A. Administrative

1. **Title:** Proposal to encode two phonetic characters and two Shona characters

2. **Requester's name:** Lorna A. Priest

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

4. **Submission date:** 2 October 2007 (revised 22 October 2007)

5. **Requester's reference (if applicable):** L2/07-334R

6. **Choose one of the following:**
   - This is a complete proposal: Yes
   - 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:** 4

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):** 3 (one combining mark)
   - Is a rationale provided for the choice? No

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](http://www.unicode.org) for such information on other scripts. Also see [http://www.unicode.org/Public/UNIDATA/UCD.html](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? No  
   If YES, explain ____________

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: ____________personal communication ____________

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) Rare  
   Reference: See comments in Section E

5. Are the proposed characters in current use by the user community? No (Shona) / Otherwise Yes  
   If YES, where? 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 YES, reference: 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? ____________  
   If YES, reference: ____________

9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? Yes  
   If YES, is a rationale for its inclusion provided? ____________COMBINING ALMOST EQUAL TO BELOW could possibly be considered COMBINING TILDE BELOW + COMBINING TILDE BELOW. ____________  
   If YES, reference: Consider U+034C COMBINING ALMOST EQUAL TO ABOVE.

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.1. Proposed Characters

<table>
<thead>
<tr>
<th>1DF</th>
<th>2C7</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>p</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>≈</td>
</tr>
<tr>
<td>E</td>
<td>☪</td>
</tr>
<tr>
<td>F</td>
<td>☪</td>
</tr>
</tbody>
</table>

### D.2. Character Names

- **1DFD** COMBINING ALMOST EQUAL TO BELOW
- **2C70** LATIN CAPITAL LETTER TURNED ALPHA
  - lowercase is 0252 ɒ
- **2C7E** LATIN CAPITAL LETTER S WITH SWASH TAIL
  - lowercase is 023F _OMEGA_LOWER_
- **2C7F** LATIN CAPITAL LETTER Z WITH SWASH TAIL
  - lowercase is 0240 _ZETA_LOWER_
D.3. Unicode Character Properties

1DFD should have a general category of Mn. Other properties for this character should match those of similar characters, such as U+0323 COMBINING DOT BELOW.

2C70, 2C7E, and 2C7F should have a general category of Lu. Other properties should match those of similar characters, such as U+0041 LATIN CAPITAL LETTER A.

We are suggesting lowercase mappings for these characters. We also would like to suggest uppercase mappings for the lowercase characters, thus:

1DFD;COMBINING ALMOST EQUAL TO BELOW;Mn;220;NSM;;;;;N;;;;;
2C70;LATIN CAPITAL LETTER TURNED ALPHA;Lu;0;L;;;;;N;;;;;0252;
2C7E;LATIN CAPITAL LETTER S WITH SWASH TAIL;Lu;0;L;;;;;N;;;;;023F;
2C7F;LATIN CAPITAL LETTER Z WITH SWASH TAIL;Lu;0;L;;;;;N;;;;;0240;
0252;LATIN SMALL LETTER TURNED ALPHA;Ll;0;L;;;;;N;LATIN SMALL LETTER TURNED SCRIPT A;2C70;2C70
023F;LATIN SMALL LETTER S WITH SWASH TAIL;Ll;0;L;;;;;N;;;2C7E;2C7E
0240;LATIN SMALL LETTER Z WITH SWASH TAIL;Ll;0;L;;;;;N;;;2C7F;2C7F

E. Other Information

E.1 Combining Marks

1DFD COMBINING ALMOST EQUAL TO BELOW: Phonetic diacritic for strident (epiglottal trill) vowels (see Figure 1). This character is used phonetically in describing Khoisan languages such as !Xõõ.

COMBINING ALMOST EQUAL TO BELOW could possibly be considered COMBINING TILDE BELOW + COMBINING TILDE BELOW. However, COMBINING ALMOST EQUAL TO BELOW is one character representing strident vowels. It does not represent creaky voiced which is what tilde below represents in the IPA. COMBINING ALMOST EQUAL TO ABOVE exists in Unicode and we believe the COMBINING ALMOST EQUAL TO BELOW should be added for linguistic usage as well.

E.2 Latin Extended

LATIN CAPITAL LETTER TURNED ALPHA is a phonetic symbol in the Americanist tradition for a voiceless, low open back rounded vowel (see Figure 2). Smalley (1989, page 392) says “Voiceless vowels are indicated by capital letters on the analogy of voiceless nasals and laterals.”

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: VI. vocoids are symbolized by capital letters: [A I Õ E …] or by large letters: [ I Æ Ō …].” (Floyd 1981, p. 37).

This character is used in teaching Americanist linguistics and should be added for linguistic usage.
LATIN CAPITAL LETTER S WITH SWASH TAIL and LATIN CAPITAL LETTER Z WITH SWASH TAIL: 
The lowercase versions of these characters are used in phonetic transcription for labialized alveolar fricatives. They are not IPA-approved, but attested in Africanist literature. The lowercase versions were added to Unicode 4.1 (see L2/03-180R).

Both uppercase and lowercase characters were historically used in the Shona (ISO 639-3:SNA) orthography from 1932 to 1955 (see Figures 3-6). The Shona number between 10-11 million in Zimbabwe and spread to Botswana and Zambia. Though Shona has a modified orthography which no longer uses these characters (the current orthography uses “sv” and “zv”), we believe these characters should be encoded in order to represent historical documents. “Books in this 1932 spelling are widely current and will continue to be so for some time…” (Fortune, 1957. p. xiv.). Note that the Shona Bible example was originally printed in 1949 with a 1999 reprint, so although this orthography is no longer officially in use, books are still available using this orthography.

sungano - citenderano.
šutu - çipura - tšeda.
Taka - ponda - tsonda.
takura - senga - twara.
tambo - rwonzi - rwodzi - bote - gau.
tarira - nangisa - ringa - penya.
tendera - bvuma - bvmira.
tenduuka - tenderuka - dzoka - pinduka.
tera - peta.
tsampa - zariwa - shatirwa - shishwa.
tsungu - simbirira - camira - kudwenga.
tendiira - pindura - sandura - tende-
tudza.
tsino - dzorera - pama - rwira.
tșingadzira - pumbudzira.
tșene - shananguruwa - yera.

Figure 3. LATIN CAPITAL LETTER S WITH SWASH TAIL and LATIN SMALL LETTER S WITH SWASH TAIL (Shona Bible, 1949, 1999 Impression, p. 253 of New Testament).
Figure 4. LATIN CAPITAL LETTER Z WITH SWASH TAIL, LATIN SMALL LETTER Z WITH SWASH TAIL, and LATIN SMALL LETTER S WITH SWASH TAIL (Shona Bible, 1949, 1999 Impression, p. 376 of Old Testament).

Sh.  Vualikana va/viri va/kuru va/kaenda kumusa kwa/vo,
(lit. Girls two big they-have-gone to-village their;
i.e. The two big girls have gone to their village).

Zigano, ziguru, zvi/vi-po,
(lit. Chairs big two they-are-there;
i.e. The two big chairs are there).

Figure 5. LATIN CAPITAL LETTER Z WITH SWASH TAIL and LATIN SMALL LETTER Z WITH SWASH TAIL (Doke, 1954, p. 52).

Figure 6. LATIN SMALL LETTER Z WITH SWASH TAIL and LATIN SMALL LETTER S WITH SWASH TAIL (Doke, 1954, p. 45).

F. References


