

TO: UTC

L2/09-270R

FROM: Deborah Anderson, SEI, UC Berkeley

DATE: 5 August 2009

RE: Kaithi Decomposition Questions for Unicode 5.2

## 1. Background

The Kaithi script will appear in Unicode 5.2. Upon a close reading of the script proposal (L2/08-194), decomposition and normalization issues have been discovered and a decision needs to be made by the UTC for 5.2. (The result of this decision will impact the Takri script, which also had forms being proposed which contained nuktas.)

## 2. Letters in Kaithi with Nukta

a. The proposal (p. 35) calls out the nukta, ◌◌ U+110BA KAITHI SIGN NUKTA, as “commonly used to distinguish” two characters, BA from VA, but does not provide a decomposition for VA. The nukta appears also in the formation of two other characters (RHA and DDDHA), which were separately encoded but have no decompositions.

With Nukta	Without Nukta
◌◌ U+110AB KAITHI LETTER VA	◌ U+110A5 KAITHI LETTER BA
◌◌ U+1109C KAITHI LETTER RHA	◌ U+1109B KAITHI LETTER DDHA
◌◌ U+1109A KAITHI LETTER DDDHA	◌ U+11099 KAITHI LETTER DDA

b. The proposal states that the letter YA does appear without a nukta. However, this un-nuktated YA is considered a stylistic variant, so is not separately proposed.

With Nukta	Without Nukta
◌◌ U+110A8 KAITHI LETTER YA	(form without nukta occurs, but is considered a variant)

c. The proposal also mentions that nukta appears under various Kaithi letters for sounds in Urdu (such as under KAITHI LETTER JA). In *Kaithi o Hindi barnamala*, the nukta appears under the KAITHI LETTERS KA, KHA, and PHA for sounds not native to Hindi phonology. Since the forms in *Kaithi o Hindi barnamala* are not generally listed in the inventories of Kaithi letters, the proposal says “[the letters with nukta] may be considered a late innovation” and the nukta should be employed to create such letters, as needed.

With Nukta	Without Nukta
(form with nukta used to indicate <i>zain</i> and other sounds not native to Hindi)	◌ U+11094 KAITHI LETTER JA
(form with nukta used to indicate sound not native to Hindi)	◌ U+1108D KAITHI LETTER KA

With Nukta	Without Nukta
(form with nukta used to indicate sound not native to Hindi)	𑂔 U+1108E KAITHI LETTER KHA
(form with nukta used to indicate sound not native to Hindi)	𑂕 U+110A4 KAITHI LETTER PHA

### 3. Questions:

- Should VA 𑂒 get a decomposition? Should RHA 𑂓 and DDDHA 𑂔 also get a decomposition (even though the proposal didn't explicitly call them out)?
- How should YA 𑂕 be documented, if the others get decompositions?
- If any of the above do get decompositions, should entries be added to CompositionExclusions.txt for Unicode 5.2, so that the *preferred* form, even in NFD, is the precomposed letter as encoded (as for the Devanagari nuktaed forms)?

