Text Editing System Using On-line Real-time Character Recognition

Hidekazu Terai* and Kazuo Nakata*

ABSTRACT: A new method of Japanese text editing is developed, using a mini-computer and data tablet. This system offers the following three advantages:

(1) The input of Japanese characters (containing "kanji"--Chinese characters--) is performed by means of recognizing the characters handwritten on the data tablet.

(2) The character to be proofread is pointed out by the touch of writing pen on the manuscript (output sheet from monitor printer) put on the tablet.

(3) Functions of proofreading are identified by recognizing the specified symbols written directly over the character.

Owing to these features, the proofreading process has become more natural and convenient compared with other systems.

1.   INTRODUCTION

In this paper we describe a new method of Japanese text editing by a mini-computer and data tablet. The text editing and proofreading include many elements which fully depend on human capability; thus in this case, the computer utility should be a man-machine type. Basic requirements for an input device in proofreading include the following: (1) Capability to input characters--Character recognition.   (2) Capability to specify position of a specific character in the manuscript--Position recognition.   (3) Capability to recognize a function to be processed--Function recognition.

Considering the data input mode, a text editing system using on-line, real-time handwriting character recognition belongs substantially to a man-machine system. Futher, the capability to accept handwritten character allows the proofreader to act as the operator. Because of characteristics of the data input devices, this system can provide the advantage of simply specifing the location information (location of characters and required process such as insertion, correction, or deletion). In this sense, the direct on-line process can provide a high cost-performance ratio for inputting data. Fig. 1 shows the flow of this direct processing.

2. HARDWARE and SOFTWARE SYSTEM DESCRIPTION

Fig. 2 shows the hardware facilities of our text editing system. The allocation of functions preset on the input tablet is shown in Fig. 3. In Fig. 3, part (A) is the function area which generates control commands for the editing system. Part (B) receives the primary text. In on-line text editing systems, normally, access to characters to be corrected is conducted on characters displayed on a CRT by a light pen and key board. In our system, however, the initial copy of manuscript print (print out of monitor printer) is placed on a tablet as in the case of conventional proof-reading, and characters to be corrected are directly accessed by the writing pen. Part (C) is used to receive the writing sheet for input
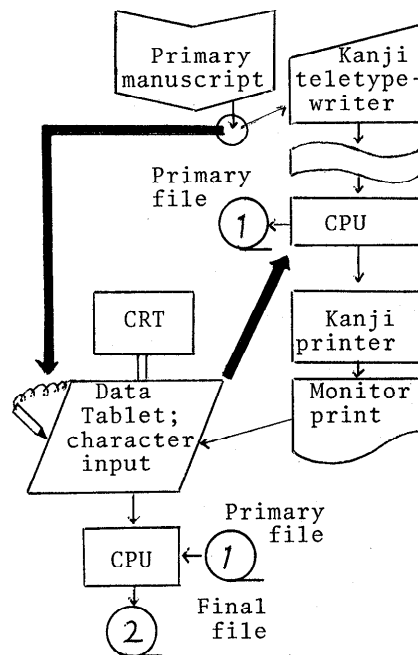

Fig.1 Flow of Text Editing System Using On-line Handwriting Character Recognition

Thick arrow ➡ indicates the flow of on-line direct input of primary manuscript without using a kanji teletypewriter.

Magnetic Drum 65K words → HITAC-10 32K words (16 bits) → Magnetic Disc 2.56M words → DTR, PTR, PTP, LP

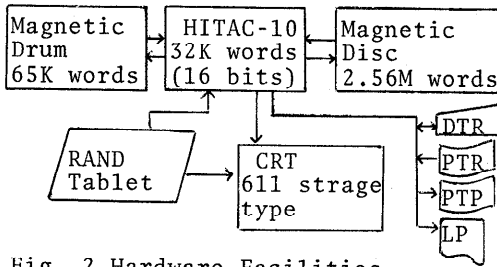RAND Tablet → CRT 611 strage type

Fig. 2 Hardware Facilities
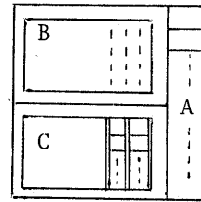of On-line Text
Editing System

Fig. 3 Allocation
of Functions
preset on Tablet

of handwritten characters.

The required commands such as correction, insertion, or deletion could be selected directly from the control command area (A) by the input pen. But it requires two steps of selection--selection of the proofreading function and specification of a character. We developed a much more natural and efficient selection process which allows single direct specification of character location and process function, and which is effected simply by marking a predetermined symbol on the character to be corrected. Then, the system recognizes the specified function and character location, and generates a required command. This method is useful in providing single-step selection and in recording the proofreader's accessing to a manuscript, and is advantageous for a subsequent check. Table 1 shows these symbols.

Our text editing system is operated as described below; (Fig. 1)
[a] A manuscript written in Japanese is initially input from a kanji teletypewriter and recorded in a primary file. In some case, the primary data may be input through on-line recognition mode of handwriting characters. The primary file is organized in a page system (ex. a page is composed of 15 rows and 30 columns), and data are printed out page by page from a monitor printer.
[b] A page of monitor print is placed on part (B) area of the tablet, and data for the page are read out from the file and displayed on a CRT display, by specifing a control area (A) using an input pen.
[c] Location of a character to be edited in the monitor print is indicated by the input pen, and a symbol which specifies the process

to be taken is made at the character location.

[d] A column which includes the character to be corrected is display-
ed on the CRT. A correct character is then handwritten in part (C)
area of the tablet.

[e] The written character is recognized and displayed.

[f] When the character is read correctly, the proofread function
starts when OK sign is input. The result is further displayed is
rewritten and step [f] is repeated.

These steps are repeated until all characters on the page are
proofread and corrected. The file is recorded by using a list
structure shown in Fig. 4.

Table 1 Proofreading Symbol

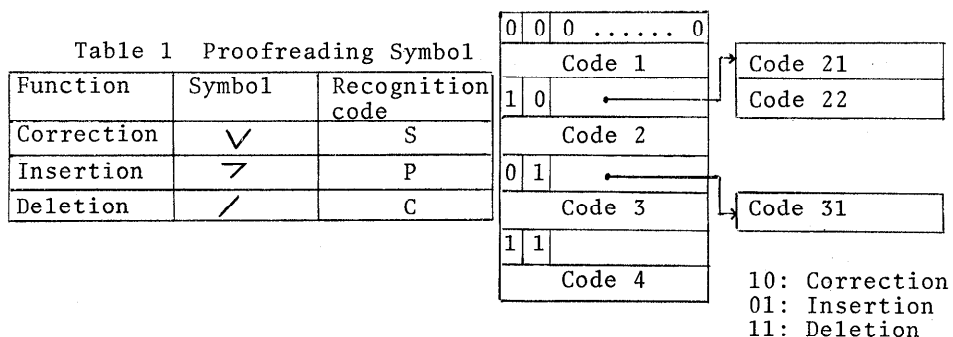| Function	Symbol	Recognition code
Correction	∨	S
Insertion	⌐	P
Deletion	/	C

Fig. 4 Strage of Character Codes

3. CHARACTER INPUT

3.1 Recognition of kanji and kana--Japanese syllabary

As has been reported, our on-line handwriting character recogni-
tion depends on structual analysis in units of strokes[1] Fundamental
stroke comprising kanji and kana are defined as listed in Table 2,
then, strokes are extracted from an unknown character (the strokes
can be easily extracted from up-down motion of the pen during the
course of handwriting), and checked for the registered strokes.
According to the number of total strokes and to certain characteris-
tic strokes (complex strokes) included in the written character, the
object to be recognized is limited to a range given by broad classi-
fication. When the number of possible characters are withen a limited

range, the positional relation of strokes and partial stroke sequence are tested by using a decision tree to obtain a correct character. We prepared a composite recognition program for 881 printed-style kanji (designated by the Ministry of Education) and 48 kana (--katakana--Japanese syllabic letter) characters.

3.2 Recognition of Alphanumerals

The recognizable characters include Arabic numerals and alphabet (capital letters). The fundamental strokes used for recognition of alphanumerals are listed in Table 3. The basic idea of recognition is identical to that of kanji, although the writing sequence is excluded from the objective of testing.

At present, the kanji and kana set is identified from the alphanumeral set by a marking on the character set selection area located in the function area (A) on the tablet.

Table 2   Fundamental Stroke of Kanji and Kana

Code	Shape of Stroke
A	— ／
B	│ │ '
C	／ '
D	＼ '
P	７ ⌐ ７ ⌐
Q	∠ ⌐
R	３ ３
S	∠ ∠ ∟ ＜
T	５
U	３
V	∠

Table 3   Fundamental Strokes of Alphanumerals

Stroke	Characters	Stroke	Characters
— ↗	Omitted	⌐	F,Y
│	Omitted	S	S
／	Omitted	Q	Q
＼		O	O,0
７ ⌐ ７	D,P,R,1,7	５	5
C ⌐	C,E,G,K	6	6
∠ L ＜	E,K,L,4	8	8
∧	A,M,N	9	9
U V	K,M,N,U,V,W,Y	N	N
2 ⌐ Z	R,Z,2	M	M
3	B,3	M	M
J	J	W	W

4.   PROOFREADING PROCESS

Fig. 5 shows an example of the correction mode. A special symbol ∨ is marked on a character in the part (B) area which requires correction. The system responds a display of a character 正 with a square ☐ around the marked character. (See Fig. 5-a) Then, the system waits for entry of the correct character(s), and enters recognition mode. After checking a display of the recognition result, if

the recognition is right, the YES function is depressed by the pen.
(See Fig. 5-b) Then, the file is corrected and correct sentence is
displayed on the CRT. (See Fig. 5-c) In case of insertion and dele-
tion, the operation is same as that of correction except writing
different proofreading symbols. Of course, in deletion mode, input
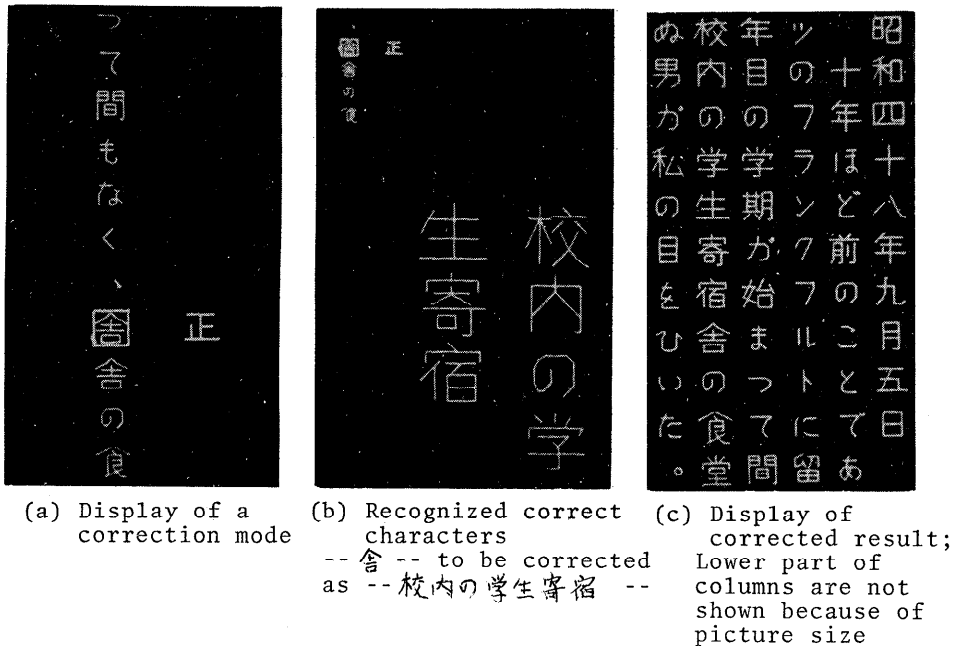of new character(s) is not necessary.

(a) Display of a   (b) Recognized correct   (c) Display of
    correction mode   characters   corrected result;
                          -- 舎 -- to be corrected   Lower part of
                      as -- 校内の学生寄宿 --   columns are not
                                                          shown because of
                                                          picture size

Fig. 5 CRT Display during Proofreading Process

5.   CONCLUSION

   These proofreading steps are very similar to ordinary "red mark-
ing", but arranged for the processing under a man-machine conver-
sational mode, so that the proofreader may operate the system
smoothly to ensure high processing efficiency.

   We express sincere gratitude to Dr. Watanabe, Director of Hitachi
Central Research Laboratory, and Dr. Numakura, Dr. Kawasaki, Manager
of our Depertment, for their close guidance and encouragement.

                              REFERENCE
(1) H. Terai & K. Nakata; On-line, Real-time Recognition of Hand-
                           writing Chinese Characters and Japanese
                           Katakana Sylabary; Trans. IECE'73/5
                           vol.56-D No.5