Comparative Analysis of current Tamil Keyboard Input Methods

Ravindran K. Paul / ravink@pop.jaring.my

Abstract

The purpose of this paper is to show the diverse requirements of Tamil Keyboard design and the diverse expressions of these concepts in currently available products. This paper will analyse the relative efficiency of 3 keyboard designs currently available and suggest further development in this area.

The use of Tamil in Personal Computers has been in existence for more than 10 years now. Development though has been at a relatively slow pace due to small market size and an absence of standards. During this period, several keyboard input methods have been created with varying degrees of efficiency and ease of use.

It will be the purpose of this paper to analyse 3 popular keyboard formats in use in Malaysia and Singapore today, namely the Tamil Typewriter Keyboard, Thunaivan Phonetic Keyboard and the I.E. Singapore Keyboard.

Despite papers submitted in the past on keyboard design, the analysis thus far has been inadequate and erroneous in some cases. A detailed explanation of the criteria used to measure efficiency as well as some indexes on how to measure keyboard speed will be made. These 3 keyboards will then be measured against these indexes to show how they fare.

Despite the differing philosophies, it will be seen that the phonetic keyboard designs are indeed better suited to speed typing. Despite speed being an important criteria, it will also be shown that ease of use should not unnecessarily be sacrificed for speed.

Overview

The primary purpose of this paper is to lay down some principles for the analysis of Tamil keyboard design. These principles will then be used to test some existing keyboard designs used in Singapore and Malaysia as well as design a possible layout optimised for speed.

The above Abstract, except for some minor changes was prepared before the actual analysis of the keyboards were done. The conclusions derived from this analysis is quite different from what is currently accepted as true to a surprising degree.

This indeed serves to show that much discussion carried out over the past 10 years has been without proper and detailed research. Even the data and analysis provided in this paper has been derived over a very short period, but this is indeed more comprehensive than what has been available over the last 10 years. It is hoped that this paper will stimulate more research in this area to come out with the appropriate conclusions on this topic.

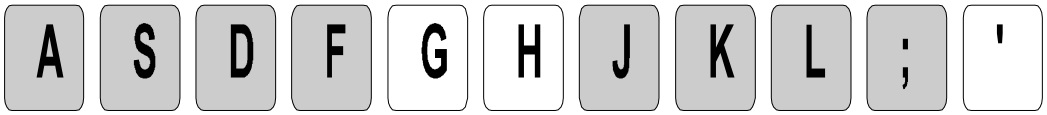
The conclusions and methodology contained herein are not to be considered as irrefutable. It is just one interpretation of the data available within the limits of time and finance. Other aspects of design have also to be considered and some of the criteria here may not be suitable. But, much of the conclusions derived are worthy of consideration and this paper is definitely a step in the right direction towards the future.

Criteria for Keyboard efficiency

Below are some criteria that are relevant to keyboard design. Although not exhaustive most of the primary criteria are contained here. More subjective criteria like common patterns in Tamil typing are beyond the scope of this paper. Of course these matters can be studied with the proper funding and time.

1. Home Keys

The characters with highest frequency of use should be on the Home Keys. In the QWERTY keyboard, these are :



2. Minimise use of SHIFT Keys

The use of SHIFT keys slows down typing as well as requires 2 fingers to be used simultaneously. In effect the use of the SHIFT key is equivalent to 2 key strokes.

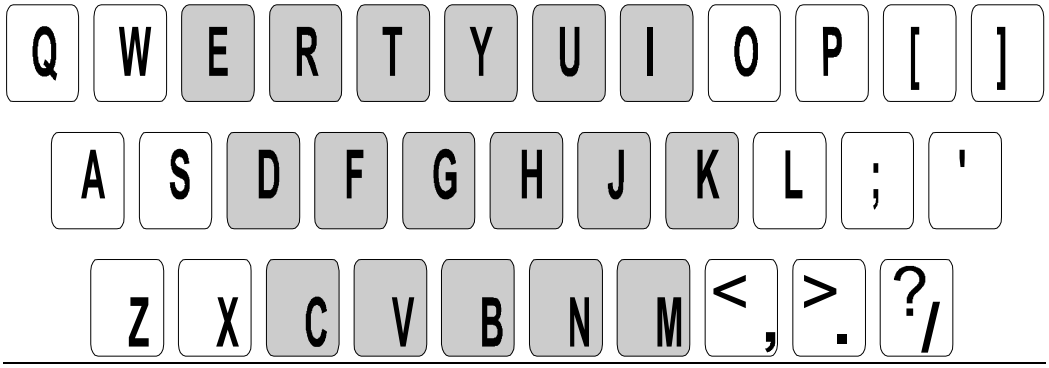
3. Minimise the number of Keystrokes

This is an obvious principle, but some qualification is necessary. Minimising keystrokes and ignoring the efficient arrangement of the keys will not be beneficial.

4. Increase use of Alternate Hands

Though not so obvious, the use of right and left hand alternately while typing should increase speed. This requires some qualification too. Although this increases speed, the close placement of common key sequences in one hand is quite often faster than alternate hand typing.

5. The index finger and the middle finger are more efficient



The index finger and the middle finger are almost equal in typing efficiency. Most non-typists, use the index and middle finger exclusively for typing.

6. Orderly arrangement of keys

Though not directly related to fast typing, an orderly arrangement of keys allows one to find the keys more easily. This is especially relevant for beginners and casual typists. At the least, the keyboard should be divided into Consonants and Vowels.

Existing Keyboard designs

Currently there are dozens of keyboard designs available on the market. For the purposes of this paper, only 2 categories will be examined. One is the Tamil Typewriter Keyboard layout which is the most commonly used layout worldwide. The other is the Phonetic layout. As one of the purposes of this paper is to show the relative efficiency of the Phonetic keyboard arrangement, 2 phonetic keyboard designs, the Thunaivan Phonetic and the I.E. Singapore design will be analysed.

Phonetic Keyboard.

The first Tamil Phonetic Keyboard to be introduced in Malaysia and Singapore is the Thunaivan Phonetic Keyboard. The term Phonetic Keyboard was probably first used in relation to the Thunaivan Phonetic Keyboard. A Phonetic Keyboard was defined as follows :

The phonetic keyboard is based on the use of only 13 keys for vowels and 17 keys for the consonants. With just these 30 keys, all the keys in the Tamil alphabet can be typed.

The principle is as follows :

Typing :

uyir elutthu = uyir elutthu

ஃ = ஃ

ஔ = ஔ

mei elutthu = mei elutthu

க = க

ட் = ட்

(Note that the definition of “mei elutthu” has been modified from க் to க)

mei elutthu + uyir elutthu = uyirmei elutthu

க	+	ஃ	=	க்
ட்	+	ஔ	=	டௌ
க	+	ஒ	=	கொ

க + ை = கை

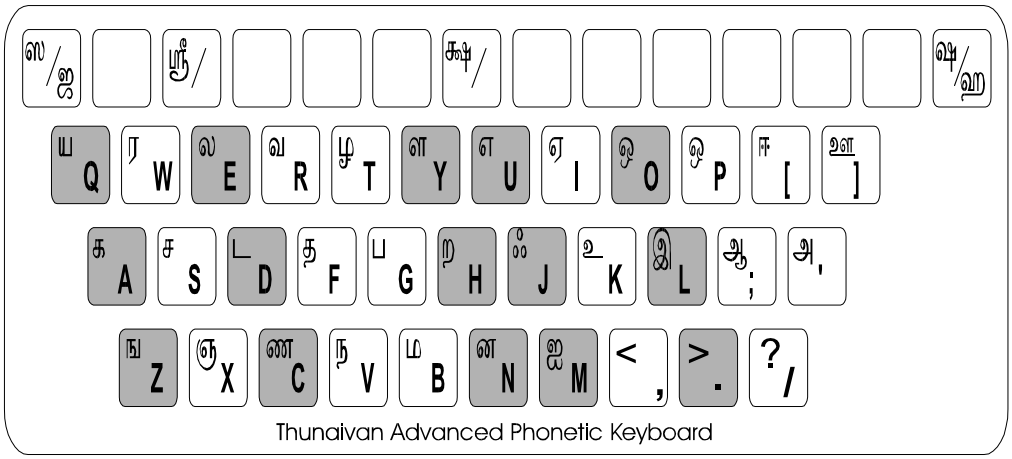
Note that a phonetic keyboard can only work with underlying programs to modify characters as they are being typed.

In this keyboard when you type **க,** **க** will appear. And similarly, when you type **ஃ,** **ஃ** will appear. But if you type **க** following it immediately with **ஃ** then the **க** will appear first on the screen and when **ஃ** is typed, the **க** will be modified into **க்**. This action is similarly repeated when, **க** is followed by **ஓ**. In this case, two keystrokes produces **கோ** which has 3 characters.

Current Phonetic Keyboards in Singapore / Malaysia

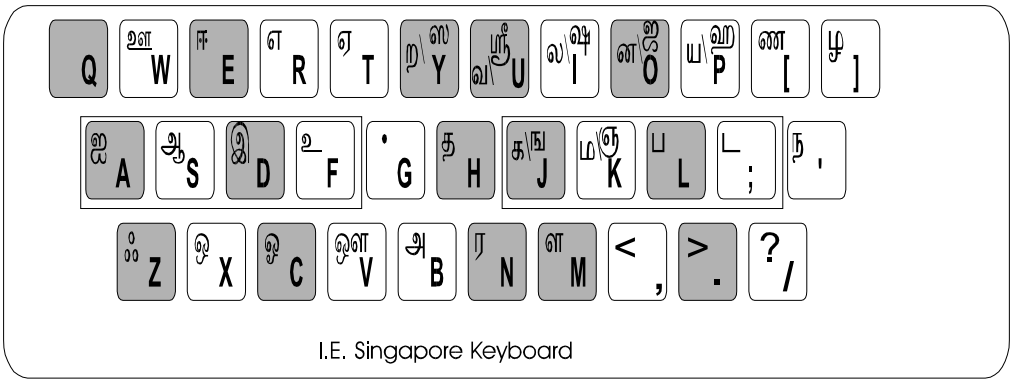
Thunaivan Phonetic Keyboard

The current form of this keyboard was introduced in 1987. The special feature of this keyboard is that it does not require the use of SHIFT key to type any Tamil words. Also it is organised in an orderly arrangement with consonants on the left and vowels on the right. Its consonants are arranged in the subdivisions of vallinam, mellinam and idaiyinam.(வல்லினம், மெல்லினம் and இடையினம்)



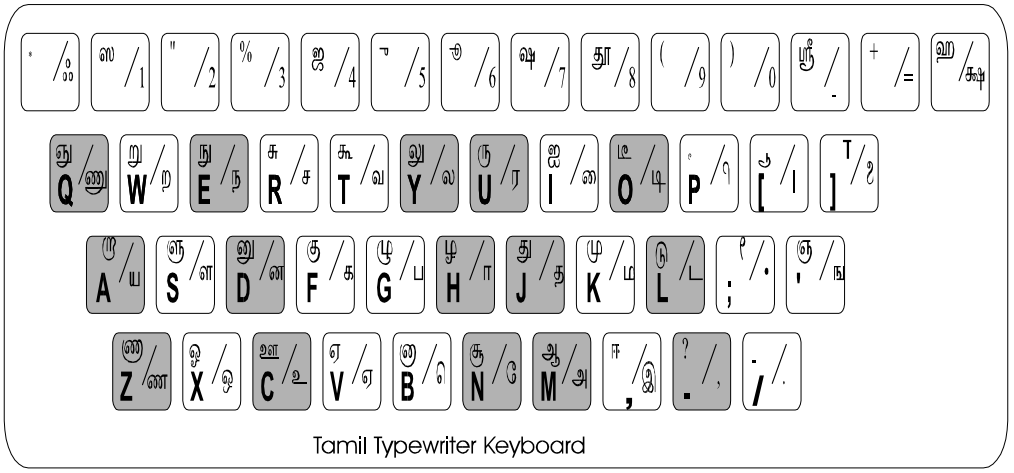
I.E. Singapore Keyboard

This keyboard was also introduced in 1987. This keyboard was designed with the objective of speed typing. It has consonants on the right and vowels on the left.



Tamil Typewriter Keyboard

The Tamil typewriter keyboard has been in existence for decades. Due to its original design for use on mechanical typewriters, it has been thought to require extensive use of the SHIFT key although the key arrangement is quite efficient. (The data provided in this paper will show that the use of SHIFT keys is actually quite minimal)



It must be noted that most of the Typewriter compatible keyboards used on computers are not an exact representation of the manual typewriter keyboard. Some of the existing layouts have reversed the sequence of use of the extension characters.

Manual Typewriter Layout : • + க = க்

Computer Typewriter Layout : க + • = க்

Other variations have tried to improve on this layout with some modification of the layout to reduce the number of keys required, but generally the principle of use is the same. Also, the use of SHIFT keys is still necessary.

Character Frequency in Tamil

To simplify analysis of Tamil keyboard design, the analysis will be limited to Tamil characters only. The Grantha characters will be completely ignored.

Following are tables depicting the frequency of Tamil characters in typical Tamil documents. The document used for analysis was from a local monthly Tamil magazine. Further more the document is a combination of various different articles so as to obtain as random a sample as possible.

Four separate tables from 4 separate documents were created to check the relative frequency of characters. Generally, all the documents showed similar characteristics, but the relative frequency of characters in various documents varied between 1 to 2 per cent and also the ascending order of character frequency varied after the 4th most frequent character in the consonants row.

Frequency of Characters in Document

	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ	ஔ	ஃ	Total	% of mei	% of keystroke
	363	89	244	10	108	9	261	35	7	95	20	0	2	1243	7.9	
க	706	97	152	6	283	47	5	45	41	71	12	0	531	1996	12.6	7.3%
ங	0	0	0	0	0	0	0	0	0	0	0	0	180	180	1.1	0.7%
ச	131	48	124	3	81	6	44	8	21	49	5	0	111	631	4.0	2.3%
ஞ	6	2	0	0	0	0	0	0	0	0	0	0	25	33	0.2	0.1%
ட	186	54	222	5	255	2	3	15	65	0	5	0	301	1113	7.0	4.1%
ண	62	6	20	3	12	0	0	0	7	0	1	0	196	307	1.9	1.1%
	390	175	258	5	430	11	23	41	109	24	46	0	335	1847	11.7	6.8%
ந	54	62	54	22	5	4	9	8	1	2	10	0	277	508	3.2	1.9%
ப	289	126	109	4	106	3	42	43	19	41	121	0	353	1256	7.9	4.6%
ம	175	119	30	10	80	20	2	27	69	2	7	0	554	1095	6.9	4.0%
ய	178	102	90	0	59	2	8	30	38	0	15	0	78	600	3.8	2.2%
ர	183	61	104	0	330	3	5	5	30	0	20	0	279	1020	6.4	3.8%
ல	113	65	81	2	41	0	2	22	87	0	12	0	439	864	5.5	3.2%
வ	385	77	235	39	59	0	19	79	38	3	10	0	7	951	6.0	3.5%
ழ	44	1	28	0	49	0	0	4	16	0	0	0	27	169	1.1	0.6%
ள	55	22	89	0	45	0	1	6	79	0	4	0	214	515	3.3	1.9%
ற	111	32	78	2	124	0	2	12	37	2	5	0	86	491	3.1	1.8%
ன	173	87	39	1	54	0	3	20	61	1	12	0	551	1002	6.3	3.7%
Uyir Total	3604	1225	1957	112	2121	107	429	400	725	290	305	0	4546	15821		
% of Uyir	22.8	7.7	12.4	0.7	13.4	0.7	2.7	2.5	4.6	1.8	1.9	0.0	28.7	100	100	53.7%
		4.5%	7.2%	0.4%	7.8%	0.4%	1.6%	1.5%	2.7%	1.1%	1.1%	0.0%	16.7%			45.0%
Total Keystrokes														அ	1.3%	100%

No. of Uyir	1243		
No. of Mei	3241	ங	180
Uyir of Uyirmei	11337 (No. of Uyirmei x 2)	ஞ	33
Mei of Uyirmei	11337		
Total Keystrokes	**27158**		

This table has been specifically formulated to analyse phonetic typing. Each cell represents the frequency of the characters found in the document. The row total represents the total of the "uyir" column and the column total represents the total of the "mei" row. Note that ஙகா, கா and கி are all considered as one character in this table. In a phonetic keyboard, since each uyirmei requries only 2 keystrokes,

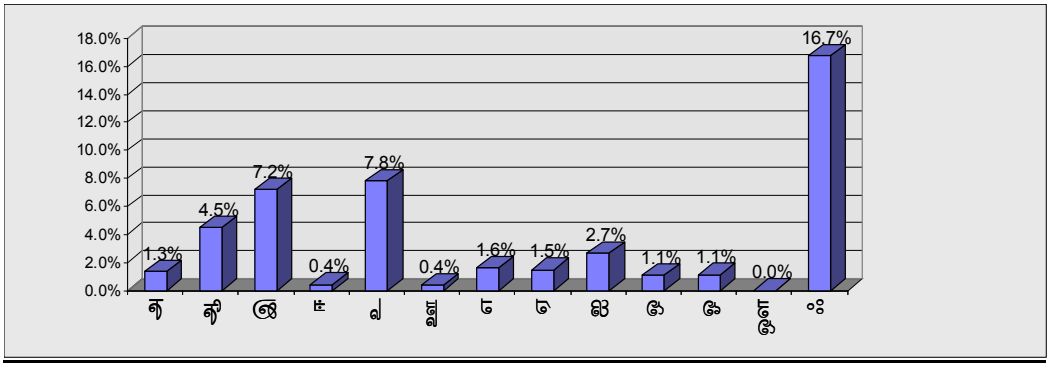
Total No. of keystrokes = (No. of Uyir) + (No. of Mei) + (No. of Uyirmei x 2)

The percentage use of each key is derived using the formula :

(Total of Uyir Column or Mei Row) / (Total No. of keystrokes)

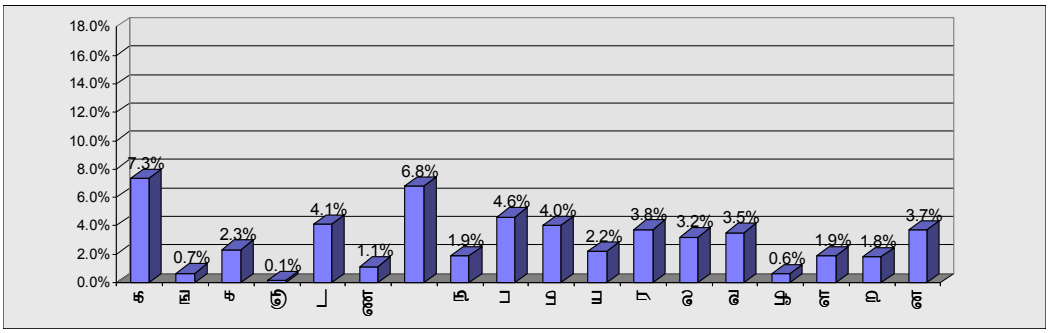
As is can be seen, the Consonant key strokes take up 53.7 % and the Vowel keystrokes take up
46.3%. (45% + 1.3%)

Bar chart on frequency of Vowel Column



Note the relatively high frequency of the vowel keystrokes over consonants below.

Bar chart on frequency of Consonant Column

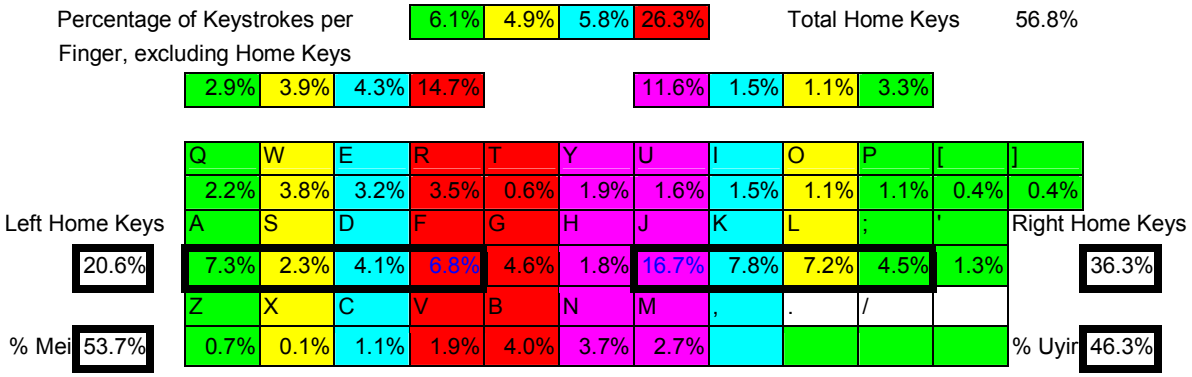


The 3 most frequent consonants in use are க, and ட. The next few characters varied in frequency depending on type of document. After this, the most frequent characters varied between ற, ம and ட.

Also note that although the highest frequency for a consonant is 7.3% for க, while the highest frequency for ஒ is more than double at 16.7 %. Similarly இ and ஐ have relatively high frequencies of 7.2% and 7.8% respectively.

It would be very important to make sure that these 3 consonants and these 3 vowels are on the home keys, especially ஒ.

Thunaivan Phonetic Keyboard Keystroke Distribution



Comments

Despite its orderly arrangement, a surprising number of key positions are in efficient locations. Keyboard efficiency can be improved in the following ways.

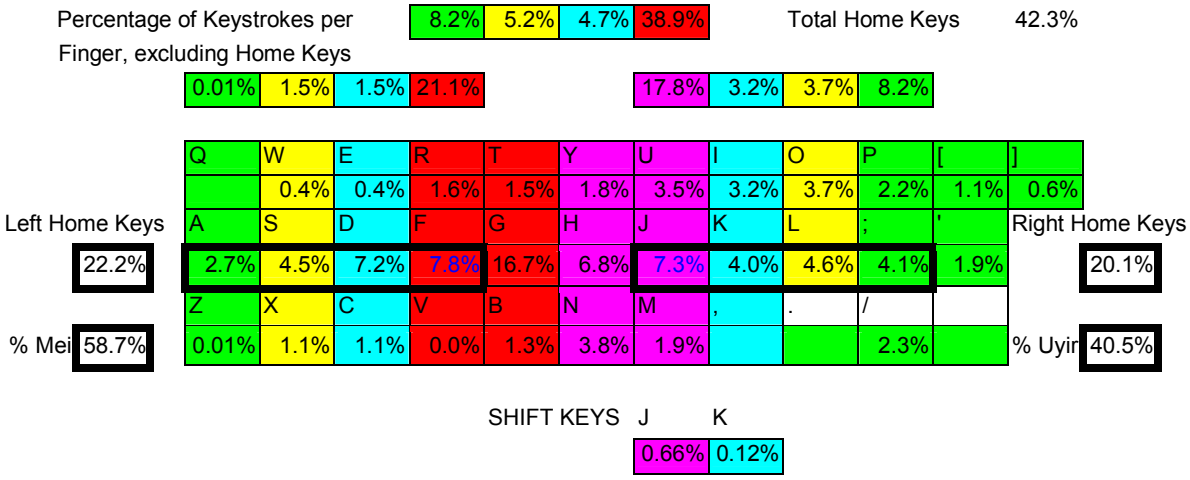
1. 2 keys of relatively high frequency are in inappropriate positions, i.e. 'Q' and 'W' which are ய and ற respectively which require the left little and ring fingers.

2. 'S' on the home keys has a relatively low frequency character ச.

3. ப on the 'G' location should be on one of the home keys.

4. Also the upper right row for the middle and index fingers are used minimally.

I.E. Singapore Keyboard Keystroke Distribution



Comments

The I.E. Singapore keyboard is relatively efficient especially due to its large concentration of keystrokes in the middle of the keyboard. Despite this it has several areas that can be rectified for greater efficiency. These include the following :

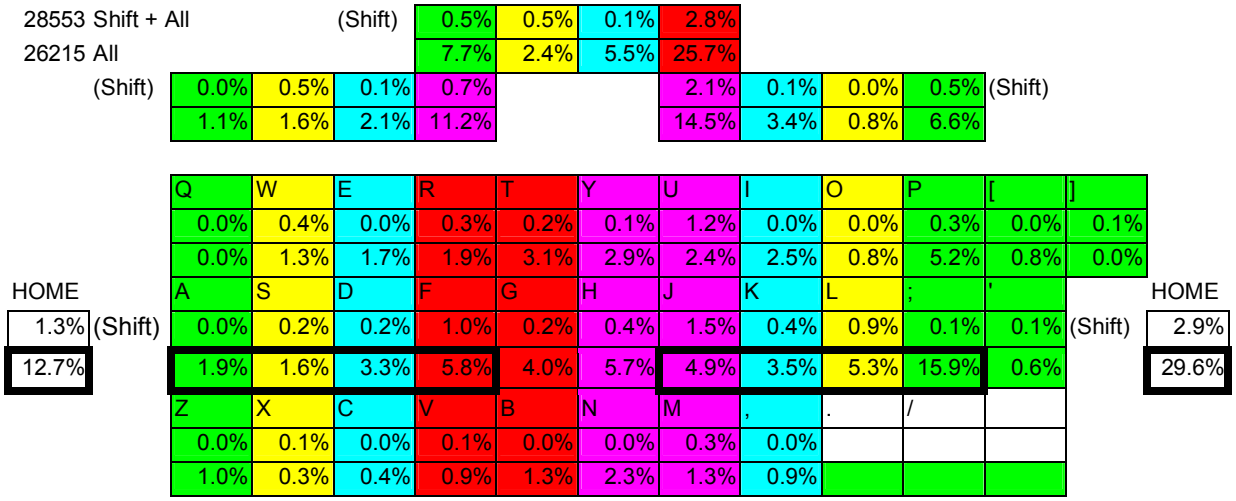
1. 3 keys are practically unused, i.e. 'Q', 'Z' and 'V', especially 'V'. This is because ஒ and ஹ are hardly used in Tamil.

2. The location of the ஃ for • has been placed out of the home keys. Considering it has the single largest frequency of use, it should be on one of the home keys.

3. A disproportionately large number of keystrokes (8.2%) are on the right little finger. In fact it is used more that the much "stronger" middle and ring fingers.

4. The "SHIFT" keys are required for ஹ and ஶ. But, the frequency use is quite minimal.

Tamil Typewriter Keyboard Keystroke Distribution



Total Shift Key	8.2%
Total Home Key	42.3%

Total Non-Shift Keys	92%

	Actual Keys	SHIFT Keys	
Total Home Key	12076		
Total Shift Home Key	1203	1203	
Total Others	11801		
Total Others SHIFT	1135	1135	
Total All Keystrokes	26215	28553	
Total Phonetic Keystrokes	27158	27158	
Difference between Keyboards		1395	5.14%

Comments

This table shows the different keystroke requirements of the Phonetic and the Typewriter keyboards. It is quite surprising to note that there are less keystrokes required for the Typewriter layout. This is because this figure (26215) does not take the SHIFT as an extra keystroke. When this is accounted for, the figure rises to 28553. Despite this, this is only a 5.14% increase in keystrokes over the Phonetic Keyboard. It would seem that there is not much difference between the Phonetic and the Typewriter keyboard.

Further analysis shows that the Home Key use at 42.3% is actually equal to the I.E. Singapore Keyboard. Of course one must take into account that this is based on a slightly higher keystroke count. Even then, it is almost equal.

There is also a high concentration of keys in the middle of the keyboard at 25.7% which is only marginally lower than the Thunaivan Phonetic Keyboard.

Within the constraints placed on its design requirement, the Tamil Typewriter Keyboard is surprisingly efficient. Still, there are several areas that would need to be rectified to increase its efficiency.

1. The location of the "hook" ஂ should preferably be a Home Key considering its relatively high frequency.
2. The keys found on the locations 'T' and 'Y' could be shifted to the 'A' and 'S' locations that are of a relatively lower frequency.
3. The location of ஜ which has taken the position of the "," and the rearrangement of the location for '.' and '?' can be quite disruptive to those used to the English Keyboards.

Table Analysing Relative Efficiency of Keyboards

		Thunaivan Phonetic	I.E. Singapore	Tamil Typewriter
	Left Home Key Use	20.6%	22.2%	12.7%
	Right Home Key Use	36.3%	20.1%	29.6%
This Figure measures the percentage of a document that can be typed without lifting your finger from the home keys.	Total Home Key Use	56.8%	42.3%	42.3%
Non - Home Key	Left Index Finger	16.5%	22.9%	11.2%
	Right Index Finger	7.1%	14.1%	14.5%
	Total Index Finger	23..6%	37.0%	25.7%
The index finger can be considered the "strongest". A higher value is better.	Total Index Finger including Home Key	47.2%	52.1%	36.5%
Non - Home Key	Left Middle Finger	4.3%	1.5%	2.1%
	Right Middle Finger	4.1%	5.1%	3.4%
	Total Middle Finger	8.5%	6.6%	5.5%
The middle finger is almost as "strong" as the index finger. A higher value is better.	Total Middle Finger including Home Key	20.4%	17.9%	12.3%

Non - Home Key	Left Ring Finger	3.9%	1.5%	1.6%
	Right Ring Finger	1.1%	3.7%	0.8%
The ring finger is relatively week. Key use should be minimal.	Total Ring Finger	4.9%	5.2%	2.4%
Non - Home Key	Left Little Finger	2.9%	0.01%	1.1%
	Right Little Finger	3.3%	8.2%	6.6%
The little finger is quite week and slow. Use of the this finger should be minimised.	Total Little Finger	6.1%	8.2%	7.7%
Total Index and Middle Finger. This shows the percentage of keystrokes that are in the middle of the keyboard. This allows you to use your stronger fingers for typing, especially for 2 finger typists.		67.5%	70.0%	48.8%

Note : The SHIFT keystrokes have been excluded from this table to facilitate comparison

As can be noted from the above table, the Phonetic keyboards are relatively more efficient on most of the indices. But, with some modifications, the disparity between the keyboards should be minimal.

Concluding Remarks

In comparing, the 3 keyboard layouts, the Thunaivan Phonetic Keyboard seems to be slightly more efficient than the I.E. Singapore keyboard and much more efficient than the Tamil Typewriter Keyboard with respect to the statistical data.

It is probable though that keyboard frequency is not the prime measure of keyboard efficiency. The orderly arrangement of the keys for easy memorization is probably more significant. With inherently efficient keyboard design, the speed of typing is more often due to the nature of the typists job as well as their psychological disposition and inherent skills.

Furthermore, 80% to 90% of those who will need to type in Tamil in this modern computer age will not professional typists. They will be teachers, writers and students who can be considered casual users. This group will never reach their maximum typing speed potential as the need to do so is not there. Thus, to the majority, ease of use should outweigh physical keyboard efficiency.

Time does not permit at the time of submission of this paper to provide an analysis of this aspect of keyboard design, but some data on this matter should be available in time for the conference.

References

1. The I.E. Singapore Keyboard is copyright of Naa Govindasamy of N.I.E. Singapore.

Speakers'/Authors' Profile

Biography of Author

Ravindran K. Paul is a computer science graduate of the National University of Singapore and the developer and author of the Thunaivan Tamil Wordprocesser and subsequently

Thunaivan for Windows. He is the designer of the first fully phonetic keyboard specifically designed for Tamil in late 1986. All of the phonetic keyboard designs created subsequently in Malaysia and Singapore are a direct result of the introduction of the Thunaivan Phonetic keyboard.

Ravindran K. Paul is a programmer by profession and runs his own software firm which is engaged in the writing of software for commercial and educational pruproses. Currently development of software utilities for Tamil usage on the Internet and for educational purposes is in progress.

Home Page : **http://itstation.com**

Organization : Micro Mart, 116M Jalan Tun Sambanthan, 50470 Kuala Lumpur, Malaysia.
Tel : 03-2742768, Fax : 03-2721281