

**Doc Type:** Working Group Document  
**Title:** Proposal to Encode Additional Short Vowel Characters for Devanagari  
**Source:** Anshuman Pandey (pandey@umich.edu)  
**Status:** Individual Contribution  
**Action:** For consideration by UTC  
**Date:** 2009-09-18

## 1 Introduction

This is a proposal to encode additional short vowel letters and signs for Devanagari in the Universal Character Set (UCS) (ISO/IEC 10646). The short vowel characters already encoded for Devanagari include ऐ U+0904 DEVANAGARI LETTER SHORT A, ए U+090E DEVANAGARI LETTER SHORT E (ँ U+0946 DEVANAGARI VOWEL SIGN SHORT E), and औ U+0912 DEVANAGARI LETTER SHORT O (ँ U+094A DEVANAGARI VOWEL SIGN SHORT O). The latter two are part of a set of characters introduced into Devanagari in the late 19th century by A. F. Rudolf Hoernle in order to represent vowel sounds found in the ‘Bihari’ languages: Bhojpuri, Magahi, and Maithili.

The set of vowel characters introduced by Hoernle are: ऐ SHORT E, ए SHORT AI, औ SHORT O, and औ SHORT AU. These letters are shortened forms of the corresponding regular Devanagari vowel and diphthong letters. They have no equivalents in standard Hindi.<sup>1</sup> In *A Comparative Grammar of the Gaudian (Aryo-Indian) Languages* (1880), Hoernle writes that “the short *ě, ō, aĩ, aũ*, 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.”<sup>2</sup> Therefore, “in order to avoid the inconvenience of two different sounds being denoted by the same sign,” Hoernle adapted the “Gurmukhí or Bangalí 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 ँ).”<sup>3</sup>

Hoernle’s characters were used by other grammarians, such as George Grierson and Samuel Kellogg, who adopted the characters in their own works, but with particular modifications as shown in Table 1. Apparently, Indian writers and scholars also adopted Hoernle’s characters. Grierson writes that all of the short vowels “have been adopted by some of the best Benares Paṇḍits”<sup>4</sup> and that “[a]ccurate writers distinguish these when writing in the Dēva-nāgarī character.”<sup>5</sup>

In addition to the four short vowel characters, Hoernle also introduced a character for representing the ‘neutral vowel’, or the articulation of the inherent vowel of the antepenultimate syllable that is typically unrealized in Hindi and related languages. It appears that Grierson carried this idea further and developed a character for the ‘sonant vowel’, which represents the articulation of the inherent vowel of the final syllable, which is generally unpronounced in these languages and also not marked by VIRAMA as absent.

The issue of encoding characters for additional Devanagari short vowels was raised previously by Jeroen Hellingman in 1998.<sup>6</sup> It appears that neither Hellingman nor others pursued the matter further, ie. through proposals submitted to the Unicode Technical Committee. This proposal includes the short vowel characters discussed by Hellingman, as well as other characters identified by the proposal author. While these characters are not used in standard Devanagari, they are necessary for representing Devanagari text found in grammars and other linguistics texts. The proposed characters will complete the repertoire of short vowel characters for Devanagari and will broaden support for historical Devanagari characters in the UCS.

<sup>1</sup> Grierson 1883: 20. <sup>2</sup> Hoernle 1880: 3. <sup>3</sup> Hoernle : 3. <sup>4</sup> Grierson 1903: 7. <sup>5</sup> Grierson 1903: 21. <sup>6</sup> Hellingman 1998.

	NEUTRAL VOWEL	SONANT VOWEL	SHORT E	SHORT AI	SHORT O	SHORT AU
Grierson	ॐ	ॐ	ॐ	ॐ	ॐ	ॐ
Hoernle	ॐ	—	ॐ	ॐ	ॐ	ॐ
Kellogg	—	—	ॐ	ॐ	ॐ	ॐ

Table 1: Comparison of forms of short vowel letters and signs. The form ॐ is a variant of अ U+0905 DEVANAGARI LETTER A that is common to the Calcutta style of Devanagari.

## 2 Characters Proposed

Seven characters are proposed for encoding as part of the Devanagari block:

- ॐ DEVANAGARI LETTER SHORT AI
- ॐ DEVANAGARI VOWEL SIGN SHORT AI
- ॐ DEVANAGARI LETTER SHORT AU
- ॐ DEVANAGARI VOWEL SIGN SHORT AU
- ॐ DEVANAGARI LETTER REVERSED E
- ॐ DEVANAGARI SIGN NEUTRAL VOWEL
- ॐ DEVANAGARI SIGN SONANT VOWEL

### 2.1 Description of Characters

**ॐ DEVANAGARI LETTER SHORT AI** This character is the shortened form of ॐ U+0910 DEVANAGARI LETTER AI. The glyph shape is derived from Hoernle. Grierson and Kellogg differ from Hoernle in their representations of this character, but the latter’s form should be taken as normative (see Table 1).. In *Seven Grammars of the dialects and subdialects of the Bihārī Language* (1883), Grierson used the glyph ॐ to represent SHORT AI. In *A grammar of the Hindī language* (1893), Kellogg used the glyph ॐ for the character. The shapes used by Grierson and Kellogg are problematic. Grierson’s use of ॐ conflicts with the existing assignment of the glyph for U+090E DEVANAGARI LETTER SHORT E. Kellogg’s form ॐ uniquely represents SHORT AI, but it is graphically disjointed from existing Devanagari characters used for representing length variations of /e/: ॐ U+090F DEVANAGARI LETTER E, ॐ U+0910 DEVANAGARI LETTER AI, and ॐ U+090E DEVANAGARI LETTER SHORT E. Given this, Hoernle’s form ॐ for SHORT AI is the most appropriate, and original, choice.

**ॐ DEVANAGARI VOWEL SIGN SHORT AI** This character is the shortened counterpart to ॐ U+0948 DEVANAGARI VOWEL SIGN AI.

**ॐ DEVANAGARI LETTER SHORT AU** This character is the shortened form of ॐ U+0914 DEVANAGARI LETTER AU.

**ॐ DEVANAGARI VOWEL SIGN SHORT AU** This character is the shortened counterpart of ॐ U+094C DEVANAGARI VOWEL SIGN AU.

**ॐ DEVANAGARI LETTER REVERSED E** This character is synonymous with ॐ U+090E DEVANAGARI LETTER SHORT E. While Hoernle preferred the shape ॐ for SHORT E, Grierson and Kellogg chose a reversed form of ॐ U+090F DEVANAGARI LETTER E. Although ॐ is already encoded for representing SHORT E in Devanagari,

as Hellingman states, the ॠ REVERSED E cannot be considered a glyph variant of ऐ SHORT E since both shapes will be required in a document that uses Grierson’s scheme for short vowels.<sup>7</sup>

◌◌ **DEVANAGARI SIGN NEUTRAL VOWEL** In Hindi and allied languages, the inherent vowel of the antepenultimate syllable is often unpronounced, but this silence is not represented orthographically. In some languages, this inherent vowel is pronounced. In order to account for this feature, Hoernle introduced the sign ◌◌. Grierson described this sign as “the shortest possible utterance”,<sup>8</sup> which may be conceived of as the shortened form of the inherent vowel. For example, a Bhojpuri word for ‘house’ is written as घरवा, in which the inherent vowel of the syllable र /ra/ is articulated: /g<sup>h</sup>arəva/ not /g<sup>h</sup>arva/. The sign ◌◌ indicates the distinction: घरवा.

◌◌ **DEVANAGARI SIGN SONANT VOWEL** In Hindi, the inherent vowel of a word-final consonant is generally not pronounced, moreover it is not marked as silent in Hindi orthography using VIRAMA, as is done for Sanskrit. In Bhojpuri, Magahi, and Maithili, this inherent vowel is often articulated, but this feature is not represented orthographically. For example, देख could represent both /dek<sup>h</sup>/ and /dek<sup>h</sup>a/ (see Figure 9). In order to compensate for this ambiguity, Grierson introduced the character ◌◌ to indicate that the inherent vowel of the consonant is articulated, eg. देख /dek<sup>h</sup>/ and देख◌◌ /dek<sup>h</sup>a/.

## 2.2 Character Properties

The properties of the characters in the Unicode Character Database format are:

```
09xx;DEVANAGARI LETTER SHORT AI;Lo;0;L;;;;N;;;;;
09xx;DEVANAGARI LETTER SHORT AU;Lo;0;L;;;;N;;;;;
09xx;DEVANAGARI LETTER REVERSED E;Lo;0;L;;;;N;;;;;
09XX;DEVANAGARI VOWEL SIGN SHORT AI;Mn;0;NSM;;;;N;;;;;
09xx;DEVANAGARI SIGN NEUTRAL VOWEL;Mc;0;L;;;;N;;;;;
09xx;DEVANAGARI SIGN SONANT VOWEL;Mc;0;L;;;;N;;;;;
```

## 2.3 Characters Not Proposed

Grierson and Kellogg used other characters for representing vowel quality in Devanagari. Some of these characters are idiosyncratic, others may be represented using characters already encoded in the UCS.

ऐ **DEVANAGARI LETTER REVERSED SHORT AI** In *A grammar of the Hindi language* (1893), Samuel Kellogg represented the short forms of E and AI with ॠ and ऐ, respectively. The basis for the two glyphs is a reversed form of ऐ U+090F DEVANAGARI LETTER E. Kellogg’s glyph for SHORT E is proposed for encoding as ॠ DEVANAGARI LETTER REVERSED E. Although ऐ could be considered atomic form of SHORT AI, it may be decomposed as ॠ + े U+0946 DEVANAGARI VOWEL SIGN SHORT E.

आ **DEVANAGARI LETTER EXTRA SHORT A** Grierson uses the character आ to represent /ă/, as shown in Figure 10. It marks length reduction of /a/ in antepenultimate syllables in languages such as Bhojpuri. Given this, it is likely that this character is actually आ U+0906 DEVANAGARI LETTER AA + ◌◌ U+0951 DEVANAGARI STRESS SIGN UDATTA, as shown in Figure 7.

◌◌ **DEVANAGARI SIGN SUPERScript AVAGRAHA** Grierson uses the character ◌◌ to mark what he calls the ‘drawled a’ /â/, as shown in Figure 13. This character is a superscript form of ऌ U+093D DEVANAGARI SIGN AVAGRAHA. Grierson shows different methods used for marking ‘drawled a’ in Figure 11: देखल, देखलऌ, देखलऌ, देखलऌ /dek<sup>h</sup>â/. While Grierson preferred the first method for indicating this sound, the normal in-line AVAGRAHA is used in modern Bhojpuri orthography.

<sup>7</sup> Hellingman 1998. <sup>8</sup> Grierson 1883: 20

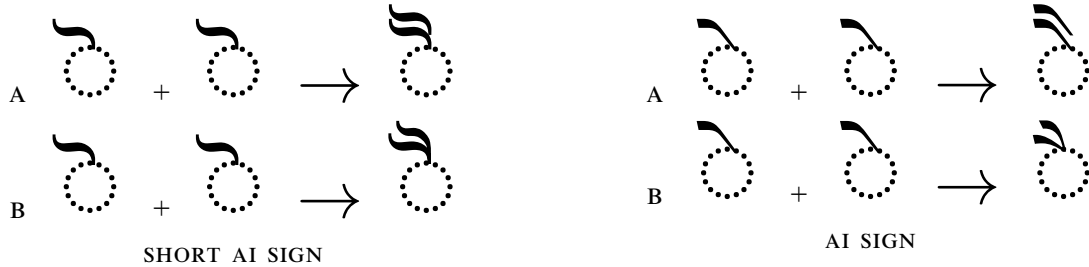
### 3 Implementation

It may be possible to represent some of the proposed characters using existing letters and signs:

$$\begin{aligned}
 \text{ऐ LETTER SHORT AI} &= \text{ऐ U+090E DEVANAGARI LETTER SHORT E} + \\
 &\quad \text{े U+0946 DEVANAGARI VOWEL SIGN SHORT E} \\
 \text{औ LETTER SHORT AU} &= \text{औ U+0912 DEVANAGARI LETTER SHORT O} + \\
 &\quad \text{े U+0946 DEVANAGARI VOWEL SIGN SHORT E} \\
 \text{ै VOWEL SIGN SHORT AI} &= \text{ै U+0946 DEVANAGARI VOWEL SIGN SHORT E} + \\
 &\quad \text{े U+0946 DEVANAGARI VOWEL SIGN SHORT E} \\
 \text{ौ VOWEL SIGN SHORT AU} &= \text{ौ U+094A DEVANAGARI VOWEL SIGN SHORT O} + \\
 &\quad \text{े U+0946 DEVANAGARI VOWEL SIGN SHORT E}
 \end{aligned}$$

Such compositions provide visually representations of the characters, but the semantic values of such characters are lost. There is no principle by which औ U+0912 DEVANAGARI LETTER SHORT O + े U+0946 DEVANAGARI VOWEL SIGN SHORT E should represent the proposed character औ DEVANAGARI LETTER SHORT AU.

Moreover, such compositions do not account for the typographical details necessary for producing aesthetically appropriate glyphs. The character ै U+0948 DEVANAGARI VOWEL SIGN AI may be logically produced by combining two instances of े U+0947 DEVANAGARI VOWEL SIGN E, but the representation requires that the two instances of े be configured in terms of position and proportion. The case for ै VOWEL SIGN SHORT AI is identical; it should not be simply considered as the combination of two instances of े U+0946 DEVANAGARI VOWEL SIGN SHORT E. In the figure below, lines A illustrate the forms of े and ै as produced by combining constituent graphical elements. Lines B show the correct glyph.



If the overwhelming attitude is that the proposed characters be rendered as composite characters, then the characters would be better composed using the proposed ै DEVANAGARI VOWEL SIGN SHORT AI, not with े U+0946 DEVANAGARI VOWEL SIGN SHORT E. This would provide a visually appealing glyph.

$$\begin{aligned}
 \text{ऐ LETTER SHORT AI} &= \text{ऐ U+090F DEVANAGARI LETTER E} + \\
 &\quad \text{ै VOWEL SIGN SHORT AI} \\
 \text{औ LETTER SHORT AU} &= \text{औ U+0906 DEVANAGARI LETTER AA} + \\
 &\quad \text{ै VOWEL SIGN SHORT AI} \\
 \text{ौ VOWEL SIGN SHORT AU} &= \text{ौ U+094A DEVANAGARI VOWEL SIGN AA} + \\
 &\quad \text{ै VOWEL SIGN SHORT AI}
 \end{aligned}$$

Despite such possible representations, the proposed characters should be encoded atomically in order to capture the semantic distinctions they were originally developed to represent.

## 4 Annotations of Short Vowels in the Names List for Devanagari

The Devanagari names list in the Unicode Standard contains annotations for characters for short vowels, which describe these characters as being used for “transcribing Dravidian”. These annotations are likely derived from ISCII 1991, which states that “the vowels ऐ and औ are used in Southern scripts for denoting vowels shorter than ए and ओ respectively.”<sup>9</sup> As evidenced in this proposal, these characters are also used for transcribing sounds in Indo-Aryan languages. Thus, the annotations for short vowel characters should be removed from the Devanagari names list or revised to indicate that they are used to represent short vowels in general.

## 5 References

- Bureau of Indian Standards. 1991. “Indian Script Code for Information Interchange.” IS13194:1991 (ISCII-91). New Delhi.
- Grierson, George A. 1883. *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: Office of the Superintendent of Government Printing, India.
- Hellingman, Jeroen. 1998. “Indian Scripts and Unicode.” <http://ldc.upenn.edu/myl/IndianScriptsUnicode.html>
- Hoernle, A. F. Rudolf. 1880. *A Comparative Grammar of the Gaudīan (Aryo-Indian) Languages: with special reference to the Eastern Hindi, 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 Rāmāyan of Tulsi Dās, also the colloquial dialects of Rājputānā, Kumāon, Avadh, Riwā, Bhojpūr, Magadha, Maithila, etc., with copious philological notes*. 2nd ed., rev. and enl. London: K. Paul, Trench, Trubner and Co.

---

<sup>9</sup> Bureau of Indian Standards 1991: 12.

*Initial:* — अ ँ आ ई इ उ ँ ए ऐ ओ ऒ ओ ऒ ऐ ऒ ऒ ऒ ऒ  
*Noninit.*: ँ — ँ ँ ँ ँ ँ ँ ँ ँ ँ ँ ँ ँ

Figure 1: The set of Devanagari short vowel letters as introduced in Hoernle (from Hoernle 1880: 3).

d. According to Hoernle and Grierson, the colloquial dialects east of Allahabad exhibit not only this short *ě*, *ō*, but also a short *āi* and *āu*, 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 :

ॠ ँ, ॡ ऒ, ओ ऒ, ओ ऒ.

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 *ě* 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 2: The set of Devanagari short vowel letters used in Kellogg (from Kellogg 1893: 4).

## THE DĒVA-NĀGARĪ ALPHABET:

### VOWELS.

अ	a	आ	ā	इ	i	ई	ī	उ	u	ऊ	ū
ऋ	ṛi	ॠ	ṛī	ऌ	ḷi	ॡ	ḷī	ॢ	e	ॣ	ě
ऐ	ai	॥	ai	ओ	o	०	ō	औ	āu	औ	au
अं	āṁ	अः	aḥ								

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

a (not expressed), ā 1; i f, i 1; u, ū 1; ri, e, ē; ai, ai; o 1, ō 1; aŭ 1, au 1.

Figure 4: The Devanagari vowel letters used in the *Linguistic Survey of India* (from Grierson 1903: 7).

Figure 5: Description by Grierson of the short vowel signs (from Grierson 1903: 7).

[illegible]

SECTION 32. The following table therefore shows the complete series of Bihári vowels as used in these

		<i>Neutral</i> ('),	<i>a</i> ,	<i>á</i> ,	<i>i</i> ,	<i>í</i> ,	<i>u</i> ,	<i>ú</i> ,	<i>ě</i> ,	<i>e</i> ,	<i>ö</i> ,	<i>o</i> .		
Initial	.	.	<i>wanting</i> ,	अ,	आ,	इ,	ई,	उ,	ऊ,	ए,	ओ,	ओ.		
Non-initial	.		ं,	.	.	ः,	ि,	ी,	ु,	ू,	ँ,	ँ,	ौ,	ौ.
							<i>aǎ</i> ,	<i>ai</i> ,	<i>aũ</i> ,	<i>au</i> .				
Initial	.	.	.	.	.	.	ऐ,	ऐ,	औ,	औ.				
Non-initial	.	.	.	.	.	.	ै,	ै,	ौ,	ौ.				

क *k*, क *ka*,† का *ká*, कि *ki*, की *kí*, कु *ku*, कू *kú*, के *kě*,  
को *ke*, को *kō*, को *ko*, कै *kāi*, कै *kai*, कौ *kāu*, कौ *kau*.

क *kā*, काँ *kā̃*, किं *kī*, कीँ *kī̃*, कुं *kū*, कूँ *kū̃*, केँ *kē̃*,  
 कैँ *kē̃*, कोँ *kō̃*, कौँ *kō̃*, कौँ *kā̃*, कैँ *kā̃*, कौँ *kā̃*, कौँ *kā̃*.

Figure 6: (from Grierson 1883: 21).

Dēva-nāgarī.		Transliteration.
Initial.	Non-initial.	
ए	ँ	e, as in एकरा <i>ek'rā</i> , तेकरा <i>tek'rā</i> .
ए	ँ	ē, as in एकर <i>ēkar</i> , तेकर <i>tēkar</i> .
ओ	ँ	o, as in ओकरा <i>ok'rā</i> , होइए <i>hoiai</i> .
ओ	ँ	ō, as in ओकर <i>ōkar</i> , लोक <i>lōk</i> .
ऐ	ँ	āi, as in ऐसनहिँ <i>āisanahī</i> , देखैतिओ <i>dekhāitiau</i> .
ऐ	ँ	ai, as in ऐसन <i>aisan</i> , देखैत <i>dekhait</i> .
औ	ँ	āū, as in औतिऐ <i>āūtiāi</i> , पौलहक <i>pāūl'hāk</i> .
औ	ँ	au, as in औताह <i>autāh</i> , पौताह <i>pautāh</i> .

Figure 7: List of all short vowels and examples of their occurrences (from Grierson 1903: 22).

SECTION 29. There is one vowel, called the neutral vowel, which has no representative in the Deva Nāgarī alphabet. It is the shortest possible vocal utterance, and is pronounced like the *o* in the word *Brighton* (*Bright'n*), or the obscure vowel sound in the final syllables of *amiable*, *centre*. This neutral vowel sound I shall represent by a dot ॠ placed after the consonant to which it refers, and in transliteration by an apostrophe: thus घरॠ *ghar'wā*, 'a house.' This neutral vowel can never commence a word, and hence it has no initial form.

Figure 8: Description of the neutral vowel (from Grierson 1883: 20).

SECTION 30. Generally, when a word ends in a consonant, that consonant is silent,—that is to say, the inherent *a* in it is not pronounced. Thus घर is pronounced *ghar*, and not *ghara*. When the inherent *a* in a final consonant is pronounced, which not unfrequently occurs in Bihārī, this will be denoted by a small circle ॠ placed after the final consonant. Thus देख *dekh*, but देखॠ *dekha*.

Figure 9: Description of the final sonant *a* (from Grierson 1883: 20).

*a*. Whenever the vowel आ *ā* finds itself in the antepenultimate syllable, i.e., the third from the end of a word, it is shortened to अ *a*. Thus, the long form of नौआ *nāūā*, a barber, is नाउआ *naūā* or नौआ *nauā*; the long form of आगि *āgi*, fire, is आगिया *āgiyā*; and the instrumental case of पानि *pāni*, water, is पानिँ *pāniē*. This अ *ā* is often written अ *a*, so that the above words would be written नउआ *naūā*, अगिया *agiyā*, and पानिँ *pāniē*, respectively.

Figure 10: Excerpt describing the short *a* (from Grierson 1903: 24).



The long, drawled *a* which I transliterate *ā*, is especially common in Bhojpuri, and its contrast with the short clear-cut *a*, which is of frequent occurrence, gives a striking piquancy to the general tone-colour of the dialect. Usually, this drawled *ā* is left unrepresented in writing but its existence is fully recognised, and various attempts are made by different people to portray it. Thus, some write the sign ' over the consonant containing this vowel. Others write ' after it. Others write : after it, and others repeat the letter अ *a* after it. For instance, the word *dēkhā-lā*, you see, is written by some देखल; by others देखलः; by others देखल:; and by others देखलअ. The first method is that adopted by the present writer.

Figure 11: Description by Grierson of the 'drawled *a*' (from Grierson 1903: 48).

The vowel *a* has four distinct sounds, which should be noted. They can be shown in Dēva-nāgarī writing, but not in the two other alphabets of Mithilā. They are as follows:—

Dēva-nāgarī.		Transcription.
Initial.	Non-initial.	
अ	...	<i>a</i> , as in अग्नि <i>agnī</i> , fire; मरैछी <i>maraichhī</i> , I am dying.
अ'	'	<i>ā</i> as in देखबह <i>dekh'bdh</i> , you will see.
आ	।	<i>ā</i> , as in आगू <i>āgū</i> , before; मारब <i>mārab</i> , I shall beat.
आ	†	<i>ā</i> , as in आगुआ <i>āguā</i> , a preccder; मारैछी <i>māraichhī</i> , I am beating.

Figure 12: Grierson's method for representing different lengths of A (from Grierson 1903: 22).

हे बाबू जी देखल हम एतना दिन से तोहार सेवा सुबित करत चलि आवत हई और कब्यों तोहार प्रको हुकुम नाहीं टरली। तबो तू हम के प्रको भेड़ी के बचो नाहीं दिहल की हम अपने सँघतिन के साथे आनन्द करी। बाकी ज्योंही तोहार ई बेटा आइल त्योंही तू ओन के खातिर भोज कइल ह। ई नाहीं सोचल की ई तोहार उहै बेटउआ हउअै जे तोहार धन कुकरम में नष्ट कै दिहलै। तब बाप बोझल

Figure 13: Excerpt showing the use of the sign to mark 'drawled *a*'. From a specimen of the Sarwaria sub-dialect of Bhojpuri spoken in south-western Gorakhpur district. Prepared in 1899 by Pandit Rām-gharīb Chaube (from Grierson 1903: 246).

पाँचम कथा ।

बान्हल काठी ओ फुजल काठी ।

एक गृहलक बेटा † सभ अपना में लड़ाइ भगंडा करैत बल । कथा सभ सँ बडत बुझौलकै परंतु केओ किछु नहिँ मानलकै । तखन गृहल अपना मन में कहलक कि, जँ एकनरा सभ कँ किछु कै देखाबी, तँ चाही जे बुझि जाय । एक दीन ओ अपना बेटा सभ कँ बजौलक आओर कहलक कि थोरैक काठी हमनरा सोभा लै आवइ । ओहि काठी सबहिँ कँ ओ गृहल जौड़ि सँ कसि कै बान्हलक । फेरि प्रत्येक सँ कहलक कि जौड़िं जनु फोअइ, आओर काठी सबहिँ कँ तोड़ि दैइ । एक एक लागि पड़ल, परंतु किछु नहिँ भेल । फेरि गृहल ओ कँ एक २ काठी देखलकै । तखन ओ सभ ओकनरा भट पट तोड़ि देलक । तखन ओकनरा सभक बाप कहलकैक, कि, बेटा, जौँ तौँ बान्हल काठी सबहिक तरंगेँ अपना में भेलि कैकँ रहबइ, तौँ सभ बैरीक दाँत चम्मत करबइ । आओर जौँ बिहुरि गेलाइ, तखन जानइ कि बिगड़ि गेलाइ ।

Figure 14: A specimen of Maithili in Devanagari showing the use of SIGN NEUTRAL VOWEL (from Grierson 1883: 33).

A B  
C जेठ बेटवा कतहूँ खेतारी में रहलै । ज जब घरे अइलै तब ई सब खुसिहाली कै बात देख के एक नोकर से पुछलैस की का भयल है । नोकर कहलैस 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 D बखरी में नाँहीं गयल । जब ई सुन के बाप बाहर आयल और मनावै लागल तब बेटवा कहलैस की तोहार धंधा ढेर दिन ले कइली और तोहरे कहलै मतिन चलली । आगे तोहार E 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 कै रंडो मुंडी के दिहलैस जैसे लौट के आयल तइसे एतवत F भोज दिहला । बाप कहलैस की बेटवा तू हमरे संगि सब दिन रहाला । जवन किछु धन और ईस्वरज ही तवन तोहरे ही । ई बेटवा हम जनली की मुद्र गयल अब हम पउली तवने से ई जलसा करै के चाहत रहल ॥

Figure 15: Specimen showing the use of all short vowel signs and REVERSED E (from Grierson 1903: 261).

ISO/IEC JTC 1/SC 2/WG 2  
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS  
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646<sup>10</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: **Proposal to Encode Additional Short Vowel Characters for Devanagari**
2. Requester's name: **Anshuman Pandey (pandey@umich.edu)**
3. Requester type (Member Body/Liaison/Individual contribution): **Individual contribution**
4. Submission date: **2009-09-18**
5. Requester's reference (if applicable): **N/A**
6. Choose one of the following:
  - (a) This is a complete proposal: **Yes**
  - (b) or, More information will be provided later: **No**

**B. Technical - General**

1. Choose one of the following:
  - (a) This proposal is for a new script (set of characters): **No**
    - i. Proposed name of script: **N/A**
  - (b) The proposal is for addition of character(s) to an existing block: **Yes**
    - i. Name of the existing block: **Devanagari**
2. Number of characters in proposal: **7**
3. Proposed category: **B.1 - Specialized (small collection)**
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**
  - (a) If available now, identify source(s) for the font and indicate the tools used: **The font is available from Anshuman Pandey.**
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 text of the proposal.**
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 are included.**

---

<sup>10</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.)? **No**
  - (a) If Yes, with whom?: **N/A**
  - 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: **The characters were used by linguists.**
4. The context of use for the proposed characters (type of use; common or rare): **Common**
  - (a) Reference: **The characters are used to represent short vowel sounds in Devanagari.**
5. Are the proposed characters in current use by the user community?: **No.**
  - (a) If Yes, where? Reference: **N/A**
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)? **No**
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? **Possibly. Please see section 3 in the body of the proposal.**
  - (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?  
**No**
  - (a) If Yes, is a rationale for its inclusion provided? **N/A**
  - i. If Yes, reference: **N/A**
11. 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)? **No**
  - (a) If Yes, is a rationale for such use provided? **N/A**
  - i. If Yes, reference: **N/A**
  - (b) Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? **No**
  - i. If Yes, reference: **N/A**
12. Does the proposal contain characters with any special properties such as control function or similar semantics? **No**
  - (a) If Yes, describe in detail (include attachment if necessary): **N/A**
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**