

# Proposal to encode the Devanagari letter and vowel sign AY

Anshuman Pandey  
 Department of Linguistics  
 University of California, Berkeley  
 Berkeley, California, U.S.A.  
 anshuman.pandey@berkeley.edu

December 7, 2015

## 1 Introduction

This is a proposal to encode two new Devanagari characters in Unicode:

Glyph	Code	Character name
ॠ	A8FE	DEVANAGARI LETTER AY
ँ	A8FF	DEVANAGARI VOWEL SIGN AY

These characters are part of a set devised in the 19th century for representing short vowel sounds of languages of north India for which native characters did not exist in Devanagari. The set consists of the letters ऐ, ऐ, औ, औ and their dependent forms े, ै, ो, ौ. They were described by their creator, the linguist A. F. Rudolf Hoernle, as follows (1880: 3):

The E[astern] H[indi] possesses fifteen vowels; a *neutral* and fourteen distinct ones. The latter consists of seven pairs, each containing a short and a long one. [...] the short ज़, ड़, ज़, ड़, are according to the usual view, unknown to the Sanskrit phonetic system, and therefore have no place in the native grammatical scheme of sounds and characters. But in order to avoid the inconvenience of two different sounds being denoted by the same sign, I have ventured to introduce into the E. H. alphabet ... new characters. For the short ज़, ड़, ज़, ड़ I shall adopt the Gurmukhī or Bangālī forms of the ordinary Nāgarī signs, which differ from the latter merely in having a serpentine form (ँ and ञ) instead of a slightly curved one (ँ and ञ).

Hoernle’s new characters were adopted by other linguists, such as George Grierson and Samuel Kellogg, who used them in their printed works, although with some modification. Also, it appears that the characters were used by Indian scholars after they were introduced. At the turn of the 20th century Grierson wrote that the characters “have been adopted by some of the best Benares Paṇḍits” and that “[a]ccurate writers distinguish these when writing in the Dēva-nāgarī character” (1903: 7, 21).

This set of Devanagari short-vowel characters is almost fully supported in Unicode, with three of the four pairs already encoded.<sup>1</sup> The first two pairs were included in Unicode 1.0, derived from ISCII 1998:

ऐ	U+090E DEVANAGARI LETTER SHORT E
े	U+0946 DEVANAGARI VOWEL SIGN SHORT E
औ	U+0912 DEVANAGARI LETTER SHORT O
ो	U+094A DEVANAGARI VOWEL SIGN SHORT O

Another pair was encoded in Unicode 6.0, but with different nomenclature:

औ	U+0975 DEVANAGARI LETTER AW
ौ	U+094F DEVANAGARI VOWEL SIGN AW

It is practical that the remaining pair ऐ and े be encoded in order to support the representation of texts that contain all of Hoernle’s short vowel characters.

## 2 Proposal History

The proposed characters and a related pair were proposed for encoding in L2/09-320 in September 2009 with the intention to supplement the existing characters ऐ / े and औ / ो:

ऐ	DEVANAGARI LETTER SHORT AI
े	DEVANAGARI VOWEL SIGN SHORT AI
औ	DEVANAGARI LETTER SHORT AU
ौ	DEVANAGARI VOWEL SIGN SHORT AU

In October 2009, the Government of India submitted a proposal to encode six characters used in a Devanagari orthography for Kashmiri (see N3710 L2/09-369). Coincidentally, औ and ौ were among these, but proposed as “KASHMIRI DEVANAGARI VOWEL BACK LOW ROUNDED O” and “KASHMIRI DEVANAGARI VOWEL SIGN BACK LOW ROUNDED O”. As L2/09-320 and L2/09-369 contained some common characters, it was requested that the UTC review both proposals in tandem (see L2/09-377). At that time, a ‘Kashmiri Ad-hoc group’ was convened during WG2 meeting #55 in Tokyo, which suggested renaming the characters in L2/09-369 (see N3727 L2/09-389). As a result, औ was designated as “DEVANAGARI LETTER AW” and ौ as “DEVANAGARI VOWEL SIGN AW”. These did not align with names assigned to graphically related characters existing in Unicode, so it was requested that the UTC instead adopt the names proposed in L2/09-320 (see L2/09-403). These requests were not considered. The UTC reviewed L2/09-320, L2/09-369, and L2/09-389 during meeting #121 in November 2009. It did not act upon L2/09-320 and requested “more input from Indic experts” (see L2/09-335R). However, it approved L2/09-369 and accepted औ and ौ as part of a repertoire of ‘Kashmiri additions to Devanagari’ with names as specified in L2/09-389. The characters औ and ौ were encoded, although within a different paradigm, but the ऐ and े were forgotten in the process.

<sup>1</sup> The अे U+090E DEVANAGARI LETTER SHORT A is another character that presumably represents a short vowel. It was encoded in Unicode 4.0. Its provenance and usage is unknown (see L2/09-321 for more information).

### 3 Current Status

At present, there is no means for representing ऐ and ै in plain text. Some reviewers of L2/09-320 suggested that these forms could be produced using combinations of existing Devanagari characters:

- ै    े U+0946 DEVANAGARI VOWEL SIGN SHORT E,  
     े U+0946 DEVANAGARI VOWEL SIGN SHORT E
- ऐ    ऐ U+090E DEVANAGARI SHORT E,  
     े U+0946 DEVANAGARI VOWEL SIGN SHORT E
- ऐ    ए U+090E DEVANAGARI LETTER E,  
     े U+0946 DEVANAGARI VOWEL SIGN SHORT E,  
     े U+0946 DEVANAGARI VOWEL SIGN SHORT E

These approaches do not produce acceptable results:

- |           |       |  |
|-----------|-------|--|
| <क, े, े> | के    | the second vowel sign is laid over the first and becomes invisible, making the output indistinct from के <क, े>.                                 |
|           | केे   | the second sign is not processed.  |
| <ऐ, े>    | ऐ     | the े is laid over the ̣ of the base glyph and becomes invisible.  |
|           | ऐे    | the े combines with the base at the top right edge, but clashes with the ̣ of the base glyph.  |
|           | ऐेे   | the े is unattached because ऐे is not a valid base for vowel signs.  |
| <ए, े, े> | ऐे    | both vowel signs are overlaid, making the output indistinct from ऐे U+090E DEVANAGARI SHORT E.   |
|           | ऐेे   | <ए, े> is replaced with ऐे U+090E DEVANAGARI SHORT E and the second े is added to the base, but it clashes with the element ̣ of the base glyph. |
|           | ऐेेे  | <ए, े> is replaced with ऐे U+090E DEVANAGARI SHORT E and the second े is left unrendered   |
|           | ऐेेेे | vowel signs are unattached because ए is not a valid base for vowel signs.  |

These sequences involve multiple vowel signs attached to a single base consonant and multiple vowel signs applied to an independent vowel. Such combinations are not recognized as valid by Devanagari shaping engines and are not provided for in fonts.

When producing text that contains short vowel characters, it is unacceptable to see ऐ displayed illegibly as ऐे, ऐेे, ऐेेे, ऐेेेे especially when it occurs alongside the visually coherent ऐे, ओ, औ.

## 4 Recommendation

As existing Devanagari characters cannot be combined to produce ऐँ and ैँ, these forms need to be encoded as separate characters in adherence to the specifications of current shaping engines and *The Unicode Standard* itself. In section 12.1 “Devanagari” (p. 443), the standard states:

Vowel letters are encoded atomically in Unicode, even if they can be analyzed visually as consisting of multiple parts. *Table 12-1* [reproduced here in figure 10] shows the letters that can be analyzed, the single code point that should be used to represent them in text, and the sequence of code points resulting from analysis that should not be used.

Based upon the above specification, the proper approach is to encode the vowel letter ऐँ and its dependent vowel sign ैँ as atomic characters, using the names and code points specified in the introduction.

**Character names** The proposed characters ऐँ and ैँ were named “DEVANAGARI LETTER SHORT AI” and “DEVANAGARI VOWEL SIGN SHORT AI” in L2/09-320. These names aligned with those assigned to the existing DEVANAGARI LETTER SHORT E, DEVANAGARI VOWEL SIGN SHORT E, DEVANAGARI LETTER SHORT O, and DEVANAGARI VOWEL SIGN SHORT O. The decision to encode औँ and ौँ as U+0975 DEVANAGARI LETTER AW and U+094F DEVANAGARI VOWEL SIGN AW, respectively, broke the convention. In L2/10-471, Shriramana Sharma suggests using the descriptor “AY” instead of “SHORT AI” to align the characters with DEVANAGARI LETTER AW. The suggestion is reasonable because the characters are graphically related.

**Allocation** As the ‘Devanagari’ block is full, the proposed characters should be encoded in the ‘Devanagari Extended’ (U+8A30) block, at the two remaining code points at U+A8FE and U+A8FF.

**Representative glyphs** The glyphs for ऐँ and ैँ are based upon Hoernle’s original forms (see figures 1, 2).

**Alternate representations** While the sign ैँ is used consistently by various authors, the letter ऐँ is not:

	ऐँ	ैँ	figures
Hoernle	ऐँ	ैँ	figures 1, 2
Grierson	ऐँ	ैँ	figures 3, 4, 6
Kellogg	ऐँ	ैँ	figure 5

These alternate letters may be handled using existing characters and as glyphic variants of the following:

Regular <i>ai</i>	Variant <i>ai</i>	Representation
ऐँ	ऐँ	DEVANAGARI LETTER SHORT E
ऐँ	ऐँ	glyphic variant of DEVANAGARI LETTER SHORT E
ऐँ	ऐँ	glyphic variant of proposed DEVANAGARI LETTER AY

The ability to represent ऐँ as a variant of ऐँ requires that DEVANAGARI LETTER AY be encoded as an atomic character.

## 5 Character Data

**Character properties** In the format of UnicodeData.txt:

```
A8FE;DEVANAGARI LETTER AY;Lo;0;L;;;;N;;;;;
A8FF;DEVANAGARI VOWEL SIGN AY;Mn;0;NSM;;;;N;;;;;
```

**Linebreaking** In the format of LineBreak.txt:

```
A8FE;AL # Lo DEVANAGARI LETTER AY
A8FF;CM # Mn DEVANAGARI VOWEL SIGN AY
```

**Syllabic categories** In the format of IndicSyllabicCategory.txt:

```
# Indic_Syllabic_Category=Vowel_Dependent
A8FF ; Vowel_Dependent # Mn DEVANAGARI VOWEL SIGN AY

# Indic_Syllabic_Category=Vowel_Independent
A8FE ; Vowel_Independent # Lo DEVANAGARI LETTER AY
```

**Positional categories** In the format of IndicPositionalCategory.txt:

```
# Indic_Matra_Category=Top
A8FF ; Top # Mn DEVANAGARI VOWEL SIGN AY
```

## 6 References

- Everson, Michael. 2009. “Consensus on Kashmiri additions for Devanagari”. N3727 L2/09-370. October 28, 2009. <http://www.unicode.org/L2/L2009/09389-n3731-kashmiri.pdf>
- Government of India. 2009. “Proposal to add Six characters in the Devanagari block for representation of Kashmiri language in Devanagari script”. N3710 L2/09-369. October 26, 2009. <http://www.unicode.org/L2/L2009/09369-n3710.pdf>
- Grierson, George A. *Seven Grammars of the dialects and subdialects of the Bihārī Language: Spoken in the Province of Bihār, in the eastern portion of the North-Western Provinces, and in the northern portion of the Central Provinces*. Calcutta: Bengal Secretariat Press.
- . 1903. *The Linguistic Survey of India*. Vol. V. Indo-Aryan Family. Eastern Group. Part II. Specimens of the Bihārī and Oṛiyā languages. Calcutta: Superintendent of Government Printing, India.
- Hoernle, A. F. Rudolf. 1880. *A Comparative Grammar of the Gauḍian (Aryo-Indian) Languages*. With special reference to the Eastern Hindī, preceded by a general introduction on the North-Indian vernaculars, and accompanied by a language-map, a comparative table of alphabets, and an index of subjects. London: Trübner & Co.
- Kellogg, Samuel H. 1893. *A Grammar of the Hindī Language*. In which are treated the High Hindī, Braj, and the Eastern Hindī of the Rámāyan of Tulsi Dās, also the colloquial dialects of Rájputáná, Kumáon, Avadh, Ríwá, Bhojpúr, Magadha, Maithila, etc., with copious philological notes. 2nd ed., rev. and enl. London: K. Paul, Trench, Trubner and Co.
- Pandey, Anshuman. 2009a. “Proposal to Encode Additional Short Vowel Characters for Devanagari”. L2/09-320. September 18, 2009. <http://www.unicode.org/L2/L2009/09320-devshortvowels.pdf>

- . 2009b. “Proposal to Deprecate DEVANAGARI LETTER SHORT A”. L2/09-321. September 22, 2009. <http://www.unicode.org/L2/L2009/09321-shortadeprec.pdf>
- . 2009c. “Comments on India’s Proposal to Add Devanagari Characters for Kashmiri (N3710 L2/09-369)”. N3725 L2/09-377. October 26, 2009. <http://www.unicode.org/L2/L2009/09377-kashmiri-add-cmt.pdf>
- . 2009d. “Comments on ‘Consensus on Kashmiri additions for Devanagari’ (N3727 L2/09-389)”. L2/09-403. October 30, 2009. <http://www.unicode.org/L2/L2009/09403-kashmiri-consensus-cmt.pdf>
- Sharma, Shriramana. 2010. “Comments on Devanagari short vowels”. L2/10-471. November 27, 2010. <http://www.unicode.org/L2/L2010/10471-dev-short-vowels.pdf>
- Unicode Consortium. 2009. “Preliminary Minutes of the UTC 121 / L2 218 Joint Meeting”. L2/09-335R. November 10, 2009. <http://www.unicode.org/L2/L2009/09335.htm>

## **7 Acknowledgments**

I thank John Hudson and Andrew Glass for their discussions about Devanagari shaping engines and the most suitable representations of the proposed characters. I hope this proposal will finally succeed in providing support in Unicode for the one remaining pair of Hoernle’s full set of short-vowel characters.

This project was made possible in part through a Google Research Award, granted to Deborah Anderson for the Script Encoding Initiative, and a grant from the United States National Endowment for the Humanities (PR-50205-15), which funds 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 Google or the National Endowment for the Humanities.

## 1. VOWELS.

3. The E. H. possesses fifteen vowels; a *neutral* and fourteen distinct ones. The latter consist of seven pairs, each containing a short and a long one. They are ँ; ऌ, ऍ; ऎ, ऑ; ऒ, ओ; औ, क; ख, ग; घ, ङ; च, छ; छ, ञ; झ, ञ. Five of these, the neutral vowel and the short ऎ, ओ, ऑ, ञ are, according to the usual view, unknown to the Sanskrit phonetic system, and therefore have no place in the native grammatical scheme of sounds and characters. But in order to avoid the inconvenience of two different sounds being denoted by the same sign, I have ventured to introduce into the E. H. alphabet, used in this treatise, five new characters. For the short ऎ, ओ, ऑ, ञ I shall adopt the Gurmukhí or Bangálí forms of the ordinary Nágari signs, which differ from the latter merely in having a serpentine form ( ` and `) instead of a slightly curved one ( ^ and ^). For the neutral vowel I shall adopt a dot (·) placed after the consonant in the same manner as a stroke (l) is placed after it to denote the long á; in transliterating I shall use the *apostrophe*. Accordingly the signs of the fifteen E. H. vowels are as follows:

*Initial:* — ँ ऌ ऍ ऎ ऑ उ ऒ उ औ क ख ग घ ङ च छ छ ञ झ ञ  
*Noninit.:* = — ँ ऌ ऍ ऎ ऑ उ ऒ उ औ क ख ग घ ङ च छ छ ञ झ ञ

*Note:* The neutral vowel requires no initial form, as it never occurs in the beginning. The short ऌ has no non-initial form, as it is inherent in the consonant, which could not be pronounced without it. When it is necessary to indicate the mere consonant, an oblique stroke, called the *viráma* or *stoppage*, is appended to the consonantal sign; thus क कँ, but क् k. The manner of writing the non-initial signs may be seen from the following examples; क k', क कँ, का कá, कि कí, की कí, कु कú, कू कú, के कé, के कé, को कó, को कó, कै कá, कै कá, कौ कá, कौ कá.

Figure 1: The set of Devanagari vowel letters and signs used by Hoernle (1880: 3). A description of the 'serpentine' signs used for short vowels is given here.

*Initial:* — ँ ऌ ऍ ऎ ऑ उ ऒ उ औ क ख ग घ ङ च छ छ ञ झ ञ  
*Noninit.:* = — ँ ऌ ऍ ऎ ऑ उ ऒ उ औ क ख ग घ ङ च छ छ ञ झ ञ

Figure 2: Detail of figure 1 with the proposed characters highlighted.

The forms of the vowels given above are the initials, and are used only at the beginning of a word ; when subjoined to a consonant they take the following forms :—

a (not expressed), ā ॐ ; i ॐ, ī ॐ ; u ॐ, ū ॐ ; ri ॐ ; e ॐ, ē ॐ ; **ai ॐ**, ai ॐ ; o ॐ, ō ॐ ; **āü ॐ**, au ॐ.

Thus क *ka*, खा *khā*, गि *gi*, घी *ghī*, चु *chu*, छू *chhū*, जू *jū*, भू *bhū*, डे *ḍe*, **ठे *ṭhāi***, डै *ḍai*, ढो *ḍho*, तो *tō*, थौ *thāü*, दौ *dau*.

Figure 3: Devanagari vowel signs used in the *Linguistic Survey of India* (from Grierson 1903: 7).

Dēva-nāgarī.		Transliteration.
Initial.	Non-initial.	
ॐ	ँ	<i>e</i> , as in प्रकरा <i>ek<sup>r</sup>rā</i> , तेकरा <i>tek<sup>r</sup>rā</i> .
ए	ँ	<i>ē</i> , as in एकर <i>ēkar</i> , तेकर <i>tēkar</i> .
ओ	ो	<i>o</i> , as in ओकरा <i>ok<sup>r</sup>rā</i> , होइए <i>hoiai</i> .
औ	ो	<i>ō</i> , as in ओकर <i>ōkar</i> , लोक <i>lōk</i> .
ॐ	ँ	<b><i>āi</i>, as in ऐसनहिँ <i>āisanahī</i>, देखैतिथी <i>dekhāitīau</i>.</b>
ए	ँ	<i>ai</i> , as in ऐसन <i>aisan</i> , देखैत <i>dekhait</i> .
ओ	ो	<i>āü</i> , as in औतिऐ <i>āütīai</i> , पौलहक <i>pāül<sup>h</sup>hāk</i> .
औ	ो	<i>au</i> , as in औताह <i>autāh</i> , पीताह <i>pautāh</i> .

Figure 4: A chart showing the short vowels and examples of their usage (from Grierson 1903: 22).

*d.* According to Hoernle and Grierson, the colloquial dialects east of Allahabad exhibit not only this short *ē*, *ō*, but also a short *āi* and *āü*, differing from the corresponding long sounds simply in quantity.\* To represent these four sounds, unknown to the Devanāgarī alphabet, Hoernle has introduced four characters from the Bangālī and Panjābī, in which he has been followed by Grierson. These are as follows, the initial form being in each case given first :

ॐ ँ *ē*, ॐ ँ *āi*, ओ ो *ō*, औ ो *āü*.

It will be observed that these differ from the corresponding long vowel characters merely in that the strokes above the line are doubly curved ; and that the initial forms of *ē* and *āi* face to the left instead of the right.† These characters will be used for the notation of these short vowels in the following pages.

Figure 5: The set of Devanagari short vowel letters used by Kellogg (from Kellogg 1893: 4).



'The signs ऋ, ऐ, औ, and औ have been introduced by European scholars in late years, and have been adopted by some of the best Benares Paṇḍits. These short vowels do not exist in Sanskrit, to which language the Dēva-nāgarī alphabet was originally confined, but do occur in the Modern Indian languages, and hence additional signs have had to be invented for them.

Figure 6: Description by Grierson of the short vowel signs (from Grierson 1903: 7). His reference to 'European scholars' is actually singular. His representation of the letter for ai follows Kellogg's ऐ and deviates from Hoernle's ऐ.

The following signs have been added to the Deva Nāgarī alphabet to express these sounds :—

	Initial.	Non-initial.
ॐ	ॐ	ॐ
ॐ	ॐ	ॐ
ai	ऐ	ऐ
au	औ	औ

SECTION 32. The following table therefore shows the complete series of Bihārī vowels as used in these Grammars :—

	Neutral ('), a, á, i, í, u, ú, é, e, ô, o.
Initial	wanting, अ, आ, इ, ई, उ, ऊ, ए, औ, ओ.
Non-initial	अ, आ, इ, ई, उ, ऊ, ए, औ, ओ.
	ai, ai, au, au.
Initial	ऐ, ऐ, औ, औ.
Non-initial	ऐ, ऐ, औ, औ.

The manner of writing the non-initial vowels is as follows :—

क k', क ka,† का ká, कि ki, की kī, कु ku, कू kú, के kē,  
 के ke, को kō, को ko, कै kǎi, कै kai, कौ kǎu, कौ kau.

When nasalized, the vowels are written as follows :—

क k̄, का k̄á, कि k̄i, की k̄ī, कु k̄u, कू k̄ú, के k̄ē,  
 के k̄ē, को k̄ō, को k̄o, कै k̄ǎi, कै k̄ai, कौ k̄ǎu, कौ k̄au.

Figure 7: Description of vowels in the Seven Grammars showing representations of the short vowel ai (from Grierson 1883: 21). The proposed DEVANAGARI VOWEL SIGN AY is highlighted.

पाँचम कथा ।

बान्हल काठी ओ फुजल काठी ।

एक गृहलक बेटा । सभ अपगना में लड़ाइ भगंडा करैत बल । कथा सभ सँ बडत बुझौलकै परंतु केओ किच्छु नहिँ मानलकै । तखन गृहल अपगना मन में कहलक कि, जँ एकगना सभ कँ किछु कँ देखाबौ, तँ चाही जे बुझि जाय । एक दीन ओ अपगना बेटा सभ कँ बजौलक आओर कहलक कि धोरक काठी हमगना सोभा लँ आवब । ओहि काठी सबहि कँ ओ गृहल जौड़ि सँ कसि कँ बान्हलक । फेरि प्रत्येक सँ कहलक कि जौड़िं जनु फोषह, आओर काठी सबहि कँ तोड़ि दँह । एक एक लागि पड़ल, परंतु किच्छु नहिँ भेल । फेरि गृहल फो कँ एक २ काठी देखलकै । तखन ओ सभ ओकगना भट पट तोड़ि देलक । तखन ओकगना सभक बाप कहलकैक, कि, बेटा, जौँ तौँ बान्हल काठी सबहि तरगँहँ अपगना में भेलि कँकँ रहबह, तौँ सभ बैरीक दाँत चम्मत करबह । आओर जौँ बिदुरि गेलाह, तखन जानह कि बिगड़ि गेलाह ।

Figure 8: Excerpt of a Maithili specimen from *Seven Grammars* showing the proposed DEVANAGARI VOWEL SIGN AY (from Grierson 1883: 33). This text also contains the other short vowel signs.

जेठ बेटवा कतहूँ खेतारी में रहलैँ । ऊ जब घरे अइलैँ तब ई सब खुसिहाली कँ बात देख के एक नोकर से पुहलस की का भयल है । नोकर कहलस की तोहार लहुरका भाय आयल ही और उन के कुसलकारी से लउटले के संती तोहार बाप खिआवत पिआवत हउएँ । ई सुन के जेठरे बेटवा के जिव में खुस आयल और बखरी में नाँहीं गयल । जब ई सुन के बाप बाहर आयल और मनाबै लागल तब बेटवा कहलस की तोहार धंधा ढेर दिन ले कइली और तोहरे कहले मतिन चलली । आगे तोहार जो कबहूँ नाँहीं भयल की एक खसी मार के लेइ अउता की अपने संगिन के खिआइत पिआइत । और ई तोहार बेटा जवन तोहार धन और दौलत बाँट के रंडो मुंडी के दिहलस जैसे लौट के आयल तइसे प्रतवत भोज दिहला । बाप कहलस की बेटवा तू हमरे संगी सब दिन रहाला । जवन किछु धन और ईस्वरज ही तवन तोहरे ही । ई बेटवा हम जनली की मुइ गयल अब हम पउली तवने से ई जलसा करै के चाहत रहल ॥

Figure 9: Excerpt of a Maithili specimen from the *Linguistic Survey of India* showing the proposed DEVANAGARI VOWEL SIGN AY (from Grierson 1903: 261). It also contains the other short vowel signs.

Table 12-1. Devanagari Vowel Letters

For	Use	Do Not Use
ॐ	0904	<0905, 0946>
आ	0906	<0905, 093E>
ई	0908	<0930, 094D, 0907>
ऊ	090A	<0909, 0941>
ॐ	090D	<090F, 0945>
ॐ	090E	<090F, 0946>
ॐ	0910	<090F, 0947>
ॐ	0911	<0905, 0949> or <0906, 0945>
ओ	0912	<0905, 094A> or <0906, 0946>
ओ	0913	<0905, 094B> or <0906, 0947>
औ	0914	<0905, 094C> or <0906, 0948>
अँ	0972	<0905, 0945>
अं	0973	<0905, 093A>
आँ	0974	<0905, 093B> or <0906, 093A>
औँ	0975	<0905, 094F>
अ्र	0976	<0905, 0956>
अ्रु	0977	<0905, 0957>

Figure 10: Table from the Unicode standard showing Devanagari vowel letters and the proper way to represent them in encoded text (from *The Unicode Standard*, version 8.0, p. 444).

**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://std.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://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

**A. Administrative**

1. **Title:** *Proposal to encode the Devanagari letter and vowel sign AY*
2. Requester's name: *Script Encoding Initiative (SEI) / Anshuman Pandey (pandey@umich.edu)*
3. Requester type (Member body/Liaison/Individual contribution): *Liaison contribution*
4. Submission date: *2015-12-07*
5. Requester's reference (if applicable):
6. Choose one of the following:
 

This is a complete proposal:	<input checked="" type="checkbox"/>
(or) More information will be provided later:	<input type="checkbox"/>

**B. Technical – General**

1. Choose one of the following:
  - a. This proposal is for a new script (set of characters):
 

Proposed name of script:	<input type="checkbox"/>
--------------------------	--------------------------
  - b. The proposal is for addition of character(s) to an existing block:
 

Name of the existing block:	<input checked="" type="checkbox"/>
<i>Devanagari Extended</i>	
2. Number of characters in proposal: *2*
3. Proposed category (select one from below - see section 2.2 of P&P document):
 

A-Contemporary	<input type="checkbox"/>	B.1-Specialized (small collection)	<input checked="" type="checkbox"/>	B.2-Specialized (large collection)	<input type="checkbox"/>
C-Major extinct	<input type="checkbox"/>	D-Attested extinct	<input type="checkbox"/>	E-Minor extinct	<input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic	<input type="checkbox"/>	G-Obscure or questionable usage symbols			
4. Is a repertoire including character names provided?
 

a. Is a repertoire including character names provided?	<input checked="" type="checkbox"/>
a. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?	<input checked="" type="checkbox"/>
b. Are the character shapes attached in a legible form suitable for review?	<input checked="" type="checkbox"/>
5. Fonts related:
  - a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
 

<i>Anshuman Pandey</i>	
------------------------	--
  - b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
 

<i>Anshuman Pandey (pandey@umich.edu)</i>	
---	--
6. References:
  - a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?
  - b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?
7. Special encoding issues:
 

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)?	<input checked="" type="checkbox"/>
---	-------------------------------------

**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 Unicode Character Database ( <http://www.unicode.org/reports/tr44/> ) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

<sup>1</sup> Form number: N4502-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, 2008-05, 2009-11, 2011-03, 2012-01)

**C. Technical - Justification**

1. Has this proposal for addition of character(s) been submitted before? If YES explain	<i>See proposal</i>	<i>yes</i>
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.)? If YES, with whom? If YES, available relevant documents:		
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	<i>See proposal</i>	
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>See proposal</i>	<i>rare</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:		<i>no</i>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	<i>See proposal</i>	<i>yes</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?		<i>yes</i>
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:		<i>no</i>
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:		<i>no</i>
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:		<i>no</i>
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference: Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference:	<i>See proposal</i>	<i>Yes</i> <i>yes</i>
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)		<i>no</i>
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:		<i>no</i>