

Universal Multiple-Octet Coded Character Set
 International Organization for Standardization
 Organisation internationale de normalisation
 Международная организация по стандартизации

Doc Type: Working Group Document

Title: Proposal to encode four characters for Oriya and Malayalam in the BMP of the UCS

Source: Michael Everson and Peter Scharf

Status: Individual Contribution

Action: For consideration by JTC1/SC2/WG2 and UTC

Replaces: N3235R section 10

Date: 2007-05-25

Introduction. This document requests the addition to the UCS of one character for Malayalam and three characters for Oriya. These characters are needed to complete the set of vocalic liquids for these two scripts. Classical Sanskrit grammar gives four independent vowels ऋ *r̄*, ॠ *r̄̄*, ॡ *l̄*, ॢ *l̄̄*, and four dependent vowel matras ृ *-r̄*, ॠ *-r̄̄*, ॡ *-l̄*, ॢ *-l̄̄*. Oriya currently has only five of these eight encoded: ୠ *r̄*, ୡ *r̄̄*, ୢ *l̄*, ୣ *l̄̄*, and ୦ *-r̄*. Malayalam has all encoded but one, ൠ *-l̄̄*.

- ൠ ORIYA VOWEL SIGN VOCALIC RR is used to mark a long vowel *r̄̄*. (Figure 1)
- ൡ ORIYA VOWEL SIGN VOCALIC L is used to mark a vocalic *l̄*. (Figure 1)
- ൢ ORIYA VOWEL SIGN VOCALIC LL is used to mark a long vocalic *l̄̄*. (Figure 2)
- ൣ MALAYALAM VOWEL SIGN VOCALIC LL is used to mark a long vocalic *l̄̄*. (Figure 3)

The current standard contains the independent vowel letters as well as vowel signs for most Brahmic scripts, particularly those which are still concerned with the Sanskrit language. In the table below, the independent vowel signs are given, along with the consonants *ra* and *la* for comparison.

	ṛ	ṝ	ḷ	ḹ	-ṛ	-ṝ	-ḷ	-ḹ	ra	la
Devanagari	ऋ	ॠ	ॡ	ॢ	ृ	ॠ	ॡ	ॢ	र	ल
Bengali	ঋ	ঠ	ড	ঢ	ৠ	ৡ	ৢ	ৣ	র	ল
Gujarati	ઋ	ૠ	ૡ	ૢ	ૣ	૤	૥	૦	ર	લ
Oriya	ୠ	ୡ	ୢ	ୣ	୦	ୠ	ୡ	ୢ	ର	ଲ
Telugu	ఋ	ౠ	ౡ	ౢ	ౣ	౤	౥	౦	ర	ల
Kannada	ಋ	ೠ	ೡ	ೢ	ೣ	೤	೥	೦	ರ	ಲ
Malayalam	ഠ	ഡ	ണ	ഢ	ൠ	ൡ	ൢ	ൣ	ര	ല
Sinhala	ආ	ඈ	ඊ	උ	ා	෦෦	ා	෦෦	ර	ල
Myanmar	ာ	ာ	ေ	ေ	့	့	့	့	ရ	လ
Balinese	ꦲ	ꦶ	ꦸ	ꦺ	ꦶ	ꦶ	ꦶ	ꦶ	ꦲ	ꦺ

Unicode Character Properties. Character properties are proposed here.

0B44;ORIYA VOWEL SIGN VOCALIC RR;Mn;0;NSM;;;;;N;;;;;
 0B62;ORIYA VOWEL SIGN VOCALIC L;Mn;0;NSM;;;;;N;;;;;
 0B63;ORIYA VOWEL SIGN VOCALIC LL;Mn;0;NSM;;;;;N;;;;;
 0D63;MALAYALAM VOWEL SIGN VOCALIC LL;Mn;0;NSM;;;;;N;;;;;

Figures.

Figure 1. ORIYA VOWEL SIGN VOCALIC RR and ORIYA VOWEL SIGN VOCALIC L in Sir Hugh McPherson’s “The Oriya Alphabet,” *Journal of Bihar and Orissa Research Society*, Mar-Jun 1924. Table I shows vowels with entries of: Roman, Nāgarī, Oriya and Remarks.

Vowels, Etc.			
Roman	Nāgarī	Oriya	Remarks.
ōm	ଓଁ	ଓଁ	This is the opening invocation
si	ସି	ସି	
dehi	ଦି	ଦି	
ra	ର	ର	
sta	ସ୍ତା	ସ୍ତା	
a	ଅ	ଅ	
ā	ଆ	ଆ	
-ā	।	।	The sign । is placed to the right of the consonant.
i	ଈ	ଈ	
-i	ୂ	ୂ	This sign ୂ is placed on the head of the consonant.
ṛ	ୠ	ୠ	
-ṛ	ୡ	ୡ	Placed to the right.
u	ଉ	ଉ	
-u	ୃ	ୃ	Placed below.
ū	ୂ	ୂ	
-ū	ୄ	ୄ	Placed below.
r̄	ୠ	ୠ	(1) This letter sounds as ru in Oriya.
-r̄	ୡ	ୡ	Placed below.
r̄̄	ୢ	ୢ	
-r̄̄	ୣ	ୣ	Placed below.
l̄	ୠ	ୠ	(2) This letter sounds as lu in Oriya.
-l̄	ୡ	ୡ	Placed below.

Figure 1

Figure 2. ORIYA VOWEL SIGN VOCALIC LL We have as yet been unable to find an image of ORIYA VOWEL SIGN VOCALIC LL. As can be seen in the chart on page 1, the typical strategy for representing the VOWEL SIGN VOCALIC L and VOWEL SIGN VOCALIC LL in a number of scripts is to reduce and subscript LETTER VOCALIC L and LETTER VOCALIC LL. There is no question but that VOWEL SIGN VOCALIC LL must exist in Oriya script for transliteration of \bar{I} ; if the glyph proves to be different from what is propose here it can be changed, as glyphs are informative.

TABLE XX - Row 0B: ORIYA

	0B0	0B1	0B2	0B3	0B4	0B5	0B6	0B7
0		ଈ	୦	ର	଼		ୱ	✓
1	ୃ		ଡ		ୂ		୧	ଓ
2	ୌ		ଣ	କ	ୃ		ୈ	
3	ୈ	ଓ	ଣ	କ	ୃ		ୈ	
4		ଐ	ଡ		ୃ			
5	ଧ	କ	ଧ	କ				
6	ଧା	କ	କ	କ		ୌ	୦	
7	ଈ	କ	ଧ	ଧ	ୈ	ୌ	୧	
8	ଈ	କ	କ	କ	ୈ		୨	
9	ଈ	ଡ		କ			୩	
A	ଈ	କ	ଧ				୪	
B	ଈ	କ	ଧ		ୈ		୫	
C	ଈ	କ	କ	ୃ	ୈ	ୈ	୬	
D		୧	ୈ	ୈ	ୈ	ୈ	୭	
E		୧	୧	୧			୮	
F	୧	୧	୧	୧		୧	୯	

G = 00
P = 00

TABLE XX - Row 0B: ORIYA

hex	Name	hex	Name
00	(This position shall not be used)	59	(This position shall not be used)
01	ORIYA SIGN CANDRABINDU	5A	(This position shall not be used)
02	ORIYA SIGN ANUSVARA	5B	(This position shall not be used)
03	ORIYA SIGN VISARGA	5C	ORIYA LETTER RRA
04	(This position shall not be used)	5D	ORIYA LETTER RHA
05	ORIYA LETTER A	5E	(This position shall not be used)
06	ORIYA LETTER AA	5F	ORIYA LETTER YYA
07	ORIYA LETTER I	60	ORIYA LETTER VOCALIC RR
08	ORIYA LETTER II	61	ORIYA LETTER VOCALIC LL
09	ORIYA LETTER U	62	<u>ORIYA VOWEL SIGN VOCALIC L</u>
0A	ORIYA LETTER UU	63	<u>ORIYA VOWEL SIGN VOCALIC LL</u>
0B	ORIYA LETTER VOCALIC R	64	(This position shall not be used)
0C	ORIYA LETTER VOCALIC L	65	(This position shall not be used)
0D	(This position shall not be used)	66	ORIYA DIGIT ZERO
0E	(This position shall not be used)	67	ORIYA DIGIT ONE
0F	ORIYA LETTER E	68	ORIYA DIGIT TWO
10	ORIYA LETTER AI	69	ORIYA DIGIT THREE
11	(This position shall not be used)	6A	ORIYA DIGIT FOUR
12	(This position shall not be used)	6B	ORIYA DIGIT FIVE
13	ORIYA LETTER O	6C	ORIYA DIGIT SIX
14	ORIYA LETTER AU	6D	ORIYA DIGIT SEVEN
15	ORIYA LETTER KA	6E	ORIYA DIGIT EIGHT
16	ORIYA LETTER KHA	6F	ORIYA DIGIT NINE
17	ORIYA LETTER GA	70	ORIYA ISSHAR
18	ORIYA LETTER GHA	71	ORIYA LETTER WA
19	ORIYA LETTER NGA	72	(This position shall not be used)
1A	ORIYA LETTER CA	73	(This position shall not be used)
1B	ORIYA LETTER CHA	74	(This position shall not be used)
1C	ORIYA LETTER JA	75	(This position shall not be used)
1D	ORIYA LETTER JHA	76	(This position shall not be used)
1E	ORIYA LETTER NYA	77	(This position shall not be used)
1F	ORIYA LETTER TTA	78	(This position shall not be used)
20	ORIYA LETTER TTHA	79	(This position shall not be used)
21	ORIYA LETTER DDA	7A	(This position shall not be used)
22	ORIYA LETTER DDHA	7B	(This position shall not be used)
23	ORIYA LETTER NNA	7C	(This position shall not be used)
24	ORIYA LETTER TA	7D	(This position shall not be used)
25	ORIYA LETTER THA	7E	(This position shall not be used)
26	ORIYA LETTER DA	7F	(This position shall not be used)
27	ORIYA LETTER DHA		
28	ORIYA LETTER NA		
29	(This position shall not be used)		
2A	ORIYA LETTER PA		
2B	ORIYA LETTER PHA		
2C	ORIYA LETTER BA		
2D	ORIYA LETTER BHA		
2E	ORIYA LETTER MA		
2F	ORIYA LETTER YA		
30	ORIYA LETTER RA		
31	(This position shall not be used)		
32	ORIYA LETTER LA		
33	ORIYA LETTER LLA		
34	(This position shall not be used)		
35	ORIYA LETTER VA		
36	ORIYA LETTER SHA		
37	ORIYA LETTER SSA		
38	ORIYA LETTER SA		
39	ORIYA LETTER HA		
3A	(This position shall not be used)		
3B	(This position shall not be used)		
3C	ORIYA SIGN NUKTA		
3D	ORIYA SIGN AVAGRAHA		
3E	ORIYA VOWEL SIGN AA		
3F	ORIYA VOWEL SIGN I		
40	ORIYA VOWEL SIGN II		
41	ORIYA VOWEL SIGN U		
42	ORIYA VOWEL SIGN UU		
43	ORIYA VOWEL SIGN VOCALIC R		
44	<u>ORIYA VOWEL SIGN VOCALIC RR</u>		
45	(This position shall not be used)		
46	(This position shall not be used)		
47	ORIYA VOWEL SIGN E		
48	ORIYA VOWEL SIGN AI		
49	(This position shall not be used)		
4A	(This position shall not be used)		
4B	ORIYA VOWEL SIGN O		
4C	ORIYA VOWEL SIGN AU		
4D	ORIYA SIGN VIRAMA		
4E	(This position shall not be used)		
4F	(This position shall not be used)		
50	(This position shall not be used)		
51	(This position shall not be used)		
52	(This position shall not be used)		
53	(This position shall not be used)		
54	(This position shall not be used)		
55	(This position shall not be used)		
56	ORIYA AI LENGTH MARK		
57	ORIYA AU LENGTH MARK		
58	(This position shall not be used)		

Group 00

Plane 00

Row 0B

TABLE XX - Row 0D: MALAYALAM

	E70	E71	E72	E73	E74	135	136	137
0		ഘി	ഠ	ര	ീ		ഈ	൧
1			ഡ	റ	ൂ		ഞ	൩
2	ം	ക	ഡ	ല	ൂ		്	൯
3	ഃ	ഓ	ണ	ള	്യ		്	൯
4		ഔ	ത	ഴ	്യ			ൽ
5	അ	ക	ഥ	വ				ൺ
6	ആ	ച	ഭ	ശ	െ		ഠ	
7	ഇ	ശ	ധ	ഷ	േ	ൗ	ഥ	
8	ഈ	ഘ	ന	സ	ൈ		ര	
9	ഉ	ങ		ഹ			ന	൯
A	ഊ	ച	ച		ൊ		ര	ൺ
B	ഈ	ഛ	ഘ		ോ		ര	ൺ
C	ൺ	ജ	ബ		ൌ		ന	ര
D		ഡ	ഭ	ഃ	ൂ		ൗ	ൽ
E	ഘി	ഞ	മ	ാ			വ	ൾ
F	ഘി	ട	യ	ി			ൺ	ൽ

G = 00
P = 00

TABLE XX - Row 0D: MALAYALAM

hex	Name	hex	Name
00	(This position shall not be used)	59	(This position shall not be used)
01	(This position shall not be used)	5A	(This position shall not be used)
02	MALAYALAM SIGN ANUSVARA	5B	(This position shall not be used)
03	MALAYALAM SIGN VISARGA	5C	(This position shall not be used)
04	(This position shall not be used)	5D	(This position shall not be used)
05	MALAYALAM LETTER A	5E	(This position shall not be used)
06	MALAYALAM LETTER AA	5F	(This position shall not be used)
07	MALAYALAM LETTER I	60	MALAYALAM LETTER VOCALIC RR
08	MALAYALAM LETTER II	61	MALAYALAM LETTER VOCALIC LL
09	MALAYALAM LETTER U	62	<i>MALAYALAM VOWEL SIGN VOCALIC L</i>
0A	MALAYALAM LETTER UU	63	<i>MALAYALAM VOWEL SIGN VOCALIC LL</i>
0B	MALAYALAM LETTER VOCALIC R	64	(This position shall not be used)
0C	MALAYALAM LETTER VOCALIC L	65	(This position shall not be used)
0D	(This position shall not be used)	66	MALAYALAM DIGIT ZERO
0E	MALAYALAM LETTER E	67	MALAYALAM DIGIT ONE
0F	MALAYALAM LETTER EE	68	MALAYALAM DIGIT TWO
10	MALAYALAM LETTER AI	69	MALAYALAM DIGIT THREE
11	(This position shall not be used)	6A	MALAYALAM DIGIT FOUR
12	MALAYALAM LETTER O	6B	MALAYALAM DIGIT FIVE
13	MALAYALAM LETTER OO	6C	MALAYALAM DIGIT SIX
14	MALAYALAM LETTER AU	6D	MALAYALAM DIGIT SEVEN
15	MALAYALAM LETTER KA	6E	MALAYALAM DIGIT EIGHT
16	MALAYALAM LETTER KHA	6F	MALAYALAM DIGIT NINE
17	MALAYALAM LETTER GA	70	<i>MALAYALAM NUMBER TEN</i>
18	MALAYALAM LETTER GHA	71	<i>MALAYALAM NUMBER ONE HUNDRED</i>
19	MALAYALAM LETTER NGA	72	<i>MALAYALAM NUMBER ONE THOUSAND</i>
1A	MALAYALAM LETTER CA	73	<i>MALAYALAM FRACTION ONE QUARTER</i>
1B	MALAYALAM LETTER CHA	74	<i>MALAYALAM FRACTION ONE HALF</i>
1C	MALAYALAM LETTER JA	75	<i>MALAYALAM FRACTION THREE QUARTERS</i>
1D	MALAYALAM LETTER JHA	76	(This position shall not be used)
1E	MALAYALAM LETTER NYA	77	(This position shall not be used)
1F	MALAYALAM LETTER TTA	78	(This position shall not be used)
20	MALAYALAM LETTER TTHA	79	<i>MALAYALAM DATE MARK</i>
21	MALAYALAM LETTER DDA	7A	<i>MALAYALAM LETTER CHILLU NN</i>
22	MALAYALAM LETTER DDHA	7B	<i>MALAYALAM LETTER CHILLU N</i>
23	MALAYALAM LETTER NNA	7C	<i>MALAYALAM LETTER CHILLU RR</i>
24	MALAYALAM LETTER TA	7D	<i>MALAYALAM LETTER CHILLU L</i>
25	MALAYALAM LETTER THA	7E	<i>MALAYALAM LETTER CHILLU LL</i>
26	MALAYALAM LETTER DA	7F	<i>MALAYALAM LETTER CHILLU K</i>
27	MALAYALAM LETTER DHA		
28	MALAYALAM LETTER NA		
29	(This position shall not be used)		
2A	MALAYALAM LETTER PA		
2B	MALAYALAM LETTER PHA		
2C	MALAYALAM LETTER BA		
2D	MALAYALAM LETTER BHA		
2E	MALAYALAM LETTER MA		
2F	MALAYALAM LETTER YA		
30	MALAYALAM LETTER RA		
31	MALAYALAM LETTER RRA		
32	MALAYALAM LETTER LA		
33	MALAYALAM LETTER LLA		
34	MALAYALAM LETTER LLLA		
35	MALAYALAM LETTER VA		
36	MALAYALAM LETTER SHA		
37	MALAYALAM LETTER SSA		
38	MALAYALAM LETTER SA		
39	MALAYALAM LETTER HA		
3A	(This position shall not be used)		
3B	(This position shall not be used)		
3C	(This position shall not be used)		
3D	MALAYALAM SIGN AVAGRAHA		
3E	MALAYALAM VOWEL SIGN AA		
3F	MALAYALAM VOWEL SIGN I		
40	MALAYALAM VOWEL SIGN II		
41	MALAYALAM VOWEL SIGN U		
42	MALAYALAM VOWEL SIGN UU		
43	MALAYALAM VOWEL SIGN VOCALIC R		
44	<i>MALAYALAM VOWEL SIGN VOCALIC RR</i>		
45	(This position shall not be used)		
46	MALAYALAM VOWEL SIGN E		
47	MALAYALAM VOWEL SIGN EE		
48	MALAYALAM VOWEL SIGN AI		
49	(This position shall not be used)		
4A	MALAYALAM VOWEL SIGN O		
4B	MALAYALAM VOWEL SIGN OO		
4C	MALAYALAM VOWEL SIGN AU		
4D	MALAYALAM SIGN VIRAMA		
4E	(This position shall not be used)		
4F	(This position shall not be used)		
50	(This position shall not be used)		
51	(This position shall not be used)		
52	(This position shall not be used)		
53	(This position shall not be used)		
54	(This position shall not be used)		
55	(This position shall not be used)		
56	(This position shall not be used)		
57	MALAYALAM AU LENGTH MARK		
58	(This position shall not be used)		

Group 00

Plane 00

Row 0D

A. Administrative

1. Title

Proposal to encode four characters for Oriya and Malayalam in the BMP of the UCS

2. Requester's name

Michael Everson and Peter Scharf

3. Requester type (Member body/Liaison/Individual contribution)

Individual contribution.

4. Submission date

2007-05-25

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

Yes.

6b. More information will be provided later

No.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

No.

1b. Proposed name of script

1c. The proposal is for addition of character(s) to an existing block

Yes

1d. Name of the existing block

Oriya, Malayalam.

2. Number of characters in proposal

4 (3, 1).

3. Proposed category (A-Contemporary; B.1-Specialized (small collection); B.2-Specialized (large collection); C-Major extinct; D-Attested extinct; E-Minor extinct; F-Archaic Hieroglyphic or Ideographic; G-Obscure or questionable usage symbols)

Category B.1.

4a. Is a repertoire including character names provided?

Yes.

4b. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?

Yes.

4c. Are the character shapes attached in a legible form suitable for review?

Yes.

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson.

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, Fontographer.

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

7. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

Yes.

8. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database <http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

See above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

Yes, some of the characters have been proposed to the UTC by the Indian National Body.

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

Yes.

2b. If YES, with whom?

Michel Angot, R. Chandrashekar, Malcolm Hyman, Susan Rosenfield, B. V. Venkatakrishna Sastry, Michael Witzel

2c. If YES, available relevant documents

Specialists in Vedic Sanskrit

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

Indologists, Indo-Europeanists, teachers, students.

4a. The context of use for the proposed characters (type of use; common or rare)

Rare; these are mostly used in Sanskrit dictionaries.

4b. Reference

5a. Are the proposed characters in current use by the user community?

Yes.

5b. If YES, where?

Scholarly publications.

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

Yes.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

Accordance with the Roadmap. Keep with other Indic characters.

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

No.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

No.

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

Yes.

11b. If YES, is a rationale for such use provided?

No.

11c. If YES, reference

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

11e. If YES, reference

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

No.

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?