A. Administrative

1. Title: Proposal to Encode Additional Latin Phonetic and Orthographic Characters

2. Requester's name: Lorna A. Priest, Peter G. Constable

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


5. Requester's reference (if applicable): L2/05-097R

6. Choose one of the following:
   This is a complete proposal: Yes
   or, More information will be provided later: No

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters): No
   b. The proposal is for addition of character(s) to an existing block: Yes

2. Number of characters in proposal: 12

3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary   x   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): 1
   Is a rationale provided for the choice? Yes
   If Yes, reference: ___________________________

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:

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, as are lower case representations (see section D)

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

1. Has this proposal for addition of character(s) been submitted before? Yes
   If YES explain The original proposal was split (L2/05-097). Characters which were controversial have been removed and may be proposed at a later time. One character (LATIN SMALL LETTER V WITH FISHHOOK) has been added.

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.)? Yes
   If YES, with whom? linguists
   If YES, available relevant documents: Email correspondence. See also References.

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Yes
   Reference: See comments in Section E

4. The context of use for the proposed characters (type of use; common or rare) Common
   Reference: Orthographic characters are used in literacy materials, liturgical books and general literature. Others are used in linguistic writing.

5. Are the proposed characters in current use by the user community? Yes
   Reference: See comments in Section E

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? Preferably
   If YES, is a rationale provided? If possible, should be kept with other related blocks in the BMP.

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

8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? No
   If YES, is a rationale for its inclusion provided?

9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? No
   If YES, is a rationale for its inclusion provided?

10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? No
    If YES, is a rationale for its inclusion provided?

11. Does the proposal include use of combining characters and/or use of composite sequences? No
    If YES, is a rationale for such use provided?
    If YES, reference:
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
    If YES, reference:

12. Does the proposal contain characters with any special properties such as control function or similar semantics? No
    If YES, describe in detail (include attachment if necessary)

13. Does the proposal contain any Ideographic compatibility character(s)? No
    If YES, is the equivalent corresponding unified ideographic character(s) identified?
    If YES, reference:

D. Proposed Characters

A code chart and list of character names are shown on a new page.
## D.1. Proposed Characters

<table>
<thead>
<tr>
<th>Character</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>xx00</td>
<td>0</td>
</tr>
<tr>
<td>xx01</td>
<td>1</td>
</tr>
<tr>
<td>xx02</td>
<td>2</td>
</tr>
<tr>
<td>xx03</td>
<td>3</td>
</tr>
<tr>
<td>xx04</td>
<td>4</td>
</tr>
<tr>
<td>xx05</td>
<td>5</td>
</tr>
<tr>
<td>xx06</td>
<td>6</td>
</tr>
<tr>
<td>xx07</td>
<td>7</td>
</tr>
<tr>
<td>xx08</td>
<td>8</td>
</tr>
<tr>
<td>xx09</td>
<td>9</td>
</tr>
<tr>
<td>xx0A</td>
<td>A</td>
</tr>
<tr>
<td>xx0B</td>
<td>B</td>
</tr>
<tr>
<td>xx0C</td>
<td>C</td>
</tr>
<tr>
<td>xx0D</td>
<td>D</td>
</tr>
<tr>
<td>xx0E</td>
<td>E</td>
</tr>
<tr>
<td>xx0F</td>
<td>F</td>
</tr>
</tbody>
</table>

## D.1. Character Names

<table>
<thead>
<tr>
<th>Character</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>xx00</td>
<td>MODIFIER LETTER RAISED UP ARROW</td>
</tr>
<tr>
<td>xx01</td>
<td>MODIFIER LETTER RAISED DOWN ARROW</td>
</tr>
<tr>
<td>xx02</td>
<td>MODIFIER LETTER RAISED EXCLAMATION MARK</td>
</tr>
<tr>
<td>xx03</td>
<td>MODIFIER LETTER RAISED INVERTED EXCLAMATION MARK</td>
</tr>
<tr>
<td>xx04</td>
<td>MODIFIER LETTER LOW INVERTED EXCLAMATION MARK</td>
</tr>
<tr>
<td>xx06</td>
<td>LATIN CAPITAL LETTER ALPHA</td>
</tr>
<tr>
<td></td>
<td>• lowercase is U+0251 ɑ LATIN SMALL LETTER ALPHA</td>
</tr>
<tr>
<td>xx07</td>
<td>LATIN CAPITAL LETTER M WITH HOOK</td>
</tr>
<tr>
<td></td>
<td>• lowercase is U+0271 ɱ LATIN SMALL LETTER M WITH HOOK</td>
</tr>
<tr>
<td>xx08</td>
<td>LATIN LETTER TRESILLO</td>
</tr>
<tr>
<td>xx09</td>
<td>LATIN LETTER CUATRILLO</td>
</tr>
<tr>
<td>xx0A</td>
<td>LATIN SMALL LETTER V WITH FISHHOOK</td>
</tr>
<tr>
<td>xx0B</td>
<td>LATIN CAPITAL LETTER W WITH HOOK</td>
</tr>
<tr>
<td>xx0C</td>
<td>LATIN SMALL LETTER W WITH HOOK</td>
</tr>
</tbody>
</table>
D.3. Unicode Character Properties

xx00..xx04 should have a general category of Lm. Other properties should match those of similar characters, such as U+02BC MODIFIER LETTER APOSTROPHE.

xx06, xx07, and xx0B should have a general category of Lu. Other properties should match those of similar characters, such as U+0041 LATIN CAPITAL LETTER A.

xx08 and xx09 should have a general category of Lo. Other properties should match those of similar characters, such as U+01C0 LATIN LETTER DENTAL CLICK.

xx0A and xx0C should have a general category of Ll. Other properties should match those of similar characters, such as U+0061 LATIN SMALL LETTER A.

E. Other Information

E.1 Modifier Letters

IPA includes two symbols to indicate tonal downstep (see Figure 1, Figure 3 and Figure 4) and tonal upstep (see Figure 1 and Figure 5). These are IPA symbols 517 and 518 respectively.

The IPA Handbook cites these, and indicates that they can be represented in Unicode using the characters U+2193 DOWNWARDS ARROW and U+2191 UPWARDS ARROW:

![Figure 1](image1.png)

However, the UCS characters U+2193 and U+2191 are full-height arrows, whereas the IPA symbols for downstep and upstep are raised, half-height arrows. IPA has full-height arrows as distinct symbols, used to represent ingressive versus egressive airflow in disordered speech; these are IPA symbols 661 and 662. These are also cited in the IPA Handbook, as shown in Figure 2. Note that in the IPA set these full arrows for indicating airflow have distinct names and numeric identifiers from the raised, half-height arrows used for downstep and upstep.

![Figure 2](image2.png)

Accordingly, we consider it necessary that separate characters for the two pairs of arrow be encoded in the UCS. The existing characters, U+2193 and U+2191, are appropriate characters for representing IPA symbols 661 and 662. (Note that Unicode includes annotations for these characters indicating their use in IPA to indicate egressive and ingressive airflow.) Thus, we propose the addition of two new characters to represent the raised, half-height arrows used in IPA for tonal downstep and upstep: MODIFIER LETTER RAISED UP ARROW and MODIFIER LETTER RAISED DOWN ARROW.

The following additional samples illustrate these modifier letters in attested usage:
Africanist linguists have their own preferred conventions for indicating upstep and downstep, which are different from the IPA-recommended symbols. Tonal downstep is indicated by Africanists using a superscript exclamation mark, as illustrated in Figure 6 to Figure 11. For upstep, an inverted exclamation mark is used; in some publications this is superscripted (see Figure 8), while in others it is subscripted (see Figures 6–7 and Figures 9–11). Because of this, we are proposing two different characters for upstep. Thus we are proposing to encode three characters (in the previous proposal we suggested encoding the upstep as one character): MODIFIER LETTER RAISED EXCLAMATION MARK, MODIFIER LETTER RAISED INVERTED EXCLAMATION MARK, and MODIFIER LETTER LOW INVERTED.
EXCLAMATION MARK. Although the usage of the africanist upstep is rare, there is clear evidence that both forms have been used and will likely continue to be used in the future.

I go on to explore the applicability of my total downstep treatment to two tonally well-documented languages which resemble Akan in having (at least) the three-way contrast [0 down]/[1 down]/[2 down] after a high tone, but which differ from Akan in also having a three-way contrast [-2 down]/[-1 down]/[0 down] after a low tone: Akan, of course, while it has the three-way contrast [H][0 down]/[H][1 down]/[H][2 down] after [H], has only the two-way contrast [L][0 down]/[LL][0 down] after [L]. The first of these two languages is Dschang, which, like Akan, is a downstep language by any definition, and the second E brié, which, although it has many of the characteristics of a downstep language, appears on the surface to be a discrete level tone language with three tone levels.

I suggest that the tone systems of these languages differ from that of Akan basically in that they do not have the tonal segment structure condition which in Akan disallows the [+high, +stepping] segment $h$, and that their [-2 down] is analysable as $LhH$ [L,H] (i = upstep) just

Pitch lowering can be contrastive (and is then termed downstep), as in Kenyan (Mamfe Bantu: Cameroon), where the site of pitch lowering is notated with $i$.

\begin{align*}
(28) & \quad \text{ebéy } mémwét \quad \text{"it hurts me"} \\
& \quad - \quad - \quad -
\end{align*}

Upstep (pitch raising) also exists. In Kimatumbi (Bantu: Tanzania), adjacent $H$ tones are separated by upstep (notated with $i$).

\begin{align*}
(29) & \quad \text{baṭįlyá kįndyé } \quad \text{"they ate the birds"}
\end{align*}
### Proposal to Encode Additional Latin Orthographic Characters

#### E.2 Latin Extended

LATIN CAPITAL LETTER ALPHA is a phonetic symbol in the Americanist tradition for a voiceless, low central vowel (see Figure 12).

---

**Figure 9.** MODIFIER LETTER RAISED EXCLAMATION MARK and MODIFIER LETTER LOW INVERTED EXCLAMATION MARK (Steward, 1993, p. 231)

| 17. | /àpà | á + ñdzwì/ | → | [àpà ñdzwì] | bag of leopard |
| 18. | /àpà | á + ǹkùwì’/ | → | [àpà ǹkùwì’] | bag of rooster |
| 19. | /àpà | á + mlbhù’/ | → | [àpà mlbhù’] | bag of dog |
| 20. | /àpà | á + ñtsìn/ | → | [àpà ñtsìn] | bag of thief |
| 21. | /lòsèn’ | é + ǹdzwì/ | → | [lòsèn’ ñdzwì] | tooth of leopard |
| 22. | /lòsèn’ | é + ǹkùwì’/ | → | [lòsèn’ ǹkùwì’] | tooth of rooster |
| 23. | /lòsèn’ | é + mlbhù’/ | → | [lòsèn’ mlbhù’] | tooth of dog |
| 24. | /lòsèn’ | é + ñtsìn/ | → | [lòsèn’ ñtsìn] | tooth of thief |
| 25. | /àpú’ | á + ñdzwì/ | → | [àpú’ ñdzwì] | arm of leopard |
| 26. | /àpú’ | á + ǹkùwì’/ | → | [àpú’ ǹkùwì’] | arm of rooster |
| 27. | /àpú’ | á + mlbhù’/ | → | [àpú’ mlbhù’] | arm of dog |
| 28. | /àpú’ | á + ñtsìn/ | → | [àpú’ ñtsìn] | arm of thief |
| 29. | /lòtsìn/ | é + ñdzwì/ | → | [lòtsìn ñdzwì] | feather of leopard |

**Figure 10.** MODIFIER LETTER RAISED EXCLAMATION MARK (Pulleyblank, 1986, p. 43)

**Figure 11.** MODIFIER LETTER RAISED EXCLAMATION MARK (Heine, 2000, p. 154)

---

*bàlà* ‘the balaphone’  
bàlà dòn ‘it’s a balaphone’  
bàlà tè ‘it’s not a balaphone’

*bàlà* dòn ‘it’s the balaphone’  
bàlà k ‘it’s not the balaphone’

*bàlà* ‘the porcupine’  
bàlà dòn ‘it’s a porcupine’  
bàlà tè ‘it’s not a porcupine’

*bàlà* dòn ‘it’s the porcupine’  
bàlà k ‘it’s not the porcupine’
Although Pike (1947) and Floyd (1981) do not illustrate the use of this symbol, both make general reference to the use of capital letters for transcribing voiceless vocoids:

“Voiceless vocoids may be written with capital letters, or letters extra-large in size, such as ‘A’ and ‘O’.” (Pike 1947, p. 5)

“Symbolization: Vl. vocoids are symbolized by capital letters: [A I O E ...] or by large letters: [ l æ ɔ ...].” (Floyd 1981, p. 37).

This character is also used orthographically in the Fe’efe’e language of Cameroon. Note that LATIN CAPITAL LETTER A is also used in this orthography, so it cannot be considered a glyph variant of LATIN CAPITAL LETTER A.

"Figure 12. LATIN CAPITAL LETTER ALPHA (Smalley, 1989 p. 392)."

"Figure 13. LATIN CAPITAL LETTER ALPHA (Tcheulahie, 1984 p. 50) [Fe’eFe’e language]"
LATIN CAPITAL LETTER M WITH HOOK is used in phonetic transcription by some in the Americanist tradition for transcribing a voiceless labiodental nasal stop.

Figure 14. LATIN CAPITAL LETTER M WITH HOOK (Pike, 1976, p. 70)

LATIN LETTER TRESILLO and LATIN LETTER CUATRILLO were used by Jesuit missionaries in the XVIth century to transcribe certain sounds in Mayan languages of Guatemala that do not occur in Spanish. (It is unclear whether the Jesuit missionaries were the inventors of these characters or if they had prior use elsewhere.) In addition to older linguistic descriptions by the Jesuits, these character are also found in recent Mayanist publications when citing data from Jesuit sources (see also: http://scripts.sil.org/RecentCuatrilloUse).

Besides the above, there are five sounds occurring in the Cakchiquel, Quiche and Tzutuhil, for which five special characters were invented, or rather adopted, by the early missionary Francisco de la Parra, who died in Guatemala, in 1566. They are the following:

Figure 16. LATIN LETTER TRESILLO and LATIN LETTER CUATRILLO (Brinton, 1885).
As can be seen in Figure 17, there may be additional characters that were used by the Jesuit linguists that will need to be added to the UCS. The need for the two proposed here is clear, however.

The IPA council recently approved the adoption of an IPA symbol for the labiodental flap. They called this a “right hook v.” Since LATIN SMALL LETTER V WITH HOOK already exists in Unicode (U+028B) and is also a distinct IPA character, we are proposing the name: LATIN SMALL LETTER V WITH FISHHOOK (an alternative would be LATIN SMALL LETTER V WITH RIGHT HOOK). This character has seen limited usage in publications since the sound it is used to represent is found in only a limited number of languages. In reference to Figure 18, the IPA Council suggested that “the design has a clearer hook, like that which appears in the alveolar flap symbol. The end of the hook should be pointing at least horizontally, if not actually downwards. This should help to make the symbol visually more distinct, even in handwritten notes” (Katerina Nikolaidis, Secretary of the International Phonetic Association, personal communication). Therefore the representative glyph used in this proposal (page 3) is slightly different than what is shown here and has been informally approved by John Esling of the IPA.

LATIN CAPITAL LETTER W WITH HOOK and LATIN SMALL LETTER W WITH HOOK are used in the orthographies of certain languages spoken in Burkina Faso: they are in current use in the orthography of the Puguli language, and the language committee for the Lobiri language also has plans to use this character in their orthography. The orthography guide for Lobiri is in the process of being finalized. These languages have a rare contrast between egressive and ingressive labial semivowels (“w”); hence, these characters are used in these orthographies contrastively with “W” and “w.”
Figure 19. LATIN SMALL LETTER W WITH HOOK (ANTBA, 2004, p. 2) [Puguli language]
F. References


