

**Title:** Preliminary Proposal to Encode the Modi Script in ISO/IEC 10646  
**Source:** Script Encoding Initiative (SEI)  
**Author:** Anshuman Pandey (pandey@umich.edu)  
**Status:** Liaison Contribution  
**Action:** For consideration by UTC  
**Date:** 2009-09-15

## 1 Introduction

This is a preliminary proposal to encode the Modi Script in the Universal Character Set (ISO/IEC 10646). The Modi script is a Brahmi-based left-to-right script that was used mainly for writing the Marathi language. Modi was developed during Maratha rule in the 17th century for administrative purposes. It was the official script of the Maratha dynasties and was used in the Indian state of Maharashtra until the middle 20th century. Thousands of Modi documents are preserved in India and hundreds in Denmark and France, owing to the Danish and French presence in Tanjore, Pondicherry, and other regions in South Asia through the 19th century.

Modi is allocated in the Supplemental Multilingual Roadmap (SMP) at the range U+11780..117DF. A total of 78 characters is required to encode a basic character set for Modi. These characters, their names, and properties are shown below.

Although based on the same model as Devanagari, Modi differs considerably from Devanagari in terms of letterforms and orthography. A comparison of the two scripts is shown in Figure 1 and 2. Apart from letterforms, the aspect that most differentiates Modi from Devanagari is the representation of consonant-vowel combinations, which are written as ligatures.

The majority of Modi documents are manuscripts, books were printed in Modi, such as William Carey's *A dictionary of the Mahratta language* (1810), excerpts of which are shown in Figure 4. There is an active scholarly community engaged in research on Modi documents.

## 2 Character Set

```
11780;MODI SIGN ANUSVARA;Mn;0;NSM;;;;N;;;;;
11781;MODI SIGN VISARGA;Mc;0;L;;;;N;;;;;
11782;MODI LETTER A;Lo;0;L;;;;N;;;;;
11783;MODI LETTER AA;Lo;0;L;;;;N;;;;;
11784;MODI LETTER I;Lo;0;L;;;;N;;;;;
11785;MODI LETTER II;Lo;0;L;;;;N;;;;;
11786;MODI LETTER U;Lo;0;L;;;;N;;;;;
11787;MODI LETTER UU;Lo;0;L;;;;N;;;;;
11788;MODI LETTER VOCALIC R;Lo;0;L;;;;N;;;;;
11789;MODI LETTER VOCALIC RR;Lo;0;L;;;;N;;;;;
1178A;MODI LETTER VOCALIC L;Lo;0;L;;;;N;;;;;
1178B;MODI LETTER VOCALIC LL;Lo;0;L;;;;N;;;;;
1178C;MODI LETTER E;Lo;0;L;;;;N;;;;;
1178D;MODI LETTER AI;Lo;0;L;;;;N;;;;;
1178E;MODI LETTER O;Lo;0;L;;;;N;;;;;
1178F;MODI LETTER AU;Lo;0;L;;;;N;;;;;
11790;MODI LETTER KA;Lo;0;L;;;;N;;;;;
```

11791;MODI LETTER KHA;Lo;0;L;;;;;N;;;;;  
 11792;MODI LETTER GA;Lo;0;L;;;;;N;;;;;  
 11792;MODI LETTER GHA;Lo;0;L;;;;;N;;;;;  
 11794;MODI LETTER NGA;Lo;0;L;;;;;N;;;;;  
 11795;MODI LETTER CA;Lo;0;L;;;;;N;;;;;  
 11796;MODI LETTER CHA;Lo;0;L;;;;;N;;;;;  
 11797;MODI LETTER JA;Lo;0;L;;;;;N;;;;;  
 11798;MODI LETTER JHA;Lo;0;L;;;;;N;;;;;  
 11799;MODI LETTER NYA;Lo;0;L;;;;;N;;;;;  
 1179A;MODI LETTER TTA;Lo;0;L;;;;;N;;;;;  
 1179B;MODI LETTER TTHA;Lo;0;L;;;;;N;;;;;  
 1179C;MODI LETTER DDA;Lo;0;L;;;;;N;;;;;  
 1179D;MODI LETTER DDHA;Lo;0;L;;;;;N;;;;;  
 1179E;MODI LETTER NNA;Lo;0;L;;;;;N;;;;;  
 1179F;MODI LETTER TA;Lo;0;L;;;;;N;;;;;  
 117A0;MODI LETTER THA;Lo;0;L;;;;;N;;;;;  
 117A1;MODI LETTER DA;Lo;0;L;;;;;N;;;;;  
 117A2;MODI LETTER DHA;Lo;0;L;;;;;N;;;;;  
 117A3;MODI LETTER NA;Lo;0;L;;;;;N;;;;;  
 117A4;MODI LETTER PA;Lo;0;L;;;;;N;;;;;  
 117A5;MODI LETTER PHA;Lo;0;L;;;;;N;;;;;  
 117A6;MODI LETTER BA;Lo;0;L;;;;;N;;;;;  
 117A7;MODI LETTER BHA;Lo;0;L;;;;;N;;;;;  
 117A8;MODI LETTER MA;Lo;0;L;;;;;N;;;;;  
 117A9;MODI LETTER YA;Lo;0;L;;;;;N;;;;;  
 117AA;MODI LETTER RA;Lo;0;L;;;;;N;;;;;  
 117AB;MODI LETTER LA;Lo;0;L;;;;;N;;;;;  
 117AC;MODI LETTER LLA;Lo;0;L;;;;;N;;;;;  
 117AD;MODI LETTER VA;Lo;0;L;;;;;N;;;;;  
 117AE;MODI LETTER SHA;Lo;0;L;;;;;N;;;;;  
 117AF;MODI LETTER SSA;Lo;0;L;;;;;N;;;;;  
 117B0;MODI LETTER SA;Lo;0;L;;;;;N;;;;;  
 117B1;MODI LETTER HA;Lo;0;L;;;;;N;;;;;  
 117B2;MODI VOWEL SIGN AA;Mc;0;L;;;;;N;;;;;  
 117B3;MODI VOWEL SIGN I;Mc;0;L;;;;;N;;;;;  
 117B4;MODI VOWEL SIGN II;Mn;0;NSM;;;;;N;;;;;  
 117B5;MODI VOWEL SIGN U;Mn;0;NSM;;;;;N;;;;;  
 117B6;MODI VOWEL SIGN UU;Mn;0;NSM;;;;;N;;;;;  
 117B7;MODI SIGN VOCALIC R;Mn;0;NSM;;;;;N;;;;;  
 117B8;MODI SIGN VOCALIC RR;Mn;0;NSM;;;;;N;;;;;  
 117B9;MODI SIGN VOCALIC L;Mn;0;NSM;;;;;N;;;;;  
 117BA;MODI SIGN VOCALIC LL;Mn;0;NSM;;;;;N;;;;;  
 117BB;MODI VOWEL SIGN E;Mn;0;NSM;;;;;N;;;;;  
 117BC;MODI VOWEL SIGN AI;Mn;0;NSM;;;;;N;;;;;  
 117BD;MODI VOWEL SIGN O;Mc;0;L;;;;;N;;;;;  
 117BE;MODI VOWEL SIGN AU;Mc;0;L;;;;;N;;;;;  
 117BF;MODI SIGN VIRAMA;Mn;9;L;;;;;N;;;;;  
 117C0;MODI DIGIT ZERO;Nd;0;L;0;0;0;N;;;;;  
 117C1;MODI DIGIT ONE;Nd;0;L;1;1;1;N;;;;;  
 117C2;MODI DIGIT TWO;Nd;0;L;2;2;2;N;;;;;  
 117C3;MODI DIGIT THREE;Nd;0;L;3;3;3;N;;;;;  
 117C4;MODI DIGIT FOUR;Nd;0;L;4;4;4;N;;;;;  
 117C5;MODI DIGIT FIVE;Nd;0;L;5;5;5;N;;;;;  
 117C6;MODI DIGIT SIX;Nd;0;L;6;6;6;N;;;;;  
 117C7;MODI DIGIT SEVEN;Nd;0;L;7;7;7;N;;;;;  
 117C8;MODI DIGIT EIGHT;Nd;0;L;8;8;8;N;;;;;  
 117C9;MODI DIGIT NINE;Nd;0;L;9;9;9;N;;;;;  
 117CA;MODI OM;Lo;0;L;;;;;N;;;;;  
 117CB;MODI HU;Lo;0;L;;;;;N;;;;;  
 117CC;MODI DANDA;Po;0;L;;;;;N;;;;;  
 117CD;MODI DOUBLE DANDA;Po;0;L;;;;;N;;;;;

### 3 References

- Carey, William. 1810. *A dictionary of the Mahratta language*. Serampore: Serampore Missionary Press.
- Dīkṣit, Moreśvar G. and V. G. Khobarekar. 1969. मराठेकालीन प्रसिद्ध व्यक्तींची हस्ताक्षरयुक्त पत्र [*Marāṭhekālīna prasiddha vyaktīmci hastākṣarayukta patre*]. पुराभिलेख विभाग — ऐतिहासिक प्रकाशन क्र० १. मुंबई: शासकीय मध्यवर्ती मुद्रणालय
- Grierson, George A. 1905. *The Linguistic Survey of India*. Vol. VII. Indo-Aryan Family. Southern Group. Specimens of the Marathi language. Calcutta: Office of the Superintendent of Government Printing, India.
- Navalkar, Ganpatrao R. 1925. *The Student's Marāṭhī Grammar*. 4th ed. Poona: Scottish Mission Press.
- Strandberg, Elisabeth [ed]. 1983. *The Moḍī Documents from Tanjore in Danish Collections*. Beiträge zur Südasienforschung, Südasien-Institut Universität Heidelberg, Band 81. Wiesbaden: Franz Steiner Verlag.

### 4 Acknowledgments

This project was made possible in part by a grant from the United States National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at the University of California, Berkeley). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.

	1178	1179	117A	117B	117C
0	ं	म	घ	ड	०
1	ः	च	ट	ठ	१
2	इ	ग	च	ा	२
3	ध	घ	च	ी	३
4	ई	उ	घ	ी	४
5	ई	उ	म	ु	५
6	ठी	छ	घ	ु	६
7	ठी	च	म	ु	७
8	ऋ	ऌ	म	ु	८
9	ऋ	ज	ठ	ु	९
A	ल	ह	उ	ु	१०
B	ल	ह	च	े	११
C	अ	उ	ळ	ॅ	१२
D	अ	ढ	प	ो	१३
E	ध	ण	श	ो	
F	ध	त	ष	ो	

क	ख	ग	घ	ङ	च	छ	ज	झ	ञ	ट	ठ	ड	ढ	ण	त	थ	द
का	खा	गा	घा	ङा	चा	छा	जा	झा	जा	टा	ठा	डा	ढा	णा	ता	था	दा
की	खी	गी	घी	ङी	ची	छी	जी	झी	जी	टी	ठी	डी	ढी	णी	ती	थी	दी
कु	खु	गु	घु	ङु	चु	छु	जु	झु	जु	टु	ठु	डु	ढु	णु	तु	थु	दु
कै	खै	गै	घै	ङै	चै	छै	जै	झै	जै	टै	ठै	डै	ढै	णै	तै	थै	दै
कौ	खौ	गौ	घौ	ङौ	चौ	छौ	जौ	झौ	जौ	टौ	ठौ	डौ	ढौ	णौ	तौ	थौ	दौ
कं	खं	गं	घं	ङं	चं	छं	जं	झं	जं	टं	ठं	डं	ढं	णं	तं	थं	दं
कः	खः	गः	घः	ङः	चः	छः	जः	झः	जः	टः	ठः	डः	ढः	णः	तः	थः	दः

क	ख	ग	घ	ङ	च	छ	ज	झ	ञ	ट	ठ	ड	ढ	ण	त	थ	द
का	खा	गा	घा	ङा	चा	छा	जा	झा	जा	टा	ठा	डा	ढा	णा	ता	था	दा
की	खी	गी	घी	ङी	ची	छी	जी	झी	जी	टी	ठी	डी	ढी	णी	ती	थी	दी
कु	खु	गु	घु	ङु	चु	छु	जु	झु	जु	टु	ठु	डु	ढु	णु	तु	थु	दु
कै	खै	गै	घै	ङै	चै	छै	जै	झै	जै	टै	ठै	डै	ढै	णै	तै	थै	दै
कौ	खौ	गौ	घौ	ङौ	चौ	छौ	जौ	झौ	जौ	टौ	ठौ	डौ	ढौ	णौ	तौ	थौ	दौ
कं	खं	गं	घं	ङं	चं	छं	जं	झं	जं	टं	ठं	डं	ढं	णं	तं	थं	दं
कः	खः	गः	घः	ङः	चः	छः	जः	झः	जः	टः	ठः	डः	ढः	णः	तः	थः	दः

Figure 1: Comparison of consonant-vowel combinations in Modi and Devanagari (from Steinberg 1983: 32–33).

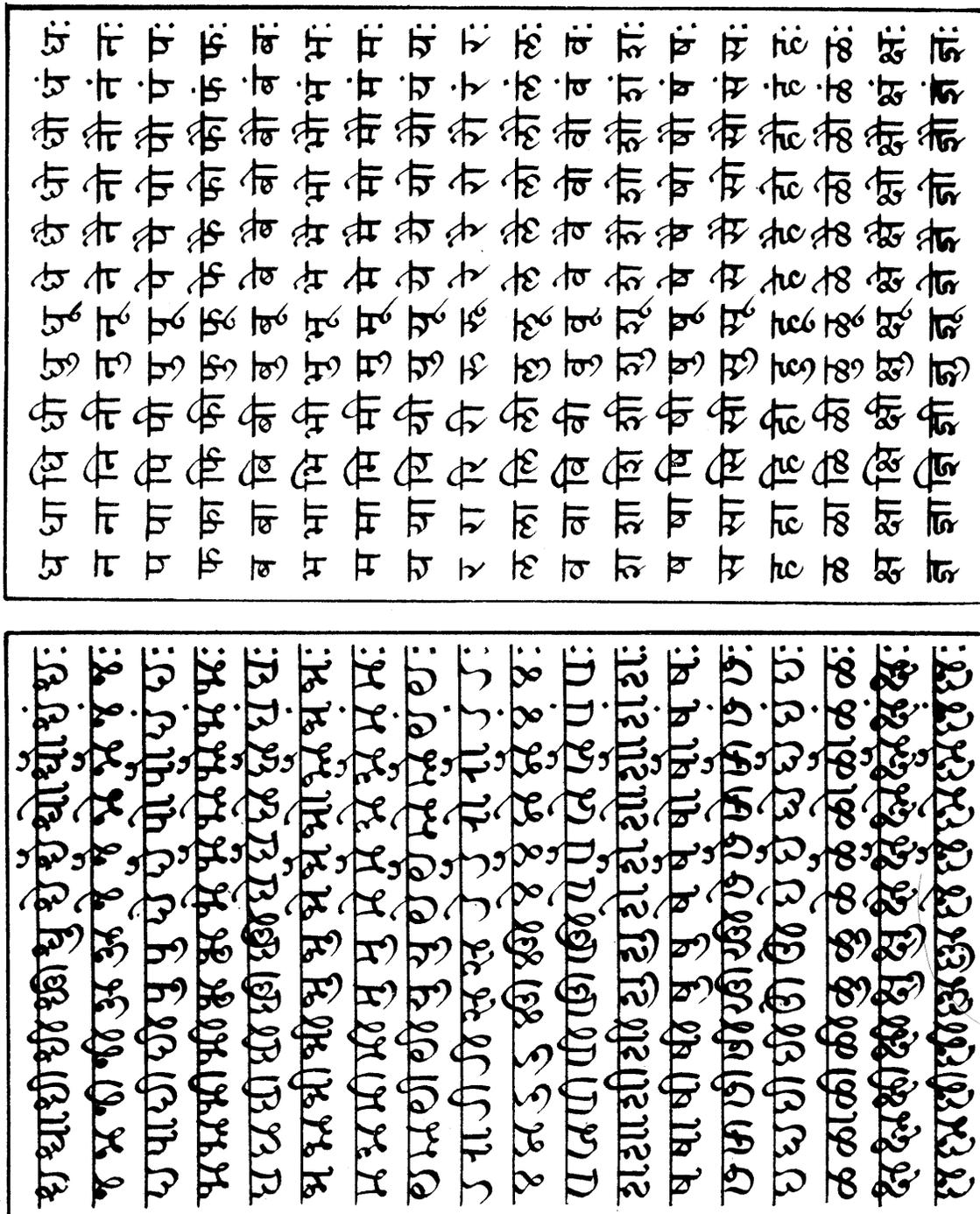


Figure 2: Comparison of consonant-vowel combinations in Modi and Devanagari (from Steinberg 1983: 34–35).



## MAHRATTA DICTIONARY.



वस.

वसणी, *a.* out of debt, not in debt.

वसक, *a.* not having a piece of the cloth which is worn round the loins tucked in behind.

वसपोमव, *a.* of an unascertained terrific form.

वसप्य, *a.* unspeakable, unfit to be spoken, obscene.

वसपनीय, *a.* unspeakable, unutterable, unfit to be spoken, obscene.

वसपल, *a.* ingenuous, frank, honest, sincere.

वसपीय, *a.* improper to be done, improper.

वसपीयेक, *a.* without bored ears, (an epithet of reproach to a Musulman.)

वसमं, *a.* useless.

वसमं, *a.* useless, unfit for any purpose, indolent.

वसमंभोर, *a.* useless, unfit for any purpose, indolent.

वसमंभोर, *a.* useless, not fit for any purpose.

वसमनणीय, *a.* unfit to be done.

## MAHRATTA DICTIONARY.



वस.

वसणी, *a.* out of debt, not in debt.

वसक, *a.* not having a piece of the cloth which is worn round the loins tucked in behind.

वसपोमव, *a.* of an unascertained terrific form.

वसप्य, *a.* unspeakable, unfit to be spoken, obscene.

वसपनीय, *a.* unspeakable, unutterable, unfit to be spoken, obscene.

वसपल, *a.* ingenuous, frank, honest, sincere.

वसपीय, *a.* improper to be done, improper.

वसपीयेक, *a.* without bored ears, (an epithet of reproach to a Musulman.)

वसमं, *a.* useless.

वसमं, *a.* useless, unfit for any purpose, indolent.

वसमंभोर, *a.* useless, unfit for any purpose, indolent.

वसमंभोर, *a.* useless, not fit for any purpose.

वसमनणीय, *a.* unfit to be done.



Marāṭhī is usually written in the so-called Bālbōdh or in the so-called Mōḍi character.<sup>1</sup>

Written character.

Bālbōdh, *lit.* 'teachable to children' is identical with Dēva-nāgari, and has been described in Vol. v. Part ii, pp. 7 and ff.

It is used in almost all printed books, and also, to a great extent, in private transactions and letters. The Mōḍi character is almost totally confined to the latter kind of writing. An example of its use will be found on pp. 259 and ff. It consists of the following signs :—

VOWELS.

अ a      ऎ ā      ई i, ī,      ए u, ū,  
 ऐ ē      औ ai      ॐ ō      ॐ au      ॐ am      यः ak

CONSONANTS.

क ka      ख kha      ग ga      घ gha      ङ ṅa  
 च cha      छ chha      ज ja      झ jha      ञ ṅa  
 ट ṭa      ठ ṭha      ड ḍa      ढ ḍha      ण ṇa  
 त ta      थ tha      द da      ध dha      न na  
 प pa      फ pha      ब ba      भ bha      म ma  
 य ya      र ra      ल la      व va  
 श śa      ष sha      स sa      ह ha  
 ङ ṅa      क्ष ksha      ण ṇya

The forms of the vowels given above are only used at the beginning of words or syllables. When the vowels follow a consonant they are expressed by means of secondary signs in the same way as in the Dēvanāgari alphabet. For the sake of teaching these signs the alphabet is disposed in *Bārākhḍīs*, or series of twelve letters, each containing a consonant combined with all possible vocalic sounds. Such Bārākhḍīs are :—

क-क-की-की-कु-कु-के-कै-को-कौ-काम-कः  
 ka kḅ ki kī ku kḅ kē kai kō kau kam kaḅ  
 ग-गा-गी-गी-गु-गु-गे-गै-गो-गौ-गं-गः  
 ga gā gi gī gu gḅ gē gai go gau gam gaḅ  
 ट-ट-ठी-ठी-टु-टु-टे-टै-टो-टौ-टं-टः  
 ṭa ṭā ṭi ṭī ṭu ṭḅ ṭē ṭai ṭō ṭau ṭam ṭaḅ

A short note on the Modi Character by B. A. Gupte will be found in the *Indian Antiquary*, Vol. xxxiv, 1905, pp. 27 and ff.

Figure 6: Chart of Modi from the *Linguistic Survey of India* (from Grierson 1905: 20).

ISO/IEC JTC 1/SC 2/WG 2  
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS  
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646<sup>1</sup>

Please fill all the sections A, B and C below. Please read Principles and Procedures Document (P & P) from  
<http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.  
Please ensure you are using the latest Form from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.  
See also <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

---

### A. Administrative

1. Title: **Preliminary Proposal to Encode the Modi Script in ISO/IEC 10646**
2. Requester's name: **University of California, Berkeley Script Encoding Initiative (Universal Scripts Project); author: Anshuman Pandey (pandey@umich.edu)**
3. Requester type (Member Body/Liaison/Individual contribution): **Liaison contribution**
4. Submission date: **2009-09-15**
5. Requester's reference (if applicable): **N/A**
6. Choose one of the following:
  - (a) This is a complete proposal: **No**
  - (b) or, More information will be provided later: **Yes**

### B. Technical - General

1. Choose one of the following:
  - (a) This proposal is for a new script (set of characters): **Yes**
    - i. Proposed name of script: **Modi**
  - (b) The proposal is for addition of character(s) to an existing block: **No**
    - i. Name of the existing block: **N/A**
2. Number of characters in proposal: **78**
3. Proposed category: **C - Major extinct**
4. Is a repertoire including character names provided?: **Yes**
  - (a) If Yes, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?: **Yes**
  - (b) Are the character shapes attached in a legible form suitable for review?: **Yes**
5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?: **Anshuman Pandey; True Type format**
  - (a) If available now, identify source(s) for the font and indicate the tools used: **N/A**
6. References:
  - (a) Are references (to other character sets, dictionaries, descriptive texts etc.) provided?: **Yes**
  - (b) Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?: **Yes**
7. Special encoding issues:
  - (a) 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; see proposal for additional details..**
8. Additional Information: 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 <http://www.unicode.org/Public/UNIDATA/UCD.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard. **Character properties and numeric information are included.**

---

<sup>1</sup> Form number: N3102-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03)

### C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?: **No**
2. 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**
  - (a) If Yes, with whom?: **Forthcoming**
    - i. If Yes, available relevant documents: **N/A**
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? **Yes**
  - (a) Reference: **Linguists and historians.**
4. The context of use for the proposed characters (type of use; common or rare): **Common**
  - (a) Reference: **Administrative documents and records from the Maratha dynasty and European countries with former presence in South Asia.**
5. Are the proposed characters in current use by the user community?: **There is an active scholarly community engaged in the study of Modi documents.**
  - (a) If Yes, where? Reference: **India and Europe.**
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?: **No**
  - (a) If Yes, is a rationale provided?: **N/A**
    - i. If Yes, reference: **N/A**
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? **Yes**
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?  
**No**
  - (a) If Yes, is a rationale for its inclusion provided?: **N/A**
    - i. If Yes, reference: **N/A**
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? **No**
  - (a) If Yes, is a rationale provided?: **N/A**
    - i. If Yes, reference: **N/A**
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?  
**Yes**
  - (a) If Yes, is a rationale for its inclusion provided? **Yes**
    - i. If Yes, reference: **See text of proposal**
11. Does the proposal include use of combining characters and/or use of composite sequences? **Yes**
  - (a) If Yes, is a rationale for such use provided? **Yes**
    - i. If Yes, reference: **See text of proposal**
  - (b) Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? **Yes**
    - i. If Yes, reference: **See text of proposal**
12. Does the proposal contain characters with any special properties such as control function or similar semantics? **Yes**
  - (a) If Yes, describe in detail (include attachment if necessary): **Virama**
13. Does the proposal contain any Ideographic compatibility character(s)? **No**
  - (a) If Yes, is the equivalent corresponding unified ideographic character(s) identified? **N/A**
    - i. If Yes, reference: **N/A**