

Request for encoding 1CF4 VEDIC TONE CANDRA ABOVE

Shriramana Sharma – jamadagni-at-gmail-dot-com

2009-Oct-11

This is a request for encoding a character in the Vedic Extensions block. This character resembles 1CD8 VEDIC TONE CANDRA BELOW, but it is placed above instead of below its base character. It also resembles 0945 DEVANAGARI VOWEL SIGN CANDRA E but being a Vedic accent it is preferably placed above the level of the ascenders of Indic vowels, consonants and vowel signs, whereas 0945, being a vowel sign, will stay below the level of Vedic accents.

ॠ

ॡ

ँ

1CD8 VEDIC TONE

*1CF4 VEDIC TONE

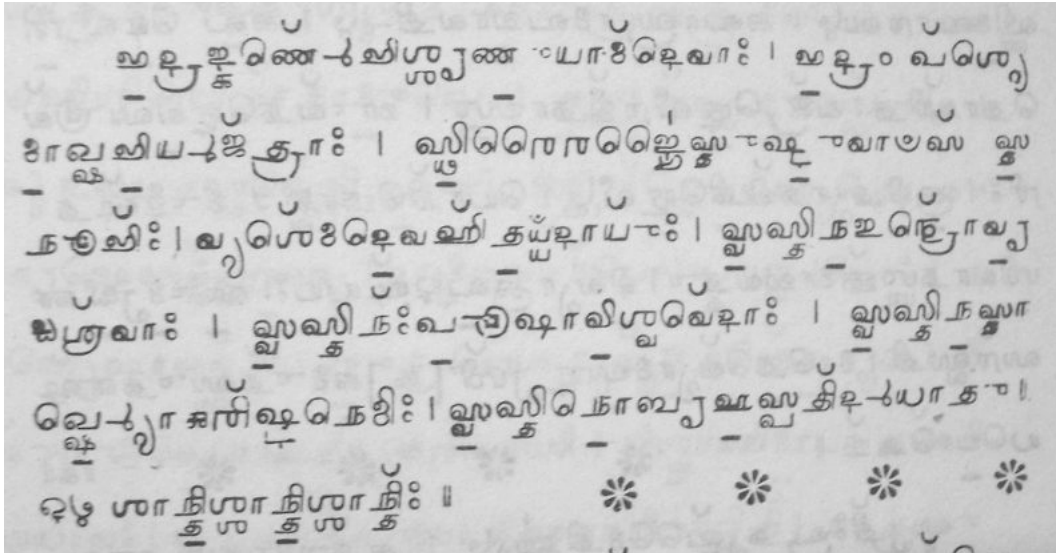
0945 DEVANAGARI VOWEL SIGN

CANDRA BELOW

CANDRA ABOVE

CANDRA E

This Vedic accent is consistently used for the svarita in books of the Yajur Veda printed in the Grantha script, as seen in the following sample from page 679 of the book Taittirīya Brāhmaṇam published by Heritage India Education Trust, Chennai in 1980:



That this character is indeed used for the svarita is evident by comparing this with the following sample of the same Yajur Vedic passage composed in the Devanagari script from <http://sanskritweb.net/yajurveda/ta-01.pdf> (retrieved 2009-Oct-11):

स्वस्ति न इन्द्रो वृद्धश्रवाः ।
 ॐ भद्रं कर्णेभिः शृणुयाम देवाः । स्वस्ति नः पूषा विश्ववेदाः ।
 भद्रं पश्येमाक्षभिर्यजत्राः । स्वस्ति नस्तार्क्ष्यो अरिष्टनेमिः ।
 स्थिरैरङ्गैस्तुष्टुवाꣳसस्तनूभिः । स्वस्ति नो बृहस्पतिर्दधातु ।
 व्यशेम देवहितं यदायुः । ॐ शान्तिः शान्तिः शान्तिः ॥

(I rearranged the image to show the text in columns for more efficient use of space here.)

As is also evident from comparing the above samples, the single upper stroke 0951 VEDIC TONE SVARITA that is used in Devanagari texts for the (normal) svarita is used in Grantha for the dirgha svarita (a special kind of svarita), which is marked in Devanagari by the double upper stroke 1CDA VEDIC TONE DOUBLE SVARITA (see third sentence on top of ŋgai).

Sometimes, even Vedic texts composed in the Devanagari script use this symbol for the normal svarita (and 0951 VEDIC TONE SVARITA for the dirgha svarita) as seen in yet another sample of the same Vedic passage from <http://sanskritweb.net/cakram/ganapatiatharvashirsham.pdf> (retrieved 2009-Oct-11):

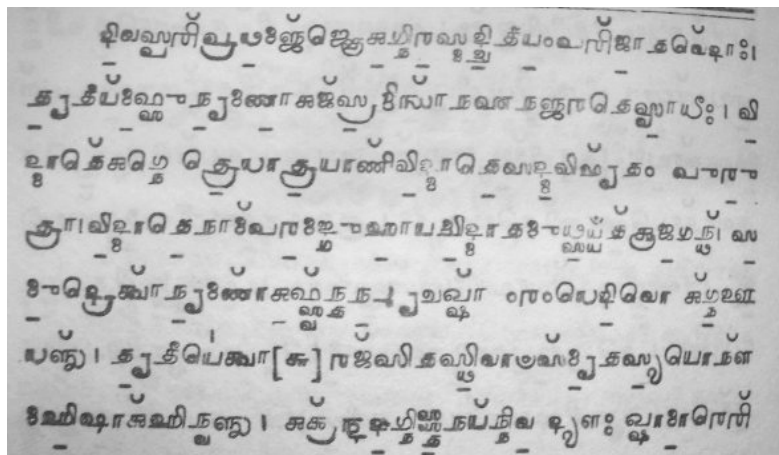
ॐ भद्रं कर्णेभिः शृणुयाम देवाः । भद्रं पश्येमाक्षभिर्यजत्राः । स्थिरैरङ्गैस्तुष्टुवाꣳसस्तनूभिः । व्यशेम
 देवहितं यदायुः । स्वस्ति न इन्द्रो वृद्धश्रवाः । स्वस्ति नः पूषा विश्ववेदाः । स्वस्ति नस्तार्क्ष्यो
 अरिष्टनेमिः । स्वस्ति नो बृहस्पतिर्दधातु ॥
 ॐ शान्तिः शान्तिः शान्तिः ॥

Usually such a choice of svarita accents for Devanagari texts is found in books or publications composed or printed in Tamil Nadu.

Here is a further sample for such usage of the proposed 1CF4 VEDIC TONE CANDRA ABOVE in Devanagari, from page 10 of the booklet Agnihotra Prayoga Prakāśa, published by Raghunātha Śrautin of Shrirangam, Tiruchi, Tamil Nadu in 2005:

दिवस्परिं प्रथमं जज्ञे अग्निस्मद्वितीयं परिं जातवेदाः । तृतीयं
 मप्सु नृमणा अजस्रं मिन्धानं एनं जरते स्वाधीः ॥ विद्या ते अग्नेत्रेधा
 त्रयाणि विद्या ते सद्य विभूतं पुरुत्रा । विद्या ते नाम परमं गुहा यद्विद्या
 तमुत्सं यत आजगन्थं ॥ समुद्रे त्वा नृमणा अप्स्वन्तर्नृचक्षा ईधे
 दिवो अग्न ऊधन् । तृतीयं त्वा रजसि तस्थिवाꣳसमृतस्य योनौ महिषा
 अहिन्वन् ॥ अक्रन्ददग्निः स्तनयन्निव द्यौः क्षामा रेरिह द्वीरुघः

The corresponding Grantha text is to be found in page 382 of Taittirīya Saṃhitā published by Heritage India Education Trust, Chennai in 1980:



It is evident that this svāra mark has consistent and unique usage to denote the svarita in texts published in Tamil Nadu. Hence it is to be considered a valid candidate for encoding.

Details of encoding

As this character has been used across scripts, it should not be encoded in a script-specific block but rather in the common Vedic Extensions block. As the present author has separately proposed a character 1CF3 ROTATED ARDHAVISARGA in the next available codepoint in that block, and since it is preferable to locate that character at 1CF3 to colocate it with its relative, 1CF2 ARDHAVISARGA, it is requested that this character be placed at 1CF4.

As there is no other character in the Vedic Extensions block or elsewhere which can be used instead for this particular purpose, this character must be encoded separately.

Argument for distinct encoding

It may be suggested that 0306 COMBINING BREVE must be used instead of encoding a new character. However, European diacritics are not used anywhere else in Indic context. Rendering engines are already bad at sequences containing the few Vedic accents encoded before Unicode 5.2. This situation should not be further complicated by introducing European diacritics into Indic text. It would only escalate the current disease of displaying dotted circles for perfectly legitimate sequences simply because the rendering engine did not expect such sequences. (I also note in passing that 0953 DEVANAGARI GRAVE ACCENT and 0954 DEVANAGARI ACUTE ACCENT were separately encoded from 0300 COMBINING GRAVE ACCENT and 0301 COMBINING ACUTE ACCENT despite real-world usage of such accents in Devanagari being almost non-existent. The present request for encoding 1CF4 VEDIC TONE CANDRA ABOVE as distinct from 0306 COMBINING BREVE is far more meaningful than *that* disunification.)

I further point out that N3366 has proposed (and got accepted in Unicode 5.2) the Vedic accents 1CD0, 1CD2, 1CD9 and 1CDD all of which could have been represented by 0302 COMBINING CIRCUMFLEX ACCENT, 0304 COMBINING MACRON, 032D COMBINING CIRCUMFLEX ACCENT BELOW and 0323 COMBINING DOT BELOW respectively. I am unimpressed by N3366's arguments that these characters should be encoded because of their "distinctive shapes". N3366 has presented no evidence that these "distinctive shapes" are not merely typographical styles or that native users would not accept the "plain" appearance of 0302 etc instead. As a native Vedic scholar I do believe that other native Vedic scholars with their "plain" pen and paper would write these characters in their "plain" form without any hesitation whatsoever. Even their figure 5.1Ga shows a "plain" round dot for 1CDD leading the authors of N3366 to quote figure 5.1Gb instead in support of their claim. Figures 4.2Ca,b also show a plain form for 1CD2. Thus this is not the proper argument for distinct encoding.

Instead the proper argument for distinct encoding of 1CD0 etc would be that these characters are intended for a distinct purpose from that of 0302 etc, just as 212A KELVIN SIGN which is typographically non-distinct from (and in fact canonically equivalent to) 004B LATIN CAPITAL LETTER K is separately encoded due to its distinct and unique purpose or as 0320 COMBINING MINUS SIGN BELOW is distinguished from 0331 COMBINING MACRON BELOW despite little or no difference in graphic form.

Apart from this, and more apposite to the present discussion, N3366 did not even consider the argument of not encoding 1CD8 VEDIC TONE CANDRA BELOW when 032E COMBINING BREVE BELOW exists. Not even the claim of distinct style was put forth!

Anyhow, by the same argument as above of distinct purpose, 1CF4 VEDIC TONE CANDRA ABOVE should be separated from 0306 COMBINING BREVE, just as 1CD8 VEDIC TONE CANDRA BELOW was separated from 032E COMBINING BREVE BELOW. I do not claim typographic distinction as the reason for separating as none is consistently observed.

The argument for separating 1CF4 VEDIC TONE CANDRA ABOVE from 0945 DEVANAGARI VOWEL SIGN CANDRA E is easier. The positioning argument (that 1CF4 should be placed *above* the level of ascenders while 0945 is *at* the level of ascenders) is not really a strong one, since it is seen even in the samples provided here that the proposed 1CF4 is often placed *at* the level of ascenders as well. Therefore the proper argument is that since 0945 is an Indic vowel sign it has the combining class of 0 with implications for normalization and hence collation, whereas 1CF4 being an accent mark should get a combining class of 230.

Thus separately encoding this character is justified.

Choice of character name

The suggested character name for 1CF4 is VEDIC TONE CANDRA ABOVE after the name of and as the counterpart of 1CD8 VEDIC TONE CANDRA BELOW. Since 1CD8 also sometimes represents a svarita, naming 1CF4 as VEDIC TONE CANDRA SVARITA would not be very appropriate.

Technical Aspects

The Unicode character properties for this character would be:

```
1CF4;VEDIC TONE CANDRA ABOVE;Mc;230;NSM;;;;;N;;;;;
```

As for collation, the general rule with Vedic accents is that it should first be determined what the svara being represented by the particular accent is. Normally Vedic scholars order the four (non-Samavedic) svara-s as follows: udatta, svarita, anudatta and pracaya. Thus, after reducing the variegated representations of these four svara-s to their semantic content, sorting must be done in the above-mentioned order.

The linebreaking property of 1CF4 is obvious, since it is a non-spacing mark.

References

1. Taittirīya Brāhmaṇam, Heritage India Education Trust, Chennai, 1980.
2. Taittirīya Saṃhitā, Heritage India Education Trust, Chennai, 1980.
3. <http://sanskritweb.net/yajurveda/ta-01.pdf>, read 2009-Oct-11.
4. <http://sanskritweb.net/cakram/ganapatiatharvashirsham.pdf>, read 2009-Oct-11.
5. Agnihotra Prayoga Prakāśa, Raghunātha Śrautin, Tiruchi, Tamil Nadu, 2005.

Official Proposal Summary Form

A. Administrative

1. Title

Proposal for encoding 1CF4 VEDIC TONE CANDRA ABOVE

2. Requester's name

Shriramana Sharma

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

Individual contribution

4. Submission date

2009-Oct-13

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

Vedic Extensions

2. Number of characters in proposal

1 (one)

3. Proposed category (A-Contemporary)

Category A.

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?

Shriramana Sharma (for this character alone).

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

Shriramana Sharma, FontForge. TTF file submitted alongwith proposal.

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.

See detailed proposal.

C. Technical – Justification

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

No.

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?

Dr Krishnamurti Shastri, former principal, Madras Sanskrit College, Chennai and Dr Mani Dravid, Madras Sanskrit College, Chennai.

2c. If YES, available relevant documents

None specifically. Mode of contact was personal conversation.

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

Vedic scholars, chiefly in Tamil Nadu, India.

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

Vedic texts, chiefly in the Grantha script and sometimes in the Devanagari script.

4b. Reference

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

Yes, occasionally.

5b. If YES, where?

Chiefly in Tamil Nadu, India.

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

The rationale is that related characters are encoded in the BMP and there is space in the block.

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

Only one character is proposed. It is logical that it is placed in the next available space in the block. As the author of this proposal has separately proposed 1CF3 ROTATED ARDHAVISARGA, this is placed at 1CF4.

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?

It resembles the character 1CD8 VEDIC TONE CANDRA BELOW and 0945 DEVANAGARI VOWEL SIGN CANDRA E.

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

Yes.

10c. If YES, reference

Its positioning and other Unicode character properties are different from those of 1CD8 and 0945.

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)?

This character is a non-spacing combining mark.

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

11c. If YES, reference

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

No.

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.

-O-O-O-