THE APPLICATION OF THE SIMILARITIES BETWEEN THE MORPHEMES OF THE ENGLISH AND CHINESE LANGUAGES TO REPRESENT CHINESE CHARACTERS PHONETICALLY WITH ENGLISH LETTERS TO FACILITATE COMPUTER APPLICATIONS MANUALLY AND BY VOICE WITH THE CHRACTER-BASED LANGUAGES CHINESE, JAPANESE AND KOREAN

Dr. Stanley K. Chan
Research Institute for Computer Application of East Asian Languages
7798 Starling Drive, Suite 207, San Diego, CA 92123, USA
Telephone: (619) 571-5285 Fax (619) 571-5992
E-mail stanleykchan@msn.com

ABSTRACT

This paper presents a new methodology for entering, manually or by voice, Chinese characters into a computer and processing them with the same ease as English words by utilizing the similarities between the morphemes of the Chinese and English languages to represent Chinese characters and radicals phonetically and unambiguously with English letters or local phonetic symbols such as Zhuyin Zimu, Hiragana and Hangul.

1. Problems With The Use Of Chinese Characters In Computer Applications

Unlike English, the written form of the Chinese, Japanese and Korean languages contains unique square-shaped symbols, called Chinese characters (actually Hanzi for Chinese, Kanji for Japanese, Hanja for Korean pronunciation. "Chinese" will be used throughout this article for clearer presentation.), that present significant obstacles to their use in computer applications. Firstly, input of such characters is daunting as it is impractical to make a keyboard with the thousands of keys that would be needed to represent the Chinese character set. Secondly, numerous instances exist in which many characters share an identical pronunciation.

Historically, the limiting factor in developing a satisfactory system for input of Chinese characters using English alphabet letters has been the existence of multiple groups of large numbers of homonymous Chinese characters. The input of a string of English letters frequently fails to provide an unambiguous reference to the desired Chinese character, generating incorrect Chinese characters for some strings of English letters.

Inadequacies in existing programs that attempt to solve this problem make them less efficient than writing Chinese characters by hand. They display all the homonymous characters, ten at a time, upon entry of a given set of letters, requiring that the typist scroll through the homonymous characters to search for the desired character. This time-consuming process is tiring to the eye and prone to errors since the typist may have to search as many as 250 homonymous characters to make a single correct selection.

2. Multiple Groups Of Numerous Homonyms

The average number of homonyms per pronunciation is 30 if we use 13,000 as a typical number of commonly-used Chinese characters. Theoretically, this problem should be reduced if intonation is taken into consideration, reducing the number of homonyms per pronunciation to 7.5. In practice, however, it is a different story. This is due to the fact that the Chinese people seem to favor pronunciations ji, qi, xi, yi, wu, yu, zhi, chi, shi, zi, ci, si, and a few others. These preferences generate over 100 groups of 40 or more homonyms[1]. This phenomenon requires a search process for the desired character when inadequate data (typically the pronunciation or pronunciation/ intonation combination of a character) is used as input.

The most effective way to address this problem, for the time being, seems to be the charactword input method discussed later in this paper.

[1] Please see Appendix 1. Not all groups with 40 or more homonyms are included. Groups with fewer than 40 are not included at all. Zhuyin Zimu with * indicates that the appropriate Zhuyin Zimu is not available in the database due to the fact that Mainland China has incorporated it into the one used in the chart.

3. Surprising Similarities Between Chinese Characters And English Words

Chinese characters can be seen to contain analogs to the prefixes and suffixes of English words. Although these components of Chinese characters are called different names and are treated differently in the rules of pronunciation, these "prefixes" and "suffixes" have been an integral part of written East Asian languages for thousands of years. Scholars call these Chinese character components "radicals," meaning "parts of a Chinese character." Radicals are systematically taught in Japan and Taiwan alongside Chinese characters and phonetic signs, namely Katakana and Hirogana in Japan and Zhuyin Zimu in Taiwan[2].

Generally, radicals fall into two categories. In one category, the radical resembles a shape or bears a meaning related to the character of which it is a part. When such a radical is related to shapes, it is called the "pictographic radical," for example, "木" , (pronounced "mu," in Chinese, meaning "wood" or "tree.") When a radical is related to meaning, it is called an "ideographic radical." We shall refer to radicals of these types as "P/I" radicals. Radicals in the second category denote the approximate or exact pronunciation of characters. These are phonetic radicals such as "每 "(pronounced "mei" in Chinese). The combination of a P/I radical and a phonetic radical makes a character. For example, combining the aforementioned radicals creates the character 梅 , meaning "plum." This combining of radicals to make a character is analogous to creating an English word by combining a prefix and suffix.

4. Application Of "Charactwords" For Manual And Voice Input Of Chinese Characters

If we use English letters to spell out the P/I radicals according to their pronunciations, using them as *prefixes*, and do the same with the phonetic radicals using them as *suffixes*, we can form "words" for the Chinese, Japanese and Korean languages. For example, combining the P/I radical 木 (spelled "mu") and the phonetic radical 每 (spelled "mei") gives us the "word" for character 梅 , spelled "mu-mei." This provides a simple methodology for using these created "words" for computer input of Chinese characters. We shall refer to these newly-created words as

[2] The following sets of phonetic signs are used by the Chinese, Japanese and Koreans respectively: Zhuyin Zimu, Hiragana or Katakana, and Hangul. Only Mainland China officially uses a Latin style alphabet to aid teaching the pronunciation of Chinese characters.

"charactwords" whose function it is to represent Chinese characters for easy computer input and processing. For voice input, all relevant parts of a charactword should be pronounced to provide all the data necessary to unambiguously represent the desired character. This method can be implemented cost-effectively as it does not require a large database, complex programming or strict noise control in the immediate working area.

5. Charactwords Are Actually Chinese Characters Expressed In The Form Of Alphabetic Letters Or Phonetic Signs

Regardless of which written phonetic signs are used to spell out Chinese characters, whether Hangul, Kana, Zhuyin Zimu, or a Latin-based alphabet, the charactwords so created remain logically and theoretically Chinese characters, with P/I radicals represented by the prefixes and phonetic radicals represented by the suffixes. The failure of existing programs to provide a satisfactory solution to the input of Chinese characters arises because such programs use insufficient data to describe the characters. These programs incorporate only one-half of the CHINESE character data available, building their databases solely upon pronunciation (and, sometimes the intonation) data while ignoring the equally important data related to the meaning of the radical and the shape of the character. By contrast, using a charactword database which contains all the relevant character data[3] yields an efficient solution. Further, with such a database, programming a computer to display the correct Chinese character upon input of the appropriate charactword becomes a simple task.

6. Markers And Intonation Indicators Further Differentiate The Charactwords From One Another

Chart 1 illustrates how a system of markers and intonation indicators differentiates the spelling of characters sharing the same P/I radical and pronunciation. Markers are input by simply repeating the last key of the charactword spelling as many times as necessary (although seldom more than four) to uniquely identify or represent the desired Chinese character. In a sense, the markers give each charactword a distinct look. Additionally, intonation indicators

3 Please refer to Chart 1 on the next page for the *complete* version of the charactword for Chinese characters used in Chinese, Japanese and Korean languages.

required for the Chinese language are incorporated in the form of the first four Chinese Mandarin consonants[4] appended to the spelling of a charactword, representing the first, second, third, and fourth intonation, respectively[5]. Together with the prefixes and suffixes, the markers and the intonation indicators can be used to create not only unique charactwords for all characters, but also charactwords for all the different ways a character can be pronounced. Please refer to subsection 9 "A Charactword Can Be Tailored For All Different Ways In Which A Chinese Character Is Pronounced" for details.

Chart 1

How Markers and Intonation Indicators Help To

Unambiguously Differentiate Complex Homonyms

Row number	The country where the characters are homony-mous	Homony-mous Chinese charact-ers	Corres-pond-ing charact-word	Intonation indicators (bold letters) added to both prefix and suffix of the charactwords	Markers (italic letters) appended to the charact-words to unambiguously differentiate them from one another
1	China	試	yan-shi	yanp-shif	yanp-shif ㄧㄢˊㄕˋ
2	China	誓	yan-shi	yanp-shif	yanp-shif*f* ㄧㄢˊㄕˋㄈ
3	China	諟	yan-shi	yanp-shif	yanp-shif*ff* ㄧㄢˊㄕˋㄈㄈ
4	China	謚	yan-shi	yanp-shif	yanp-shif*fff* ㄧㄢˊㄕˋㄈㄈㄈ
5	China	謚	yan-shi	yanp-shif	yanp-shif*ffff* ㄧㄢˊㄕˋㄈㄈㄈㄈ
6	China	讖	yan-shi	yanp-shif	yanp-shif*fffff* ㄧㄢˊㄕˋㄈㄈㄈㄈㄈ
7	Japan	杠	ki-ko	not applicable	ki-ko き* こ
8	Japan	杞	ki-ko	not applicable	ki-ko*o* き* こ こ

[4] In Mandarin Chinese (the official dialect of Mainland China, Taiwan and Singapore,) no pronunciation of a Chinese character ends with one of the first four consonants, so the presence of one of these consonants at the end of the spelling of a charactword can be readily recognized as an indicator of intonation.

[5] The indicator for the first intonation is traditionally omitted, therefore only the second, third and fourth consonants are used for the corresponding intonations.

9	Japan	杭	ki-ko	not applicable	ki-k*ooo* き* こ こ こ
10	Japan	枯	ki-ko	not applicable	ki-k*oooo* き* こ こ こ こ
11	Korea	杞	muk-ki	not applicable	muk-ki 목*기
12	Korea	枝	muk-ki	not applicable	muk-k*i* 목*기
13	Korea	棋	muk-ki	not applicable	muk-k*ii* 목*기
14	Korea	機	muk-ki	not applicable	muk-k*iii* 목*기

7. Explanation Of Chart 1

Rows 1 - 6 contain six Chinese characters with an identical P/I radical, pronunciation and intonation. Column 6 shows how markers, indicated in italics, unambiguously differentiate all six charactwords. Similar results are shown for Japanese characters in rows 7 -10, and for Korean in rows 11 - 14. National or local phonetic symbols are hand written alongside the Latin-based alphabetic letters in column 6.

The rule for assigning markers is based upon the number of pen strokes comprising a character: the character with the least pen strokes receives no marker, and the number of markers increments one at a time in relation to the number of additional pen strokes comprising a character, with the most markers assigned to the character made up of the most pen strokes. When two or more characters have the same number of pen strokes, the order of appearance for the Chinese language is based upon their order of appearance in Kang Xi Zi Dian (Emperor Kang Xi's Dictionary, 康熙字典), which is the case for the characters in rows 3 and 4. Similarly, The Modern Reader's Japanese-English Dictionary (最新漢英辭典) is used for Japanese, and for Korean we have selected 新活 用玉篇that does not have an English title. The number of markers becomes a non-issue with our careful software design. Please refer to subsection 11 for details.

8. Input Of The Prefix Is Efficient

Many of the 214 P/I radicals can be input with one or two keystrokes. For example, 木 will

appear on the screen by typing just ㄇ (the full spelling is ㄇㄨ) in Zhuyin, or "m" (the fullspelling is "mu") in the Pinyin system. Voice recognition capability could further simplify charactword input. An advantage of the charactword system is that most of the P/I radicals are commonly-used Chinese characters, and all have associated pronunciations. Only 38 of the 214 P/I radicals are not phonetically recognized by a typical high school graduate in Taiwan[6]. In practical computer input applications, any necessary reference can be provided by a large card at the computer terminal that displays all 214 radicals and highlights the 38 uncommon radicals.

9. A Charactword Can Be Tailored For All Different Ways In Which A Chinese Character Is Pronounced

There are occasions when a character can be pronounced in more than one way. An example is the Chinese character, 好, which is ordinarily pronounced "hao*m*[7]," meaning "good." When it is pronounced as "hao*p*[8]" preceding an adjective, it means "very." It can also be pronounced "hao*f*[9]", meaning "fond of" or "hobby, "depending on the context. When "nyu" (or "nu", the pronunciation of the P/I radical, 女, for 好,) and "haom", "nyu" and "haop" or "nyu" and "haof" are entered, 好 will be displayed. This is because in our database, either pronunciation (hence the charactword) is linked to the same character.

The same technique can be applied to the Japanese language where some Kanji characters can be pronounced in as many as 9 different ways. For an example, 鸚can be pronounced as "ryu, ru, bo, hyo, mu, kyo, gu, ryo and hibari." By collecting all different charactwords representing all the different pronunciation in our database, character, 鸚, will be displayed when one of the following sets is typed into the computer equipped with our software: "tori[10]" and "ryu"; "tori" and "ru"; "tori" and "bo"; "tori" and "hyo"; "tori" and "mu"; "tori" and "kyo"; "tori" and "gu"; "tori" and "ryo"; "tori" and "hibari".

6 According to our random survey of the Taiwanese students studying in the San Diego area.

7 Letter "*m*" is the indicator for the third intonation. Please refer to footnote 5 in subsection 6 for details.

8 Letter "*p*" is the indicator of the second intonation. Please refer to footnote 5 in subsection 6 for details.

9 Letter "*f*" is the indicator for the fourth intonation. Please refer to footnote 5 in subsection 6 for details.

10 "Tori" is the Japanese pronunciation of P/I radical 鳥, the radical of character 鸚。

10. Charactwords Inherit All The Data Of Chinese Characters

As discussed earlier, there are four kinds of data in most Chinese characters: 1. pronunciation; 2. intonation; 3. meaning; and 4. shape or construction.

It is easy to see that a charactword might contain the first three of the four types of data, but it is a little bit difficult to see how a string of alphabetic letters can denote the "shape" of a character, until we run into Complex Homonyms (CH).

Complex Homonyms are characters that share the identical P/I radicals, pronunciation and intonation, such as 誓試謚 諟 謚識 . As all of us can see, the clever ancient Chinese used the "shape difference" to unambiguously differentiate each character from the others. We can do the same with a charactword by adding marker(s) to it, one at a time, making each and every charactword look different from one another, much as English words "knight" and "night".

11. The Wonders Of Software Eliminate The Guess Work: No Need To Depend On Memory Or Dictionaries

This leads to a couple of questions: how can I remember how many markers there are in the charactword for character "諟"? Do I have to refer to my Chinese dictionary all the time? The answer? You do not need to memorize the number of markers for any charactword at all, nor do you need to use your dictionary. That is because through our software design, all members of a group of Complex Homonyms can be displayed on the screen once the charactword for a member of the group is entered. In this extreme example of 誓試謚諟 謚 識 , once "yanp" and "shi" are typed in the environment of the software, ***all*** of these characters are displayed. Suppose you want to use "諟", after typing "yanp" and "shi", all you need to do is to repeatedly type "i" three more times until the highlight moves onto it. (This does not take much time since your right middle finger is already on that key). Then hit the space bar, and, " 諟" will be in your text. All Complex Homonyms will be handled this way in our software. Although there are groups of CHs, there are not as many members in each group as shown in this

extreme example, especially with just the commonly-used characters. Leave your dictionary on the shelf when you use our software. You don't need it.

12. Typing Of The Charactwords Can Be Simplified To Empower Non-Career Typists To Achieve The Typing Speed Of A Professional

At first glance, the numerous "markers" in the charactwords seem to require a lot of typing. Closer examination reveals this is not so. A marker is created by the repetition of the last key of the charactword, so the typist does not need to move his/ her finger- he/ she can just keep pressing the last key, moves the highlight to a different character, once per keystroke. The examples in Chart 1 are rare and extreme. They are intended to demonstrate the usefulness of markers in creating unique charactwords to unambiguously represent each one of a group of complex homonymous characters, which is otherwise very difficult to accomplish. Our experience shows that more than 90% of the characters in Chinese can be unambiguously differentiated with just the regular charactwords. Here are some examples: each of the characters 基(ji),欺(qi),稀(xi),醫(yi)[11] has over 140 homonyms[12], but by first typing their P/I radicals, 土(tu),欠(qian),禾(he), 酉(you)[13], respectively,then the Pinyin of the characters, you will have precisely these four characters on the screen, eliminating more than 140 homonyms for each of them. The time it takes to type the markers, the intonation indicators and even prefixes will become a non-issue when users become well acquainted with the charactwords and move to the next stage of inputting described below.

With careful software design and a broad database, the input of charactwords can be greatly simplified. For an example, most of the Chinese idioms (成語) containing four or more characters can be input with an average of just 1 to 1.25 keystrokes for each character, using the acronyms of their corresponding suffixes. For the speech parts such as nouns, pronouns, verbs, adjectives and adverbs consisting of two characters, a modified but theoretically sound acronym

11 The English letters in the parenthesis are actually the Pinyin letters denoting the pronunciation of these characters.

12 The actual numbers are over 220, 140, 180 and 250 respectively. Please refer to Appendix 1 for details.

13 The English letters in the parenthesis are actually the Pinyin letters denoting the pronunciation of these P/I radicals. Work has begun to reduce the number of keys required for each P/I radicals. It looks promising that the number of keys for the most commonly used P/I radicals, such as 人, 心, 手, 金, 木, 水, 火, 土, 日, 月, 石, 艸 etc. can be reduced to just one keystroke.

system of their suffixes and prefixes can be used to input them at an average of 4.5 to 5 keystrokes per character. These can also be applied to voice input. Whether inputting manually or by voice, an average East Asian can enter more than 1,000 characters per hour into a computer, a satisfactory speed for most of us[14]. When people become accustomed to the charactwords, the typing speed will increase to about 3,000 characters per hour, rivaling the speed of today's career typists[15].

13. Typing Speed Of Non-career Typists Can Be Calculated

I wish to express my gratitude to the members of ROCLING X Program Committee for bringing to my attention that the contention that the average Chinese can type 3,000 characters per hour is controversial. Because it is common knowledge that non-career typists can write Chinese characters much faster than they can type, further analysis is called for to accurately model the typing speed ranges. Our latest investigation revealed that an experienced[16] American typist can type 45 words per minute. An average English word consists of 6[17] alphabetic letters. This translates into 270 keystrokes per minute.

A charactword consists of an average of 4.5 to 5 Zhuyin Zimu[18]. Assuming an average "experienced" Chinese can type as fast as his/her American counterpart, it will mean that he/ she can type about 54 Chinese characters per minute. The hourly speed would, therefore, be 3,240 characters.

All skills improve with time and practice. Typing is no exception. The speed of "typing" Chinese characters via charactword-based input method will increase as time goes by because all impeding elements are eliminated.

Since our prototype uses Zhuyin Zimu, we are currently unable to establish the speed for typing the Pinyin alphabets. We shall conduct a test to determine the speed for typing Pinyin

[14] An average person is one who is semi-familiar with keyboard layout and can type 1.75 keys per second. At an average rate of 5 keystrokes per character, more than 1,260 characters can be entered into a computer per hour by an average person.

[15] Most of the people can type an average 4.5 to 6 keys per second after using the keyboard consistently for about 3 to 6 months. At that speed, one can enter more than 3,600 characters into a computer per hour.

[16] "Experienced" means being familiar with the keyboard layout *and* having 480 hours of practice in typing.

[17] Most experts say that the average letters in an English word is 7. I use 6 in order to be conservative.

[18] This takes into account of the abbreviation of the P/I radicals and two-character vocabularies, but not the four-character idioms.

once our database for Pinyin is complete. It is, however, expected to be somewhat slower than typing Zhuyin Zimu (when typing individual characters,) because some of the latter frequently requires fewer keystrokes to represent characters. The typing speed will be the same for both Zhuyin and Pinyin when inputting vocabularies of two or more characters.

14. The Advantages Of Our Software

. The most important aspect of our software is the theory upon which it is based. As previously mentioned, our software is based on charactwords that contain all relevant data needed to describe Chinese characters including the various ways to pronounce them. This means that the pronunciation, intonation, meaning, and even the shape of any given Chinese character is taken into consideration in the making of a charactword. A Chinese character is unique because of its appearance, only one of the four data used to create a charactword. This is why a charactword can unambiguously represent the character or its various pronunciations.

This completeness of data in the charactwords makes our software faster in producing the desired character and easier to use by the general public. It is faster, because it eliminates searching for the desired character among scores or even hundreds of homonyms. It is easier to use for two reasons. One, the charactwords resemble the characters in three ways: the pronunciation, the intonation and the meaning. Two, charactwords are more logically organized, with the meaningful part as the prefixes and the phonetic part as the suffixes. By using markers, the charactwords can even have unique shapes. This built-in familiarity to the East Asian public is especially true in Taiwan because the P/I radicals are systematically taught beginning in elementary school.

As discussed earlier, our software also has a very complete database that includes special, yet logical and easy-to-understand coding of two, three, four or multiple-character vocabularies, that can be input with far fewer keystrokes than many individual characters. It also incorporates coding of various ways a character can be pronounced as discussed in subsection 9.

The most popular Chinese software in the US is Twinbridge, according to my research. Twinbridge is a versatile software product that provides many ways of entering a Chinese

character. The one that compares closest to ours is Zhuyin. In entering individual characters, it takes an average of 22.5 seconds to produce a commonly-used character with an "untrained[19]" version of the software. It took only 2.1 seconds with our "prototype" software operated by an "inexperienced" typist. Twinbridge fared better with two-character vocabularies input than single character input, but ours can do better[20]. With idioms of four or more characters, ours can generate a character with less than one second per character[21] theoretically.

Appendix 2 summarizes the comparison of these two software products that can be categorized into two separate classes, differentiated by their underlying methodologies and the degree to which they make use of all available character description data. Please note that the typing speed with our software seem to be the same, 2.1 seconds for all three categories, they are actually different: the first category is for a *single* character; the second *doubles* the number of characters; the third *quadruples* the number of characters. The reason for this "coincidence"? The number of keystrokes for all three categories are about the same: four to five.

15. A Vehicle And An Opportunity For The East Asians

For the vast majority of the East Asians who choose not to use memory-intensive input methods, the charactword-based input system can be a viable alternative, because it is easy to learn and use without depending on memory. It will also make computers easier to use, and more *useful*, to the residents of this region. All in all, the charactword-based input method could be an ideal vehicle for the 1.5 billion Chinese, Japanese and Korean people to access the information highway alongside the Americans and Europeans, providing East Asia an opportunity to make another powerful contribution to humanity again.

[19] This product has a mechanism that moves the typed Chinese characters to the front of their homonyms. So it is easier to find the Chinese characters after using the product for a while. At that point, the product is "trained."

[20] While our database for this feature is not yet complete, we known that one set of this type of vocabularies can be entered by typing just four or five keys.

[21] The database for this type of input is not complete at the moment, but we know that over 90% of the four-character idioms can be inputted with just four keys. The rest of these type of idioms can be entered with just five or six keys, but it will be more likely five than six keys. It really does not matter whether it will be five or six keys, because the keys after the 4th one will be the repetition of the 4th key.

Appendix 1 **Homonym groups and the approximate number of homonyms belonging to each group**

Pronunciation	Number of homonyms	Pronunciation	Number of homonyms	Pronunciation	Number of homonyms	Pronunciation	Number of homonyms
bi ㄅㄧ	120	lian ㄌㄧㄢ	50	jing ㄐㄧㄥ	60	chen ㄔㄣ	50
bao ㄅㄠ	40	lin ㄌㄧㄣ	40	ju ㄐㄩ	120	cheng ㄔㄥ	50
bei ㄅㄟ	40	ling ㄌㄧㄥ	60	jun ㄐㄩㄣ	40	chu ㄔㄨ	50
biao ㄅㄧㄠ	40	lei ㄌㄟ	50	qi ㄑㄧ	140	shi ㄕ	120
bo ㄅㄛ	80	gan ㄍㄢ	40	qian ㄑㄧㄢ	90	shan ㄕㄢ	50
pi ㄆㄧ	70	ge ㄍㄜ	60	qiang ㄑㄧㄤ	40	shen ㄕㄣ	50
pu ㄆㄨ	40	gu ㄍㄨ	70	qiao ㄑㄧㄠ	50	shu ㄕㄨ	40
di ㄉㄧ	80	gui ㄍㄨㄟ	50	qiu ㄑㄧㄡ	60	zi ㄗ	70
du ㄉㄨ	40	ke ㄎㄜ	50	qu ㄑㄩ	70	ci ㄘ	40
ta ㄊㄚ	40	kui ㄎㄨㄟ	50	quan ㄑㄧㄢ	40	si ㄙ	60
ti ㄊㄧ	50	he ㄏㄜ	60	xi ㄒ	180	suo ㄙㄨㄛ	40
tu ㄊㄨ	40	han ㄏㄢ	60	xian ㄒㄧㄢ	110	sui ㄙㄨㄟ	40
tong ㄊㄨㄥ	50	hao ㄏㄠ	40	xiang ㄒㄧㄤ	40	yan ㄧㄢ	150
tang ㄊㄤ	40	hu ㄏㄨ	90	xiao ㄒㄧㄠ	60	yang ㄧㄤ	50
mao ㄇㄠ	40	hui ㄏㄨㄟ	40	xing ㄒㄧㄥ	40	yao ㄧㄠ	80
mi ㄇㄧ	50	huo ㄏㄨㄛ	40	xu ㄒㄩ	80	you ㄧㄡ	80
mei ㄇㄟ	40	huan ㄏㄨㄢ	60	xun ㄒㄩㄣ	60	yu ㄩ	200
mo ㄇㄛ	60	huang ㄏㄨㄤ	50	zhi ㄓ	160	yue ㄩㄝ*	40
fan ㄈㄢ	50	hung ㄏㄨㄥ	60	zhan ㄓㄢ	50	yuan ㄩㄢ	70
fen ㄈㄣ	50	ji ㄐㄧ	220	zhen ㄓㄣ	70	yun ㄩㄣ	50
fei ㄈㄟ	50	jia ㄐㄧㄚ	70	zhou ㄓㄡ	40	wu ㄨ	120
feng ㄈㄥ	40	jian ㄐㄧㄢ	120	zhu ㄓㄨ	100	wei ㄨㄟ	130
fu ㄈㄨ	170	jiao ㄐㄧㄠ	80	chi ㄔ	90	wan ㄨㄢ	50
ni ㄋㄧ	50	jie ㄐㄧㄝ*	100	chan ㄔㄢ	60	e ㄜ	70
li ㄌㄧ	160	jin ㄐㄧㄣ	60	chou ㄔㄡ	40	ao ㄠ	40
liao ㄌㄧㄠ	40					an ㄢ	40

Appendix 2 Comparison of Twinbridge And Our Software

Name	Single character input			2-character vocabulary input			4-character idiom input			Fonts
Name of software	Input method	Is the result unique and what is the implication?	Average time to generate a character	Input method	Is the result unique and what is the implication?	Average time to generate a vocabulary	Input method	Is the result unique and what is the implication?	Average time to generate an idiom	Stability of fonts and the reason why
Twinbridge	Pronunciation of the character	No. Searching for the desired character is required	22.5 seconds	Continuous typing of: 1. the pronunciation of the two characters, but not necessarily all of the alphabets, or 2. acronyms of the pronunciation	No. Searching for the desired character is required	12.8 seconds	Typing the acronyms of the pronunciation	Not enough data to determine. It is likely to need to search if two or more idioms share the same acronym	Unable to determine due to insufficent data	Poor, because it operates in the English Windows where the fonts are single-byte while the Chinese fonts are double-byte.
Asian Language Software Solution (Ours)	Character-word based input	Yes. You get exactly the character you desire.	2.1 seconds or less (Zhuyin Zimu input only)	Typing of the acronym and marker(s)	Yes. You get exactly the vocabulary you desire.	2.1 seconds or less (Good for both Zhuyin and Pinyin)	Typing of the acronym and marker(s)	Yes. You get exactly the idiom you desire.	2.1 seconds less (Good for both Zhuyin and Pinyin)	Excellent font stability because it operates in the Chinese Windows.