

BHASHA IME

(Free software)

User Reference

Version 7.3.0

<http://sites.google.com/site/bhashaime/>

Send feedback and update notification request to

bhashaime@gmail.com

Author : Venkatesh

Table of Contents

Overview.....	3
Phonetic Input Scheme.....	5
Devanagari/Grantha.....	5
Infrequent Input (Devanagari, Grantha, Kannada (GodaGr font), Telugu (Pothana2000Vedic).....	5
Vedic Input - Devanagari, Grantha, Kannada (GodaGr Font), Telugu (Pothana2000Vedic).....	5
PUA symbols in Siddhanta Font (Vedic mode).....	6
Support for Alternate forms of svara representation.....	6
Kannada/Malayalam/Telugu.....	7
Tamil.....	7
Devanagari symbols with non-Devanagari (Kannada/Telugu etc.).....	7
IAST/ISO 15919.....	8
IAST and ISO15959 in Upper Case.....	8
ZWJ, ZWNJ, and ZWS (for all scripts).....	8
User configurable Parametrs.....	9
Auxiliary features (for all scripts).....	9
Transliteration.....	10
OCR Training.....	10
Configuring Windows.....	10
Configuring Applications.....	10

Overview

The IME helps creation of Unicode text content in Devanagari , e-Grantha, Kannada, Malayalam, Telugu, Tamil, and Roman (IAST/ISO15919) scripts; Vedic input in Devanagari and e-Grantha inclusive. Besides, the IME also supports transcription of Unicode-to-Unicode and Non-Unicode-to-Unicode.

Platform: Windows XP SP3 (developed & tested), Win7 and Vista

Target Applications: All Windows applications in general; Openoffice, LibreOffice, Microsoft office suite, IE, Firefox, Chrome , NotePad, WordPad for eg.

Text Input Scheme: Phonetic (default), and INSCRIPT (menu selectable)

Unicode Support: Subset of Ver 6.3

Script Support: Devanagari, e-Grantha, Kannada, Malayalam, Tamil, Telugu, and Roman with Diacritics (IAST and ISO15919)

Installation: The application is portable. No installation required; just run the exe.

Usage: Run the exe. The IME shows an empty icon in the system tray. Open/switch-to the target application, left/right-click the icon, and choose a script from the pop-up menu. The icon will change to reflect the selected script. You could also choose the desired script using the short-cut indicated in the system tray menu instead. You can start typing in the application. The IME can be turned on/off from the menu/shortcut.

Document Tracking: The IME tracks each open doc/window and associated script. Switching to a doc/window will cause IME to automatically switch to the last used script in the doc/window. (Of course, each doc could have text in multiple scripts).

Fonts: Any Unicode complaint font for classical texts. For Vedic texts, Siddhanta (Devanagari) and e-Grantha (Grantha) are most suited. For Roman with Diacritics, any font which supports it (eg. DejaVu, Linux Libertine, Gentium Plus).

Phonetic Input Scheme

+ESC means pressing Escape key after the preceding keys. (Omit '+')

Devanagari/Grantha

a	aa/A	i	ii/I	u	uu/U	R	RR	L	LL	ऐ = E	e	ai	औ = O	o	au	M	H
				k	kh	g	gh	G		c	ch	j	jh	J			
				T	Th	D	Dh	N		t	th	d	dh	n			
				p	ph	b	bh	m									
y	r	l	v	sh/S	shh/Sh	s	h	x = ऌ	᳚ = z	l = .	ll = ..	ॐ=om+ESC					

eg. श्रीमान् वेङ्कटनाथार्यः == shriimaan veGkaTanaathaaryaH Devanagari only

Infrequent Input (Devanagari, Grantha, Kannada (GodaGr font), Telugu (Pothana2000Vedic))

ॠ/᳚ = gm+ESC	ॡ/᳛ = gg+ESC	यँ = y+w	वँ = v+w	लँ = l+w
× (upadhmaniya) = HH (alternately, H + ESC)				
× (jihvamuliya) = HHH (alternately, H + ESC + ESC)				
◌̣ (anunasika) = w				

Vedic Input - Devanagari, Grantha, Kannada (GodaGr Font), Telugu (Pothana2000Vedic)

Switch to Vedic Mode by pressing the two SHIFT keys simultaneously. The Tray Icon will change to indicate Vedic Mode. Press the SHIFT keys again to revert to classic (non-Vedic) mode.

◌̣ (svarita) = ' (single quote)	◌̣̣ (anudatta) = / (forward slash)
" (dirgha svarita) = ' + ' (2 successive quotes) or " (a double quote)	

The GodaGr font supports Taittiriya and Sama Svaras using Graphite Technology. See the font document for more details.

PUA symbols in Siddhanta Font (Vedic mode)

: (visarga) = Alt+Ctl+Shift+h		
' (svarita) = Alt+Ctl+' (quote)	" (dirgha svarita) = ' + ' (2 quotes) or " (a double quote) or Alt+Ctl+" (2 quotes) or Alt+Ctl+Shift+" (a double quote)	
— (long top bar) = Alt+Ctl++ (plus)	- (short) = Alt+Ctl+- (hyphen)	` (mini) = Alt+Ctl+Shift+- (hyphen)

Support for Alternate forms of svāra representation

Siddhanta font presents svarita and dirgha-svarita in a syllable with leading non-base consonant “Ra” ('reph' in Opentype terminology), above the 'reph' form (र्व, र्वी). It may be desirable to have the svāra along side the reph, for more clarity. This can be effected by pressing **ESC** following the original form.

Also every successive ESC increases the distance between the 'reph' and the svāra.

Be mindful that the new form uses codepoints from PUA rather than the Unicode-defined.

र्व + ESC => र्व ¹	र्व ¹ + ESC => र्व ¹	र्व ¹ + ESC => र्व ²	र्व ² + ESC => र्व ²
-------------------------------	--	--	--

Kannada/Malayalam/Telugu

a	aa/A	i	ii/I	u	uu/U	R	RR	L	LL	e	ee/E	ai	o	oo/O	au	M	H
			k	kh	g	gh	G			c	ch	j	jh	J			
			T	Th	D	Dh	N			t	th	d	dh	n			
			p	ph	b	bh	m										
				y	r	l	v	sh/S	shh/Sh	s	h	x [= ಷ್/ಞ್/ಞ್]	w [= ಁ (anunasika)]				

Tamil

a	aa/A	i	ii/I	u	uu/U	e	ee/E	ai	o	oo/O	au	M	H
	k				G			c		j		J	
	T				N			t				n	
	p				m								
y	r	l	v	z	L/x = ள்	R/(r + ESC)	n + ESC	sh/S	shh/Sh	s	h		

Devanagari symbols with non-Devanagari (Kannada/Telugu etc.)

l (danda) = Alt+Ctl+.	ll (double danda) = Alt+Ctl+..
-----------------------	--------------------------------

IAST/ISO 15919

a	aa/A	i	ii/I	u	uu/U	R	RR	L	LL	ě = E/e	e/ee/E	ai	ö = O/o	o/oo/O	au	M	H
				k	kh	g	gh	G		c	ch	j	jh	J			
				T	Th	D	Dh	N		t	th	d	dh	n			
				p	ph	b	bh	m									
y	r	l	v	sh/S	shh/Sh	s	h	x [= <u>ḷ</u> ! (ळ्/ॡ/ँर)]					w [= <u>ṽ</u> (anunasika)]				
H + ESC [= k (Tamil ஃ)]				z [= <u>ḷ</u> ! (Tamil ழ)]				r + ESC [= ṛ (Tamil ற)]				n + ESC [ṇ (Tamil ண)]					

IAST ISO

IAST and ISO15959 in Upper Case

To type in upper case, keep the CAPS LOCK key on.

Also there is help for occasional change of case (as in beginning of a sentence). IME supports change of case for the last typed character with Shift+ESC key-combo.

a + Shift + ESC = A	Ḥ + Shift + ESC = ḥ
ḷ + Shift + ESC = Ḹ	and so on ...

ZWJ, ZWNJ, and ZWS (for all scripts)

ZWJ == Shift+Space	ZWNJ == Alt+Ctl+Space	ZWS == Alt+Ctl+Shift+Space
--------------------	-----------------------	----------------------------

User configurable Parameters

A number of parameters are available to be set by user. These can set in User_Config.txt in the K_ folder.

Auxiliary features (for all scripts)

- To make typing syllables like ग् (गु), ज्ञ (गुञ्) more intuitive, the IME converts the ग् (गु) to ग् (गु) and ज्ञ (गुञ्) respectively when followed by क्/क varga or च्/च varga. Thus the above syllables can be typed as 'mga' and 'mca' instead of Gga and Jca.
- The above conversion however can be avoided for the syllable by pressing **ESC** key after m. Thus 'mESCga' would result in ग्ग (गुगु)
- Key combos like 'aa', 'ii', 'uu', 'ai', 'au', 'kh', 'ch', ... can be broken by keying an **ESC** between the keys. Thus, while key sequence 'aa' would result in आ, aESCa would be अअ. Similarly, 'iESCi' would be इइ, 'kESCh' would be क्क्ह.
- In Devanagari and Grantha, the period key (.) obtains the danda char (।). To obtain the period char (.), type period key (.) followed by the **ESC** key. (। + **ESC** -> .)
- In Tamil, ந்/ந்/நா/நி/நீ/நு/நூ/நெ/நே/நை/நொ/நோ/நௌ followed by **ESC** key results in ண்/ண/ணா/ணி/ணீ/ணு/ணூ/ணெ/ணே/ணை/ணொ/ணோ/ணௌ. The feature works both ways (ie. ண்/... + **ESC** ---> ந்/...)
- Also, in Tamil, ற் and ற் can be muted with the **ESC** key, just as above.
- 'q' works as an alternate for **ESC** key, useful with applications like GMail which block **ESC**; also, just in case it's more comfortable.
- 'f' is used only for OCR training (detailed below) and generally useless for normal text entry.
- To manifest phonetic operation, the IME depends on being able to keep track of keys typed so far. Moving the cursor away from an incomplete syllable (except backspacing), (using arrow keys, page-up/page-down/mouse-click etc.) would make IME loose track and start a new syllable.

Transliteration

Refer to Transliteration-Guide.pdf.

OCR Training

Training OCR software (eg. Tesseract) requires the characters like ா, ி, etc., (in Tamil) to be typed . The chars can be obtained by typing ஆ, இ, etc., followed by 'f'.

OCR char	ா	ி	ீ	ு	ூ	ெ	ே	ை	ொ	ோ	ௌ	ஃ
Keys	aaf	if	iif	uf	uuf	ef	eef	aif	of	oof	auf	\f

Configuring Windows

To enable support for Bharatiya scripts in Windows XP,

- Go to 'Control panel', and click on 'Regional and language options'
- Go to 'Languages' Tab
- Check select 'Install Files for complex script and right-to-left languages [including Thai]'
- Press OK

Configuring Applications

Some applications need to be tune-configured for IME to work with them.

LibreOffice

- Go to 'Tools'->'AutoCorrect Options'->'Word Completion'
- Uncheck 'Enable Word Completion' and 'Collect Words'

Microsoft Office 2003

- Go to Start->All Programs->Microsoft Office->Microsoft Office Tools->Microsoft Office 2003 Language Settings
- In 'Enabled Languages' Tab, choose 'Indic' from 'Available languages' combo

- Select all scripts you need from the list below and Press 'Add >>' button
- Press OK

Microsoft Office Word 2003

- Go to 'Tools'->'Options'->'Edit'
- Uncheck 'Smart cut and paste'
- Press OK