**A. Administrative**

1. **Title:** Proposal to add additional phonetic characters to the UCS

2. **Requester's name:** INCITS/L2, UTC, SIL International

3. **Requester type (Member body/Liaison/Individual contribution):** member body (L2), liaison (UTC), expert (SIL)

4. **Submission date:** 2004-4-19


6. **Choose one of the following:***
   - This is a complete proposal: **yes**
   - More information will be provided later:

**B. Technical – General**

1. **Choose one of the following:***
   - a. This proposal is for a new script (set of characters): **No**
   - b. The proposal is for addition of character(s) to an existing block: **Yes (partially)**
     - Name of the existing block: Latin Extended-B and Phonetic Extensions. One character is to be allocated to a block approved by WG2 for addition in Amendment 1: Combining Diacritical Marks Supplement. Sixty-four characters are to be allocated to a new block: the proposed new block is: Phonetic Extensions Supplement (1D80–1DBF).

2. **Number of characters in proposal:** 73

3. **Proposed category (select one from below - see section 2.2 of P&P document):**
   - A-Contemporary
   - B.1-Specialized (small collection)
   - B.2-Specialized (large collection)
   - C-Major extinct
   - D-Attested extinct
   - E-Minor extinct
   - F-Archaic Hieroglyphic or Ideographic
   - G-Obscure or questionable usage symbols

4. **Proposed Level of Implementation (1, 2 or 3) (see Annex K in P&P document):** 3
   - If Yes, reference: proposal includes one combining mark

5. **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

6. **Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?** SIL International
   - If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: TrueType font is currently available from SIL International (http://scripts.sil.org/)

7. **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

8. **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)? Yes: suggested character properties are included (see §0.3)

9. **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.

C. Technical - Justification

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Has this proposal for addition of character(s) been submitted before?</td>
<td>Not to WG2</td>
</tr>
<tr>
<td>If YES explain</td>
<td>proposal has been submitted to and approved by UTC; first submission to WG2</td>
</tr>
<tr>
<td>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.)?</td>
<td>Yes</td>
</tr>
<tr>
<td>If YES, with whom?</td>
<td>linguists</td>
</tr>
<tr>
<td>If YES, available relevant documents:</td>
<td>see samples in §E and references in §F</td>
</tr>
<tr>
<td>3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?</td>
<td>Yes</td>
</tr>
<tr>
<td>Reference:</td>
<td>linguists, from several sub-disciplines (see discussion in §E and references in §F)</td>
</tr>
<tr>
<td>4. The context of use for the proposed characters (type of use; common or rare)</td>
<td>Characters are used in linguistic descriptions (books, journal publications, etc.), dictionaries and similar linguistic documents. Some of the characters are used in many parts of the world; others are associated with particular regions of the world.</td>
</tr>
<tr>
<td>Reference:</td>
<td>see discussion in §E and references in §F</td>
</tr>
<tr>
<td>5. Are the proposed characters in current use by the user community?</td>
<td>yes</td>
</tr>
<tr>
<td>If YES, where? Reference:</td>
<td>throughout the world; some are more common in some regions (see §E)</td>
</tr>
<tr>
<td>6. After giving due considerations to the principles in the P&amp;P document must the proposed characters be entirely in the BMP?</td>
<td>preferably, yes</td>
</tr>
<tr>
<td>If YES, is a rationale provided?</td>
<td>yes</td>
</tr>
<tr>
<td>If YES, reference:</td>
<td>living characters; if possible, should be kept with other phonetic symbols in the BMP</td>
</tr>
<tr>
<td>7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?</td>
<td>not req’d</td>
</tr>
<tr>
<td>8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?</td>
<td>no</td>
</tr>
<tr>
<td>If YES, is a rationale for its inclusion provided?</td>
<td>n/a</td>
</tr>
<tr>
<td>If YES, reference:</td>
<td>n/a</td>
</tr>
<tr>
<td>9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?</td>
<td>possibly</td>
</tr>
<tr>
<td>If YES, is a rationale for its inclusion provided?</td>
<td>yes</td>
</tr>
<tr>
<td>If YES, reference:</td>
<td>encoding as atomic characters is preferable; see discussion in §E.3 and §E.5.1</td>
</tr>
<tr>
<td>10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? The character LATIN SMALL LETTER C WITH STROKE is similar in appearance to U+00A2 CENT SIGN</td>
<td></td>
</tr>
<tr>
<td>If YES, is a rationale for its inclusion provided?</td>
<td>yes</td>
</tr>
<tr>
<td>If YES, reference:</td>
<td>see discussion in §E.5.1</td>
</tr>
<tr>
<td>11. Does the proposal include use of combining characters and/or use of composite sequences?</td>
<td>yes</td>
</tr>
<tr>
<td>If YES, is a rationale for such use provided?</td>
<td>yes</td>
</tr>
<tr>
<td>If YES, reference:</td>
<td>one combining character is proposed; see discussion in §E.5.6</td>
</tr>
<tr>
<td>Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?</td>
<td>no</td>
</tr>
<tr>
<td>If YES, reference:</td>
<td>n/a</td>
</tr>
<tr>
<td>12. Does the proposal contain characters with any special properties such as control function or similar semantics?</td>
<td>no</td>
</tr>
<tr>
<td>If YES, describe in detail (include attachment if necessary)</td>
<td>n/a</td>
</tr>
<tr>
<td>13. Does the proposal contain any Ideographic compatibility character(s)?</td>
<td>no</td>
</tr>
<tr>
<td>If YES, is the equivalent corresponding unified ideographic character(s) identified?</td>
<td>n/a</td>
</tr>
</tbody>
</table>
D. Proposal

D.1 Character chart

The characters in this proposal have been approved for encoding by UTC, and given tentative code-position assignments in multiple blocks. The relevant columns are shown here, with previously-assigned characters also shown to provide context. The characters proposed here are shown with a yellow highlight; characters already approved by WG2 for inclusion in Amendment 1 (M44.1, M44.5) are shown with a green highlight; characters already encoded in ISO 10646:2003 and Unicode 4.0 have no highlight.

Note that there are apparent gaps left in two of the columns, at U+0237 and U+1D77–U+1D7A. These code positions are intentionally left blank here, and have been tentatively assigned by UTC to other characters that will be presented to WG2 in separate proposals.

The character chart is presented on a new page.
<table>
<thead>
<tr>
<th>Decimal</th>
<th>023</th>
<th>1D6</th>
<th>1D7</th>
<th>1D8</th>
<th>1D9</th>
<th>1DA</th>
<th>1DB</th>
<th>1DC</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>⁰</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
</tr>
<tr>
<td>1</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
</tr>
<tr>
<td>2</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
</tr>
<tr>
<td>3</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
</tr>
<tr>
<td>4</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
</tr>
<tr>
<td>5</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
</tr>
<tr>
<td>6</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
</tr>
<tr>
<td>7</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
</tr>
<tr>
<td>8</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
</tr>
<tr>
<td>9</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
</tr>
<tr>
<td>A</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
</tr>
<tr>
<td>B</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
</tr>
<tr>
<td>C</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
</tr>
<tr>
<td>D</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
</tr>
<tr>
<td>E</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
</tr>
<tr>
<td>F</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
<td>ø</td>
</tr>
</tbody>
</table>

N2740 – Proposal to Encode Additional Phonetic Characters in the UCS
D.2 Names list

Latin Extended-B

0238 LATIN SMALL LETTER DB DIGRAPHER
0239 LATIN SMALL LETTER QP DIGRAPHER
023A LATIN SMALL LETTER C WITH STROKE

Phonetic Extensions

1D7B LATIN SMALL CAPITAL LETTER I WITH STROKE
1D7C LATIN SMALL LETTER IOTA WITH STROKE
1D7D LATIN SMALL LETTER P WITH STROKE
1D7E LATIN SMALL CAPITAL LETTER U WITH STROKE
1D7F LATIN SMALL LETTER UPSILON WITH STROKE

Phonetic Extensions Supplement

1D80 LATIN SMALL LETTER B WITH PALATAL HOOK
1D81 LATIN SMALL LETTER D WITH PALATAL HOOK
1D82 LATIN SMALL LETTER F WITH PALATAL HOOK
1D83 LATIN SMALL LETTER G WITH PALATAL HOOK
1D84 LATIN SMALL LETTER K WITH PALATAL HOOK
1D85 LATIN SMALL LETTER L WITH PALATAL HOOK
1D86 LATIN SMALL LETTER M WITH PALATAL HOOK
1D87 LATIN SMALL LETTER N WITH PALATAL HOOK
1D88 LATIN SMALL LETTER P WITH PALATAL HOOK
1D89 LATIN SMALL LETTER R WITH PALATAL HOOK
1D8A LATIN SMALL LETTER S WITH PALATAL HOOK
1D8B LATIN SMALL LETTER SH WITH PALATAL HOOK
1D8C LATIN SMALL LETTER V WITH PALATAL HOOK
1D8D LATIN SMALL LETTER X WITH PALATAL HOOK
1D8E LATIN SMALL LETTER Z WITH PALATAL HOOK
1D8F LATIN SMALL LETTER A WITH RETROFLEX HOOK
1D90 LATIN SMALL LETTER ALPHA WITH RETROFLEX HOOK
1D91 LATIN SMALL LETTER D WITH HOOK AND TAIL
1D92 LATIN SMALL LETTER E WITH RETROFLEX HOOK
1D93 LATIN SMALL LETTER OPEN E WITH RETROFLEX HOOK
1D94 LATIN SMALL LETTER REVERSED OPEN E WITH RETROFLEX HOOK
1D95 LATIN SMALL LETTER SCHWA WITH RETROFLEX HOOK
1D96 LATIN SMALL LETTER I WITH RETROFLEX HOOK
1D97 LATIN SMALL LETTER OPEN O WITH RETROFLEX HOOK
1D98 LATIN SMALL LETTER SH WITH RETROFLEX HOOK
1D99 LATIN SMALL LETTER U WITH RETROFLEX HOOK
1D9A LATIN SMALL LETTER EZH WITH RETROFLEX HOOK
1D9B MODIFIER LETTER SMALL TURNED ALPHA = <super> 0252
1D9C MODIFIER LETTER SMALL C = <super> 0063
1D9D MODIFIER LETTER SMALL C WITH CURL = <super> 0255
1D9E MODIFIER LETTER SMALL ETH = <super> 00F0
1D9F MODIFIER LETTER SMALL REVERSED OPEN E = <super> 025C
1DA0 MODIFIER LETTER SMALL F = <super> 0066
1DA1 MODIFIER LETTER SMALL DOTLESS J WITH STROKE = <super> 025F
1DA2 MODIFIER LETTER SMALL SCRIPT G = <super> 0261
1DA3 MODIFIER LETTER SMALL TURNED H = <super> 0265
1DA4 MODIFIER LETTER SMALL I WITH STROKE = <super> 0268
1DA5 MODIFIER LETTER SMALL IOTA = <super> 0269
1DA6 MODIFIER LETTER SMALL CAPITAL I = <super> 026A
1DA7 MODIFIER LETTER SMALL CAPITAL I WITH STROKE = <super> 1D7B
1DA8 MODIFIER LETTER SMALL J WITH CROSSED-TAIL = <super> 029D
1DA9 MODIFIER LETTER SMALL L WITH RETROFLEX HOOK = <super> 026D
1DAA MODIFIER LETTER SMALL L WITH PALATAL HOOK = <super> ID85
1DAB MODIFIER LETTER SMALL CAPITAL L = <super> 029F
1DAC MODIFIER LETTER SMALL M WITH HOOK = <super> 0271
1DAD MODIFIER LETTER SMALL TURNED M WITH LONG LEG = <super> 0270
1DAE MODIFIER LETTER SMALL N WITH LEFT HOOK = <super> 0272
D.3 Unicode character properties

Character properties for the proposed characters should be as follows:

- U+0238–U+023A, U+1D7B–U+1D9A: 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+0061 LATIN SMALL LETTER A).

- U+1D9B–U+1DBF: All of the proposed characters should have a general category of Lm. Compatibility decompositions should be as shown above. Other properties should match those of similar characters, such as U+02E1 MODIFIER LETTER SMALL L.

- U+1DC2: This character should have a general category of Mn, and a canonical combining class of 230. Other properties should match those of similar characters, such as U+0323 COMBINING DOT BELOW.

E. Other information

Supporting information for the proposed characters is presented below, organized according to the following broad categories of characters:

- phonetic symbols with palatal hook: U+1D80–U+1D8F
- phonetic symbols with retroflex hook: U+1D8F–U+1D9A
- phonetic modifier letters: U+1D9B–U+1DBF
- other phonetic symbols: U+0238–U+023A, U+1D7B–U+1D7F, U+1DC2
E.1 Phonetic symbols with palatal hook

E.1.1 Background: transcription conventions for palatalization

In phonetic transcription, consonant letters with palatal hook are generally used to represent consonant phonemes with palatalized articulation. Since 1989, the representation recommended by the International Phonetic Association has been to use superscript j; that is, the UCS character U+02B2 MODIFIER LETTER SMALL J.

Prior to 1989, however, IPA practice allowed for the use of palatal hook on consonant symbols. The older representation is still documented in the IPA Handbook (IPA 1999), and is often referred to in general books on phonetics.

Figure 1. From IPA (1949), p. 13.

Figure 2. From Catford (1988), p. 222.

Within the linguistics tradition for study of the Russian language, use of characters with palatal hook has been common practice.

Figure 3. Consonants with palatal hook used for Russian (Boyanus and Jopson 1939, p. xiv).

---

2 Characters with palatal hook are not, in fact, used in that publication, however.
specially designed letters. In this book we adopt largely the second method: palatalized consonants are indicated by a normal consonant letter with a small j attached to it at the lower right-hand corner. Thus, ș represents ‘palatalized d’, ș represents ‘palatalized s’, ș represents ‘palatalized ș’ and so on.

Figure 4. Consonants with palatal hook used for Russian (Jones and Ward 1969, p. 82).

Characters with palatal hooks have been used in relation to other languages as well, however:

Figure 5. Examples of characters with palatal hook used in relation to Australian languages (Evans 1995, p. 744).

Figure 6. Examples of characters with palatal hook used in relation to African languages (Tucker 1971, p. 648).

It is in relation to Russian that the widest selection of symbols with palatal hook is used, however, and the inventory proposed here is based on the requirements for Russian. An inspection of a reasonably representative sampling of the linguistics literature suggests that this may be a complete inventory of required palatal-hook characters: apart from the characters proposed here and those already encoded in the UCS (e.g. U+01AB LATIN SMALL LETTER T WITH PALATAL HOOK), no clear attestation of any other phonetic symbols using palatal hook has been found. There is a pair of marginal cases, c and ezh, for which use of palatal-hook forms has not been clearly attested, but for which evidence indicates a need to encode palatal-hook forms may possibly arise in the future; these will be described below. Beyond these, however, no additional candidates for palatal-hook forms are currently known.

Various authors have used typographic approximations for palatal hook when the selection of type available to them has not been extensive enough. This can be seen in Figure 6, above, in
which a comma is used; others have used a cedilla:

3. All other consonant sounds may be either hard or soft. A comma is placed beneath those that are soft to indicate how they should be pronounced. The following examples are given to demonstrate this same basic phenomenon in English. Remember that very few of the soft and hard English consonants are exactly like their equivalents in Russian.

![Figure 7. Cedilla as typographic approximation of palatal hook (Clark 1983, p. xx).](image)

One other convention used by Slavicists is to indicate palatalization using a modifier letter apostrophe; e.g., /t/. A sample following this convention can be seen in Figure 12, below.

The fact that approximations such as comma are used as a fallback when adequate type is not available can be seen in cases where conventions are mixed:
In Figure 8, the use of true palatal-hook characters for the Russian palatalized consonants in all cases but ɡ and v demonstrate clearly that this was the author’s preferred practice for representing palatalization. And it is clear from other examples involving Russian (see, for instance, Figure 7) that g-comma and v-comma are intended to represent palatalized consonants. We can only conclude that the author did not use palatal-hook typeforms in these two cases because they were not available to him.

### E.1.2 Inventory of proposed palatal-hook characters

The inventory of proposed characters corresponds to palatalized consonant phones of Russian. The most commonly-encountered palatal-hook symbols can be seen in the sample from Jones and Ward (1969) shown in Figure 9:
Figure 9. Russian palatalized consonant phonemes (Jones and Ward 1969, p. 299).

This set of thirteen characters with palatal hook is consistently corroborated by several authors. (Note that one of these, t-palatal hook, is already encoded in the UCS. Hence, this accounts for twelve of the fifteen palatal-hook characters proposed.)

Other sources use additional characters with palatal hook in order to transcribe phonetic surface forms in Russian. Thus, the occurrence of palatal-hooked variants for ɡ, esh, and x in Figure 11 below; the ɡ-palatal hook can also be seen in Figure 4 above, and the x-palatal hook is also seen in Figure 10:

Figure 10. Character x with palatal used for Russian (IPA 1949, p. 14).
The inventory from Boyanus and Jopson (1939) in Figure 11 with the exception of esh-palatal hook is corroborated by Ward (1966), by Clark (1983) (see Figure 7) and by Dawson et al (1964). This inventory is also corroborated by Wade (2000) (see Figure 8), though that author uses the IPA symbol esh-curl (U+0286) rather than esh-palatal hook.

### E.1.3 Marginal cases: c, ezh

Some descriptions of Russian also make reference to palatalized post-alveolar voiced fricative and voiceless affricate, as shown in Figure 12:

![Figure 12](image)

**Figure 12.** Russian palatalized consonants, including post-alveolar fricatives and affricate (Halle 1994, p. 42).
Note that, in the chart in Figure 12, the author presents a complete inventory of palatalized consonants but is using the alternate convention of indicating palatalization by means of a modifier letter apostrophe, mentioned above. Also, this author is using the hacek diacritic to represent post-alveolar sounds: š, ž and č rather than š, ž and ť. Thus, the palatalized post-alveolar voiced fricative and voiceless affricate are represented as ž' and č' respectively. These phones are also attested using the comma representation described above:

Again, it appears that, in such situations, the author has used a comma approximation of palatal-hook forms simply because adequate type that included characters with palatal hooks was not available.

The implication of this is that, were the type available, the author might have used c-palatal hook “č” to represent the palatalized post-alveolar voiceless affricate. Also, with an author that used ezh rather than z-hacek for the voiced post-alveolar affricate, it seems possible that ezh-palatal hook “ž” might have been used to represent the palatalized variant of that sound.

Potential use of c-palatal hook is also suggested from the following sample from Africanist literature, in which c-comma is used for a palatalized consonant (Figure 6, repeated here for convenience as Figure 14):

In this work, the author is presenting various representations for phones of sub-Saharan languages. The second column of his table is labeled “I.P.A.”, and his practice in other cases of palatalized consonants is to use a palatal-hook form, as seen here in the case of z-palatal hook. Thus, it appears that c-comma is being used due to a lack of type for c-palatal hook.

Therefore, in addition to those characters proposed here, there is evidence that suggests the possibility of eventually needing to represent c-palatal hook “č” and z-palatal hook “ž” in the UCS. In the samples shown, however, various alternate representation conventions were used, and not the palatal-hook variants of c and z. In the absence of clear attestation for these
characters, therefore, they are not included in this proposal. They are documented here, however, to show what the full extent of required palatal-hook characters might eventually be.

E.2 Phonetic symbols with retroflex hook

Twelve of the proposed characters are symbols with retroflex hook. In phonetic transcription, the retroflex hook is used with vowel symbols and also with consonant symbols, but with slightly different functions, and with differences in usage and attestation. These two categories are discussed separately.

E.2.1 Vowel symbols with retroflex hook

Nine of the proposed characters 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 15. Samples of symbols with retroflex hook: IPA (1999), p. 173.](image1)

Vowel symbols with retroflex hook are still occasionally used by linguists in current publications, as seen in Figure 16:

![Figure 16. Latin small i with retroflex hook: Evans (1995), p. 740.](image2)

Current publications may also use these characters for purposes of citing historic practice, as illustrated in Figure 15.

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 was approved by the
International Phonetic Association in 1946, as shown in the following figures:

Figure 17. IPA vowel symbols with retroflex hook: IPA (1946), p. 16.

Figure 18. 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, no other phonetic vowel symbols using retroflex hook have been encountered, except for the lone instance of inverted small-capital r with retroflex hook shown in Figure 15, which is considered here to be anomalous.

### E.2.2 Consonant symbols with retroflex hook

Three of the proposed characters 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 ascending stem to indicate implosive stops (c.f. U+0257 LATIN SMALL LETTER D WITH HOOK). This speech sound is rare but is attested in at least the Parkari language (Hoyle 2001).

Figure 19. From IPA (1999), p. 179.

Figure 20. 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 pu-će poće on poći u-a m woľe sač-o-h poći.

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 21. 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 “ʃ” and LATIN SMALL LETTER EZH WITH RETROFLEX HOOK “ʒ” 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:

Figure 22. From Laver (1994), p. 559.

Norwegian (Southeastern, Larvik)

[vaːlːəkkəŋ] ‘overalls’

[bʰɑːfɑːrəːŋ] ‘hair colours’

[sʰyːdʰɪŋ] ‘as thin as an awl’

[sʰuːfʰɪŋ] ‘sweet and sour’

Figure 23. From Laver (1994), p. 560.
E.3 Representation of symbols with palatal or retroflex hooks as sequences with U+0321, U+0322

Question 9 of section C above asks whether any of the characters can be encoded as a character sequence. The proposed characters discussed in §E.1 and §E.2 could possibly be viewed as sequences involving the characters U+0321 COMBINING PALATALIZED HOOK BELOW and U+0322 COMBINING RETROFLEX HOOK BELOW respectively. It is suggested that this would be inappropriate, however, and that encoding using atomic characters is very much to be preferred.

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. In particular, the characters U+0321 COMBINING PALATALIZED HOOK BELOW and U+0322 COMBINING RETROFLEX HOOK BELOW should not be used in a productive manner.

There simply are only certain base characters that can sensibly be modified with a palatal hook or 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+01AB LATIN SMALL LETTER T WITH PALATAL HOOK, U+0321 COMBINING PALATALIZED HOOK BELOW >, or a character sequence < U+0290 LATIN SMALL LETTER Z WITH RETROFLEX HOOK, U+0322 COMBINING RETROFLEX HOOK BELOW >. In practice, there are very limited inventory of characters that are used with palatal hook or 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+0321 COMBINING PALATALIZED HOOK BELOW or for U+0322 COMBINING RETROFLEX HOOK BELOW: the way in which a base character is modified using a palatal or 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+0321 COMBINING PALATALIZED HOOK BELOW and 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 the UCS itself in that existing characters such as U+01AB LATIN SMALL LETTER T WITH PALATAL HOOK and U+0290 LATIN SMALL LETTER Z WITH RETROFLEX HOOK do not have any decomposition. The combining marks U+0321 COMBINING PALATALIZED HOOK BELOW and U+0322 COMBINING RETROFLEX HOOK BELOW are not currently used in any decomposition, even 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+0321 COMBINING PALATALIZED HOOK BELOW and U+0322 COMBINING RETROFLEX HOOK BELOW is not recommended, and insofar as existing characters with palatal hook and retroflex hook are not considered presentation forms of existing sequences, it is argued that the characters proposed here are likewise not to be considered presentation forms of existing sequences.

E.4 Phonetic modifier letters

In general, modifier letters are used in phonetic transcription to represent secondary aspects of articulation. Secondary articulations may involve aspects of simultaneous articulation that are considered to be in some sense less dominant to the basic sound (for instance, nasalized vowels are typically conceived in terms of their oral counterparts but with the additional secondary articulation of nasalization); or they may involve a transitional articulation of a type that might otherwise be considered a complete speech sound in its own right but for various reasons is interpreted by the linguist as a secondary element in a complex speech sound (for instance, diphthongs, or nasal onset of oral stop consonants). In some situations, the recommended transcription using the International Phonetic Alphabet would not involve a modifier letter; thus, many of the proposed characters are not officially-approved IPA notation. Nevertheless, the use of these modifier letters is fairly commonplace among linguists, even those that advocate the use of IPA.

The proposed modifier letters include those used in phonetic transcription to represent vowel-like sounds, and those used to represent consonantal sounds. These two groups will be discussed separately.

E.4.1 Vowel modifier letters

Vowel modifier letters are often used by linguists in transcribing diphthongs. Diphthongs are speech sounds involving two distinct but sequentially-contiguous vocalic gestures—two vowel targets. For instance, whereas the Spanish phoneme /e/ is typically spoken with a single vowel target, [e], the English phoneme /e/ is very often spoken with two vowel targets, [e] and [i]. Following the conventions of IPA strictly, the English phoneme could be transcribed as [ei] or [ɛi]. Occasionally, though, linguists will transcribe such a diphthong as [ɛi] or [i], according to which component is considered to be secondary—an “on-glide” or an “off-glide”: 
Vowel modifier letters are also sometimes used to transcribe syllables that have a marginally-vocalic nucleus or a vocalic nucleus of very short duration, such that the vowel component of the syllable seems suppressed in relation to the consonantal components.

There are already a number of vowel modifier letters encoded in the UCS. Most of these were added in ISO/IEC 10646-1:2000 AMD2 and Unicode 4.0 and are in the Phonetic Extensions block:

<table>
<thead>
<tr>
<th>Code</th>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1D43</td>
<td>ᡃ</td>
<td>MODIFIER LETTER SMALL A</td>
</tr>
<tr>
<td>1D44</td>
<td>ᡣ</td>
<td>MODIFIER LETTER SMALL TURNED A</td>
</tr>
<tr>
<td>1D45</td>
<td>ᡤ</td>
<td>MODIFIER LETTER SMALL ALPHA</td>
</tr>
<tr>
<td>1D46</td>
<td>ᡥ</td>
<td>MODIFIER LETTER SMALL TURNED AE</td>
</tr>
<tr>
<td>1D49</td>
<td>ᡧ</td>
<td>MODIFIER LETTER SMALL E</td>
</tr>
<tr>
<td>1D4A</td>
<td>ᡩ</td>
<td>MODIFIER LETTER SMALL SCHWA</td>
</tr>
<tr>
<td>1D4B</td>
<td>ᡪ</td>
<td>MODIFIER LETTER SMALL OPEN E</td>
</tr>
<tr>
<td>1D4C</td>
<td>ᡫ</td>
<td>MODIFIER LETTER SMALL TURNED OPEN E</td>
</tr>
<tr>
<td>1D4E</td>
<td>ᡬ</td>
<td>MODIFIER LETTER SMALL TURNED I</td>
</tr>
<tr>
<td>1D52</td>
<td>ᡪ</td>
<td>MODIFIER LETTER SMALL O</td>
</tr>
<tr>
<td>1D53</td>
<td>ᡥ</td>
<td>MODIFIER LETTER SMALL OPEN O</td>
</tr>
<tr>
<td>1D54</td>
<td>ᡨ</td>
<td>MODIFIER LETTER SMALL TOP HALF O</td>
</tr>
<tr>
<td>1D55</td>
<td>ᡩ</td>
<td>MODIFIER LETTER SMALL BOTTOM HALF O</td>
</tr>
<tr>
<td>1D58</td>
<td>ᡪ</td>
<td>MODIFIER LETTER SMALL U</td>
</tr>
<tr>
<td>1D59</td>
<td>ᡦ</td>
<td>MODIFIER LETTER SMALL SIDEWAYS U</td>
</tr>
<tr>
<td>1D5A</td>
<td>ᡧ</td>
<td>MODIFIER LETTER SMALL TURNED M</td>
</tr>
<tr>
<td>2071</td>
<td>ᡮ</td>
<td>SUPERSCRIPT LATIN SMALL LETTER I</td>
</tr>
</tbody>
</table>

Table 1. Vowel modifier letters already encoded in the UCS
This covers those vowel sounds that are most commonly encountered in the world’s languages. This list does not include all vowel symbols used in phonetic transcription, however. In principle, any vowel gesture may potentially be one of the targets in a diphthong. Ladefoged and Maddieson (1996, p. 322) comment, “The kinds of vowels that occur as targets in diphthongs are no different from those that occur as single vowels.”

When combined with modifier letters already encoded in the UCS, the vowel modifier letters proposed here cover most of the vowel symbols from the IPA and Americanist traditions:

![Figure 26. IPA vowels (IPA 1999, p. ix).](image)

![Figure 27. Americanist vowels (Pullum and Ladusaw 1996, p. 298).](image)

3 While IPA is increasingly prevalent, the Americanist tradition is still in use, and the use of superscripts to transcribe diphthongs may be more prevalent among those that use Americanist conventions. Some vowels in the Americanist system use diacritics, but it is assumed that combining marks can be used in sequences with modifier letters as well as with other letters. Capital vowel letters are used by some in the Americanist tradition to transcribe voiceless vocoids, but this proposal does not include modifier-letter counterparts to Latin capital vowel letters. We are not aware at the present time of a user need for capital vowel modifier letters in order to transcribe a voiceless, secondary component of a diphthong using Americanist conventions.

4 There is some variation within Americanist usage. Whereas Pullum and Ladusaw show a small capital I for...
The vowel portion of the overall proposal is summarized in Table 2, which includes an index to samples illustrating each one:

<table>
<thead>
<tr>
<th>Character</th>
<th>Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1D9B</td>
<td>Figure 41</td>
</tr>
<tr>
<td>1D9F</td>
<td>Figure 33</td>
</tr>
<tr>
<td>1DA4</td>
<td>Figure 32, Figure 36</td>
</tr>
<tr>
<td>1DA5</td>
<td>Figure 40</td>
</tr>
<tr>
<td>1DA6</td>
<td>Figure 33, Figure 39</td>
</tr>
<tr>
<td>1DA7</td>
<td>Figure 38</td>
</tr>
<tr>
<td>1DB1</td>
<td>Figure 33</td>
</tr>
<tr>
<td>1DB6</td>
<td>Figure 32</td>
</tr>
<tr>
<td>1DB7</td>
<td>Figure 35</td>
</tr>
<tr>
<td>1DB8</td>
<td>Figure 40</td>
</tr>
<tr>
<td>1DBA</td>
<td>Figure 30, Figure 36</td>
</tr>
</tbody>
</table>

Table 2. Proposed vowel modifier letters

The following samples serve to illustrate the use of vowel modifier letters in general to transcribe diphthongs, and also to demonstrate attestation of the vowel modifier letters proposed. The samples will contain vowel modifier letters that are already encoded as well as those being proposed; those already encoded will be highlighted in blue; those being proposed, in red. Note that some of these samples show modifier letters with diacritical marks; it is assumed that these diacritical marks can be encoded using combining characters.

Figure 28. Vowel modifier letters: a, schwa, u (Czaykowska-Higgins and Willett 1997, p. 408).

the front unrounded lower-high vocoid, many represent this vocoid using small iota. Also, some use a small v with hook for the back round lower-high vocoid, rather than the small capital u shown here. Barred iota and barred v-hook for central lower-high vowels are not used, however.
Figure 29. Vowel modifier letter: schwa (Bessell 1998, p. 5).

Figure 30. Vowel modifier letter: turned v (Laver 1994, p. 560).

Figure 31. Vowel modifier letters: e, open-e, u (Malone 1999, p. 353).

Figure 32. Vowel modifier letters: i, i-bar, e, open-e, u, u-bar, o, open-o (Lojenga 1994, p. 90).
Figure 33. Vowel modifier letters: o-bar, small capital I, schwa, reversed open e (Laver 1994, p. 559).

Figure 34. Vowel modifier letter: open-o (Brink et al 1998, p. 99).

Figure 35. Vowel modifier letter: upsilon (Cuttenden 2001, p. 133).
Figure 36. Vowel modifier letters: open-e, schwa, i, i-bar, o, u, turned-v, (Wayland and Allard 2001, p. 76).

listed in the earlier table, the initial voiced stops are unexploded. These are not sequences of the form (d t) but are simply homorganic pairs of stops, with the first member being voiced and unreleased, and the second being voiceless and, on some occasions, also ejective and affricated.

Figure 37. Vowel modifier letter: schwa (Ladefoged and Maddieson 1996, p. 80).

Figure 38. Vowel modifier letter: small capital i-bar (Bailey 1985, p. xxv).
The vowel modifier letters in the following table are those that would be needed to provide complete coverage for IPA and Americanist vowel symbols as shown in Figure 26 and Figure 27 but for which attestation has not been found.
Table 3. Vowel modifier letters not currently proposed for encoding

\[\oplus\] MODIFIER LETTER SMALL AE
\[\ominus\] MODIFIER LETTER SMALL REVERSED E
\[\ominus\] MODIFIER LETTER SMALL CLOSED REVERSED OPEN E
\[\upnu\] MODIFIER LETTER SMALL RAMS HORN
\[\ominus\] MODIFIER LETTER SMALL O WITH STROKE
\[\upalpha\] MODIFIER LETTER SMALL LIGATURE OE
\[\upalpha\] MODIFIER LETTER SMALL CAPITAL OE
\[\upnu\] MODIFIER LETTER SMALL CAPITAL U WITH STROKE
\[\upnu\] MODIFIER LETTER SMALL CAPITAL Y

E.4.2 Consonant modifier letters

Consonant modifier letters are often used to transcribe articulatory modifications that may apply to a wide variety of consonantal sounds, such as aspiration (typically transcribed as \( [ʰ]\)) or labialization (typically transcribed as \( [ʷ]\)). Consonantal modifier letters can also be used to transcribe sounds that involve a secondary consonantal articulation in addition to the dominant consonant, either simultaneously or as a transitional effect, such as a lateral release (typically transcribed as \( [ˡ]\)).

The most commonly-used consonant modifier letters are already encoded in the UCS. Several others are also in use, however. The inventory that seems to be needed includes nasals (e.g. to transcribe nasal onset or release of oral stops), fricatives (for fricative release of stops), approximants and some stops. Modifier counterparts for other symbols, such as clicks and trills, are not required. The samples shown below demonstrate attestation of most of the proposed inventory. The proposed consonant modifiers are listed along with an index to the samples illustrating each one in Table 4 to Table 7.

Note that a modifier counterpart to small c is proposed. The small letter c is used to represent a palatal stop. In fact, the modifier that is attested (see Figure 48) is c-cedilla, which represents a palatal fricative. It is assumed that that a voiceless affricate with a secondary palatal fricative component can be represented using a sequence < modifier letter small c, combining cedilla >. This requires, though, that the modifier letter small c be encoded.

Note also that modifier letters l-palatal hook and t-palatal hook are proposed. While the use of palatal hook for indicating palatalization is no longer an IPA recommendation, l-palatal hook and t-palatal hook are proposed here because they are attested, as seen in the samples.
### Table 4. Proposed nasal consonant modifier letters and figures that illustrate them

<table>
<thead>
<tr>
<th>Character</th>
<th>Description</th>
<th>Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1DAC</td>
<td>MODIFIER LETTER SMALL M WITH HOOK</td>
<td>Figure 45</td>
</tr>
<tr>
<td>1DAE</td>
<td>MODIFIER LETTER SMALL N WITH LEFT HOOK</td>
<td>Figure 42, Figure 43, Figure 44, Figure 46, Figure 47</td>
</tr>
<tr>
<td>1DAF</td>
<td>MODIFIER LETTER SMALL N WITH RETROFLEX HOOK</td>
<td>Figure 43, Figure 44</td>
</tr>
<tr>
<td>1DB0</td>
<td>MODIFIER LETTER SMALL CAPITAL N</td>
<td>Figure 43, Figure 44</td>
</tr>
</tbody>
</table>

The only IPA fricative symbol for which attestation of a corresponding modifier letter was not found is small h with stroke “h”.

### Table 5. Proposed fricative consonant modifier letters and figures that illustrate them

<table>
<thead>
<tr>
<th>Character</th>
<th>Description</th>
<th>Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1D9C</td>
<td>MODIFIER LETTER SMALL C (base for c-cedilla)</td>
<td>Figure 48, Figure 52</td>
</tr>
<tr>
<td>1D9D</td>
<td>MODIFIER LETTER SMALL C WITH CURL</td>
<td>Figure 53, Figure 57</td>
</tr>
<tr>
<td>1D9E</td>
<td>MODIFIER LETTER SMALL ETH</td>
<td>Figure 49</td>
</tr>
<tr>
<td>1DA0</td>
<td>MODIFIER LETTER SMALL F</td>
<td>Figure 48, Figure 50, Figure 58</td>
</tr>
<tr>
<td>1DA8</td>
<td>MODIFIER LETTER SMALL J WITH CROSSED-TAIL</td>
<td>Figure 49</td>
</tr>
<tr>
<td>1DB2</td>
<td>MODIFIER LETTER SMALL PHI</td>
<td>Figure 48</td>
</tr>
<tr>
<td>1DB3</td>
<td>MODIFIER LETTER SMALL S WITH HOOK</td>
<td>Figure 42, Figure 48</td>
</tr>
<tr>
<td>1DB4</td>
<td>MODIFIER LETTER SMALL ESH</td>
<td>Figure 48, Figure 51, Figure 52, Figure 57, Figure 58</td>
</tr>
<tr>
<td>1DBB</td>
<td>MODIFIER LETTER SMALL Z</td>
<td>Figure 49, Figure 51, Figure 57, Figure 58</td>
</tr>
<tr>
<td>1DBC</td>
<td>MODIFIER LETTER SMALL Z WITH RETROFLEX HOOK</td>
<td>Figure 42, Figure 49</td>
</tr>
<tr>
<td>1DBD</td>
<td>MODIFIER LETTER SMALL Z WITH CURL</td>
<td>Figure 42, Figure 53, Figure 57</td>
</tr>
<tr>
<td>1DBE</td>
<td>MODIFIER LETTER SMALL EZH</td>
<td>Figure 49, Figure 52, Figure 58</td>
</tr>
<tr>
<td>1DBF</td>
<td>MODIFIER LETTER SMALL THETA</td>
<td>Figure 48, Figure 51, Figure 54, Figure 55, Figure 56</td>
</tr>
</tbody>
</table>
The only IPA approximant symbol for which attestation of a corresponding modifier letter was not found is small turned y “\u200c”.

A small glottal stop modifier “\u200c” (the mirror counterpart to U+02E4 MODIFIER LETTER SMALL REVERSED GLOTTAL STOP) is also attested. Potentially, however, the modifier small glottal stop can be unified with U+02C0 MODIFIER LETTER GLOTTAL STOP. A new character is not proposed at this time.

In the samples below, modifiers that are already encoded will be highlighted in blue, while those being proposed will be highlighted in red.

![Figure 42. Consonant modifier letters: s-hook, z-curl, z-retroflex hook, n-lefthook, small capital L (Laver 1994, p. 559).]

![Figure 43. Consonant modifier letters: n-left hook, n-retroflex hook, small capital n (Laver 1994, p. 583).]
Figure 44. Consonant modifier letters: n-left hook, n-retroflex hook, small capital n (Laver 1994, p. 584).

sequence (ibid., p. 16). Sequences spanning a morpheme boundary furnish many examples of two-phoneme sequences. The difference is shown by the words *ki.tu* “circumcised boy” and *kar.kan-.fi* “kitehawk-erg.”

Figure 45. Consonant modifier letters: m-hook, n-left hook (Burquist 2001, p. 118).

Figure 46. Consonant modifier letter: n with left hook (Evans 1995, p. 732).

Figure 47. Consonant modifier letter: n with left hook (Pigott 1997, p. 469).

Figure 48. Consonant modifier letters: phi, f, theta, esh, s-hook, c-cedilla (Laver 1994, p. 581).

Figure 49. Consonant modifier letters: eth, z, ezh, z-retroflex hook, crossed-tail j (Laver 1994, p. 581).
Figure 50. Consonant modifier letter: f (Kraehenmann 2001, p. 139).

Figure 51. Consonant modifier letters: theta, esh, z (Laver 1994, p. 559).

Affrication is transcribed in any of three ways. One way is to write a small superscript homorganic fricative symbol after the stop symbol, as in dι. The alternative modes of transcription involve writing both components on the line and either joining them with a linker diacritic, as in [ts], or joining the two symbols physically together, as in [ts].

Figure 52. Consonant modifier letters: esh, ezh, c-cedilla (Laver 1994, p. 364).

otherwise, however, the conventional way of transcribing preplaced nasals, i.e. a nasal preceded by its homorganic stop [bm, dŋ, jŋ], has been retained in the present material.

Second, as also acknowledged by Bishop (1996:235) for Kensiu, the preplaced nasals are historically and cognitively developments from simple nasals and have simple nasal reflexes in other Mon-Khmer languages. Importantly, reduplications of preplaced nasals are always realised as the simple nasal counterpart: [sɔmsʰn] / sɔsmi ‘to buzz around a nest’, [hɔnθŋ] / hɔnθ / ‘to devour’, [jɪIDADE] / jɪjɛn / ‘to dream’, [rɪŋŋ] / ɪŋŋ / ‘wide’. Furthermore, Malay loanwords which originally have final nasals are usually realised with the preplaced counterpart: [buaŋhɔŋ] from Malay aŋ ‘poultry’, [bulaŋhɔŋ] from Malay bula ‘moon’, [kuaŋhɔŋ] from Malay kucing ‘cat’.

Figure 53. Consonant modifier letters: c-curl, script g, dotless j-stroke, z-curl (Burenhult 2001, p. 35).

25. panasaut ?o ęg? ęal?ews scekton, sprinkle self with that its scale the salmon

26. ni? xwəstal?iŋ?ay? , ste ?aw?niis they were stuck on that be like The little things would stick on, just like a garment.

Figure 54. Consonant modifier letter: theta (Hukari et al, p. 43).

D-Curly-tail-Z Voiced alveolo-palatal affricate
ligature Superseded by 104+183

Superscript theta Voiceless dental fricative release
Superscript schwa Mid central vowel release

Figure 55. Consonant modifier letter: theta (IPA 1999, p. 179).
and Polish. As shown in (3), the stem-final consonants /t d/ in Polish (Rubach 1984) are affricated into the alveolo-palatales [t* d*] when followed across a morpheme boundary by the locative singular /e/, verbalising /e i/ or the feminine suffix /i* a/, by virtue of Coronal Palatalisation.

(3) **Polish Coronal Palatalisation** (Rubach 1984)

<table>
<thead>
<tr>
<th>nom sg</th>
<th>loc sg</th>
<th>verbalising</th>
<th>fem</th>
</tr>
</thead>
<tbody>
<tr>
<td>brat ‘brother’</td>
<td>loc sg /e/</td>
<td>brac + ie</td>
<td>fem /i* + a/</td>
</tr>
<tr>
<td>cud ‘miracle’</td>
<td></td>
<td>cudz + ie</td>
<td></td>
</tr>
<tr>
<td>lot ‘flight’</td>
<td>verbalising /e/</td>
<td>lec + ie + ć</td>
<td></td>
</tr>
<tr>
<td>brud ‘dirt’</td>
<td>verbalising /i/</td>
<td>brudz + i + ć</td>
<td></td>
</tr>
<tr>
<td>kot ‘cat’</td>
<td>fem /i* + a/</td>
<td>koc + ic + a</td>
<td></td>
</tr>
</tbody>
</table>

Another source of sibilant affricates is Strident Assimilation, as in Polish (Rubach 1994). The anterior obstruents /t d/ in Polish are optionally affricated before sibilant fricatives or affricates within a lexical item or across word boundaries, as shown in (4).

(4) **Polish Strident Assimilation** (Rubach 1994)

<table>
<thead>
<tr>
<th></th>
<th>[t*s]</th>
<th>[t*]</th>
<th>[d*g]</th>
<th>[d*]</th>
</tr>
</thead>
<tbody>
<tr>
<td>odsoboty</td>
<td>‘since Saturday’</td>
<td>‘harder’</td>
<td>‘distinguish’</td>
<td>‘budget’</td>
</tr>
<tr>
<td>twardszy</td>
<td></td>
<td></td>
<td>‘distinguish’</td>
<td>‘budget’</td>
</tr>
<tr>
<td>odcedzic</td>
<td></td>
<td></td>
<td>‘distinguish’</td>
<td>‘budget’</td>
</tr>
<tr>
<td>swiadczyc</td>
<td></td>
<td></td>
<td>‘distinguish’</td>
<td>‘budget’</td>
</tr>
<tr>
<td>odznaczyc</td>
<td></td>
<td></td>
<td>‘distinguish’</td>
<td>‘budget’</td>
</tr>
<tr>
<td>budzet</td>
<td></td>
<td></td>
<td>‘distinguish’</td>
<td>‘budget’</td>
</tr>
<tr>
<td>przedzwonkiem</td>
<td></td>
<td></td>
<td>‘distinguish’</td>
<td>‘budget’</td>
</tr>
<tr>
<td>oddzielici</td>
<td></td>
<td></td>
<td>‘distinguish’</td>
<td>‘budget’</td>
</tr>
</tbody>
</table>

Figure 57. Consonant modifier letters: c-curl, esh, z, z-curl (Kim 2001, p. 93).
Similarly, there is a sound change from Proto-Bantu to Mvumbu which also shows plosive assimilation before the high vowels /i/ and /u/. As shown in (14a), the plosives /b d t g k/ in Proto-Bantu were affricated in Mvumbu, to /d' t' g' k'/ before /i/ or to /b' p' m' n'/ before /u/. But plosives before non-high vowels in Proto-Bantu were not affricated in Mvumbu, as in (14b) (from Ohala 1983, after Guthrie 1967–71).

(14)  

<table>
<thead>
<tr>
<th>Proto-Bantu</th>
<th>Mvumbu</th>
</tr>
</thead>
<tbody>
<tr>
<td>*-buma</td>
<td>b'umo</td>
</tr>
<tr>
<td>*-dib-</td>
<td>d'wo</td>
</tr>
<tr>
<td>*-dut</td>
<td>b'ure</td>
</tr>
<tr>
<td>*-tiito</td>
<td>t'ir</td>
</tr>
<tr>
<td>*-tud-</td>
<td>p'ule</td>
</tr>
<tr>
<td>*-gida</td>
<td>m-a'ic</td>
</tr>
<tr>
<td>*-gubo</td>
<td>m-b'u'u</td>
</tr>
<tr>
<td>*-kingo</td>
<td>t'un</td>
</tr>
<tr>
<td>*-kuba</td>
<td>p'owo</td>
</tr>
<tr>
<td>b. *-bod</td>
<td>-buo</td>
</tr>
<tr>
<td>*-di</td>
<td>-di</td>
</tr>
<tr>
<td>*-toog</td>
<td>-tuog</td>
</tr>
<tr>
<td>*-gada</td>
<td>-kala</td>
</tr>
<tr>
<td>*-konde</td>
<td>-kwande</td>
</tr>
</tbody>
</table>

In contrast, the underlying plosives /t d/ in Quebec French are usually affricated into [t' d'] only before high front vowels. As shown in (15), the consonants /t d/ are affricated before the high front vowel /i/, the high front rounded vowel /y/, the palatal glide /j/ or the high front rounded glide /u/ within a morpheme (Charbonneau & Jacques 1972, Cedergren et al. 1991, Ostiguy & Tousignant 1993, Papen 1998).

(15)  

<table>
<thead>
<tr>
<th>Standard French</th>
<th>Quebec French</th>
</tr>
</thead>
<tbody>
<tr>
<td>pe[t'i:t]</td>
<td>pe[t'i:te]</td>
</tr>
<tr>
<td>[ti]pe</td>
<td>[ti:pe]</td>
</tr>
<tr>
<td>[tj]ens</td>
<td>[tj]ens</td>
</tr>
<tr>
<td>[tu]Jer</td>
<td>[tu]Jer</td>
</tr>
<tr>
<td>[dj]x</td>
<td>[dj]x</td>
</tr>
<tr>
<td>[di:]rc</td>
<td>[di:]rc</td>
</tr>
<tr>
<td>[dj]eu</td>
<td>[dj]eu</td>
</tr>
<tr>
<td>[dy:]rer</td>
<td>[dy:]rer</td>
</tr>
<tr>
<td>[ty] viens le matin</td>
<td>[ty] viens le matin</td>
</tr>
<tr>
<td>il est plain[t][f]</td>
<td>il est plain[t][f]</td>
</tr>
</tbody>
</table>

Figure 58. Consonant modifier letters: f, esh, z, ezh (Kim 2001, p. 91).

**Alveolo-palatal affricates in Kurdish (Suleimaniya accent)**

[te] ‘where’  
[je] ‘ear’

Figure 59. Consonant modifier letter: turned h (Laver 1994, p. 365).
Now, if you add voicing, you can pronounce “Obama Obama.” Do you remember the Igbo people of Western Nigeria (Biafra)? Their tribal name was usually spelled Ibo in American newspapers since outsiders seldom correctly pronounce the double consonant. You can pronounce it correctly if you say “Ibo” — be careful not to just say Ibo, or Ig-bo.

Figure 60. Consonant modifier letter: script g (Brewster and Brewster 1976, p. 275).

Labiodentization, which can be marked with a superscript [ç], is quite common as an extralinguistic idiosyncrasy of particular individuals. In English, it is sometimes heard as a segmental feature modifying [s] and [z], and is not uncommon as a modification of [t].

Figure 61. Consonant modifier letter: v-hook (Laver 1994, p. 323).

Most languages of the Iwaidjan family have a series of complex segments that have been described as “lateral flaps” (Pym and Larrimore 1979) or “prelateralized stops” (Handelmann 1991). In all four languages apico-alveolar and apico-postalveolar complex segments /lt/ and /lt/ exist; fuller investigation of these languages may reveal palatal /t/ as well. The complex segments contrast with simple laterals /l/ and /l/, and with true clusters /lt/, /lt/ which span two syllables. Prelateralized stops pattern phonologically like single phonemes. Unlike clear clusters, they can be syllabic- and word-initial, as in Amurak /tank/ “dingo” and /awat/ “water”, and in slow syllabifications

Figure 62. Consonant modifier letters: l-retroflex hook, l-palatal hook (Evans 1995, p. 735).

Note that there is a typographic anomaly in the sample shown in Figure 62: retroflex (right-turning) hooks have been used on the t and modifier l, but the author was clearly discussing palatalization. What the author was intending, then, was a modifier l-palatal hook. It is not clear whether this was merely a typographic error or an attempt to approximate the palatal hook to compensate for an incomplete selection of type; it is clear, though, that the appropriate character to encode in this case is modifier l-palatal hook.

Figure 63. Consonant modifier letter: t-palatal hook (Halle 1971, p. 71).

Figure 64. Consonant modifier letters: turned-m with long leg (Golston and Kehrein 1998, p. 323).

E.5 Other phonetic symbols

E.5.1 LATIN SMALL LETTER C WITH STROKE
The character LATIN SMALL LETTER C WITH STROKE is often used to represent a voiceless alveolar affricate, particularly by Americanist linguists.
Note that this character has similar appearance to one of the glyph variants of U+00A2 CENT SIGN. That character has other glyph variants, however, such as “¢”, that are not acceptable for phonetic transcription. Moreover, phonetic symbols often are adopted for orthographic uses, potentially along with a case pair. The character properties of U+00A2 (e.g. General Category Sc) are not appropriate for phonetic characters, given that potential for orthographic use. For these reasons, unification with U+00A2 is not recommended; a distinct character is preferable.

Also, question 9 of section C above asks whether any of the proposed characters can be encoded as a character sequence. LATIN SMALL LETTER C WITH STROKE might be conceived as being represented as a sequence involving the overlay character U+0338 COMBINING LONG SOLIDUS OVERLAY. It is suggested that this would be inappropriate, however, and that encoding using an atomic character is very much to be preferred.

Apart from certain mathematical operators that decompose into sequences using this overlay character, there is a clear precedent for Latin characters not to represent characters such as LATIN SMALL LETTER C WITH STROKE using sequences involving U+0338: there are several Latin characters with stroke encoded in the UCS, but none of them has a decomposition involving U+0338:
Therefore, insofar as existing characters with overlaid stroke are not considered presentation forms of existing sequences, LATIN SMALL LETTER C WITH STROKE likewise should not be considered a presentation form of some existing sequence.

E.5.2 The characters LATIN SMALL LETTER DB DIGRAPH and LATIN SMALL LETTER QP DIGRAPH

These characters are used to represent labiodental stops, which are known to occur in some Bantu languages. These characters have been used primarily by Africanists in language descriptions, but are also attested in general works on phonetics and phonology.

Figure 68. From Doke (1950), p. 17.

21. The languages of this zone are notable for the different voiced labial sounds that occur in them. In MANDA (11) there is a labio-dental semi-vowel, e.g. -ilik- ‘put’, where the first consonant appears to be distinct from -w-. In TODGA (15) there is a labio-dental plosive which is distinct from the bilabial plosive, e.g. -bar- ‘shine’, -bar- ‘give birth to’. In POKA (21b) there is a ‘v’ without friction, which is

Figure 69. From Guthrie (1967), p. 61.

Figure 70. From Ladefoged and Maddieson (1996), p. 18.
E.5.3 The character LATIN SMALL LETTER IOTA WITH STROKE

This character is used by Slavic linguists in descriptions of Russian:

15.9 Before unstressed the velarization effect is not very strong and may conveniently be ignored for present purposes. It should be noted that the word был, quoted as an example above, is very often unstressed and pronounced бул.

Figure 71. From Jones and Ward (1969), p. 81.

The following symbols from the IPA are used in the Introduction for the phonetic transcription of Russian words.

Vowels

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>IPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>as in ил</td>
<td>[il]</td>
</tr>
<tr>
<td>i</td>
<td>as in пыл</td>
<td>[il]</td>
</tr>
<tr>
<td>t</td>
<td>as the first vowel in игла</td>
<td>[t'gla]</td>
</tr>
<tr>
<td>ɛ</td>
<td>as the first vowel in дыра</td>
<td>[d'ra]</td>
</tr>
</tbody>
</table>

Figure 73. From Wade (2000), p. 2.5

E.5.4 The characters LATIN SMALL CAPITAL LETTER I WITH STROKE, LATIN SMALL CAPITAL LETTER U WITH STROKE and LATIN SMALL LETTER UPSILON WITH STROKE

The characters LATIN SMALL CAPITAL LETTER I WITH STROKE and LATIN SMALL CAPITAL LETTER U WITH STROKE are used by some Americanist linguists to represent central lower-high vocoids:

---

5 In this sample, the region of intersection where the stroke crosses the stem of the iota appears blank. This is a font / rendering error and should be ignored.
The barred small capital I is also used in some recent Oxford dictionaries (though with a different meaning), as is the barred upsilon:

In addition to these transcriptions of recent developments in RP, the two composite symbols, [ɪ] and [ʊ], are used to represent [ɪ] or [a] and [ʊ] or [a] respectively (see p. x above and the discussion of vowel reduction below, p. xvii). The fol-

**beautiful**

BR 'bjʊtɪf(ə)l  
AM 'bjudəf(ə)l

**beautifully**

BR 'bjʊtɪf(ə)lɪ,  
'bjʊtɪflɪ  
AM 'bjudəf(ə)lɪ

**E.5.5 LATIN SMALL LETTER P WITH STROKE**

In the Americanist tradition, barred stop symbols are often used to represent fricatives, with barred-p representing a voiceless bilabial fricative.
The COMBINING SNAKE BELOW is used by some in the Americanist tradition to indicate lenis (weak) articulation.

\[ /p/ \text{ (fortis) is produced with tight lip closure and no aspiration.} \]
\[ /p'\text{ (lenis) is produced with a slightly looser lip closure and is released with a minimum of aspiration, yet it is much weaker than fully aspirated } /p'/. \]

Figure 83. From Floyd (1981), p. 117.
F. References


Kraehenmann, Astrid. 2001. “Swiss German stops; geminates all over the word.” *Phonology* 18.109–45.


