

# Revised Proposal to Encode Phonetic Symbols with Retroflex Hook in the UCS

*Date:* 2004-02-01

*Author:* Peter Constable, Microsoft

*Address* One Microsoft Way  
Redmond, WA 98052  
USA

*Tel:* +1 425 722 1867

*Email:* petercon@microsoft.com

---

## A. Administrative

- |     |                                  |                                                                    |
|-----|----------------------------------|--------------------------------------------------------------------|
| 1.  | Title                            | Proposal to Encode Phonetic Symbols with Retroflex Hook in the UCS |
| 2.  | Requester's name                 | SIL International (contact: Peter Constable)                       |
| 3.  | Requester type                   | Expert contribution                                                |
| 4.  | Submission date                  | 2003-05-30                                                         |
| 5.  | Requester's reference            | L2/03-170                                                          |
| 6a. | Completion                       | This is a complete proposal                                        |
| 6b. | More information to be provided? | Only as required for clarification.                                |
- 

---

## B. Technical—General

- |     |                                                 |                                |
|-----|-------------------------------------------------|--------------------------------|
| 1a. | New Script? Name?                               | No                             |
| 1b. | Addition of characters to existing block? Name? | Yes — Phonetic Extensions      |
| 2.  | Number of characters in proposal                | 12                             |
| 3.  | Proposed category                               | A                              |
| 4.  | Proposed level of implementation and rationale  | 1 (no combining marks or jamo) |
| 5a. | Character names included in proposal?           | Yes                            |

5b. Character names in accordance with guidelines?	Yes
5c. Character shapes reviewable?	Yes
6a. Who will provide computerized font?	SIL International
6b. Font currently available?	Yes
6c. Font format?	TrueType
7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?	Yes
7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?	Yes
8. Does the proposal address other aspects of character data processing?	Yes, suggested character properties are included (see section E).

---

### **C. Technical—Justification**

---

1. Has this proposal for addition of character(s) been submitted before?	An earlier proposal (L2/03-170) was reviewed by UTC (meeting #95) along with other proposals, including one for additional miscellaneous phonetic symbols (L2/03-190). The latter included three consonant symbols with retroflex hook. It was asked that those characters be merged into this proposal.
2a. Has contact been made to members of the user community?	No
2b. With whom?	n/a
3. Information on the user community for the proposed characters is included?	Linguists.
4. The context of use for the proposed characters	Linguistic descriptions (books, journal publications, etc.); dictionaries.
5. Are the proposed characters in current use by the user community?	These were more often used several decades ago, though some are attested in recent publications.
6a. Must the proposed characters be entirely in the BMP?	Preferably
6b. Rationale?	If possible, should be kept with other phonetic symbols in the BMP.

- |     |                                                                                                                      |                                              |
|-----|----------------------------------------------------------------------------------------------------------------------|----------------------------------------------|
| 7.  | Should the proposed characters be kept together in a contiguous range?                                               | Preferably                                   |
| 8a. | Can any of the proposed characters be considered a presentation form of an existing character or character sequence? | Possibly (see discussion in section F below) |
| 8b. | Rationale for inclusion?                                                                                             | See discussion in section F below.           |
| 9a. | Can any of the proposed characters be considered to be similar (in appearance or function) to an existing character? | No                                           |
| 9b. | Rationale for inclusion?                                                                                             | n/a                                          |
| 10. | Does the proposal include the use of combining characters and/or use of composite sequences?                         | No.                                          |
| 11. | Does the proposal contain characters with any special properties?                                                    | No.                                          |
- 

---

#### **D. SC2/WG2 Administrative**

---

1. Relevant SC2/WG2 document numbers
2. Status (list of meeting number and corresponding action or disposition)
3. Additional contact to user communities, liaison organizations, etc.
4. Assigned category and assigned priority/time frame

Other comments

---



---

#### **E. Proposed Characters**

---

A code chart and list of character names are shown on a new page.

## E.1 Code Chart

	xx0
0	ɑ̣
1	ɑ̥
2	ɑ̜
3	ɛ̣
4	ɛ̥
5	ɛ̜
6	ə̣
7	ɪ̣
8	ɪ̥
9	ɪ̜
A	ụ
B	ʊ̣
C	
D	
E	
F	

## E.2 Character Names

xx00	LATIN SMALL LETTER A WITH RETROFLEX HOOK
xx01	LATIN SMALL LETTER ALPHA WITH RETROFLEX HOOK
xx02	LATIN SMALL LETTER D WITH HOOK AND TAIL
xx03	LATIN SMALL LETTER E WITH RETROFLEX HOOK
xx04	LATIN SMALL LETTER OPEN E WITH RETROFLEX HOOK
xx05	LATIN SMALL LETTER REVERSED OPEN E WITH RETROFLEX HOOK
xx06	LATIN SMALL LETTER SCHWA WITH RETROFLEX HOOK
xx07	LATIN SMALL LETTER I WITH RETROFLEX HOOK
xx08	LATIN SMALL LETTER OPEN O WITH RETROFLEX HOOK
xx09	LATIN SMALL LETTER ESH WITH RETROFLEX HOOK
xx0A	LATIN SMALL LETTER U WITH RETROFLEX HOOK
xx0B	LATIN SMALL LETTER EZH WITH RETROFLEX HOOK

### E.3 Unicode Character Properties

All of these characters should have a general category of Ll; no case mapping for these characters is proposed. Other properties should match those of similar characters (e.g. U+0273 LATIN SMALL LETTER N WITH RETROFLEX HOOK).

## F. Other Information

### F.1 Vowel symbols with retroflex hook

Nine of the twelve characters proposed are vowel symbols modified with retroflex hook.

In phonetic transcription, vowel symbols with retroflex hook are generally used to represent vowel phones with rhoticity (“r-colouring”). Since 1989, the representation recommended by the International Phonetic Association has been to use the rhotic hook; that is, the UCS characters U+025A LATIN SMALL LETTER SCHWA WITH HOOK and U+025D LATIN SMALL LETTER REVERSED OPEN E WITH HOOK, and otherwise a character sequence of a vowel sign followed by U+02DE MODIFIER LETTER RHOTIC HOOK.

Prior to 1989, however, IPA practice was to use a retroflex hook on vowel symbols. The older representation is still cited in the IPA Handbook (IPA 1999):

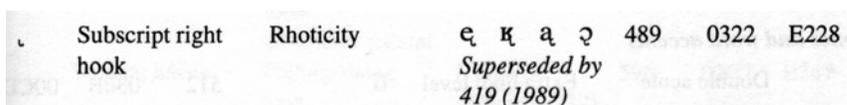


Figure 1. Samples of symbols with retroflex hook: IPA (1999), p. 173.

Vowel symbols with retroflex hook are still occasionally used by linguists in current publications, as seen in Figure 2:

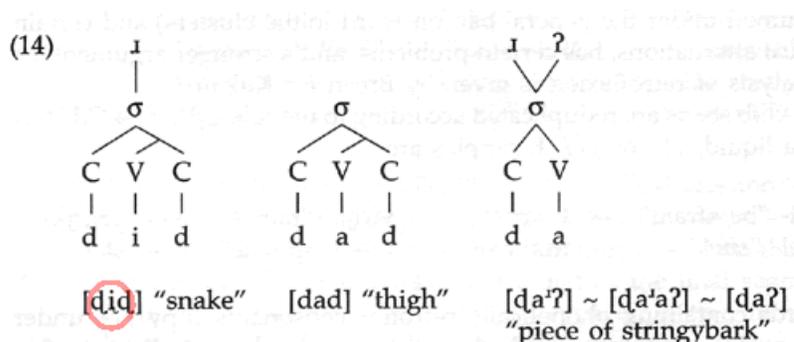


Figure 2. Latin small i with retroflex hook: Evans (1995), p. 740.

Current publications may also use these characters for purposes of citing historic practice, as illustrated in Figure 1.

Insofar as the current IPA recommendation is to use rhotic hook, it is suggested that the `NamesList.txt` file in the Unicode Character Database include an annotation to that effect.

The inventory of characters for vowel symbols proposed is that which were approved by the International Phonetic Association in 1946, as shown in the following figures:

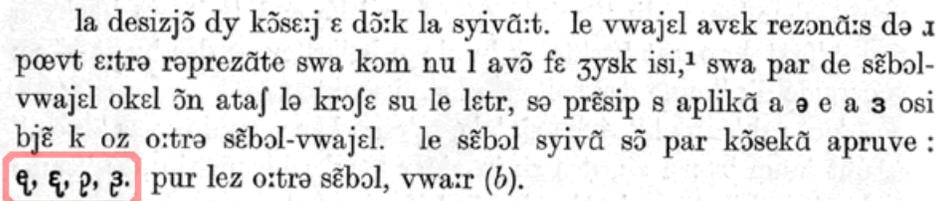


Figure 3. IPA vowel symbols with retroflex hook: IPA (1946), p. 16.

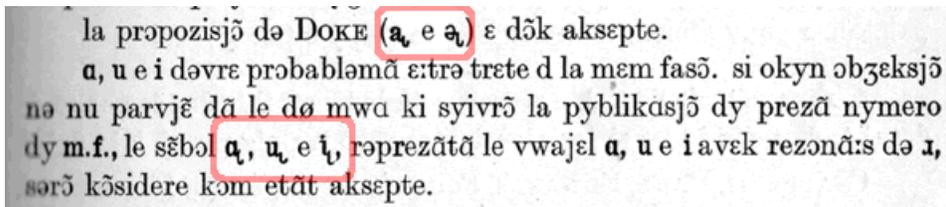


Figure 4. IPA vowel symbols with retroflex hook: IPA (1946), p. 16.

An inspection of a reasonably representative sampling of the linguistics literature suggests that this is a complete inventory: apart from the characters proposed here and already encoded in the UCS (e.g. U+025A LATIN SMALL LETTER SCHWA WITH HOOK), I have not encountered any other phonetic vowel symbols using retroflex hook, except for the lone instance of inverted small-capital r with retroflex hook shown in Figure 1, which I take to be anomalous.

## F.2 Consonant symbols with retroflex hook

Three of the twelve characters proposed are consonant symbols modified with retroflex hook.

The character LATIN SMALL LETTER D WITH HOOK AND TAIL is used to represent a voiced retroflex implosive stop. It is not explicitly IPA-approved, but it is listed in the IPA Handbook (IPA 1999) and is consistent with IPA conventions of using a retroflex hook to indicate retroflexion and a hooked stem to indicate implosive stops (c.f. U+0257 LATIN SMALL LETTER D WITH HOOK). This speech sound is rare but that is attested in at least the Parkari language (Hoyle 2001).

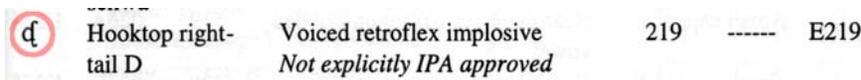


Figure 5. From IPA (1999), p. 179.

Labial	Dental	Dent/Al	Alveol	Post-al	R'flex	Palatal	Velar	Uvular	Total
[b]	ɖ	ɗ ~ ɗ̥	ɗ	ɗ̥	ɗ̥̥	ʃ	g	ʁ]	

Figure 6. From Laver (1994), p. 582.

The subject of a Parkari Nonfinal form is normally the same as the subject of the main verb. However, there are exceptions, e.g.

“My farmwork”

- 11 pur-e pər-e ən pəh u-a m wəle **sɑɫ-ō-h** pəŋi.  
fill in-nonfinite result-nonfinite and then that-G in again **leave-P-pres** water  
After filling them in then **I let** the water into it.

Figure 7. From Hoyle (2001), p. 254.

The name LATIN SMALL LETTER D WITH HOOK AND TAIL is proposed rather than LATIN SMALL LETTER D WITH HOOK AND RETROFLEX HOOK as the repetition of “hook” in the latter is confusing, and the former provides similarities with the related characters U+0256 LATIN SMALL LETTER D WITH TAIL and U+0257 LATIN SMALL LETTER D WITH HOOK.

The characters LATIN SMALL LETTER ESH WITH RETROFLEX HOOK and LATIN SMALL LETTER EZH WITH RETROFLEX HOOK are used to represent retroflex counterparts to the palato-alveolar fricatives esh “ʃ” and ezh “ʒ”. These symbols are not IPA-approved, and their appropriateness is questioned by some linguists since the sounds represented by esh and ezh are “usually regarded as having the blade of the tongue raised towards the hard palate,” a gesture that would “preclude tongue tip retroflexion” (Peter Ladefoged, personal communication). Nevertheless, these symbols are, in fact, used by some linguists:

*Hindi*  
[n̪ɐɾ̪əŋ̪t̪ʰɜɾ̪] ‘continuously’  
[pɑɾ̪ɐ̃j̪ːɔ̃m̪] ‘labour, hard work’  
[kʰɑɾ̪ɐ̃t̪ʰ] ‘reason (for doing something)’  
[h̪w̪ɔ̃l̪i] ‘a (particular) festival’

Figure 8. From Laver (1994), p. 559.

*Norwegian (Southeastern, Larvik)*  
[vɑ̃t̪ɛzkr̪ɑŋɔ̃] ‘overalls’  
[h̪w̪ɔ̃r̪fɑɾ̪vɛɾ̪] ‘hair colours’  
[s̪w̪yːl̪w̪t̪w̪ɔ̃n̪] ‘as thin as an awl’  
[s̪w̪t̪w̪ɔ̃t̪ɔ̃] ‘sweet and sour’

Figure 9. From Laver (1994), p. 560.

	Bilab	LaDe	InDe	Alv	Ret	PoAl	RePo	AlPal	Pal
Plos	p b			t d	ʈ ɖ			ʈ ɖ	c ɟ
Aff	pʰ bʱ	pf bv	tʰ dʰ	ts ɖ	tʂ ɖʂ	tʃ dʒ	tʃ dʒ	tʃ dʒ	cç ɟʝ
Fric	ɸ β	f v	θ ð	s z	ʃ ʒ	ʃ ʒ	ʃ ʒ	ç ʒ	ç ʝ
Nas	ɱ m	ɱ ɱ		ɳ n	ɳ ɳ			ɳ ɳ	ɳ ɳ

Figure 10. From Diehl (1995), p. 1.

### F.3 Representation as sequences with U+0322

Question 8a of section C above asks whether these characters can be considered a presentation form of an existing character or character sequence. They could possibly be viewed as sequences involving U+0322 COMBINING RETROFLEX HOOK BELOW, but I suggest that this would be inappropriate and is irrelevant. While combining marks in general are assumed to be applicable to arbitrary characters in a generative manner, allowing dynamic representation of text elements such as *Latin small a with bridge below*, there are certain combining marks for which this is not appropriate, one of these being U+0322 COMBINING RETROFLEX HOOK BELOW. This view has been expressed on the Unicore discussion list, and some of the rationale provided here has been expressed by others on that list.

There simply are only certain base characters that can sensibly be modified with a retroflex hook, both in a linguistic sense as well as a typographic sense. For instance, it would be silly for both linguistic and typographic reasons to encode a character sequence < U+0290 LATIN SMALL LETTER Z WITH RETROFLEX HOOK, U+0322 COMBINING RETROFLEX HOOK BELOW >. In practice, there is a very limited inventory of characters that are used with retroflex-hook modification.

Also, whereas it is feasible to create font/rendering implementations that can productively display sequences involving arbitrary base characters followed by a combining mark such as U+0300 COMBINING GRAVE ACCENT using mechanisms such as glyph attachment points, this is not feasible for U+0322 COMBINING RETROFLEX HOOK BELOW: the way in which a base character is modified using a retroflex hook is dependent on the particular base character involved. Font implementations must assume a specific inventory of retroflex-hook forms.

Thus, in terms of usage requirements and the realities of implementation, dynamic composition using U+0322 COMBINING RETROFLEX HOOK BELOW is not a good choice, and should be avoided.

Note that this view is corroborated by existing characters in Unicode itself in that characters such as U+0290 LATIN SMALL LETTER Z WITH RETROFLEX HOOK do not have any

decomposition. The combining mark U+0322 COMBINING RETROFLEX HOOK BELOW is not currently used in any decomposition, though there are a number of potential candidates for such decompositions existing in the UCS.

Therefore, since there are good reasons why productive use of U+0322 COMBINING RETROFLEX HOOK BELOW is not recommended, and insofar as existing characters with retroflex hook are not considered presentation forms of existing sequences, it is suggested that the characters proposed here are likewise not to be considered presentation forms of existing sequences.

---

## G. G. References

---

- Diehl, Lon. 1995. "IPA phonetic symbols for describing languages of East and Southeast Asia." Ms.
- Evans, Nick. 1995. "Current issues in the phonology of Australian languages." *The handbook of phonological theory (Blackwell handbooks in linguistics, 1)*, ed. by John A. Goldsmith, 723–61. Oxford: Blackwell Publishers.
- Hoyle, Richard A. 2001. *Scenarios, discourse and translation: The scenario theory of Cognitive Linguistics, its relevance for analysing New Testament Greek and modern Parkari texts, and its implications for translation theory*. University of Surrey Roehampton PhD thesis.
- International Phonetic Association (L'association Phonétique Internationale). 1946. "parti administrativ." *Le maître phonétique* 85:1, p. 16–7.
- . 1999. *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.
- Laver, John. 1994. *Principles of phonetics*. (Cambridge textbooks in linguistics.) Cambridge: Cambridge University Press.