performs the following functions:

1. Distinguishing between
   (a) character identifying inputs, e.g. phonetics in Western alphabet form or phonetics markings of a particular ideographic language
   (b) control function commands, e.g. delete character scroll up, execute, etc.
   (c) alphanumeric symbols appearing in a text of ideographic language.
2. Routing these signals to their appropriate destinations. Thus, character identifying inputs are passed onto the interface 70 for subsequent processing by control unit 25. Simple edit commands (e.g. delete character) and alphanumerics are processed and passed to video store and generator 50C. The more complicated commands (e.g. cut, paste) are passed to control unit 25 via interface 70.
3. Organising the contents of the display store and generator 50C. This includes putting the contents of a

dictionary file (obtained from store 30 via control unit 25) into the bottom lines of the screen, moving the operator-selected character to the upper area of the screen, and processing simple editing commands, e.g. re-organise the video store in 50C to remove a character to be deleted, etc.

4. Accepting and carrying out commands from control unit 25. These commands arise when the operator puts in some complicated editing functions, for example, on "scroll down" where control unit 25 signals the command decode unit 50' to remove the last text line from the screen, and move all the remaining lines down by one line position. The control unit 25 will also cause an appropriate line to be fed from the user text store in 30 and passed on to the decode unit 50', which will put this line into the top of the screen and hence complete the scroll down operation.

The command decode unit 50' can be realised by comparator type logic 50A to perform function 1, above, and buffer and logic elements 50B to carry out functions 2, 3 and 4. These components may be of the integrated circuit type, e.g. series 7400 logic circuits.

To assist the command decode unit an interface unit 70 handles all the communication between decode unit 50' and the rest of the system. Typically, the interface unit 70 has 8 data lines for incoming and outgoing data, one busy line to indicate whether the unit is busy and thus cannot accept or send data, one strobe line through which control unit 25 signals that the data on the 8 data lines are valid, and 6 address lines. These address lines are mentioned because a multi-station version of this invention is envisaged whereby several operators, each having their own keyboard and display unit, can be controlled by a single control unit 25. Therefore, the address lines are used to distinguish one station from another.

The interface unit 70 may also accept information directly from the store 30 through the store control device 32 (data path shown dotted). This bypasses the control unit 25 and is used in cases where an appreciable quantity of information is to be passed from store to display without any need for processing by control unit 25, e.g. filling up a blank screen with text previously generated by the user.

The interface unit 70 can be realised by buffer and logic elements as well as line drivers. Line drivers are provided because, in the multi-station configuration, some keyboard and display units may be located at a considerable distance away from the main system, e.g. 100 metres.

The control unit 25 performs the following functions:

1. Accepting keyboard input via units 50' and 70.

2. If the input is a character identifier component e.g. a phonetic, it will decode the input by comparing it with a list of valid inputs and then instruct store control device 32-40 to effect a search in store 30 for the correct dictionary file. The information, if found, is transmited to unit 50' via unit 70. If the input is not valid, an error message will be sent to the display via unit 50' to alert the operator and ask for a repeat input.

3. If the keyboard input received is an edit command and cannot be handled by unit 50', control unit 25 will carry out the function (in conjunction with unit 30 if necessary). As an example: If the operator specifies a "cut" function, the passage "cut out" from the text will be stored temporarily by unit 25, either inside the unit itself or in a specially reserved area in 30. Later on, when the operator wants to "paste" the passage back into some other part of the text (or a completely different text), the passage is fetched and passed back to the display via unit 50'.

4. Organising the printout of text files as specified by the user. When this facility is activated, unit 25 fetches the correct file from 30 and presents it to printer drive 90A. This can be done on the 'handshake' principle whereby 90A signals that it is free to accept information and 25 then sends a predetermined amount of data to 90A which then indicates that it is busy and goes ahead to cause printer 90B to print the text. It will signal 'free' again when the information has all been printed.

5. Outputs to telex and/or telegraph via drive unit 100. Unit 25 organises text output as in the case of the printer. An important difference is that the user may prefer to output standard telegraph code instead of actual character patterns. Unit 25 therefore fetches the standard telegraph code representation of the text and sends it accordingly. The 'handshake' principle may again be used between units 25 and 90A.

6. Outputs to mainframe computers via interface 120. This is similar to printing or telegraphy and the user has a choice of outputting in code or in character patterns.

7. Outputs to photosetter or equivalent via drive unit 130. This is for typesetting purpose and the output mechanism is similar to the above except that the layout format may be more critical.

8. Maintaining a repertoire of error messages and error recovery precedures so that when an invalid command or specification is received, the operator is alerted with a message and provided with a list of procedures to recover from or rectify the error.

9. Testing various other parts of the system immediately after switch on by sending them appropriate data and instructions and checking responses. Failure may be indicated on the screen of the visual display unit.

10. Maintaining system integrity by performing various 'housekeeping' functions, e.g. allocating file numbers to user-generated text files to prevent duplication and confusion, checking storage 30 to see if the user is running out of storage space, and performing checks on user identity (especially in multi-station configurations) so that unauthorised operators are not allowed to inrterfere with other operators' texts.

11. Performing configuration checks, e.g. asking the operator to specify, via the display and keyboard, what type of printer is attached to the system, or what type of computer mainframe is attached, or how many work-stations 80 are attached, etc., so that the correct output protocols are used and all the active work-stations are allowed access to the system.

12. Deciding which transfers from store 30 to other units (70, 90A, 100, 120, 130) do not need processing and can therefore be transferred direct without going through unit 25. On appropriate instructions, to initiate these transfers and check on completion that the transfer is correct.

The dictionary storage device 30 (and 30A & B) may be of floppy disc type (available, for example, from Shugart Associates) and the store control device 32 (and/or 40) may constitute the MOS floppy disc controller FD 1771B from Western Digital.

The file structure 36, 38 and file tag detector 44 are as described in Figure 1.

The store or store part 30B is where user generated texts are stored. These texts are arranged in files 37 identified by the file number (and in some cases file names). Each file may possess additional information in the file header, e.g. originator, date first created, data modified, read/write access control, length etc.

The storage medium used for 30B is ideally of a travelling medium type, e.g. a floppy disc, magnetic tape or cassette, cartridge disc, etc., so that the stored user text can be taken to other systems by physically removing and relocating the storage medium.

The video store and generator can be realised by random access memory, say type 4116 from MOSTEK, together with buffers, logic elements and video drive circuits if the display 50D is of video display type.

A convenient way of organising the display store, electronically, is to store an exact image of the display content in 50D in store 50C so that the information displayed to the operator is in fact a copy of part of the contents of 50C. In this way, changes to be displayed information can be made very rapidly by altering the contents of 50C.

An area is also provided for use as temporary storage by 50B. This area stores, temporarily, the date and instructions on which the contents 50B are being processed and also provides work space for unit 50B to put intermediate results etc.

The construction of the video generator will depend on the type of display used and, typically in the case of a VDU, it may contain two buffers, one of which contains one screen line of information and is used on a current basis to update the screen. The other is being filled in by 50B with the information for the next screen line so that, on completion of updating the current line, the roles of the two buffers are interchanged.

If the display 50D is a standard TV monitor the video drive circuitry will consist of blanking information onto which synchronising signals are superimposed.

The printer drive 90A may be constructed in a similar way to the display drive, i.e. two line buffers where one is used for output while the other is being loaded. Apart from these buffers, logic elements are used to interpret layout commands in the user generated text, e.g. new paragraph, new line, new page, title etc. These cause drive 90A to output to the printer sufficient spaces and new line commands to effect the layout. The printer may be of an electro-static type (from say Versatec) or impact type (say from Wang or Pragma).

The telex/telegraph drive can take one of several forms viz:

1. 80 - 0 - 80 volts or lower voltage types driving telex or telegraph lines direct;
2. driving data communication equipments which in turn drive the telex lines, e.g. British Post Office DCE 3A
3. driving a paper tape device and the paper tape produced can be used to drive a telex/telegraph terminal in the conventional way.

Although the physical construction of telex/telegraph drive 100 follows one of the above forms, its functions are the same in every case, viz:

(i). output the standard telegraph code or character pattern as desired;

(ii). conform and support all timing signals necessary, e.g. time-out signals, answer-back signals in telex, dialling signals in telex, etc.

The mainframe computer interface 120 is constructed in similar fashion to other output devices. The main differences are:

(i) the interface will conform to that used by the mainframe computer, e.g. CCITT V 24 standards or similar;

(ii) the protocol for communication used by the mainframe computer will have to be observed;

(iii) all timing signals required by the computer will have to be supplied and observed.

This unit can be realised by buffers, logic elements and line drivers.

The photosetter drive 130 will be subjected to similar design constraints as the interface 120, and all signals must now conform to what is used by the photosetter or other equivalent typesetting machines. Layout commands are decoded and translated to a form suitable for that particular photosetter. Either standard codes or character bit patterns can be used during output.

Device 60 is the overall control and timing unit for co-ordinating the activities of the various components of the system and includes a timer for issuing timing signals, say every 10 milliseconds, start up and run down signals during system switch-on and switch-off, etc.

CLAIMS

1. Ideographic character selection apparatus comprising a first means having a plurality of individually-operable signal generating devices each corresponding to a different one of a predetermined plurality of character identifier components and each operative to produce a unique signal; second means responsive to a combination of said unique signals resulting from several device operations of the first means for causing

12. Deciding which transfers from store 30 to other units (70, 90A, 100, 120, 130) do not need processing and can therefore be transferred direct without going through unit 25. On appropriate instructions, to initiate these transfers and check on completion that the transfer is correct.

The dictionary storage device 30 (and 30A & B) may be of floppy disc type (available, for example, from Shugart Associates) and the store control device 32 (and/or 40) may constitute the MOS floppy disc controller FD 1771B from Western Digital.

The file structure 36, 38 and file tag detector 44 are as described in Figure 1.

The store or store part 30B is where user generated texts are stored. These texts are arranged in files 37 identified by the file number (and in some cases file names). Each file may possess additional information in the file header, e.g. originator, date first created, data modified, read/write access control, length etc.

The storage medium used for 30B is ideally of a travelling medium type, e.g. a floppy disc, magnetic tape or cassette, cartridge disc, etc., so that the stored user text can be taken to other systems by physically removing and relocating the storage medium.

The video store and generator can be realised by random access memory, say type 4116 from MOSTEK, together with buffers, logic elements and video drive circuits if the display 50D is of video display type.

A convenient way of organising the display store, electronically, is to store an exact image of the display content in 50D in store 50C so that the information displayed to the operator is in fact a copy of part of the contents of 50C. In this way, changes to be displayed information can be made very rapidly by altering the contents of 50C.

An area is also provided for use as temporary storage by 50B. This area stores, temporarily, the date and instructions on which the contents 50B are being processed and also provides work space for unit 50B to put intermediate results etc.

The construction of the video generator will depend on the type of display used and, typically in the case of a VDU, it may contain two buffers, one of which contains one screen line of information and is used on a current basis to update the screen. The other is being filled in by 50B with the information for the next screen line so that, on completion of updating the current line, the roles of the two buffers are interchanged.

If the display 50D is a standard TV monitor the video drive circuitry will consist of blanking information onto which synchronising signals are superimposed.

The printer drive 90A may be constructed in a similar way to the display drive, i.e. two line buffers where one is used for output while the other is being loaded. Apart from these buffers, logic elements are used to interpret layout commands in the user generated text, e.g. new paragraph, new line, new page, title etc. These cause drive 90A to output to the printer sufficient spaces and new line commands to effect the layout. The printer may be of an electro-static type (from say Versatec) or impact type (say from Wang or Pragma).

The telex/telegraph drive can take one of several forms viz:

1. 80 - 0 - 80 volts or lower voltage types driving telex or telegraph lines direct;
2. driving data communication equipments which in turn drive the telex lines, e.g. British Post Office DCE 3A
3. driving a paper tape device and the paper tape produced can be used to drive a telex/telegraph terminal in the conventional way.

Although the physical construction of telex/telegraph drive 100 follows one of the above forms, its functions are the same in every case, viz:

(i). output the standard telegraph code or character pattern as desired;

(ii). conform and support all timing signals necessary, e.g. time-out signals, answer-back signals in telex, dialling signals in telex, etc.

The mainframe computer interface 120 is constructed in similar fashion to other output devices. The main differences are:

(i) the interface will conform to that used by the mainframe computer, e.g. CCITT V 24 standards or similar;

(ii) the protocol for communication used by the mainframe computer will have to be observed;

(iii) all timing signals required by the computer will have to be supplied and observed.

This unit can be realised by buffers, logic elements and line drivers.

The photosetter drive 130 will be subjected to similar design constraints as the interface 120, and all signals must now conform to what is used by the photosetter or other equivalent typesetting machines. Layout commands are decoded and translated to a form suitable for that particular photosetter. Either standard codes or character bit patterns can be used during output.

Device 60 is the overall control and timing unit for co-ordinating the activities of the various components of the system and includes a timer for issuing timing signals, say every 10 milliseconds, start up and run down signals during system switch-on and switch-off, etc.

CLAIMS

1. Ideographic character selection apparatus comprising a first means having a plurality of individually-operable signal generating devices each corresponding to a different one of a predetermined plurality of character identifier components and each operative to produce a unique signal; second means responsive to a combination of said unique signals resulting from several device operations of the first means for causing

display of a corresponding unique set of possible characters; and third means operable for selection of any particular one character from the displayed set.

2. Character selection apparatus according to claim 1, wherein said second means is operative after less than 7 device operations of the first means.

3. Character selection apparatus according to claim 1 or claim 2, wherein at least most of the different possible sets of characters, the sets individually capable of being caused to be displayed by device operations of the first means, each have less than 20 characters.

4. Character selection apparatus according to any preceding claim, wherein the first means includes a keyboard-type unit as hereinbefore defined, with said devices corresponding to character identifier component keys.

5. Character selection apparatus according to claim 4, wherein the keyboard device has further keys and decoding means constituting the third means.

6. Character selection apparatus according to claim 4, or claim 5, wherein the keyboard unit has further keys and decoding means for causing display and controlling manipulation of selected characters of a multi-character message.

7. Character selection apparatus according to any one of the preceding claims, arranged such that each combination of unique signals to which said second means is responsive, and produced by said first means, is to be representative of at least substantially the sound of each character of the corresponding unique set of characters, each actuated signal generating device producing a signal representing a different, phonetic, character identifier component.

8. Character selection apparatus according to claim 7, wherein the second means is responsive to operation of the devices of the first means in the appropriate sequence required to produce the combination of unique signals representative of the desired sound.

9. Character selection apparatus according to any one of the preceding claims, the characters required to be selected being Chinese characters, and symbols capable of representing character identifier components, corresponding to signal generating devices, comprising the Chinese phonetic symbols.

10. Character selection apparatus according to any one of the preceding claims, symbols capable of representing character identifier components, corresponding to signal generating devices, comprising at least one alphabet of a written form of a language.

11. Character selection apparatus according to claim 10, the characters required to be selected being Japanese Kanji characters, and an alphabet being the Japanese Katakana alphabet.

12. Character selection apparatus according to claim 10, or claim 11, the characters required to be selected being Japanese Kanji characters, and an alphabet being the Japanese Hiragana alphabet.

13. Character selection apparatus according to claim 10, the characters required to be selected being Chinese characters, and an alphabet being the Korean alphabet.

14. Character selection apparatus according to any one of claims 11 to 13, an alphabet being of a written form of a language not including the ideographic characters required to be selected by the apparatus.

15. Character selection apparatus according to claim 14, the alphabet being the English alphabet.

16. Character selection apparatus according to claim 15, wherein the first means includes a conventional alpha-numeric keyboard-type unit as hereinbefore defined.

17. Character selection apparatus according to any one of the preceding claims, wherein the first means includes a keyboard-type unit as hereinbefore defined, with said devices corresponding to at most 49 character identifier component keys.

18. Character selection apparatus according to any one of the preceding claims, wherein the first means is arranged to employ selectively one of a plurality of different possible encoding input modes, the appropriate encoding input mode being selected in response to the actuation of means under the control of an operator, and the signal generating devices selectively correspond to the character identifier components of each of the different encoding input modes.

19. Character selection apparatus according to any one of the preceding claims, wherein the successive display of a plurality of unique sets of possible characters is caused by said second means after the production of the plurality of corresponding combinations of unique signals by the first means, and to which combinations of signals said second means is capable of being responsive to cause the display of the plurality of unique sets of characters.

20. Character selection apparatus according to any one of the preceding claims, wherein the means to display the selected unique sets of characters comprises a visual display unit.

21. Character selection apparatus according to any one of the preceding claims, comprising a word-organised binary digital store containing words each unique to one possible character and organised into sets of such words capable of causing said display, and wherein the second means includes decoding means responsive to device operations of the first means to assemble a binary digital address word identifying a said set according to device operations of the first means.

22. Character selection apparatus according to claim 21, comprising further storage means for actual character display data which further storage means is addressable by the binary words of the binary digital store to supply display data to a display unit visible to an operator of the first means.

23. Character selection apparatus according to claim 21, or claim 22, comprising a display unit means whereby the second means is operative to address the store, after device operations of the first means, for