

NEW VIETNAMESE INPUT METHOD

Jin Sugano

Fujitsu Limited

1015 Kamikodanaka Nakahara-ku, Kawasaki 211, Japan

Abstract

Conventional Vietnamese keyboards, being Roman-letter-based, require large numbers of keystrokes to enter because a part of single syllables are represented by long character strings.

The method we propose for Vietnamese-language entry uses 32 keys -- 30 for characters and two thumb-shift keys -- to type syllables in a maximum of three keystrokes.

1. Overview of the Vietnamese Language

Syllables in Vietnamese basically consist of initials, a medial, vowels, and finals with tone indication.

The 19 initials are b, c (k or q), ch/tr, d/gi/r, đ, g (gh), h, kh, l, m, n, ng (ngh), nh, ph, s/x, t, th, v, and an unsounded consonant.

Characters in parentheses have the same phonetic value as the characters preceding them -- c, k, and q, for example, all have the same phonetic value, [k]. Such characters indicate the type of vowel following the initial or that the initial is followed by the medial. The letter c mainly represents phonetic [k] for example. The letter k itself is used when the vowel following [k] phonetic begins with e, ê, i, or y. The letter q is used when phonetic [k] is followed by the medial. These rules are determined by Vietnamese orthography.

Characters following a slash (/) have the same phonetic value as the characters preceding them in a common language based on the Hanoi, or northern, dialect. Such pronunciation is sometimes differentiated in the middle and southern dialects, with Vietnamese orthography having different spellings representing the same syllable -- cha and tra, for example, both representing the syllable [ca], da, gia, and ra representing the syllable [za]; and sa and xa representing the syllable [sa].

The unsounded consonant, although having no representative character, is generally counted phonemically as an initial.

The single medial in Vietnamese is u (o). The vowel following the medial determines whether u or o is used.

The 14 vowels in Vietnamese are ă, a, â, e, ê, i (y), ô, o', o, u, u', ia (iê, ya, yê), ua (uô), and u'a (u'o').

The 8 finals are p, t, c (ch), m, n, ng (nh), i (y), and u (o).

Five of the six tones in Vietnamese are represented by symbols, with only the mid level tone not shown, as in Table 1.

Table 1. Tones and tone symbols in Vietnamese

Tone	Symbol
Mid level tone	None
Low falling tone	`
Low rising tone	ˆ
High broken tone	~
High rising tone	ˊ
Low broken tone	ˋ

2. Problems in Conventional Character Entry

Keys used for entering Vietnamese text from the conventional Vietnamese keyboard are shown in Figure 1.

Typing the syllable "nghiêng" ("slope"), for example, using these keys requires eight strokes -- n, g, h, i, ^, e, n, and g.

The inconvenience of the conventional Vietnamese input method is the number of keystrokes. And the solution to reduce the keystrokes is to use phoneme combinations. Since this usually involves more keys than alphabetic entry, however, the key to success in phoneme entry depends on how we reduce the number of key strokes, and, at the same time, reduce the number of required keys. This requires finding a trade off between the two.

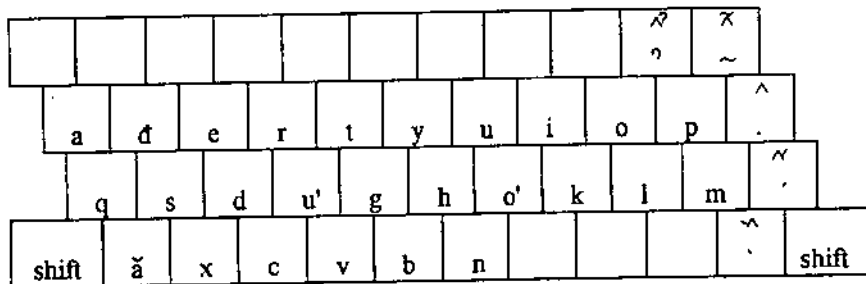


Figure 1. The layout of the conventional Vietnamese keyboard

3. Principles Behind the Proposed Scheme

Two main principles lie behind the language input method we developed:

- Reducing the number of keystrokes
- Reducing the number of keys on the keyboard.

Strong parallels between Vietnamese characters and Chinese Pinyin suggested to us that certain technical concepts involving Pinyin entry would also apply in Vietnamese-language entry.

Two such concepts are:

- Phoneme combinations in key entry and key assignment
- Multipurpose character key implementation and its control

3.1 Key Entry in Phoneme Combinations

Vietnamese phonemes have elements similar to their Chinese counterparts, but Vietnamese has more vowels, finals, and tones than Chinese.

Pinyin character entry uses a thumb-shift keyboard to enter each syllable in two keystrokes [1]. This would be impracticable with Vietnamese, however, attempting to enter the initial and medial in the first keystroke, then the vowel, final, and tone in the second keystroke, would require that the second stroke cover 672 different input elements, i.e., 14 vowels multiplied by 8 finals times six tones -- far too many to be efficiently assigned even to several dozen keys.

We considered an approach that enters a syllable in two or three keystrokes.

To enter a syllable in three keystrokes, we organized Vietnamese phoneme elements into three types of input -- input element 1, consisting of initials and the medial, input element 2, consisting of vowels and tones, and input element 3, consisting of finals.

We combined vowels and tones into the same input element because tone symbols are usually written over the vowel.

Every syllable in Vietnamese contains input elements 1 and 2, but not all contain input element 3, i.e., finals.

Thus, syllables without a final can be entered in two keystrokes, and those with a final in three keystrokes.

Mentioning the keyboard layout to enter each input element with one keystroke, the model keyboard is shown in Figure 2.

The thumb-shift keyboard now commonly used to enter Japanese, Chinese, and Korean [2] features two keys used by the thumbs. Other keys are used for character entry.

As mentioned before, Vietnamese orthography lets the same syllable be represented by different spellings that share the same phonetic value. Input elements using different spellings are assigned to different keys or key codes at need to identify characters representing a syllable entered in up to three keystrokes.

Assume that the keys used to enter input element 1 -- initials and medials -- are arranged as shown in Figure 3.

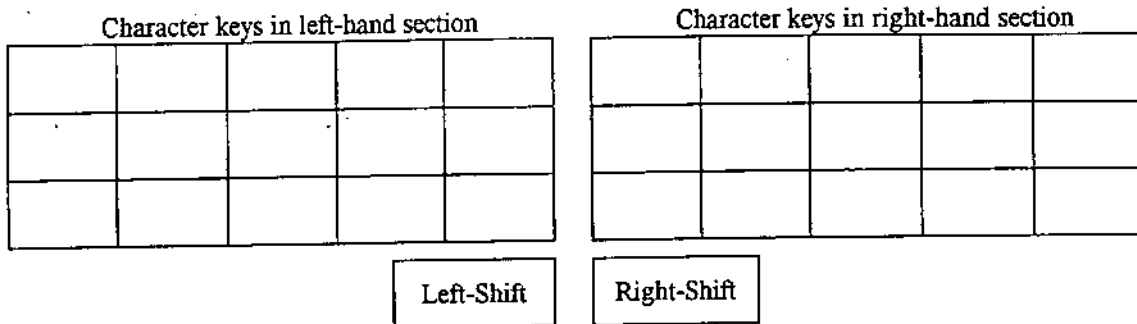


Figure 2. The thumb-shift keyboard

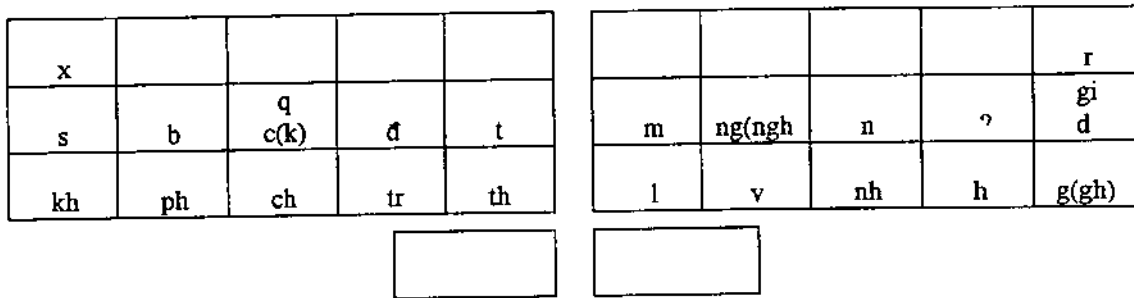


Figure 3. The keys for initials and medials

This enables 22 keys to be used to input element 1. The ʔ key represents the unsounded consonant, so ʔ is not displayed or printed when this is pressed. Entering an initial, such as t, with the medial u (tu) requires that the character and a thumb-shift key -- the left one in this case -- be pressed together. For an initial without the medial, only the character key -- marked t -- in Figure 3 -- is pressed.

Next, assume that the keys used to enter input element 2 -- vowels and tones -- are arranged as shown in Figure 4.

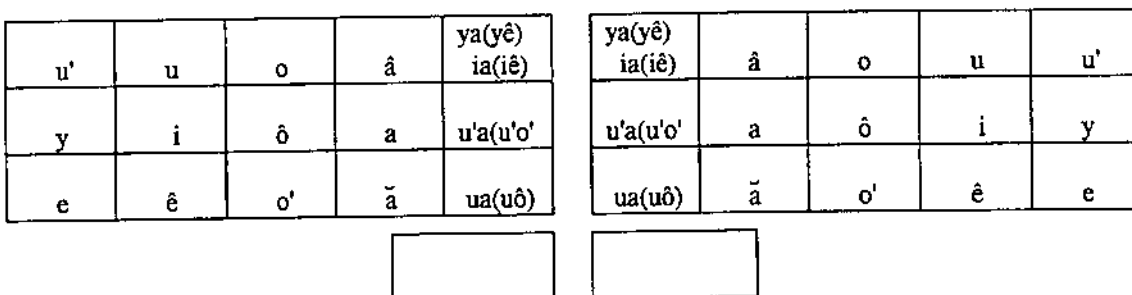


Figure 4. The keys for vowels and tones

Table 2. The method for vowels and tones keystroke

vowels with mid level tone	Press a key in the right-hand section.
vowels with high rising tone	Press a key in the right-hand section together with the right thumb-shift key.
vowels with low rising tone	Press a key in the right-hand section together with the left thumb-shift key.
vowels with low falling tone	Press a key in the left-hand section.
vowels with low broken tone	Press a key in the left-hand section together with the left thumb-shift key.
vowels with high broken tone	Press a key in the left-hand section together with the right thumb-shift key.

This enables 30 keys to be used to input element 2.

The left and right halves of the keyboard each contain a separate set of vowels. Character and thumb-shift keys are combined to enter vowels as shown in Table 2.

Pinyin entry groups initials with medials and vowels with tones.

As explained above, our method differs from the Pinyin method in that a third keystroke is used to enter input element 3 – finals – as shown in Figure 5.

	p	ch c	i(y)	t

m	nh ng	n	u(o)	

Figure 5. The keys for finals

This enables 8 keys to be used to input element 3.

For keys marked with two finals, the lower one is entered by pressing the character key and the upper one by pressing the character and thumb-shift keys together.

3.2 Multipurpose Character Key Implementation

The use of thumb-shift keys explained in Section 3.1 considerably reduces the number of character keys required for entering Vietnamese, but some 60 keys would still be required if the keys in Figures 3 to 5 were assigned separately.

A way to further reduce the number of keys would be to use individual keys for more than one input element.

To do this, we combined the layouts in Figures 3, 4, and 5 into that in Figure 6.

x-u'	u-p	-ch o-c	â-i	ia-t
s-y	b-i	q- c-ô	đ-a	t-u'a
kh-e	ph-ê	ch-o'	tr-ă	th-ua

ia-m	-nh â-ng	o-n	u-u	r-u'
m-u'a	ng-a	n-ô	?-i	gi- d-y
l-ua	v-ă	nh-o'	h-ê	g-e

Figure 6. The layout of new Vietnamese keyboard

Note that input elements 1 and 2 are assigned to keys without shading and input elements 2 and 3 to keys with shading.

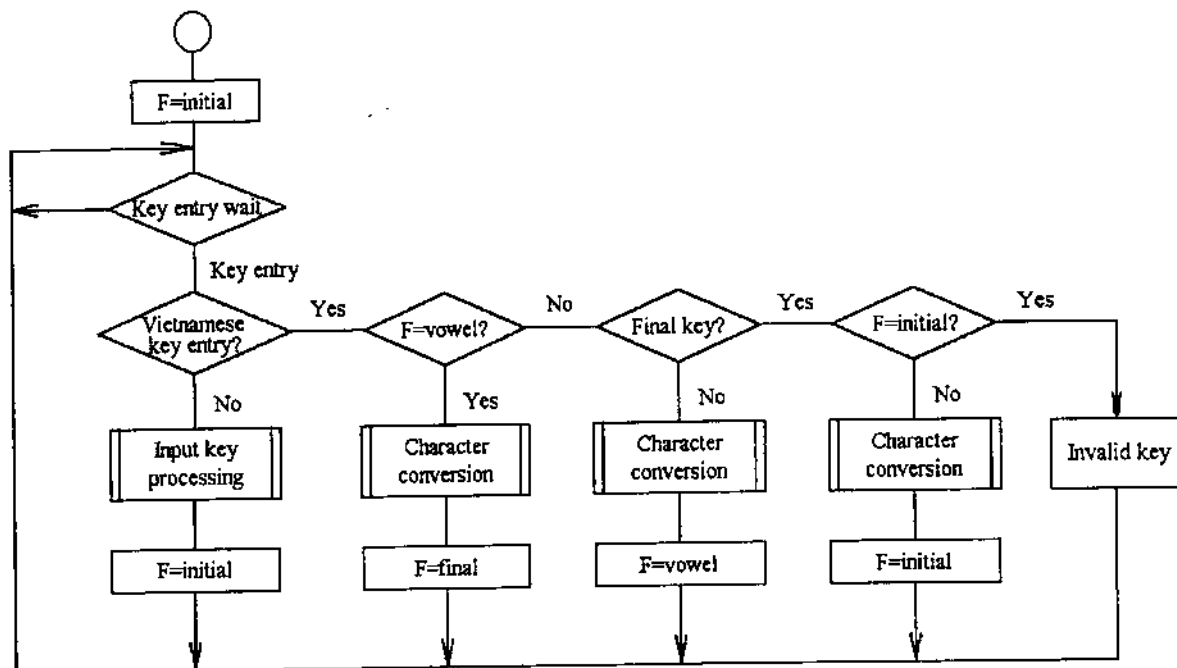
Keys with shading are used to enter finals, so we call them final keys.

Multipurpose keys are enabled using an input mode flag in the input processing program and controlling the flag based on the following logic (Figure 7).

The user presses keys, to which the required characters are assigned, in the sequence initial, medial, then vowel, tone, then final. In short, syllables without a final require two keystrokes and those with it require three keystrokes.

Character conversion in the flowchart in Figure 7 involves searching the table (Table 3) based on an input key code(s) and the value of the input mode flag, converting the input key to the corresponding Vietnamese character codes.

Part of the input processing program corrects the spelling of entered characters, based on Vietnamese orthography.



Note: Input mode flags are indicated by the letter F.

Figure 7. The flowchart of character conversion

As mentioned before, more than one character in Vietnamese may have the same phoneme value at the phoneme level. If these characters can be identified uniquely based on the phoneme they are followed by, they can be assigned to the same input key under the method we propose so that their spelling is corrected by an orthography routine in the input processing program.

An exhaustive description of the orthographic routine is beyond the scope of this paper, and has been limited here to that related to determining initials.

Three pairs of characters -- c and k, g and gh, and ng and ngh -- are assigned to identical character keys and generate identical key codes when entered.

When one of the keys these pairs are assigned to is pressed, the input processing program first displays the most frequently used first character -- c, g, or ng. The program then checks for a following vowel that begins with e, ê, i, or y and, if it finds one, replaces c, g, or ng with the second character -- k, gh, or ngh -- then displays the character with the vowel.

This is only a very small part of how the orthographic routine operates.

4. Conclusions

To what extent, then, did we realize the two principles given in Section 3?

The first was reducing the number of keystrokes required for Vietnamese character entry.

The example of syllable "nghiêng" entry (Section 2) conventionally required eight keystrokes. Our method enables this to be entered using three keystrokes -- ngh (ng-a key), iê (ia-m key), and ng (â-ng key). Syllables without a final, such as "ngựa" ("horse"), conventionally required five keystrokes -- n, g, , u', and a. Our method reduces this to two -- ng (ng-a key) and ỳ'a (t-u'a with left-shift key). Further, syllables in Vietnamese were conventionally separated by blanks, requiring the user to press the space key at the end of each syllable. Our method has assigned this to be done automatically by the program (not explained here due to space considerations) -- reducing the number of required keystrokes even further.

The second principle was reducing the number of keys on the keyboard.

Our method combines phonemes to realize the first principle. Theoretically, phoneme entry involves far more keys than alphabetic entry. We avoid this by effectively using thumb-shift keys, multipurpose keys, and an orthographic routine to reduce the number of required keys to 32 -- 30 for characters and two thumb-shift keys. This is about the same number used conventionally. Assigning character keys in three stages also makes entry easier.

Table 3. The table of character conversion

Character key	Initial or final			Vowel		
	Pressed	Pressed together with right shift key	Pressed together with left shift key	Pressed	Pressed together with right shift key	Pressed together with left shift key
Upper line in right-hand section	ia-m	-m		ia	ía	ia
	â-ng/nh	-ng	-nh	â	á	â
	o-n	-n		o	ó	ô
	u-u	-u		u	ú	û
Middle line in right-hand section	r-u'	r-		u'	ú'	û'
	m-u'a	m-		u'a	ú'a	û'a
	ng-a	ng-	ngu-	a	á	â
	n-ô	n-	nu-	ô	ó	ô
	?-i	?-	u-	i	í	ï
Lower line in right-hand section	d/gi-y	d-	du- gi-	y	ý	ÿ
	l-ua	l-	lu-	ua	úa	úa
	v-ã	v-		ã	á	ã
	nh-o'	nh-	nhu-	o'	ó'	ô'
	h-ê	h-	hu-	ê	é	ê
Upper line in left-hand section	g-e	g-	gu-	e	é	ê
	ia-t	-t		ia	ía	ia
	â-i	-i		â	á	â
	o-c/ch	-c	-ch	ô	ó	ô
	u-p	-p		ù	û	u
Middle line in left-hand section	x-u'	x-	xu-	ù'	û'	u'
	t-u'a	t-	tu-	u'a	ú'a	û'a
	đ-a	đ-	đu-	à	á	a
	c/q-ô	c-	qu-	ô	ó	ô
	b-i	b-		i	í	i
Lower line in left-hand section	s-y	s-	su-	ý	ÿ	y
	th-ua	th-	thu-	úa	úa	úa
	tr-ã	tr-	tru-	ã	á	ã
	ch-o'	ch-	chu-	ò'	ó'	ô'
	ph-ê	ph-		è	é	ê
kh-e	kh-		kh-	è	é	

Acknowledgement

I would like to thank my coworker, Ms. Meichen Lee, for checking this paper and assisting in Vietnamese typewriter surveys.

Reference

- [1] Jin Sugano. "Chuugokugo Nyuuryoku Houhou no Hanashi (Input methods and keyboards for Chinese)". Asahi, Tokyo, 1991, pp.148-169
- [2] Jin Sugano. "Application of Chinese key entry principles to Hanguk entry", Proceedings of 1992 International Conference on Chinese Information Processing, Beijing, Oct. 26-28, 1992.