Six-Digit Stroke-based Chinese Input Method

Lai-Man Po, Chi-Kwan Wong, Yiu-Ki Au, Ka-Ho Ng and Ka-Man Wong

Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong

Email: eelmpo@cityu.edu.hk

***Abstract*—During the last three decades, more than one thousand Chinese input methods have been developed. However, people are still looking for better input methods in terms of easy to use, easy to remember, high input speed and small keypad implementation on handheld devices. The well-known stroke-based Chinese input method using only five basic stroke types could achieve low learning curve and small numeric keypad implementation but its input speed is limited for complex Chinese characters with a lot of strokes. To tackle this problem, simplified stroke-based Chinese character and phrase coding methods using (3+3) rules are proposed in this paper. The proposed method only uses the first 3 stroke codes and the last 3 stroke codes to represent the first and last radical information of the character for achieving lower average code length and higher hit rate of first character on the candidate list. To further enhance the input speed, a very user-friendly (3+3) phrase coding rule is also proposed for inputting Chinese phrases in terms of 2-character, 3-character and long-character phrases. Three special key assignment designs are developed for practical implementation of the proposed Chinese character and phrase input method using conventional QWERTY keyboard, PC's numeric keypad and mobile phone 12-key keypad. Experimental results have shown that the proposed character coding can achieve lower average code length and higher Hit Rate of First Character as compared with conventional stroke-based method and some well-known Chinese input methods. The proposed coding rules are also very easy to use and remember.**

Keywords*—Chinese Input Method, Chinese Phrase, Text Entry, Stroke Input Method, Six-Digit.*

I. Introduction

Chinese is an ideograph-based language with thousands of complex characters. One of the major subjects in "Chinese Computing" is about how to input Chinese characters into the computer efficiently. These methods are known as Chinese input methods [1-7] and over a thousand of Chinese input methods have been developed. Each method has its own strengths and weaknesses in terms of learning curve, input speed and handheld device applications. They can mainly be classified into four categories: (1) Arbitrary Codes, (2) Pronunciation-based, (3) Shape-based and (4) Numeric keys based.

A. Arbitary Codes Chinese Input Methods

In arbitrary codes methods, some arbitrary codes are assigned to the characters. For example, regional position code method (区位) which inputs Chinese characters using the Row-Cell numbers in GB2310-82 and Telex code input method (电报码) simply uses China's telex code developed in 1911. These methods require user to memorize the codes for each character and are obviously not good for general use.

B. Pronunciation-based Chinese Input Methods

The two most popular pronunciation-based Chinese input methods in mainland China and Taiwan are Pinyin (拼音) [3] and Zhuyin (注音), respectively. The Pinyin method simply uses the Pinyin table as its conversion dictionary. This method is very easy to learn for people who are already familiar with Putonghua (or Mandarin). For inputting the Chinese character of "汉", we can enter its Pinyin letters of "han" into the computer for starting the character input. However, we will get a list of 30 or more candidates to choose from, as there are a lot of Chinese characters with identical Pinyin string. It means Pinyin has very high identical code rate (or character-to-code collision rate). Statistics showed that the collision rate of Pinyin input method ranges from 3 to over 100 with an average of 17. Such high collision rate had limited the input speed of Pinyin method. To tackle this problem, several approaches were developed to make Pinyin-based input methods more efficient. As Putonghua has 4 tones, one approach is to specify the tone of the character to reduce the number of candidates. Another well-known approach to shorten the candidate list is to input more than one character at a time, for example, typing in phrases.

The second well-known pronunciation-based method is Zhuyin, which is a Taiwan counterpart of Pinyin. It is because Zhuyin symbols are commonly used to record Chinese characters in Taiwan. In Hong Kong, most people only speak Cantonese. Thus, there is also a Cantonese pronunciation-based Chinese input method (粤拼). In general, pronunciation-based methods are easy to learn if the phonetic symbols have been learnt and users are familiar with the pronunciations, but the input speed of pronunciation-based Chinese input method is still relatively low.

C. Shape-based Chinese Input Methods

To achieve high input speed, radical coding methods were developed, in which Chinese characters are structured in terms of radicals and strokes. They usually use fewer keystrokes than pronunciation-based input methods. The fastest input method of this kind is Wubi (五笔), which is very popular in Mainland China for simplified Chinese characters input. Wubi can encode some frequently used Chinese characters with just 2 codes and a maximum of 4 codes. Moreover, most of the common Chinese characters have unique Wubi codes, thus no further selection process is required after the input of Wubi codes. This is a very attractive feature of the Wubi for providing very high input speed. Advanced methods were extended to multiple characters (phrases) input with a maximum of 4 codes allowing a well-trained user to input over 100 characters per minute.

The counterpart of the Wubi for traditional Chinese character set is Cangjie (仓颉), which is very popular in Taiwan and Hong Kong. Similar to Wubi, Cangjie also uses the Roman letter keys to map the pre-defined Chinese character's radicals and strokes. The input speed of Cangjie is also very high for well-trained user but the major problems of these shape-based input methods are the difficulty in learning and the easiness to forget. Many users have experienced hardship when they start to learn these methods, since it is really hard to memorize the keystroke combinations and the complicated coding rules. To release this problem, Jianyi (简易/速成) was developed, which is a simplified Cangjie input method. Jianyi just uses the first and last radical codes of Cangjie and therefore significantly simplifies the coding rule of Cangjie. But it has a major drawback of much higher collision code rate. That means the users need to choose a desired character from a longer candidate list. In practice, whether Jianyi is really easier to learn is controversial, since it is difficult to figure out the Jianyi codes of many characters without understanding on the Cangjie decomposition rules.

D. Numeric Keys Based Chinese Input Methods

To improve the learning curve of the shape-based input methods as well as using small keypad of handheld devices such as mobile phones, numeric keys based methods were developed. The three most well-known examples of this type of input method are Q9 (九方), CKC (纵横) [4] and T9 Bihua Method (笔画). The Q9 input method uses 9 numeric keys (1 to 9) to input Chinese characters. It is considered as an easy to learn method but its input speed is quite low. The CKC also uses the numeric keys but with 10 keys (0 to 9) and a maximum of 4 digits to represent a Chinese character, which can achieve a much higher input speed than Q9. In CKC, all possible shapes of strokes are classified into 10 groups, each represented by one of the ten possible digits 0-9. Chinese characters can then be input by following the order in which the strokes are identified at the 4 corners of the character. Although CKC only uses ten keys, to remember all these groups of patterns are not easy because many radicals or strokes are mapped to a single key. In addition, users need to learn a new character structure decomposition method.

The most popular Chinese input method on mobile phones is the T9 Bihua input method, which is really easy to learn, as it is purely a stroke-based coding method. T9 Bihua has only five basic stroke types and its coding rule is based on the hand-written stroke order. However, quite a lot of Chinese characters have the same first radical, which makes some of the complex Chinese characters require a lot of keystrokes to input. This problem limits the conventional stroke-based input speed. In this paper, a new stroke-based Chinese character and phrase coding methods with at most 6 digits called diGi-6 (G6 or 六码笔画) method is proposed for speeding up the T9 Bihua. The new idea is very similar to the Jianyi that uses the first and last radical codes for enhancing the learning curve of Cangjie but the purpose of G6 is to speed up T9 Bihua. In addition, Chinese phrase coding rule is also proposed to speed up the multiple characters input further. The learning curve of G6 is very low and can achieve relatively high input speed. The details of these coding methods will be described in the following sections.

II. STROKE-BASED CHINESE INPUT METHOD

Before introducing the new G6 coding, we first review the conventional five basic stroke types that form the foundation of many stroke-based Chinese input methods. In which, the 24 strokes of simplified Chinese character set and 41 strokes of traditional Chinese character set are grouped into 5 basic types of strokes based on their common characteristics. These five stroke types are

(1) Horizontal strokes [横]: 一
(2) Vertical strokes [竖] : 丨
(3) Left-Falling strokes [撇]: 丿
(4) Right-Falling strokes or Dot strokes [点捺]: 丶
(5) Turning strokes [折]: フ

The detailed definitions of each stroke type are described as follows:

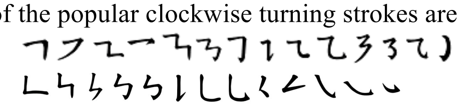
1. Horizontal strokes [一]：A stroke with its written direction from left to right horizontally or flicking up rightward slightly are classified as horizontal stroke type. The two popular horizontal strokes are "一 ㇀".

2. Vertical strokes [丨]：A stroke with its written direction from top to bottom is defined as a vertical stroke and the popular vertical strokes are "丨 亅". The reason to include the 亅stroke with small clockwise turning hook at the end of a normal vertical stroke is for avoiding the confusion of some of the Chinese characters with different vertical stroke written styles with or without hook at the end such as 示 and 示 .

3. Left-falling strokes [丿]：A stroke with its written direction from top right to bottom left is defined as a left-falling stroke. This type of strokes can be either long or short like "丿 ㇓".

4. Right-falling or Dot strokes [丶]：As both a right-falling stroke and a dot stroke have the same written direction from top left to bottom right, they are classified as the same type of stroke. Some of the most popular strokes of this type are "丶丶㇏㇏".

5. Turning strokes [フ]： A turning stroke has at least one turn no matter its direction is clockwise or anti-clockwise. Some of the popular clockwise turning strokes are



These five basic stroke types are easy to understand and remember. In practice, each of these stroke types is assigned to one of the keys on the small mobile devices keypad. Normally, they are assigned to numeric keys 1 to 5. The coding rule of the conventional stroke-based input method is based on the stroke order of the handwritten Chinese character. These five basic stroke types and handwritten stroke order of character are well defined in Chinese language. People familiar with Chinese language should know these five types of strokes and characters' stroke order when they were taught to write Chinese characters. That reduces the conventional stroke-based Chinese input method's learning curve significantly. However, for some complex Chinese characters with many strokes, it can be a time consuming job and the typing speed is highly decreased. For examples, each character of "输入" and "健康" can be encoded as

- 输：丨フ一丨一丿丶一丨フ一一丨丨
- 入：丿丶
- 健：丿丨フ一一一一丨フ丶
- 康：丶一丿フ一一丨丶一丿丶

For simple Chinese characters with a few strokes such as 入, the T9 Bihua is very efficient but for complex characters with many strokes such as 输, it requires a lot of keys input. This is because T9 Bihua uses variable code-length for encoding the characters that considerably reduce its input speed for complex characters with many strokes.

III. G6 Character And Phrase Encoding Rules

To solve the long code-length problem for encoding complex Chinese characters in conventional stroke-based input method, the proposed G6 method uses at most 6 code-length to encode Chinese character and phrase. In Taiwan, 5,401 Chinese characters are listed as frequently used characters. For code-length of 6 with base 5 number this could provide code space of 19,530 ($=5+5^2+5^3+5^4+5^5+5^6$), which have sufficient coding space for frequently used characters in both Traditional and Simplified Chinese character sets. To achieve higher input speed with a maximum code-length of 6, the proposed G6 Chinese input method uses (3+3) rule for both character and phrase coding. The character (3+3) rule refers to the first 3 and the last 3 strokes of a character. The first 3 strokes are the first, second and third strokes, and the last 3 strokes are third from last, second from last and the very last stroke. On the other hand, the phrase (3+3) rule uses the first 3 strokes of the first and last characters of the phrase. The details of the G6 characters and phrases encoding are described in the below two subsections.

A. *G6 Character Coding*

To encode Chinese characters, G6 divides Chinese characters into two groups with total number of strokes not greater than 6 and greater than 6. For the Chinese characters with the total number of strokes not greater than 6, the character is encoded like an ordinary stroke-based input method by entering all the stroke's codes. For example, the following are the G6 codes of some simple characters with the number of strokes not greater than 6：

- 大 ：一丿丶
- 小 ：丨丿丶
- 禾 ：丿一丨丿丶
- 己 ：フ一フ

If the number of strokes of a Chinese character is greater than 6, the (3+3) rule is used. This means to enter the first 3 and last 3 strokes. The following are the examples of the G6 codes of some characters with the number of strokes greater than 6：

- 健: 丿丨フ 丨フ丶
- 康：丶一丿 一丿丶
- 输：丨フ一 一丨丨

Apparently, we can see that G6 is very efficient in both complex and simple Chinese characters input.

Basically, the first 3 strokes are used to identify the side radical of the complex Chinese characters. It is found that most of the side radicals of the Chinese characters can be distinct by using 3 strokes, especially for the simplified Chinese characters set. In addition, the last 3 strokes are used to identify the characters with similar or same side radical. Suppose some complex Chinese characters with the same side radicals have about 400 characters. As we use the last 3 strokes with 5 types of stroke, we can achieve about 3 ($=400/5^3$) characters code collision. With this feature, it is easy to provide user-friendly interface to list the Chinese characters with all the identical code's characters for the final selection. This feature of the (3+3) rule can make G6 single Chinese character achieves relatively high input speed.

B. *G6 Phrase Coding*

To increase input speed, most of the Chinese input systems provide the associate character list when the first character is entered and then choose the associate characters from that list. In practice, the associate characters approach may not be very efficient as there may be a long list of associate characters for users to choose. Another way of improving the input speed is to input more than one character a time especially when typing in a phrase. G6 divides the Chinese phrases into 2-character, 3-character and multi-character phrases. The encoding method of G6 phrase is also very easy, which uses the (3+3) rule again. However, the first (3) refers to the first 3 strokes of the first character in the phrase, and the second (3) refers to the first 3 strokes of the last character.

2-Character Phrases

As two characters form most of the Chinese phrases, G6 use a separate database for all 2-character phrases. With use of the (3+3) rule the 2-character phrase coding is very simple by using the first 3 strokes for the first character and the first 3 strokes of the second character. Here are some 2-character Chinese phrases encoded by this rule:

- 输入: 一フ丨 丿丶
- 中国: 丨フ一 丨フ一
- 香港：丿一丨 丶丶一
- 健康：丿丨フ 丶一丿

Since the minimum length for the G6 single Chinese character coding is one, the minimum length of the 2-character phrase code is 2 and the maximum length is 6. In practice, a special phrase key is required for indicating that it is a phrase input instead of a character input.

3-Character Phrases

Since 3-character phrases are also very popular in Chinese language, G6 includes a 3-character phrases input feature with the same (3+3) coding rule as the following examples:

- 中國人: 丨フ一丿丶
- 输入法 : 一フ丨 一フ丶
- 经济学 : フフ 一丶丶丿

Long-Character Phrase

Any Chinese phrase composes of more than 3 characters is considered as a long phrase and here are some long-character Chinese phrases encoded by (3+3) phrase coding rule:

- 一模一样 : 一 一丨丿
- 身体健康 : 丿丨フ 丶一丿
- 中华人民共和国 : 丨フ一 丨フ一

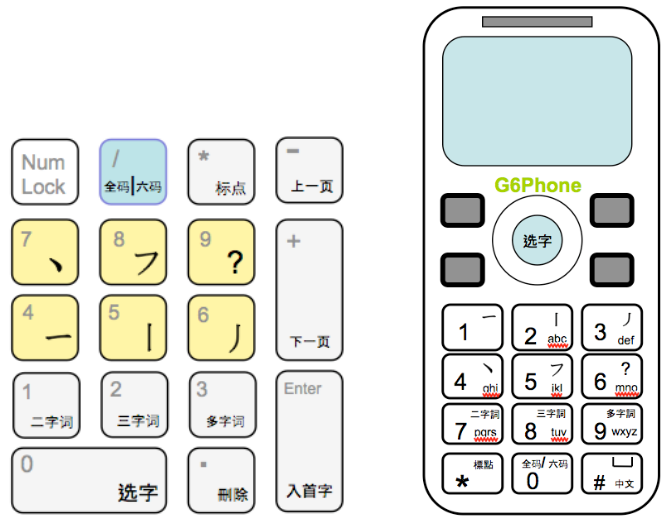
The G6 phrase input feature significantly increases the input speed. It can be seen that 36 codes are required to input the 7 characters phrase of "中华人民共和国". With G6 phrase input, it can be reduced to 6 codes, which can achieve less than one codeword per character. In addition, this phrase encoding method is very easy to use and remember as only one rule for all three lengths of phrases. In CKC, different phrase coding rules are defined for different length of phrases. On the other hand, the first 3 strokes of first and last characters of the phrase coding rule is very easy to use, as the first 3 strokes of Chinese character is very easy to identify and remember.

IV. Implementations of G6 Input Method

In practice, G6 Chinese input method can be implemented in both small numeric keypad and Roman keys of the QWERTY keyboard. In this section, the numeric keypad, mobile phone keypad and Roman keys implementation of G6 are proposed.

A. G6 Numeric keypad Design

The most convenient way to implement G6 is to use the personal computer's numeric keypad and a recommended G6 numeric keypad design is shown in Fig. 1(a). The five basic stroke types are assigned to keys "4" to "8" and the key "9" is assigned as a "Wild Card Key" with symbol of "?" for unknown stroke input. The "Enter" key is assigned for selecting the first candidate of the characters list (入首字). It is because the proposed G6 have relatively higher hit rate of first character which will be discussed in section V, then user can use this big key to input the top character of the candidate list easily. If the target character is not listed at the top, the "0" key is assigned as "Candidate Selection Key" (选字). User can press this key to select any character listed on the candidate list using of the numeric keys of 1 to 9 or using the



(a) (b)

Fig. 1: G6 Keys Assignment for (a) Personal Computer Numeric Keypad and (b) Mobile Phone 12-key Keypad.

Fig. 2: G6 keys assignment for QWERTY keyboard.

key of "+" (下一页) and "-" (上一页) to page down and up of the candidate list. In addition, the "0" key also works as "Space" key for input a space between characters. Moreover, the "*" key is assigned as punctuation (标点) and the "." key is assigned as "Delete last code key" (刪除). To distinguish the G6 phrase codes from the character codes and length of the phrases, the "1", "2" and "3" keys are assigned to "2-Character Phrase" (二字詞), "3-Character Phrase" (三字詞) and "long-Character Phrase" (多字詞), respectively as shown in Fig. 1(a). Finally, as an enhancement of the full strokes Chinese character input method, the G6 input system can also support the full code input (input all the stroke codes), thus a "mode swap" key is assigned to "/" (全码|六码) for swapping between full code and G6 code of the first 3 and last 3 coding rule. Similarly, G6 can also be applied to 12-key mobile phone keypad as shown in Figure 1(b). Positions of the keys assignment are very similar to the numeric keypad but the key's numbers are different. In addition, the "Character Selection Key" is normally assigned to control buttons center key. This design can let users input the Chinese character and phrase with one hand.

B. G6 QWERTY Keyboard Design

On the other hand, G6 is also very efficient in QWERTY keyboard using Roman keys assignment and the layout of G6 QWERTY keyboard design is shown in Fig. 2. In which the basic five strokes keys and the "Wild Card Key" are assigned to keys of J, K, L, U, I and O, which can be pressed by the more agile three fingers (index, middle and ring fingers) of the right hand. Moreover, the major control keys of "Select Character", "Mode Swap", "Punctuation", "2-Character Phrase", "3-Character Phrase", and "Long-Character Phrase" are assigned to "Space Bar", Y, P, H, N and M keys, thus they can also be accessed with just the right hand. Such keys arrangement can let users enter Chinese character easily with single hand. For unpopular Chinese characters, users can use the numeric keys of "1 to 9" on the top of the keyboard to select character from the candidate list with identical G6 code.

C. G6 Implementation on MS-Windows Platforms

Recently, G6 is implemented on MS-Windows OS platform using C++ language and it is available for download in [11]. This implementation supports both Roman keys and numeric keys assignments as shown in Fig. 1(a) and Fig. 2. In addition, the software supports both simplified and traditional Chinese characters with 20,902 and 13,117 characters respectively. Two examples of the G6 user interface for entering the character of "树" and "翡" are shown in Fig. 3. In which, the character of "翡" has unique G6 code but it cannot use 6 code-length to uniquely identify this character using conventional stroke order with the first 6 stroke codes. The G6 phrase coding were also implemented with 65,536 2-character phrases, 37,677 3-character phrases and 52,502 long-character phrases for simplified Chinese phrase input. For traditional Chinese input, the system consists of 65,536 2-character phrases, 37,675 3-character phrases, and 52,567 long-character phrases. Three G6 phrases input examples are shown in Fig. 4 for the phrases input of "检验", "经济学" and "中华人民共和国". These examples illustrated that some of the frequently used phrases can be listed on the first page of the candidate list, making phrase input very efficient.

V. PERFORMANCE EVALUATIONS

In order to evaluate the efficiency of G6 (六码笔画) input method, we have performed three types of performance comparison with T9 Bihua (T9笔画), Pinyin (拼音), Wubi (五笔), Zhuyin (注音), Cangjie (仓颉), Jianyi (简易) and CKC (纵横) in both simplified and traditional Chinese character sets. In the simplified Chinese characters set evaluation, G6 was compared with T9 Bihua, Pinyin, Wubi and CKC using the most frequently used 3,000 characters created by Peking University of China with ordering based on their usage rate. For traditional Chinese characters set, G6 was compared with T9 Bihua, Zhuyin, Cangjie and Jianyi using the most frequently used 3,000 traditional Chinese characters generated by the Chinese University of Hong Kong [8]. The first comparison is based on the average length of all input codes as defined by:

$$ALIC = \frac{\text{The sum of the length of all Chinese character's input code}}{\text{The sum of all Chinese characters in code table}} \quad (1)$$

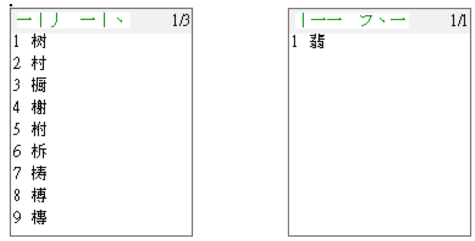


Fig. 3: The user interface of the G6 MS-Windows implementation for entering the characters of "树" and "翡".

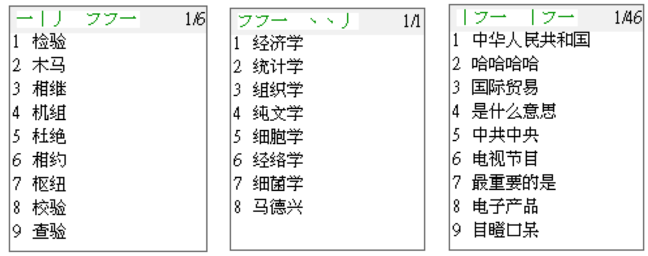


Fig. 4: The user interface of the G6 MS-Windows

The ALIC means the average length of all input codes, so the lower the ALIC is, the faster the input method should be. The second column of Table 1 listed the ALIC of the input methods used in our analysis, which shows that G6 can reduce the ALIC from 9.388 of simplified characters set and 11.473 of traditional characters set to 5.798 and 5.859, respectively. G6's ALIC is longer than the Pinyin, Zhuyin, Wubi, Cangjie and CKC by 2 but the ALIC may not be a good parameter for comparing the input speed of G6 with these non-stroke-based Chinese input methods. It is because the actual number of keys pressed to input the character couldn't be reflected by ALIC if the character-to-code collision rate is very different.

In order to achieve better input speed estimation, the second evaluation used the HRFC (Hit Rate of First Character) [9]. The HRFC is the probability of the target Chinese character appears in the first place in the input method's candidate list. The HRFC is computed as follows:

$$HRFC = \frac{\text{The sum of the inputted characters in the first place}}{\text{The sum of inputted characters}} \quad (2)$$

The higher the probabilities of the expected characters in the first place are, the more convenient the user is, the memorizations of the characters' position can be reduced and the press of the selection key may be saved. The third evaluation was to compare the KSPC (Key Strokes Per Character) [10], which is the number of keystrokes, on average, to generate each character of text in a given language using a given text entry technique. The KSPC is computed as follows:

$$KSPC = \frac{\sum K_c \times F_c}{\sum C_c \times F_c} \quad (3)$$

where K_c is the number of keystrokes required to enter a character, $C_c = 1$ (the size of each character), and F_c is the frequency of the character in the corpus. It is the average number of key presses required for entering a character, which can directly represent the efficiency of the input method and commonly used for input speed method performance evaluation. In our experiments, T9 Bihua and G6 both input the strokes one by one due to the dynamically result displaying characteristic, until the maximum code length have been reached. When the target Chinese character has appeared in the first page of the resulting character candidate list, it will be assumed a selection key is pressed to select the target character. In which 9 characters per page is used for displaying the candidate list. While the other compared input methods were simply entered the full code directly. If full code was entered but the target character was not on the first page of the resulting candidate list, additional page down keys will be added until the target character has been found. The experiments only considered single characters input and assumed that all the special features, like associate character and phrase input, were turned off. Since the maximum code length of T9 Bihua can be very long, we limited it to 6 for making the analysis fairer and more accurate. Because Wubi, CKC, Pinyin and Zhuyin have multiple records for a single character which would affect the KSPC value, so only the first record have been counted when calculating KSPC, but ALIC and HRFC would be calculated as normal.

From Table 1, we can find that G6 can achieve a much higher HRFC than T9 Bihua and the two pronunciation-based input methods of Pinyin and Zhuyin. G6's HRFCs are 12% and 14.5% better than T9 Bihua in simplified and traditional Chinese character sets, respectively. The KSPCs of G6 and T9 Bihua are very similar while G6 is a little bit better. Compared to Pinyin and Zhuyin methods, G6 can also improve HRFC by about 50% and 21% for simplified and traditional character sets, respectively. The KSPCs of these methods are very similar. These comparisons show that G6's input speed is good as compared to these low learning curve Chinese input methods. On the other hand, comparing to shape-based Chinese character input methods or high learning curve methods, G6 achieves about half HRFCs of Cangjie and Wubi, but its KSPC is only 10% higher than Cangjie and 29% higher than Wubi. Comparing to CKC, G6's HRFC is 23.1% higher but with higher KSPC. Finally, G6's HRFC is 33.5% higher Jianyi but its KSPC is higher. These results show that G6's input speed should lower than Wubi, Cangjie and CKC but it may achieve higher input speed comparing to Jianyi due to much higher HRFC.

VI. CONCLUSION

In this paper, new Chinese character and phrase coding methods using the conventional 5 basic stroke types are proposed for improving the input speed of the conventional stroke-based input method using variable code length. The new coding is called G6 as at most six digits are used to encode both Chinese characters and phrases using two (3+3) coding rules. The G6's (3+3) single character coding rule is defined by using the first 3 strokes and the last 3 strokes of the Chinese character. While the G6's (3+3) phrase coding rule is defined by using the first 3 strokes of the first and last characters of the phrases. The G6 character coding achieves a much higher HRFC than T9 Bihua in both simplified and traditional Chinese character sets evaluations. The G6's phrase coding can achieve very low keystrokes per character by multiple characters input. In addition, both of the character and phrase coding rules are very easy to remember and use. The proposed G6 coding scheme can be applied on both PCs with large QWERTY keyboard and handheld devices with small 12-key numeric keypad. A G6 based input system was successfully implemented on MS-Windows OS platform with satisfactory performance.

Table 1. The comparison results of ALIC, HRFC and KSPC

Simplified Chinese Character Set			
Input Method	ALIC (codes)	HRFC	KSPC (keys)
G6 (六码笔画)	5.798	61.4%	5.186
T9 Bihua (T9笔画)	9.388	49.4%	5.401
Pinyin (拼音)	3.063	11.8%	5.221
Wubi (五笔)	3.341	99.2%	4.017
CKC (纵横)	3.001	38.3%	4.240
Traditional Chinese Character Set			
Input Method	ALIC (codes)	HRFC	KSPC (keys)
G6 (六码笔画)	5.859	52.8%	5.405
T9 Bihua (T9笔画)	11.473	38.3%	5.952
Zhuyin (注音)	3.019	31.8%	4.416
Cangjie (仓颉)	3.891	98.3%	4.917
Jianyi (简易)	1.992	19.3%	3.107

ACKNOWLEDGMENTS

We would like to express our earnest appreciation to our undergraduate students, Ka-Hang Ma, Pak-Yu Yau, Hiu-Yi Li, Shu-Pui Wan, Cheuk-Yin Tang, Yan-Fat Liu, Kwok-Wai Lam, Sze-Ho Pang, Suet-Ying Chan, Chun-Chung Ng and Siu-Lun Tang, for helping us on analyzing G6 performance and implementing the G6 software in different platforms.

REFERENCE

[1] L. Wellington and C.P. Yu et al., A historical advancement of Chinese language computing, Computer Processing of Chinese and Oriental Languages vol. 4 (No. 1) (1988), pp. 57–81.

[2] Z Chen and K-F Lee, "A new statistical approach to Chinese Pinyin input," *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, pp. 241-247, 2000.

[3] T. Ooka and M.H. Chien, The practical application of the input method converting PINYIN to characters, Proceeding of International Conference on Text Processing with a Large Character Set (1983), pp. 131–136.

[4] Q Zhu, P Li, P Gu, P Qian, "A Chinese Mobile Phone Input Method Based on the Dynamic and Self-study Language Model," LECTURE NOTES IN COMPUTER SCIENCE, 2006 – Springer.

[5] H.Y. Gu, A Chinese-character inputting system using a new type of phonetic keyboard and a compound Markov language model, Proceeding of ROCLINGC (1994), pp. 253–262.

[6] J Qiao, Y Qiao and S Qiao, "Six-Digit coding Method," Communication of ACM, vol. 33, No. 5, pp. 491-494, May 1990.

[7] S C Hsu, "A Flexible Chinese Character Input Scheme," UIST'91, pp. 195-201, Nov. 1991.

[8] http://humanum.arts.cuhk.edu.hk/Lexis/lexi-can/faq.php

[9] P. F. Li, P. Gu, Q. M. Zhu, "A Dynamic and Self-study Language Model Oriented to Chinese Characters Input", Proceedings of the Seventh ACIS International Conference, 2006.

[10] I. Scott MacKenzie, "KSPC (Keystrokes per Character) as a Characteristic of Text Entry Techniques".

[11] http//www.ee.cityu.edu.hk/~g6code/