

入力効率を考慮したタミル語入力メソッド

1 J-5

シバスタラン スハルナン, 唐澤 博
山梨大学 工学部 電子情報工学科

E-Mail: (suharnan, karasawa)@opal.esi.yamanashi.ac.jp

1 はじめに

タミル語は、南インドからマレーシアに至る地域で使用されており、近年、コンピュータ処理化の計画が進んでいる。本研究では、従来から慣用的に使用されてきたタミル語入力メソッド（R.Sundaram 方式、G.Swaminathan 方式）に対して、文字の隣接関係を考慮し、音読みに忠実なローマ字綴りに基づいた入力メソッドを新たに開発した。これはシフト・キーを全く使わない方法である。ここでは従来の方式と入力効率および使用感について比較評価した結果を報告する。

2 入力方式の種類

タミル語は付録1に示しているように、文字数が247種類と多く、そのうち発音が微妙に違う文字が40%あり、子音が文章中に文字として存在することなどで、タミル語入力方式の開発を難しくしている。そこで、一昨年開発したタミル語入力方式[1]を更に改良しタミル語入力の際シフト・キーを全く使わない方式を考案した。表2に我々が開発した入力方式を示している。

この入力方式では表2の母音に対応しているアルファベットを押すことにより母音が入力され、表2の子音に対応しているアルファベットを押すことにより子音が入力されるようにしている。また母音と子音の組合せの文字を入力する際には、順番に子音に対応しているアルファベットを母音に対応しているアルファベットに組み合わせることで入力することにより入力可能としている。しかしその時において子音を入力する時、押す必要となっていた“/”を入力する必要がない。R.Sundaram 方式を表3に示している。R.Sundaram 入力方法は本入力方式と似ている。しかし母音、子音と対応しているアルファベットは異なっている。また R.Sundar

am 方式ではシフト・キーが使われている。表4に示しているのは G.Swaminathan 方式である。G.Swaminathan 方法で母音入力にはタミル文字母音対応表、子音入力にはタミル文字子音対応表、そして音の組合せからなる文字には順番に子音母音組合せ表の第一表を第二表に組み合わせることにより入力できるようになっている。

表1: 本提案方式タミル文字入力対応表

母音文字入力対応表

a	aa	i	ii	u	uu	e	ee	ai	o	oo	au
அ	ஆ	இ	ஈ	உ	ஊ	ஏ	ஈ	ஐ	ஓ	ஔ	ஔ

子音文字入力対応表

k/	c, s/	d	th, t/	p, b/	rr/	ng/	ny/	nnn/
க	ச	ட	த	ப	ர	ங	ந	ண
n/	m/	nn/	y/	r/	l/	v, w/	ll/	ll/
ந	ம	ன	ய	ர	ல	வ	ழ	ள

表2: R.Sundaram 方式によるタミル文字入力対応表

母音文字入力対応表

a	aa	i	ee	u	oo	e	E	ai	o	O	au
அ	ஆ	இ	ஈ	உ	ஊ	ஏ	ஈ	ஐ	ஓ	ஔ	ஔ

子音文字入力対応表

k	c	t	th	p	R	NG	ny	_N
க	ச	ட	த	ப	ர	ங	ந	ண
n	m	_n	y	r	l	v	zh	L
ந	ம	ன	ய	ர	ல	வ	ழ	ள

表3: G.Swaminathan 方式のタミル文字入力対応表

母音文字入力対応表

aa	aA	ii	il	uu	uU	ee	eE	aY	oo	oO	aw
அ	ஆ	இ	ஈ	உ	ஊ	ஏ	ஈ	ஐ	ஓ	ஔ	ஔ

子音文字入力対応表

kk	ss	tt	dd	p	RR	ff	gg	NN
க	ச	ட	த	ப	ர	ங	ந	ண
bb	mm	nn	yy	rr	ll	vv	zz	LL
ந	ம	ன	ய	ர	ல	வ	ழ	ள

母音と子音の組合せのタミル文字入力対応表

第一表

k	s	t	d	p	R	f	g	N
க	ச	ட	த	ப	ர	ங	ந	ண
b	m	n	y	r	l	v	z	L
ந	ம	ன	ய	ர	ல	வ	ழ	ள

第二表

a	A	i	I	u	U	e	E	Y	o	O	w
அ	ஆ	இ	ஈ	உ	ஊ	ஏ	ஈ	ஐ	ஓ	ஔ	ஔ

An Efficient New Input Method for Tamil Language Characters
Masayuki Harada, Hiroshi Karasawa
Yamanashi University
4-3-11 Takeda, Kofu, Yamanashi 400, Japan

3 各入力方式を用いての入力実験結果

評価実験では”チルクラル”という一般的なタミル語テキストの400文字分を入力することにした。同じ文章を1回目2回目と各入力方式で入力し得られた結果を表4,5,にまとめている。表4のAve1はシフト・キーの重みを”0”とした時の一タミル文字表に使われた平均打鍵数でAve2はシフト・キー評価実験からの重み”0.57”を使つての平均打鍵数である。表5のAveは一タミル文字表示に使つた平均時間を示す。

表4: 総合打鍵数

Per. No	Suharnan		Sundaram		Swaminathan	
	1st	2nd	1st	2nd	1st	2nd
1	1148	974	1135	1009	954	997
2	813	780	904	808	785	769
3	899	892	1078	920	959	911
4	989	898	997	896	996	1022
5	976	1039	885	—	870	952
6	907	829	950	1070	964	946
7	913	906	966	924	901	919
8	945	837	961	958	911	900
9	865	845	1027	1024	859	844
10	941	1062	1033	865	916	889
Ave ₁	2.30		2.42		2.28	
Ave ₂	2.30		2.50		2.36	

表5: 総合打鍵時間

Per. No	Suharnan		Sundaram		Swaminathan	
	1st	2nd	1st	2nd	1st	2nd
1	25'26"	19'24"	30'49"	23'39"	36'56"	37'49"
2	17'29"	14'26"	25'10"	19'22"	25'33"	24'01"
3	18'30"	13'38"	21'22"	14'12"	28'15"	19'07"
4	25'29"	17'38"	33'09"	20'07"	27'19"	16'47"
5	22'46"	18'56"	27'02"	—	36'45"	39'48"
6	11'05"	10'15"	20'11"	17'56"	40'42"	30'17"
7	17'56"	15'38"	24'22"	23'08"	39'23"	37'15"
8	16'00"	14'24"	22'08"	21'08"	37'42"	35'19"
9	9'46"	9'30"	15'52"	15'39"	39'37"	37'06"
10	19'07"	18'37"	18'40"	17'25"	40'47"	37'58"
Ave	2.52sec.		3.24sec.		5.01sec.	

4 考察

1. 表5より、一文字当たりの打鍵数を比較すると本方式は2.30と打鍵数が一番低く次に2.36,2.5とG.Swaminathan方式,R.Sundaram方式が続く。これは我々の方式で打鍵数が多いタミル文字が一般タミル語文章で出てくる確率が少ないことが理由である[2]と考えられる。

2. 表6より一文字当たりの打鍵時間を比較しても本方式の一文字当たりの入力時間2.52秒と一番低く次に3.24秒,5.01秒とR.Sundaram方式,G.Swaminathan方式が続く、これは本入力方式は他の入力方式より音読みに忠実なローマ字綴りに基づいた入力メソッドになっていたことやシフト・キーを全く使わない方法であることが理由と考えられる。

3. 被験者の感想として、今回開発した方法は他の方法よりはるかに覚え易いという意見があった。

5 おわりに

本提案方式は今まで使われていたタミル語入力方式より効率が良いということを評価実験が示している。しかし発音が似ている文字を入力する時には打鍵数が多くなっている。この問題は、日本語変換辞書を日本語の入力に使うのと同じように、タミル語辞書をタミル語入力の際に用いて発音の組合せを調べることにより解決できると考える。

参考文献

- [1] 武富敬, シバスタランスハルナン, 松永一美: タミル言語の統計的性質, 情報処理学会全国大会講演論文集(3), PP.81-82(1994)
- [2] Taketomi, H and Suharnan, S: Design and Implementation of Tamil Processing System, ICCPOL94, PP.149-154(1994)

付録1: タミル文字表

தமிழ் எழுத்து அட்டவணை

அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஔ	஌	஍
க	கா	கி	கீ	கு	கூ	கெ	கே	கை	கொ	கோ	கௌ	கௌ
ச	சா	சி	சீ	சு	சூ	செ	சே	சை	சொ	சோ	சௌ	சௌ
ட	டா	டி	டீ	டு	டூ	டெ	டே	டை	டொ	டோ	டௌ	டௌ
த	தா	தி	தீ	து	தூ	தெ	தே	தைய	தொ	தோ	தௌ	தௌ
ப	பா	பி	பீ	பு	பூ	பெ	பே	பைய	பொ	போ	பௌ	பௌ
ற	றா	றி	றீ	று	றூ	றெ	றே	றைய	றொ	றோ	றௌ	றௌ
ங	ஙா	ஙி	ஙீ	ஙு	ஙூ	ஙெ	ஙே	ஙைய	ஙொ	ஙோ	ஙௌ	ஙௌ
ஞ	ஞா	ஞி	ஞீ	ஞு	ஞூ	ஞெ	ஞே	ஞைய	ஞொ	ஞோ	ஞௌ	ஞௌ
ண	ணா	ணி	ணீ	ணு	ணூ	ணெ	ணே	ணைய	ணொ	ணோ	ணௌ	ணௌ
ந	நா	நி	நீ	நு	நூ	நெ	நே	நைய	நொ	நோ	நௌ	நௌ
ம்	மா	மி	மீ	மு	மூ	மெ	மே	மைய	மொ	மோ	மௌ	மௌ
ன்	னா	னி	னீ	னு	னூ	னெ	னே	னைய	னொ	னோ	னௌ	னௌ
ய	யா	யி	யீ	யு	யூ	யெ	யே	யைய	யொ	யோ	யௌ	யௌ
ர்	ரா	ரி	ரீ	ரு	ரூ	ரெ	ரே	ரைய	ரொ	ரோ	ரௌ	ரௌ
ல்	லா	லி	லீ	லு	லூ	லெ	லே	லைய	லொ	லோ	லௌ	லௌ
வ்	வா	வி	வீ	வு	வூ	வெ	வே	வைய	வொ	வோ	வௌ	வௌ
ழ்	ழா	ழி	ழீ	ழு	ழூ	ழெ	ழே	ழைய	ழொ	ழோ	ழௌ	ழௌ
ள்	ளா	ளி	ளீ	ளு	ளூ	ளெ	ளே	ளைய	ளொ	ளோ	ளௌ	ளௌ