TR-88-18

EVALUATION OF CHINESE INPUT METHODS BY S-P CHART ANALYSIS

Evaluation of Chinese Input Methods By S-P Chart Analysis

C. M. Tseng[*], S. P. Yang[*], M. S. Hwu[**], and K. Y. Cheng[**]

[*] Institute of Applied Mathematics, National Chung Shing University
Taichung, Taiwan, R. O. C.
and
[**] Department of Computer Science and Information Engineering
National Taiwan University, Taipei, Taiwan, R. O. C.

ABSTRACT

This paper presents a new approach which employs the concept of S-P chart analysis in the evaluation of keyboard-based Chinese input methods. This approach is based on a behavior-oriented evaluation model. The behavior-oriented evaluation is based on examining a sequence of tests. It will show when a sample of n characters is given to m operators for decoding, the test scores can be recorded into an m x n S-P chart. A static operational behavior pattern concerning an input method can be obtained by performing S-P chart analysis on the test. Also, a dynamic learning pattern concerning the method can be obtained by performing global chart analysis on several tests.

I. INTRODUCTION

There are a few papers which discussed and presented the evaluation procedures for Chinese input methods[2,3,9,11,14]. They can be classified into two categories: method-oriented and behavior-oriented. The method-oriented evaluation puts focus on the interior process of how the method is functioning, while the behavior-oriented evaluation puts focus on the exterior measurement of how effective the method will be. Therefore, two different concepts were developed, the method-oriented is for the design of a better machine and the behavior-oriented is for the comparasion of different existing machines, both are from user's point of view. Because of this difference between them, the method-oriented regards the input action of an operator as visual routines[9] and the evaluation procedures are based on the parsing of these routines, while the behavior-oriented emphasizes on macro view of many operators using different input systems and thus the evaluation procedures are based on statistical measurement.

In this paper, we propose a new approach for the evaluation of keyboard-based Chinese input methods by using chart analysis[18,18,22]. It is a modification of S-P chart analysis proposed by Takahiro Sato[18,19]. S-P chart analysis is primarily for the evaluation of CAI courseware as well as that of student learning behavior. An m row and n column zero-one chart is formed by giving n problems for m students to solve and recording the (i,j)-element as 1 if the i-th student solve the j-th problem correctly else 0. A S-P chart is obtained by row and column permutation of the (0,1)-matrix so that the students are ordered according to the score or row sum from high to low and the problems are ordered

according to the number being answered or column sum from high to low. Some evaluation parameters can be obtained from S-P chart analysis to understand the student learning behavior, the problem property, and the performance of a particular student to a particular problem. Since S-P chart analysis is based on a statistical model, so our approach is primarily behavior-oriented.

In the following presentation, we shall divide our discussion into sections. Section II presents an overall system architecture for the evaluation of keyboard-based Chinese input methods. Section III introduces some important service items which are obtained from S-P chart analysis. Section IV describes some important parameters in the evaluation of Chinese input methods using keyboard. Section V demonstrates the relation between evaluation parameters and S-P chart service items.

II. SYSTEM ARCHITECTURE

We shall restrict our discussion on the evaluation of Chinese input methods using standard keyboard only. The Chinese input using standard keyboard in the existing systems includes using internal codes, phonetic symbol codes, stroke symbol codes, radical symbol codes, or their mixed codes as keys to decompose the Chinese character[1,6,7,8,12,13,15,17]. The effectiveness of a particular method depends on some measurable parameters such as number of primitive code symbols, number of meta rules, average number of keys depressed, collision rate, ..., etc. and learning behavior such as average learning time, average input speed, average error rate, ..., etc. The value associated with each measurable parameters can be directly obtained from the meta rules defined in

3

the method. However, the performance measure on learning behavior can be obtained only from performing the statistical analysis.

Our evaluation procedures are based on the statistical analysis. A set of testing samples(some selected Chinese characters) are given to a group of people for a test. The result is recorded in a chart with testing samples as columns, persons as rows, and element (i,j) = 1 if the i-th person answers correctly to the j-th sample else element (i,j) = 0. After performing the appropriate row and column permutations, a resultant chart, called a S-P chart, is obtained. The statistical analysis is based on the resultant S-P charts. Fig. 2.1 shows the system architecture for the evaluation of Chinese input methods using chart analysis.

Designer   Persons under test

Visual Programming Environment for the definition of a Keyboard-based Chinese Input Method → Chinese character Input Test

Charts Data Base

Computaion of Evaluation Parameters   Analysis of S - P Chart

Merging

Global Analysis

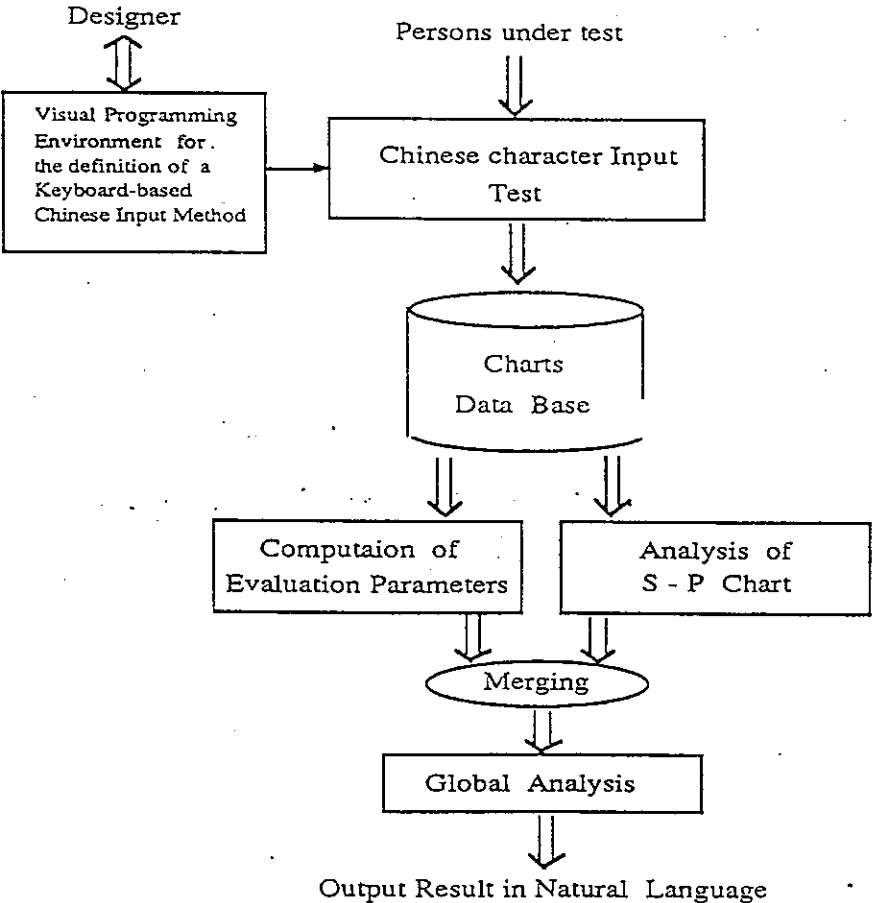Output Result in Natural Language

Fig. 2.1. The system architecture.

4

As can be seen from Fig. 2.1, there is a visual programming environment[10] which is provided for a Chinese input designer to define his/her input method visually. This method is then under a test by providing n Chinese character samples to m persons for the test. The testing result is stored in the Charts Data Base. Two paths for the evaluation are processed. One path is the computaion of some evaluation parameters and the other path is to perform the S-P chart analysis. After that, an advanced analysis by integrating these two paths, namly the evaluation parameters and the analysis result of S-P charts, is performed to obtain the final evaluation results. These evaluation results are presented to designer through a natural language output[20]. However, in the following we shall not discuss the visual programming environment and the natural language output, instead, we focus our discussion on the evaluation by using chart analysis only. Interested readers are encouraged to consult[10] for a description of how to design such a visual programming environment and that of how to output natural language statements from chart analysis[20].

III. ANALYSIS OF S - P CHART

As stated previously, our evaluation approach is mainly based on analyzing the S-P chart. However, the chart analysis adopted in our approach is much more complicated as the S-P chart analysis proposed by Takahiro Sato, because some modifications of the S-P analysis are added to suit for the evaluation of Chinese input methods using standard keyboard.

The S-P chart analysis is first described in the following. Giving n Chinese characters to m persons for the test of a particular Chinese input method, an m x n (0,1)-matrix can be obtained, where the (i,j)-element is 1 for a right answer or 0 for a wrong answer. This (0,1)-matrix is an original S-P chart . For example,

Fig. 3.1 shows an original S-P chart for 30 persons(or students) to test 31 characters(or problems).

```
S----30
P----31                    PROBLEM NUMBER

                       1111111111222222222233
STUDENT     1234567890123456789012345678901
NUMBER      -------------------------------
        1   0111110111011101001100000011000
        2   1111111110110110111111111000001
        3   1011011111111111111111000011000
        4   0111110100111111101111000000000
        5   1111111100011111111110010100000
        6   1100111111110011001011000011100
        9   1111011110011111110110000000000
       10   1111111000011110010011111111101
       11   1111111111111111111100100011100
       12   1111111100011111111101111111100
       13   1111101100010000001101000000000
       15   1110101111111111100011010000000
       16   1111111111111111111111111111111
       18   1111111111111111111011111100000
       19   1111111111111111111111000000000
       20   1111111111111111011101000000000
       21   1101111110011111111110000100010
       22   1111111111111111111110000000000
       24   1111111111111111111111110111110
       26   1111101100011100001110000011111
       27   1111111111111101111111111110011011
       28   1111111111101111111011110000010
       29   1111011011111101111100001011010
       30   1111111111111111111011110010100
       32   1111111111111111111111010000000
       34   1111101111111110111111100000000
       35   1111111111111111011100000000000
       36   1111111111111111101111101011111
       37   1010101111010000000011000001110
       38   1111111111111110101110011001000000
```

Fig. 3.1. An example of a matrix for 30 students on 31 problems. (Courtesy of Takahiro Sato of C&C Systems Research Laboratories NEC Corporation)

This chart can be rearranged according to the descending order of the student scores in rows and the descending order of the problems being correctly answered in columns. The permutation result is a S-P chart. There are two boundary curves in the S-P chart, the S-curve and the P-curve. They are formed as shown in Fig. 3.2.

```
                                                  Test score
                                                  (Correct answers)
              PROBLEM  NUMBER
        easy ────────────────→ difficult          Caution signal
         1    1  1 111  2211211222223223             Caution Index
STUDENT  2712383453669478178281783499361
NUMBER──────────────────────────────────CA   %    C.S
      16  1111111111111111117111111111111 31  100  0.00
      24  1111111111111111111111111111110 10 29  94  0.04
      36  1111111111111111111110111191110 1 28  90  0.41
      27  1111111111111101111111111118011 27  87  0.30
      38  1111111111111-118111111911180000 25  81  0.22
      13  1111111111110111111100188110  25  81  0.33      S-curve
      12  1111111111110101011011111191    23  81  0.73
      11  111111111111111101101111110100 24  77  0.27
       2  1111111111101111011181091101 24  77  0.44
      32  11111111111111111111111089108888 23  74  0.02
      17  11111111111111111111000808980 22  71  0.00      P-curve
      23  11111111111011111188112111008 22  71  0.25
       3  1110111101111111111111118800008 22  71  0.30
      18  11111011111108100100110111111811 22  71  1.13
      34  111111111110111111911118810000000 21  68  0.15
      30  11111111111011111111108088800100 21  68  0.17
      29  1111101100111111181190011908 21  68  0.63
      35  11111111111111110018088000000 20  65  0.01
      22  1111111111111110110110088000000 20  65  0.04
      20  111111111111111100111990000000 20  65  0.05
       5  111111111111111811621890018808 20  65  0.30
      21  01119111111111111118100000001018 20  65  0.43
      15  111111111011010111119100010000 18  58  0.29
       6  1111910011110101811811100000 18  58  0.56
      26  111111111000110119090011001001 18  58  0.73
       9  11111110111911111008100000000 17  55  0.11
       4  109111111111111110100100000000 17  55  0.18
       1  100111111011111800188811000000 16  52  0.36
      37  111011001090018181181000101099 13  42  0.73
      13  11111191109010810010000000000 11  35  0.10
──────────────────────────────────────CA   %  AT C.S
```

(Correct answers) CA

```
2222222222222222222222211111         ←── Number of students
98888877753553443221893321090876          getting correct

%
9999999999988888877776644433322      ←── Passing rate
733333000833330873307733073073            of each problem

C.P              ··1·  ··    ·  ··    ←── Caution signal
88900100000800900800900000000000
................................
8445790301235558624623460256344
68895187331430009614526499394

TOTAL MEAN  68.0%    ←──── Average passing rate

          MEAN    SD    ATT.MEAN
STUDENT   21.3    4.4    0.31
PROBLEM   20.6    7.4    0.49

03=0.46    ←──── Disparity coefficient
```

Fig. 3.2. The S-P chart and the S-P curve analysis. (Courtesy of Takahiro Sato of C&C Systems Research Laboratories NEC Corporation)

The shapes of S-P curves are useful for identifying unusual response patterns behind the method under test. For example, when both curves sharply tend to upper-left corner may indicate that the method is very difficult for most persons(Fig. 3.3(a)), on the other hand, when both curves sharply tend to lower-right corner may indicate that the method is easy (Fig. 3.3(b)), and when both curves alternate between the diagonal line drawn from upper-right to lower-left corners as shown in Fig. 3.3(c) may indicate a normal (not easy not hard) method. Also, new information may be obtained by measuring local shape of the S and the P curves in order to have further diagnosis on specific skill understanding of an operator. For example, a large caution index as shown in Fig. 3.2 indicates an operator's unusual response in the phenomena that he can solve a few difficult characters but can't solve considerably many easy characters (#10 student with two dots in caution signal).

(a) a very difficult method   (b) an easy method   (c) a normal method
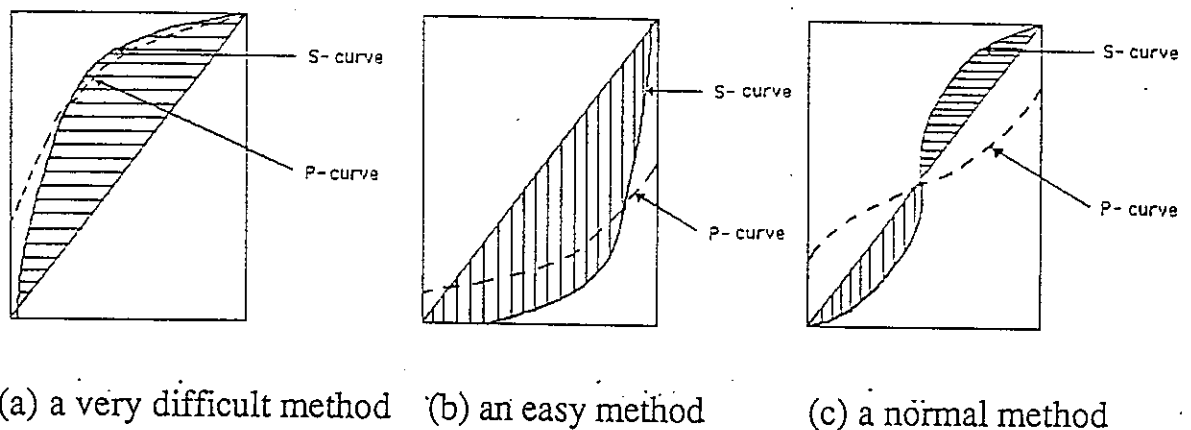
Fig. 3.3. Examples of S-P Charts.

. In addition to caution index, the disparity coefficient, is used in the diagnosis of the learning behavior of all students in decomposing characters concerning a particular method. The disparity coefficient D^* is used to measure the degree of non-homogeneity for the persons under test or the characters for the test. Its value is in proportion to the area surrounded by S-curve and

P-curve. A large value of D^* may indicate that the persons under test are heterogeneous, i.e., their capability in decomposing the characters is varied or the set of characters for the test is improper, i.e., the character set consists of many difficult vocabulary (new words) for most persons. We shall call the items such as caution index, disparity coefficient, ..., etc., which can be directly obtained from analyzing the S-P chart and its related information, as the service items.
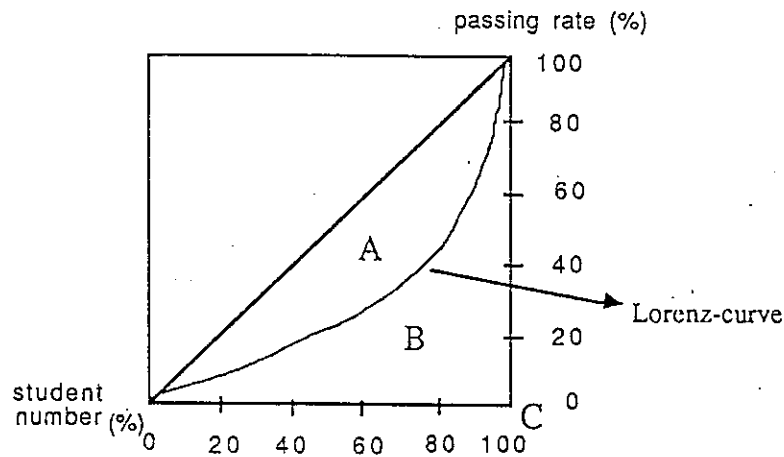
Fig. 3.4. The Lorenz-curve.

The S-P chart related information such as Lorenz-curve[22] (as shown in Fig. 3.4) provide another service item, the Gini coefficient $GINI$, which can be computed as

$$GINI = \frac{\text{area of part A}}{\text{area of part A} + \text{area of part B}}$$

The Gini coefficient is used to measure the distribution of the test scores . The little $GINI$ value, i.e., Lorenz-curve is closed to diagonal line, indicates a normal distribution of the tested scores, otherwise, i.e., Lorenz-curve is closed to the point C , indicates an abnormal distribution.

IV. COMPUTATION OF EVALUATION PARAMETERS

After a keyboard-based Chinese input method is defined, several evaluation parameters such as the number of primitive code symbols, the collision rate, the average number of keys depressed for each character, ..., etc. of the method can be obtained from method-oriented evaluation[1,3,9,14]. These parameters are classified as static. To some dynamic evaluation parameters such as the average learning time, the average input speed, and the average error rate, ..., etc. for an operator can be measured directly[2,11]. So, we shall not discuss them. However, in the following, we shall discuss yet another important dynamic evaluation parameter, the degree of learning in certain Chinese input methods such as Tsang-Chi, Three-Corner, Simplest, ..., etc.

The input process of a Chinese character in these Chinese input methods is primarily the character decomposition and code symbols selection. Any Chinese character input to an operator's visual mind can be regarded as a sensory icon being processed by an operator's perception mechanism. The conceptual graph[5] in the operator's perception is a decomposition graph which is used to guide the decomposition and performing a proper selection. Although different methods produce different decomposition graphs for each Chinese character, but all of the graphs can be defined uniquely as $G_d = (V, E)$, where V is the set of nodes representing the code symbols and E is the set of edges representing the part-of predicate and the spatial relations between code symbols. Fig. 4.1 shows the decomposition graph of the character 腦 in Tsang-Chi method[9,21], where the visited terminal nodes in a counterclockwise traverse of the graph are the primitive code symbols corresponding to the radicals 月 , 女 , 女 , 田 . In other words, the decomposition process may be viewed as a deduction.
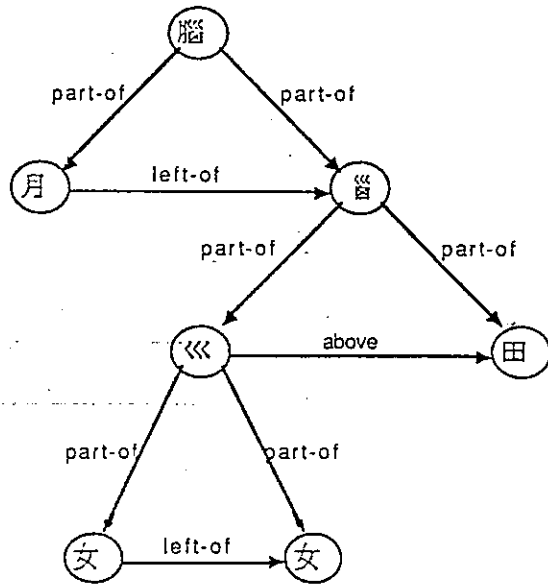
Fig. 4.1. The decomposition graph of character 腦 .

In scoring the deduction problem, we usually give a higher score to the student who reasons correctly to as deeper as possible. Similarly, in scoring the Chinese character input using keyboard, a person gives more correctly decomposed terminal nodes, a higher score should be given to them. In the normalized case, we give the score one to the person who completely decomposes the character, because this means that he entirely know how to decompose this character. On the contrary, the score zero indicates none of the terminal nodes is correctly answered, or the person does not know how to decompose this character at all. In general, to measure the degree of learning, we may use the fuzzy logic[4].

Let $P = \{ p_i \}$, i = 1, 2, ..., R be the set of primitive code symbols being adopted in a keyboard-based Chinese input method and $S_{ti} = \{ s_j \}$, j = 1, ..., n be the collection of Chinese characters given for the i-th person to test. Suppose U is the universal set of P or $U = 2^P$, then S_{ti} is a proper subset of U. Let $\mu_i(s_j)/s_j$

denote that each s_j in S_{ti} is assigned with a function $\mu_i(s_j)$, called the membership function of s_j, to indicate the partial score that person i can have in decomposing character j. The set of partial score for each Chinese character under consideration is a fuzzy subset of U

$$S_{ti} = \sum_{j=1}^{n} \mu_i(s_j) / s_j$$

where μ_i is a mapping

$$\mu_i : U \rightarrow [0, 1].$$

When s_j is a terminal node of the decomposition graph, i.e., $s_j \in P$, the value of $\mu_i(s_j)$ is 1 if s_j is correctly picked otherwise the value of $\mu_i(s_j)$ is 0. For each nonterminal node, its partial score is the average of its children nodes. Hence, we use the following formula

$$\mu_i(s_j) = \begin{cases} 1 & s_j \in P \text{ and the user picts } s_j \text{ (correct)} \\ 0 & s_j \in P \text{ and the user does not picts } s_j \text{ (incorrect)} \\ \sum_{k=1}^{X} \mu_i(s_k)/X & s_k \in \text{Child}(s_j), \ s_j \notin P \end{cases}$$

where X is the number of children of s_j in the decomposition graph. Fig. 4.2(a) shows a possible testing result of the character 膾 by the i-th person, where $\checkmark$ indicates a correct pick and $\times$ an incorrect pick. The corresponding value of

$\mu_i(s_j)$ for s_j = 腦 , 月 , 甾 , 巛 , 女 , 女 , 田 are shown in Fig. 4.2(b). That is

the character 腦 decomposed as 月 水 女 大 get a score of $5/8$.
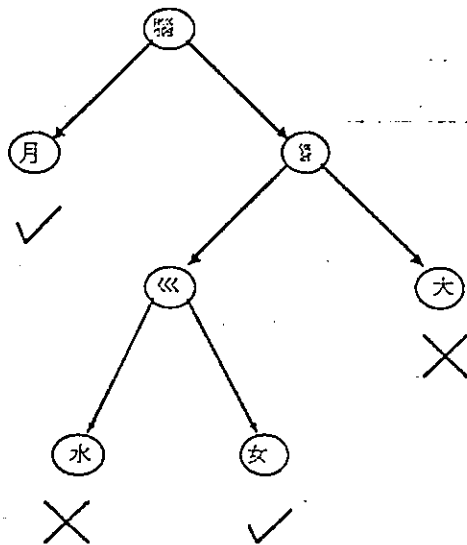
Fig. 4.2(a). A possible testing result   Fig. 4.2(b). The corresponding values
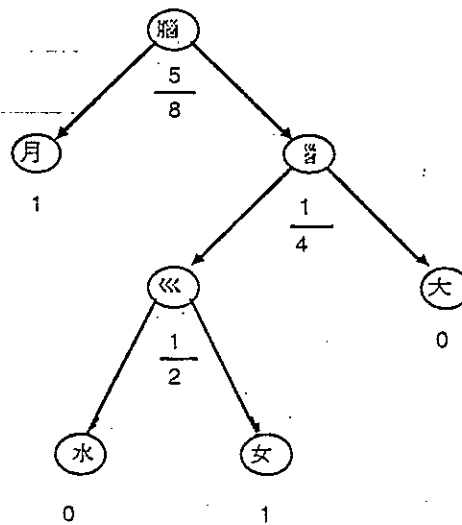
        of character 腦 .   of $\mu_i(s_j)$ result from Fig.

                                            3.2(a).

The computation using fuzzy logic also obtains a chart, an m x n matrix, with each element a value between 0 and 1. As will be seen in the next section, this chart can be combined with the S-P chart as a global chart for understanding of the learning behavior to a particular method.

V. GLOBAL CHART ANALYSIS

A global chart is the combination of an (0,1) S-P chart and a non-(0,1) chart obtained from computing fuzzy logic. Let S_c be the total scores obtained from

the (0,1) S-P chart. Suppose $C_t = \{c_j\}$, $j = 1, ..., n$ is the collection of Chinese characters to be tested. Let $F(C_t)$ be the summation of $\mu_i(c_j)$, $i = 1,..., m$ and $j = 1, ..., n$, $\mu_i(c_j)$ is the (i,j)-th element of the non-(0,1) chart, i.e.,

$$F(C_t) = \sum_{i=1}^{m} \sum_{j=1}^{n} \mu_i(c_j)$$

Then, a global chart consists of a S-curve, a P-curve, and a $F(C_t)$-curve. Notice that $F(C_t) - S_c$ is the summation of the values $\bar{\mu}_i(c_j)$ with $0 < \mu_i(c_j) < 1$, $i = 1, ...,$ m and $j = 1, ..., n$, a value indicates the total score of all partially correct answers. This value can be combined with some service items obtained from S-P chart analysis to explain some phenomena occurred in the method. For example, three possible cases are shown in Fig. 5.1 which shows the combination of the disparity coefficient D^* obtained from S-P chart analysis and the value $F(C_t) - Sc$.

| case 1 | case 2 | case 3 |
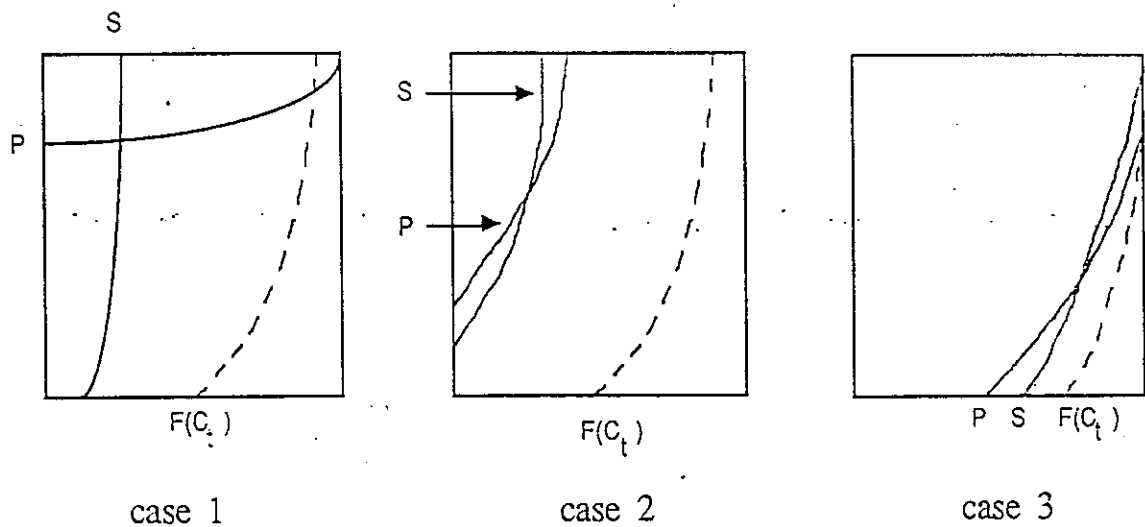
Fig. 5.1. Three illustarted cases for global chart analysis.

In case 1, the value of $F(C_t) - S_c$ is considerablly high which indicates only few percentage of the characters can be decomposed correctly. The value of $F(C_t)$ is also considerably high which seems to indicate a high passing rate may be expected in the next test. However, since the value of D^*, the disparity coefficient, is also high which indicates an abnormal situration occurred in the test. Thus, we can not sure whether the passing rate will be improved in the next test. In case 2, since the value of D^* is low which indicates a normal state is reached, thus contrary to case 1, we are sure in normal situation the passing rate will be improved in the next test. In case 3, since the values D^* and $F(C_t) - S_c$ are low which indicates almost every person can correctly decompose all the characters. Thus the method is considered a good one. We may conclude that a good Chinese input method is the one which under several tests can reach to the global chart of case 3 in a very short time.

More specifically, the global chart of D^* and $F(Ct) - Sc$ can be used in the dynamic analysis for understanding the process to reach to a steady state of a particular method. For example, Fig. 5.2 shows the dynamic diagram of two methods by plotting the values obtained from global chart analysis in a sequence of tests.
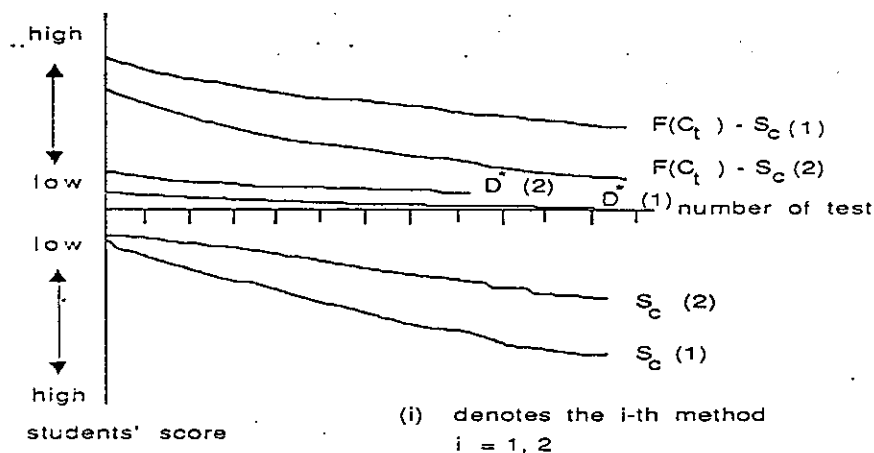

Fig. 5.2. The dynamic variation of D^* and $F(C_t)$-S_c chart analysis.

In this example, the first Chinese input method is better than the second one because its D^* value is always low, its S_c value increases faster, and its $F(C_t) - S_c$ value decreases more sharply.

Other service item like Gini coefficient *GINI* can be used in the dynamic analysis also. For example, Fig. 5.3 shows the variation of Gini coefficients corresponding to four different methods. Since the Gini coefficient curves of the first and the fourth methods tend to zero more rapidly than the other two, they are considered to be better methods. Because when Gini coefficient tends to zero the test score tends to a normal distribution, which indictes a steady state is reached.

(i)   denotes the i-th method, i = 1, 2, 3, 4
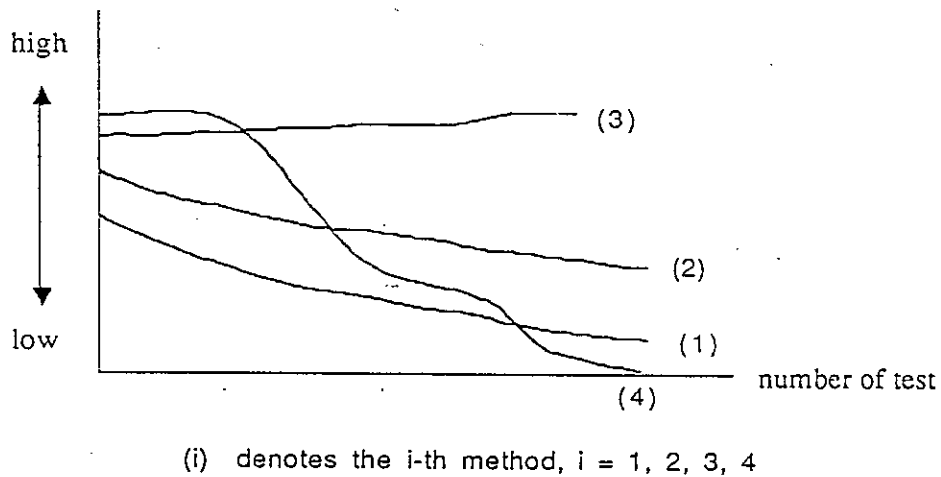
Fig. 5.3. Dynamic variation of the Gini coefficients corresponding to four different methods.

VI. CONCLUSION

A new approach for the evaluation of keyboard-based Chinese input methods by combining the analysis of S-P chart and the computation of evaluation parameters into as the global chart analysis is presented. We show

that this behavior-oriented approach is useful to a Chinese input designer for supplying the information about user's friendlyness in the aspect of human-machine interface.

REFERENCES

[1] C. C. Hsieh, Y. W. Hwang, S. Lin and K. M. Hsu, "The Analysis and Design of Chaio taung Radical System," *Proceedings of 1st International Symposium on Computer and Chinese, I/O System*, pp.49-62, 1973.

[2] C. K. Chen and R. W. Gong, "Evaluation of Chinese Input Methods," *Computer Processing of Chinese & Oriental Languages*, No. 4, pp.236-247, 1984.

[3] C. Wei, "Evaluation of Chinese Character Keyboards," *EEE Computer*, pp.54-59, January, 1985.

[4] H. J. Zimmermann, *Fuzzy Set Theory and its Applications*, Kluwer-Nijhoff Publishing.

[5] J. F. Sowa, *Conceptual Structures*, Addison-Wesley Publishing Company, 1984.

[6] K. M. Hsu, "A Simplifications for The Input of Chinese Radical System," *Journal of National Chiao Tung University*, Vol. 1, pp.1-18, 1976.

[7] K. T. Huang, "The Input and Output of Chinese and Japanese Characters," *IEEE Computer* , pp.18-24, January, 1985.

[8] K. Y. Cheng and F. K. Yu, "On Disambiguous Chinese Phonetic Input," *Proceedings of International Conference on Chinese and Oriental Language Computing*, pp.2-8, 1987.

[9] K. Y. Cheng, M. S. Hwu, S. Y. Hsu, "Evaluation of Reduced-Radical Chinese Input Methods from Parsing Visual Routines," Technical Report, Inst. of Information Science, Academia Sinica, 1988.

[10] K. Y. Cheng, M. S. Hwu and S. Y. Hsu, "Some Remarks on Developing Applications in Visual," Technical Report, Inst. of Information Science, Academia Sinica, 1988.

[11] L. C. Chen, *1982 Chinese Computer Market Report*, III Technical Report C20, pp.19-35, 1983.

[12] L. R. Hu, Y. W. Chang and K. T. Haung, "The Three Corner Coding Method," *Proceedings of 1979 International Conference of Chinese Language Computer Society*, pp.45-61, 1979.

[13] M. W. Du, C. C. Hsieh, S. H. Cji and S. C. Chu, "Character Composition and

Methods to Represent Radicals," *Proceedings of 1st International Symposiun on Computer and Chinese, I/O Systems*, pp63-78, 1973.

[14] S. Y. Lo, "A Scientific Model for Comparing Various Methods of Inputting Chinese Characters into Computer," *Computer Processing of Chinese & Oriental Language*, Vol.2, No.1, pp.36-57, 1985.

[15] S. Y. Lo, "Theoretical Limit on the Shortest Possible Input Codes for Chinese Characters," *Proceedings of 1985 International Conference on Computer Computing*.

[16] T. Y. Kiang, "Construction of Chinese Characters in Terms of Basic Strokes," *Proc. Int. Conf. of Chinese Language Computer Society*, 1982.

[17] T. Y. Kiang and T. H. Cheng, "A New Chinese Indexing System Based on Separation of Chacter Main and subordinate Components," *Proceedings of International Computer Symposium*, pp.130-141, 1982.

[18] Takahiro Sato, "The state of art on S-P analysis activites in Japan," C&S Systems Research Labs., NEC Corp., May.1984.

[19] Takahiro Sato, "The S-P chart analysis," C&C Systems Research Labs. NEC Corp., April.1986 .

[20] W. R. Shieh, M. S. Hwu, S. P. Lu, and K. Y. Cheng, "Natural Language Output for S-P Chart Analysis," Proceedings of National Computer Symposium, pp.221-229, R. O. C., 1987.

[21] 倉頡第三代中文輸入法, 全華科技 , 9-th edition, July, 1988.

[22] 郭婉容 , 總體經濟學 , 1985.