Assignment of Tamil Characters to the Telephone Keypad

V. Mahinthan* and S. Niranjayan**
*Department of Electrical and Computer Engineering, University of Waterloo, University of Waterloo, <Email: mveluppi@engmail.uwaterloo.ca>;
**Department of Electrical and Computer Engineering, National University of Singapore, Singapore <Email: g0203523@nus.edu.sg>

Abstract

Nowadays, text input in telecommunication devices is becoming more important with the increasing amount of text-based services. Easy, correct and efficient text input, search and retrieval via the telephone keypad is a basic user requirement. In order to use Tamil in telecommunication devices, a Tamil keypad input scheme should be standardized. In this paper, we study various keypad schemes of Tamil alphabet. The phonetic property of Tamil language is considered in the assignment of the Tamil characters. The analysis and comparison of the proposed schemes are based on the quantifiable factors but, there are human-engineering factors to be considered in order to choose the optimal scheme. Those are, the character searching time of the user and the level of easiness in handling various key press operations by hand. The searching time depends on the way how the Tamil alphabets are remembered and correlated by average users.

1. Introduction

The original reason for assigning letters to the rotary dial pad and later to the numeric telephone keys was to provide alphabetic "aliases" for digits, as mnemonics in dialing. Nowadays, text input in telecommunication devices is becoming more important with the increasing amount of text-based services. Easy, correct and efficient text input, search and retrieval via the telephone keypad is a basic user requirement. Finding the characters necessary to enter a name in the phonebook, searching for a name, knowing the bus schedule, writinge a short message or logging on to the internet are most common.

The standard available, addressing assignment of characters to the 12-key telephone keypad, is limited to the assignment of the basic 26 Latin letters (a to z) [1]. Recently, ETSI Standardized the European languages in [2] based on both language-independent and language-specific keypad assignment. Furthermore, maximum number of languages that are presently supported by a specific telecommunications device or service is less than 50. Unfortunately, Tamil is not in the list. This motivates us to study the assignment of Tamil characters to the keypad.

2. Assignment of characters

Tamil language has 247 characters including 12 ëuyirí, 18 ëmeií, 216 ëuyirmeií and 1 ëauthamí. If we assign all the characters in the keypad, one key contains approximately 21 characters. It is ambiguous and the average keystroke per character is high. To avoid this situation, phonetic property of the Tamil characters is used in our proposed scheme as in phonetic Tamil keyboard input scheme [3]. It means each ëuyirmeií can be represented by ëmeií + ëuyirí. The phonetic property of the language reduces the number of characters which is equivalent to 31 characters including 12 ëuyirí, 18 ëmeií and 1 ëauthamí. Rest of the paper, we consider only these 31 characters to implement the Tamil language.

Firstly, frequency of occurrence (FoO) of Tamil letters is calculated from the observation of [4]. In most of the languages the ëspaceí character has very high FoO, therefore the FoO of other characters are recalculated including the space character. FoO of space is given by

$$FoO_{space} = 100 / (1 + AWL)\ \% \qquad (1)$$

Where   is the average word length of tamil language. The word length analysis in [5] yields that . ëvadaí letters and punctuations are not considered in the analysis. Since we have 32 character including the space to assign in the keypad, each key contains approximately 3 characters. The characters are ranked in **table 1** according to the occurrence among all characters, among ëuyirí characters and among ëmeií characters.

3. Proposed schemes

Efficiency of the keypad input scheme depends on both number of keystroke required per character and the time taken to distinguish the characters for a non-expert user. Average keystroke required per character is given by [6],

$$KSPC = \sum_{i=1}^{31} f_i n_i \qquad (2)$$

where   is the number of keystroke to type to   character.

To minimize the KSPC, most appeared characters should have less number of keystrokes. On the other hand, it increases the character ambiguity. There is a tradeoff between ambiguity and number of keystroke. But, it is very difficult to quantitatively characterize the ambiguity.

Various schemes are analyzed and tabulated according to the KSPC in **table 2**. The FoO scheme is based on the frequency of occurrence and it is the most ambiguous one. Furthermore, FoO scheme uses 12 keys to achieve less KSPC. Ambiguity of the rest of the schemes is basically reduced by separating the ëuyirí and ëmeií characters. The scheme that is based on alphabetic order which uses 10 keys and each key contains 3 characters. ëAuthamí can be assigned as fourth character in any of those keys. Alphabetic order scheme has less ambiguity and high KSPC.

Tamil Alphabet	FoO in words,	Position in Alphabet	Position in 'Uyir'	Position in 'Mei'	FoO in sentences (with 'Space') f_i
space	-	-	-	-	10.34
அ	14.10%	1	1		12.64
ஆ	4.34%	8	4		3.89
இ	7.26%	3	3		6.51
ஈ	0.56%	27	10		0.50
உ	7.30%	2	2		6.55
ஊ	0.50%	28	11		0.45
எ	2.10%	18	6		1.88
ஏ	1.31%	22	7		1.17
ஐ	2.83%	15	5		2.54
ஒ	1.20%	23	8		1.08
ஓ	1.01%	24	9		0.91
ஔ	0.01%	30	12		0.01
ஃ	0.00%	31	-	-	0.00
க்	6.33%	5		2	5.68
ங்	0.80%	26		17	0.72
ச்	1.96%	19		13	1.76
ஞ்	0.31%	29		18	0.28
ட்	2.79%	16		11	2.50
ண்	1.56%	20		14	1.40
த்	6.54%	4		1	5.86
ந்	2.31%	17		12	2.07
ப்	4.00%	10		6	3.59
ம்	4.67%	7		4	4.19
ய்	3.51%	12		8	3.15
ர்	4.20%	9		5	3.77
ல்	3.34%	14		10	2.99
வ்	3.91%	11		7	3.51
ழ்	1.01%	25		16	0.91
ள்	1.44%	21		15	1.29
ற்	3.41%	13		9	3.06
ன்	5.36%	6		3	4.81

Table 1. Frequency of occurrence of Tamil characters ('Uyir', 'Mei' and 'Autham')

Scheme I uses first six keys for ëuyirí which is placed in alphabetical order. The rest of the keys are assigned for ëmeií as shown in figure 1 where characters are arranged according to the FoO. ëAuthamí is assigned in the third position of the sixth key. In scheme II, ëuyirí is assigned for first four keys and ëmeií is assigned in the other numbered keys (5-9,0) as shown in **figure 2**. Both ëuyirí and ëmeií are arranged in the keys according to the FoO. ëAuthamí is assigned in the fourth position of the fourth key. The KSPC of scheme II is little bit verse compared to than that of scheme I, but, it utilizes only 10 keys rather than 12 keys. In scheme III, ëUyirí is assigned as first position in all keys through alphabetical order and ëmeií takes second and third position of the first nine keys as given in the table 2. The ëmeií is assigned according to the FoO. Further, ëauthamí get the second position of the pound (#) key.

Scheme (keys used)	Position of characters												KSPC
	1	2	3	4	5	6	7	8	9	0	*	#	
FoO (12)	-	அ	உ	இ	த்	க்	ன்	ம்	ஆ	ர்	ப்	வ்	1.32
	ய்	ற்	ல்	ஐ	ட்	ந்	எ	ச்	ண்	ள்	ஏ	ஒ	
	ஏ	ழ்	ங்	ஈ	ஊ	ஞ்	ஔ	ஃ					
Scheme I (12)	அ	இ	உ	எ	ஐ	-	க்	த்	ப்	ம்	ர்	ன்	1.38
	ஆ	ஈ	ஊ	ஏ	ஒ	ஓ	ட்	ந்	ய்	ல்	வ்	ற்	
					ஃ	ஔ	ங்	ச்	ஞ்	ண்	ழ்	ள்	
Scheme II (10)	அ	ஆ	இ	உ	க்	த்	ப்	ம்	ர்	ன்	-		1.40
	எ	ஏ	ஐ	ஒ	ட்	ந்	ய்	ல்	வ்	ற்			
	ஓ	ஔ	ஈ	ஊ	ங்	ச்	ஞ்	ண்	ழ்	ள்			
				ஃ									
Scheme III (12)	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	-	ஔ	1.69
	க்	ட்	த்	ந்	ப்	ம்	ய்	ர்	ல்		ஓ	ஃ	
	வ்	ற்	ன்	ங்	ச்	ஞ்	ண்	ழ்	ள்				
Alphabet order (10)	அ	ஈ	எ	ஒ	க்	ஞ்	த்	ம்	ல்	ள்			1.89
	ஆ	உ	ஏ	ஓ	ங்	ட்	ந்	ய்	வ்	ற்			
	இ	ஊ	ஐ	ஔ	ச்	ண்	ப்	ர்	ழ்	ன்			
	ஃ												
Multiple planes (12)	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	-	சி1	சி2	2H - 1.01
	க்	ச்	ட்	த்	ப்	ற்	ய்	ர்	ழ்	ஒ	ஓ	ஔ	1H - 1.53
	ங்	ஞ்	ண்	ன்	ம்	ந்	வ்	ள்	ல்	ஃ			

Table 2. Proposed schemes and average keystroke per character (‘Uyir’, ‘Mei’ and ‘Autham’)

The multiple plane scheme uses the star (*) key and the pound (#) key, named as the combination keys C1 and C2, to set two different character planes in the first nine keys (i.e from 1Ö9). In the first 9 keys, a single press will represent the first character, simultaneous pressing of any key with C1 will represent the second character and simultaneous pressing of C2 will represent the 3rd character assigned to the key. In this technique, in two handed operation (suitable for mobile phones) every character assigned in the first 9 keys needs approximately a single key stroke time. Different to the first 9 keys, the characters assigned to the last 3 keys are accessed by repeated key presses. But for one handed operation each simultaneous press operation is taken as 2 key strokes. For 2 handed operations the arrangement of characters in the first 9 keys is independent of their FoO. Anyhow, the arrangement as in table 2 gives ergonomics to the user and reduces key searching time.

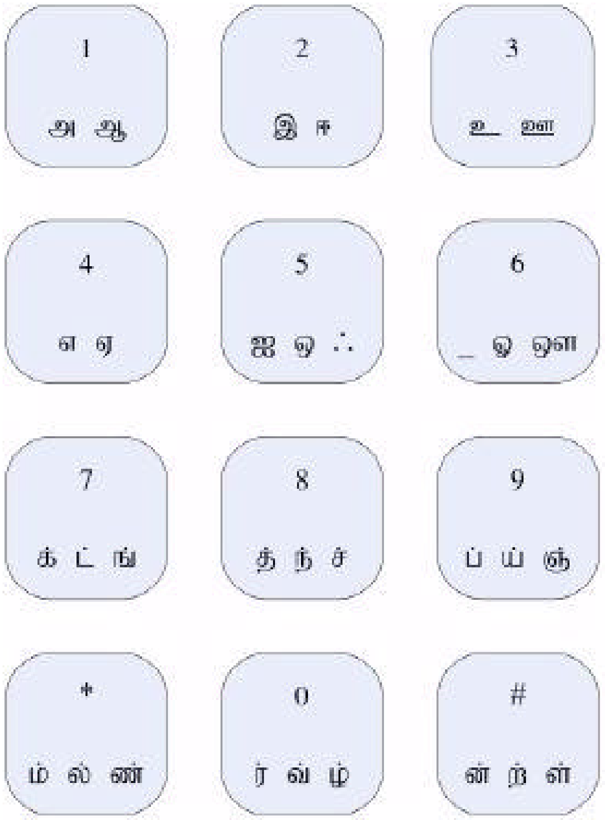

Fig. 1. Keypad configuration of scheme I (12 keys are used)

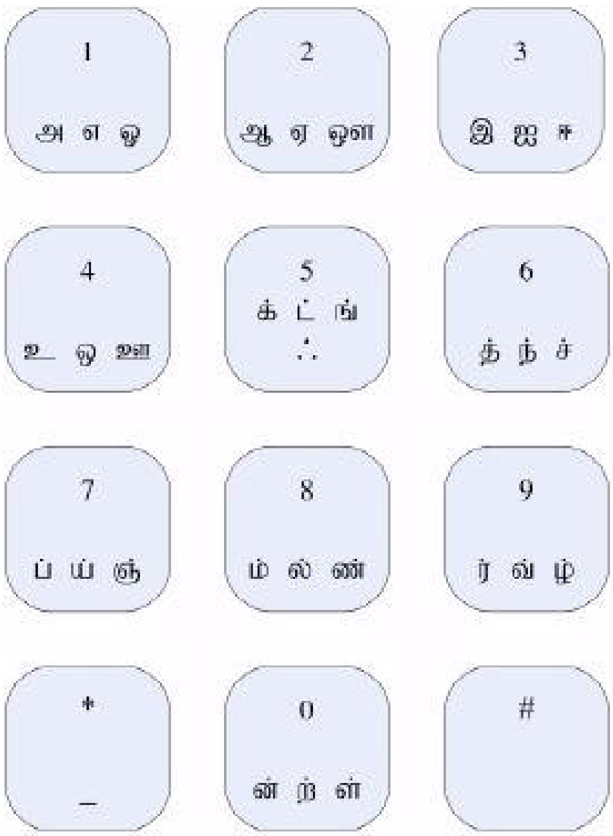

Fig. 2. Keypad configuration of scheme II (10 keys are used)

4. Discussion and Conclusions

It should be noted that in order to reduce KSPC, the device should do intelligent input processing in accepting the sequential character inputs. If an ëuyirí follows a ëmeií both should be combined and taken as a ëuyir-meií. Also in the multiplane scheme, after pressing an ëuyirí the device can automatically change to one of the ëmeií planes to reduce the key stroke time. 10 keys schemes are more appropriate than 12 keys schemes to use in standard telephone because it requires few control keys. Normally (*) key and (#) key are used as control keys. In contrast, special keys are used in mobile phone for the purpose of control.

From above results for single hand operation, we can observe that KSPC of the scheme I and the scheme II much lower than that of the scheme III and the alphabetical order scheme. In addition, the ambiguity of both schemes better than the FoO scheme. Scheme II is utilized 10 keys rather than 12 keys of scheme I, we recommend the scheme II for single hand keypad text input scheme for Tamil.

This analysis and comparison is based on the quantifiable factors, but as we mentioned previously there are human-engineering factors to be considered in order to choose the optimal scheme. Those are, the character searching time of the user and the level of easiness in handling various key press operations by hand. The searching time depends on the way how the Tamil alphabets are remembered and correlated by average users. These two factors can be compared amongst the schemes by user surveys like [7].

References:

[1] Arrangement of digits, letters and symbols on telephones and other devices that can be used for gaining access to a telephone network, ITU-T Recommendation E. 161, Jan. 2001.

[2] Human Factors (HF); User Interfaces; Character repertoires, ordering rules and assignments to the 12-key telephone keypad, ESTI, ES. 202 130, v 1.1.1, Sep. 2003.

[3] ìKeystroke Sequence for the Phonetic Keyboard,î TamilNet 99 Keyboard Standard http://www.tamilvu.org/Tamilnet99/annex2.htm

[4] Gift Sironmany, ìEntropy of Tamil prose,î Information and control, Vol. 6. No. 3, Sep. 1963.

[5] Bharati, Akshar, Prakash Rao K, Rajeev Sangal, S M Bendre, ìBasic Statistical Analaysis of Corpus and Cross Comparision among Corporaî, Proceedings of ICON-2002, Mumbai, India, Dec. 2002.

[6] Rick W. K. Ha, Pin-Han Ho and X. Sherman Shen, ìSIMKEYS: An efficient approach in text entry for mobile communications,î IEEE Consumer Communication and Networking Conference, Jan. 2004.

[7] S. Ronkainen, and J. Marila, ìEffect of Auditory Feedback on Multitap on Text Input Using Standard Telephone Keypad,î Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan, July 2002.