THE DIAGONAL SYMBOL ENCODING SYSTEM (DSES)– A SIMPLIFIED 3-CORNER CODING METHOD AND ITS FEASABILITY FOR CJK ADOPTION

*Jack Kai-tung Huang**

Abstract

Hesitation in adopting the CJK-vernacular character computer as a tool still exists in library management and office automation. The Diagonal Symbol Encoding System (DSES), a simplified 3-Corner Coding Method, is designed for the need and convenience of CJK CASUAL USERS. Four digits represent an entire CJK vernacular character, encoding the two symbols (2 digits per symbol) which occupy the diagonal corner or diagonal position of a given CJK vernacular character. Anyone can learn the DSES to input CJK vernacular characters within 30 minutes. After 60 hours of on the job training and practice, the user may reach an input speed of 27 characters per minute, which is the average speed of Chinese handwriting.

1. Introduction

The complexity of Chinese characters has presented great obstacles to modern international communication. The lack of a system by which the more than 50,000 monosyllabic and non-alphabetic scripts could be manipulated and entered into automated instruments such as computers, typewriters, printing,

*Jack Kai-tung Huang, Prof. & Head, Dept. of Computer Science, Ming-Chuan College, ROC.

presses, and the telex has led to many problems in this modern world of the rapid transmittal of information. In other words, a system is desperately needed whereby the written language of a huge segment of the world population could be controlled, and used efficiently in the world communication network. Japanese Kanji and Korean Haja have more or less the same problem as Chinese characters. The major problem encountered in computerizing Chinese characters, Japanese Kanji and Korean Haja for commercial use has proved to be encoding.

During the past century many scholars and experts in this field have struggled to devise new systems. More than 100 methods of ordering Chinese characters, Japanese Kanji and Korean Haja have been released in their homelands and abroad, and most have served as important contributions to Chinese Japanese and Korean automation.

As computer technology became more and more widely used, the intellect of the Chinese, Japanese and Korean which had originally produced such a rich and sublimely beautiful script, was baffed as to how to reduce it to a simple code, suitable for translation into binary form. To date, over 30 Chinese input systems have been devised in Taiwan, R.O.C. alone.

2. The need of a simplified input coding system

The current trend is that the use of the CJK-language computer in various applications is increasing. Nevertheless, greater reluctance to use the CJK-language computer as a managerial tool still exists. It is a fact that most of the existing systems are not completely satisfactory to their users because of certain common shortcomings[1].

The design of the Diagonal Symbol Encoding System (DSES), a simplified input system of the Three Corner Coding Method[2], is based on considerations of the need and convenience of the CJK CASUAL USERS, and the greater efficiency and productivity

of library management through the implementation of office automation. These considerations are as follows:

(1) Origins of Chinese characters, Japanese Kanji and Korean Haja may be traced through their symbolic representation[3].

(2) Number of strategies applied for coding Chinese characters, Japanese Kanji or Korean Haja must not exceed a total of five.

(3) Capability of instantly and positively identifying the code number without hesitation when one reads a Chinese character. Japanese Kanji, or Korean Haja. Nevertheless, maximum faulty-tolerance functions must also be featured.

(4) Least possible effort spent on learning and memorization.

(5) Coding spaces for necessary special characters and professional graphics must be preserved.

(6) Least possible duplication of characters, Kanji or Haja occuring within a semi-closed CJK character set (frequently used characters[4,5]), and a positive solution to this problem must be provided for data processing and communication.

(7) Both manual and machine-readable index application should be compatible and interchangeable, and both must be consistent.

(8) Fixed length code for unique formations. Between the characters, and for space bar operation should not be required.

(9) One-handed numerical key-pad operation capability for easier turning of pages by another hand and better visual focus on source documents.

(10) Compatible with any standard ASCII keyboard or numerical key-pad alone, without modification of the key tops, or

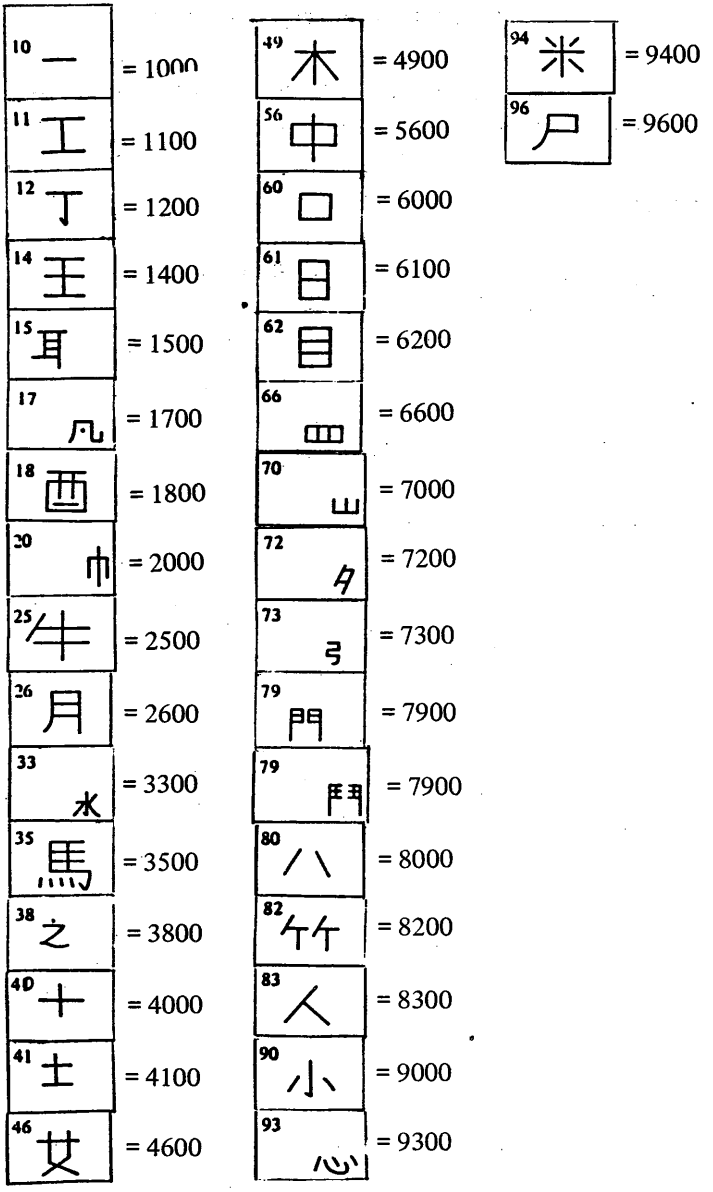
alternation of the arrangement of the standard ASCII keyboard. This is essential for reduced cost of keyboard manufacture and for international marketing.

3. The diagonal symbols

The radical (symbol) is the most important element of a compound, since characters are traditionally arranged under the 214 radicals of the K'ang-hsi dictionary, published in A.D. 1716. Through analysis and statistical research, however, it was discovered that modern Chinese characters are not limited by the scope of the K'ang 'hsi radicals. Some are derived instead from the 540 radicals of Hsu-Shen's Shuo-wen Chieh-tzu (A.D. 100), the 542 radicals of Liang Ku Yeh-wang's Yu-Pien (A.D. 534). and from various roots of modern Chinese characters in daily use.

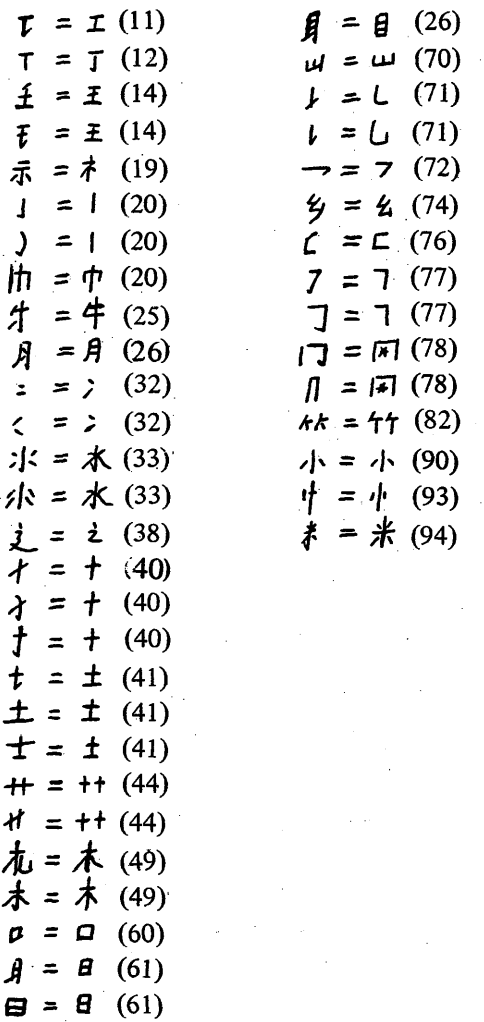
Of the 78 diagonal symbols, 60 are identical to the radicals from Mei Yung-tso's 214-radical system of A.D. 1615, originate from Hsu Shen's 540-radical system, and 1 oriented from Ku Yeh-wag's 542-radical system. Approximately less than one-fourth of the symbols (17 in all) are root originated from roots composing one part of the radicals in Ku Yeh-wang's 542-radical system, or originated from roots composing parts of Chinese characters in daily use.

Of the 78 diagonal symbols, 34 are identical to the Chinese characters in daily use. Thus, in length, the digonal code is a unique formation of four-digit figure code number, while in coding, the last two digits are filled with "00" for these 35 diagonal symbols to have a unique formation of complete four-digit figure code number. These 34 diagonal symbols that identical to the Chinese characters in daily use are:


10 一 = 1000
11 工 = 1100
12 丁 = 1200
14 王 = 1400
15 耳 = 1500
17 凡 = 1700
18 西 = 1800
20 巾 = 2000
25 牛 = 2500
26 月 = 2600
33 水 = 3300
35 馬 = 3500
38 之 = 3800
40 十 = 4000
41 土 = 4100
46 女 = 4600
49 木 = 4900
56 中 = 5600
60 口 = 6000
61 日 = 6100
62 目 = 6200
66 皿 = 6600
70 山 = 7000
72 夕 = 7200
73 弓 = 7300
79 門 = 7900
79 = 7900
80 八 = 8000
82 = 8200
83 人 = 8300
90 小 = 9000
93 心 = 9300
94 米 = 9400
96 尸 = 9600

4. Equivalent forms of diagonal symbols

Many diagonal symbols have several different but equivalent forms. Although they may slightly differ in writing styles, but the code remains the same. These equivalent forms of diagonal symbols are:



5. An on-screen secondary indexing system is provided

The DSES is not a one character–one code system. To avoid input error, an on-screen secondary indexing system must be provided and used. For indication in selecting the right one from the duplicated CJK characters. If there is an English letter "Z" points at the specific character on the screen that indicates this is a sole code for that particular character, Kanji or Haja.

6. Different styles of CJK characters

There are various styles of some CJK characters. Many such cases occur with the most frequently used CJK characters. In DSES, the code number is genereated according to the shape of the CJK characters. If two different styles of a certain CJK character occurs, there must be two different code numbers for them.

7. DSES plus 3-Corner Coding Method

CJK-language computers with the DSES processing capability[6], besides the capability of generating new CJK characters are also featured with the processing capability of 3-Corner Coding Method. The purpose for this design is to provide another optional input method for CJK PROFESSIONAL USERS.

The 3-Corner Coding Method is able to process 9,899 non-Chinese characters and 990,000 CJK characters. So far as number of characters concerns, it is sufficient for any CJK PROFESSIONAL USER[7].

8. The problem of CJK communication

A physical problem remained in CJK information communication because of the incompatibility of external and internal codes existing between different makes of CJK computer in the world. This situation isolates each individual CJK computer system from every other.

Now, however, the creation of ISO 646 and 2022 based Chinese Character Code for Information Interchange (CCCII)[8] and Cross-Reference Data Base for CCCII[9] system have radically solved the problems of CJK character interchange between CJK computer systems. If CCCII and its Cross-Reference Data Base are properly installed and implemented, these two universal tools will eliminate incompatibility and make CJK information flow and communication between CJK computer systems possible.

9. The structure of Diagonal Symbol Encoding System

(1) The organization of DSES.

There are 49 major symbols (radicals) and 29 minor symbols (radicals). This total of 78 symbols (radicals) is selected from those most often found in frequently-used Chinese characters, Japanese Kanji and Korean Haja are sorted into a logical system, which may be represented in tabular form by a 10x10 square (in which 42 units are blanks reserved for special uses.) Each symbol (radical) carries two numbers derived from the 58 horizonatal and vertical axes as 58 units of the table (DSES requires 58 keys or a numerical key-pad). Minor symbols (radicals) carry the same number as the major symbols (radicals) to which they are related.

In coding, two digits representing each of the major or minor symbols (radicals) that occupied at the diagonal corner or position of a given Chinese character, Japanese Kanji or Korean Haja encoding order may follow one of the 5 strategies in accordance as how the character, Kanji or Haja is structured. The

The Diagonal Symbol Encoding System (DSES)

tabular form is as follows:

Prof. Jack Kai-tung Huang

Symbols Chart Code

DIAGONAL SYMBOL ENCODING SYSTEM (DSES)–
A SIMPLIFIED 3-CORNER CODING METHOD

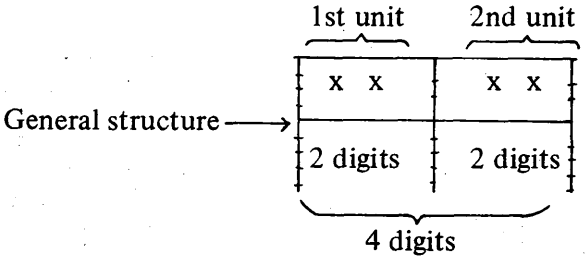
黃 克 東 *Jack Kai-tung Huang*

00	01	02	03	04	05	06	07	08	09
10	11	12	13	14	15	16	17	18	19
20	21	22	23	24	25	26	27	28	29
30	31	32	33	34	35	36	37	38	39
40	41	42	43	44	45	46	47	48	49
50	51	52	53	54	55	56	57	58	59
60	61	62	63	64	65	66	67	68	69
70	71	72	73	74	75	76	77	78	79
80	81	82	83	84	85	86	87	88	89
90	91	92	93	94	95	96	97	98	99

Large-major symbols 49, small-minor symbols 29
Total 78 fundamental symbols, 58 codes

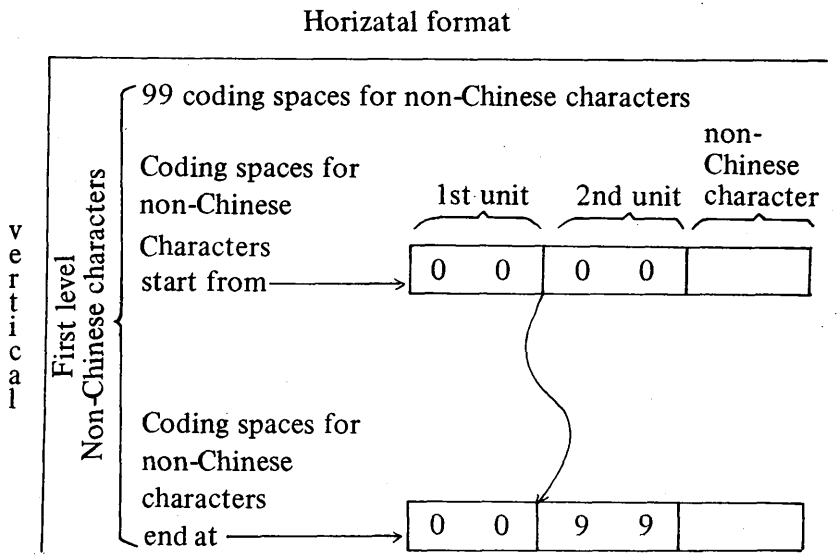
大一主符號 49 個，小一副符號 29 個
共計 78 個基本符號， 58 組編號。

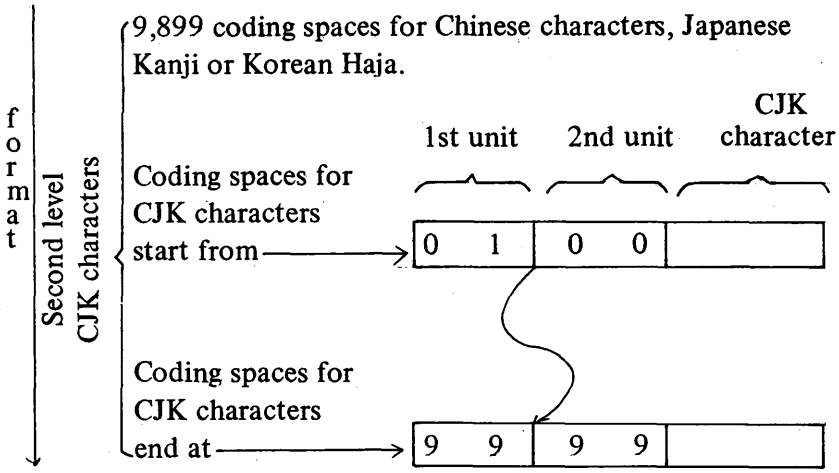
(2) The architecture of DSES.



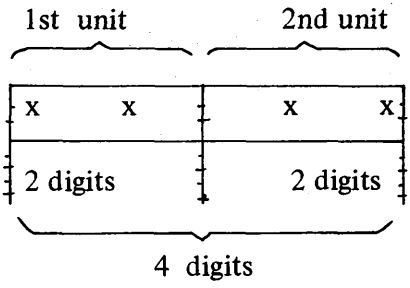
* x = 1 digit

The architecture of the structure of the Diagonal Symbol Encoding System (DSES) may be described as two formats in two dimensions; the horizontal format and the vertical format (divided into two(2) levels; first level are coding spaces for non-Chinese characters, second level are coding spaces for Chinese characters, Japanese Kanji or Korean Haja) are as follows:



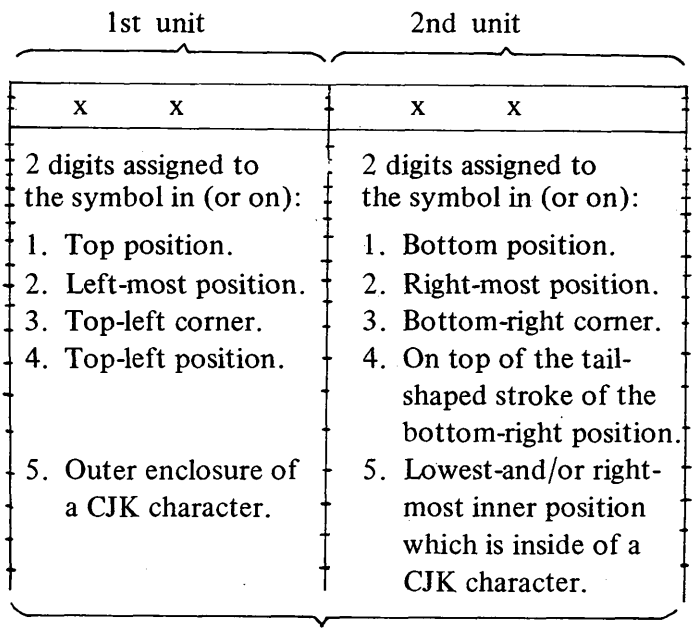


A. In the horizontal format: the architecture is divided into 2 units as follows:



* x = 1 digit

When the Diagonal Symbol Encoding System is applied in encoding Chinese characters. Japanese Kanji or Korean Haja, the format is as follow:



* x = 1 digit

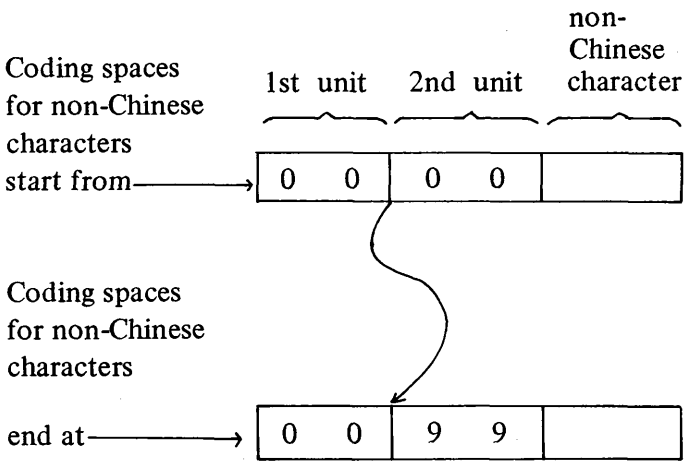
a. The sequence of the 4 digits of the diagonal code does not follow an overall consecutive order, but the 2 digits assigned to each diagonal symbol are given in consecutive order.

b. Within each unit of the 2 units, 2 digits are consecutively ordered by the numerical code which is assigned according to the diagonal symbol that occupies the corresponding corner or position of the Chinese character, Japanese Kanji or Korean Haja.

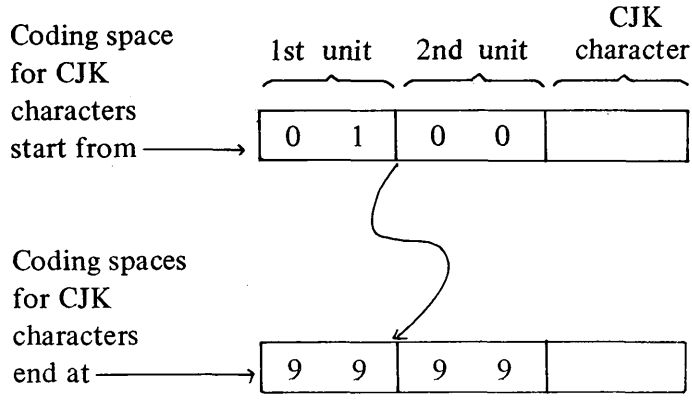
B. In the vertical format, the architecture of coding spaces is divided into two(2) levels; first level are coding spaces for non-Chinese characters, second level are coding spaces for Chinese characters, Japanese Kanji or Korean Haja,

the detailed explanation is as follows:

a. First level: 0000 to 0099 reserves coding spaces for 99 non-Chinese characters (to be released).



b. Second level: 0100 to 999 reserves 9899 coding spaces for Chinese characters within a semi-closed CJK character set.



(3) The 5 coding strategies of DSES.

Four digits represent an entire Chinese character or Japanese Kanji, or Korean Haja in the encoding of the two symbols (2 digits for a symbol) which occupy the diagonal corner or diagonal position of a given Chinese character, Japanese Kanji or Korean Haja. The encoding order must follow one of five strategies in accordance with how the character, Kanji or Haja is structured.

A. Strategy 1:
Encode the symbol (radical) occupying the top and bottom positions of a Chinese character, Japanese Kanji or Korean Haja.

As: 

Example:

Chinese character	Symbols in diagonal positions ↓	
	Symbol in the top position 亠	Symbol in the bottom position 乂
文	Code for symbol in the top position 01	Code for symbol in the bottom position 43
(Wen. literature)	Diagonal code ↑	

B. Strategy 2:
Encode the symbol (radical) occupying the left-most and right-most positions of a Chinese character, Japanese Kanji, or Korean Haja.

As: 

Example:

Chinese character	Symbols in diagonal positions	
	Symbol in the left-most position 亻	Symbol in the right-most position 壬
任	Code for symbol in the left-most position 22	Code for symbol in the right-most position 14
(Jen. to let)	Diagonal code ↑	

C. Strategy 3:
Encode the symbol (radical) occupying the top-left and bottom-right corner of a Chinese character, Japanese Kanji, or Korean Haja.

As: 

Example:

Chinese character	Symbols in diagonal corners ↓	
	Symbol in the top-left corner	Symbol in the bottom-right corner
許 (Hsu, to permit)	Code for symbol in the top-left corner: 01	Code for symbol in the bottom-right corner: 40
	Diagonal code ↑	

D. Strategy 4:

Encode the symbol (radical) occupying the top-left position and the symbol (radical) that is on the top of the tail-shaped stroke of the bottom-right position of a Chinese character, Japanese or Korean Haja.

As: 

Example:

Chinese character	Symbol in diagonal positions ↓	
	Symbol in the top-left position	Symbol on top of the tail-shape stroke of the bottom right position
颱 (Tai, typhoon)	Code for symbol in the top-left position: 17	Code for symbol on top of the tail-shape stroke of the bottom right position: 60
	Diagonal code ↑	

E. Streategy 5:

Encode the symbol (radical) occupying the outer enclosure and the lowest-and/or right-most inner position which is inside a Chinese character, Japanese Kanji or Korean Haja.

As: 

Example:

Chinese character	Symbols at disgonal positions ↓	
	Symbol on the outer enclosure	Symbol in the lowest-and/or-right most inner position
員 (Yuan, round)	Code for symbol on the outer enclosure: 63	Code for symbol in the lowest-and/or-right most inner position: 80
	Diagonal code ↑	

10. Conclusion

The primary goal in designing the Diagonal Symbol Encoding System is to meet the need for a convenient system for theCJK CASUAL USERS, therefore, input speed is not emphasized as a design consideration.

The DSES is not a one character–one code system. To avoid input error, an on-screen secondary indexing system must be provided and used. For indication in selecting the right one from the duplicated Chinese characters, Japanese Kanji or Korean Haja.

With the aid of a printed Diagonal Symbols Chart, anyone can learn to input Chinese characters, Japanese Kanji or Korean Haja within 30 minutes. Using the Chinese phonetic value input system (which is based on the National Phonetic Symbols System), which also needs a secondary indexing system, as a basis for comparison, we estimate that after 60 hours of on the job training and practice, input speed may fairly match the average speed of Chinese handwriting, which is 27 characters per minute.

References

1. (1) C. T. Chang, C. C. Hsieh, C. C. Yang, Jack K. T. Huang and K. Mei. *The Development of Software Tools For Sinology. Part 2 (Computer Technology and Application in Chinese Information Systems) of International Aspects of Information System Productivity, ASIS-83,* Washington D.C., Oct. 4, 1983.

   (2) Jack Kai-tung Huang, The Input and Output of Chinese and Japanese Characters, p. 28–24, COMPUTER, IEEE Computer Society, N.Y., N.Y. 10017, USA, Jan. 1985.

   (3) Joseph D. Becker, Typing Chinese, Japanese and Korean, p. 27–34, COMPUTER, IEEE Computer Society, N.Y., N.Y. 10017, USA, Jan. 1985.

2. (1) Three-Corner Coding Method (TCCM) invented by L. R. Hu, Y. W. Chang and Jack Kai-tung Huang, 1969, released 1977.

(2) A) L. R. Hu, Y. W, Chang and Jack Kai-tung Huang, *Three-Corner Coding Method. The Digitalized Chinese Dictionary.* (Taipei: Northern Gate Books Co., 1977).

B) *Training Manual for the Three-Corner Coding Method* (Chinese and English ed.) authors and address same as above, Apr. 1979, pp. 125–127.

C) *Chinese Indices Interchange Table,* authors and address same as above, Dec. 1981.

3. Jack Kai-tung Huang, Considerations prior to the Designing the Three-Corner Coding Method (TCCM) and the Origin of the Fundamental Symbols of the TCCM. The First Asian–Pacific Conference on Library Science, March 13–19, 1983; pp. 311–343, Taipei, Cultural and Social Centre for the Asian and Pacific Region, Seoul, Korea.

4. *Table of Frequently Used and Standard Chinese Characters,* Edited by Ministry of Education, Cheng Chung Books Co., Taipei, Taiwan, ROC, 1979.

5. Japanese Industrial Standard, Code of the Japanese Graphic Character Set for Information Interchange, Level 1, JIS C 6226, 1978. Established 1978-01-01. Investigated by Japanese Industrial Standard Committee. Published by Japanese Standard Association, Tokyo, Japan, 1978.

6. Computer companies that have decided to install DSES:
   (1) Dai-E Systems Inc. (on Apricot computer systems), 29783 Town Center Loop West, Wilsonville, OR 97070, USA, Tel. (503 682-3231.
   (2) Teltec Electronic Systems Ltd, (on FuTec Chinese Composer), 68, Sung Wong Toi Rd., Freder Centre 1/c, Kowloon, Hong Kong, Tel. 3-338893.

7. Jack Kai-tung Huang, The Input and Output of Chinese and Japanese Character, and The Three-Corner Coding Method: Design and Application, p. 18–24, COMPUTER, IEEE Computer Society, N.Y., N.Y. 10017, USA, Jan. 1985.

8. Jack Kai-tung Huang, Ching-chun Hsieh, Chung-tao Chang, Chen-chau Yang, The CCCII and Its Feasibility for International Adoption, International Cooperation in Chinese Bibliographical Automation, Australian National University, Canberra, Australia, Aug. 29–Sept. 1, 1982.

9. S. S. Tseng, T.C. Gau, C.C. Hsieh, C.C. Yang, C.T. Chang, Jack K. T. Huang, A Multi-Lingual Coding System Based on CCCII, The Second Asian-Pacific Conference on Library Science, May 20–24, 1985, Seoul, Korea, Cultural and Social Centre for the Asian and Pacific Region, Seoul, Korea.