

**ISO/IEC JTC 1/SC 2/WG 2  
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS  
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646<sup>1</sup>**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

**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

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

Proposed name of script: \_\_\_\_\_

b. The proposal is for addition of character(s) to an existing block: Yes

Name of the existing block: Combining Diacritical Marks Supplement, Latin Extended-D

2. Number of characters in proposal: 4

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 (one combining mark)

Is a rationale provided for the choice? No

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.

<sup>1</sup> Form number: N2652-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11)

**C. Technical - Justification**

1. Has this proposal for addition of character(s) been submitted before? If YES explain _____	<u>No</u>
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.)? If YES, with whom? <u>linguists</u> If YES, available relevant documents: <u>personal communication</u>	<u>Yes</u>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference: <u>See comments in Section E</u>	<u>Yes</u>
4. The context of use for the proposed characters (type of use; common or rare) Reference: <u>See comments in Section E</u>	<u>Rare</u>
5. Are the proposed characters in current use by the user community? If YES, where? Reference: <u>See comments in Section E</u>	<u>No (Shona) / Otherwise Yes</u>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference: <u>If possible, should be kept with other related blocks in the BMP</u>	<u>Preferably</u>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<u>Preferably together with other related blocks</u>
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference: _____	<u>No</u>
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference: <u>Consider U+034C COMBINING ALMOST EQUAL TO ABOVE.</u>	<u>Yes</u> <u>COMBINING ALMOST EQUAL TO BELOW could possibly be considered COMBINING TILDE BELOW + COMBINING TILDE BELOW.</u>
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? If YES, is a rationale for its inclusion provided? If YES, reference: _____	<u>No</u>
11. Does the proposal include use of combining characters and/or use of composite sequences? 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: _____	<u>No</u>
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary) _____	<u>No</u>
13. Does the proposal contain any Ideographic compatibility character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? If YES, reference: _____	<u>No</u>

### D.1. Proposed Characters

	1DF	2C7
0		Ɔ
1		
2		
3		
4		
5		
6		
7		
8		
9		
A		
B		
C		
D		
E		Ɔ
F		Ɔ

### D.2. Character Names

1DFD COMBINING ALMOST EQUAL TO BELOW

2C70 LATIN CAPITAL LETTER TURNED ALPHA

- lowercase is 0252 ɔ

2C7E LATIN CAPITAL LETTER S WITH SWASH TAIL

- lowercase is 023F Ɔ

2C7F LATIN CAPITAL LETTER Z WITH SWASH TAIL

- lowercase is 0240 Ɔ

### 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:

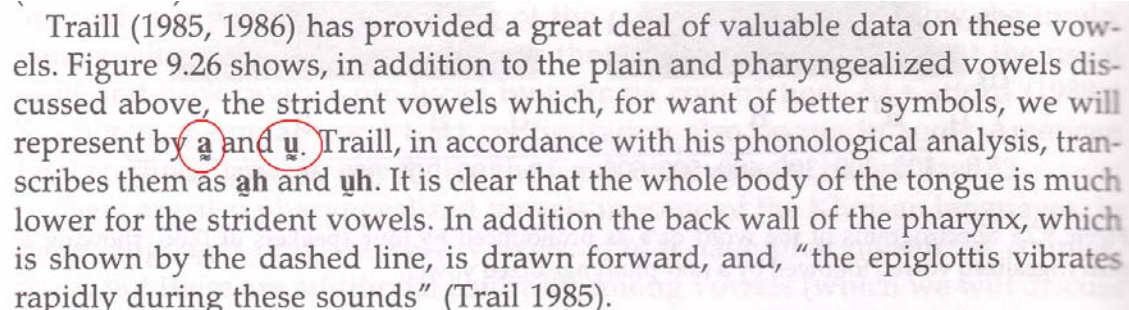
```
1DFD;COMBINING ALMOST EQUAL TO BELOW;Mn;220;NSM;;;;;N;;;;;
2C70;LATIN CAPITAL LETTER TURNED ALPHA;Lu;0;L;;;;;N;;;;;0252;
2C7E;LATIN CAPITAL LETTER S WITH SWASH TAIL;Lu;0;L;;;;;N;;;;;023F;
2C7F;LATIN CAPITAL LETTER Z WITH SWASH TAIL;Lu;0;L;;;;;N;;;;;0240;
0252;LATIN SMALL LETTER TURNED ALPHA;Ll;0;L;;;;;N;LATIN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

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.



Trail (1985, 1986) has provided a great deal of valuable data on these vowels. Figure 9.26 shows, in addition to the plain and pharyngealized vowels discussed above, the strident vowels which, for want of better symbols, we will represent by a and u. Trail, in accordance with his phonological analysis, transcribes them as ah and uh. It is clear that the whole body of the tongue is much lower for the strident vowels. In addition the back wall of the pharynx, which is shown by the dashed line, is drawn forward, and, "the epiglottis vibrates rapidly during these sounds" (Trail 1985).

Figure 1. COMBINING ALMOST EQUAL TO BELOW (Ladefoged, 1996, p. 310).

#### 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 O E ...] or by large letters: [I Æ O ...]." (Floyd 1981, p. 37).

This character is used in teaching Americanist linguistics and should be added for linguistic usage.

[ɒ] lower-low back unrounded  
 oral vowel: Sammy 264; 261ff.  
 [ɒ̃] long 269ff. [ɒ̃] nasal-  
 ized 333ff. [ɒ̃] laryngealized  
 398f. [ɒ̃] retroflexed 417ff.  
 [ɒ̃] breathy 420ff. [ɒ̃] pharyn-  
 gealized 441ff. [ɒ̃^] raised  
 370ff. [ɒ̃^] lowered 370ff.  
 [ɒ̃^] fronted 370ff. [ɒ̃^]  
 backed 370ff. [ɒ̃^] with slight  
 rounding 370ff. [ɒ̃^] voiceless  
 392ff.

Figure 2. LATIN CAPITAL LETTER TURNED ALPHA (Smalley, 1989, p. 310).

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.

Seta - şipura - pşoda.

Taka - ponda - tsonda.

takura - senğa - twara.

tambo - rwonzi - rwodzi - 6ote - gavi.

tarira - nangisa - ringa - penya.

tendera - bvuma - bvumira.

tendeovuka - tenderuka - dzoka - pinduka

tera - peta.

tsamğa - zarirwa - shatirwa - shishwa.

tsungğa - simbirira - camira - kudwenğa.

tsindikira - pindura - sandura - tende-

udza.

tsiva - dzorera - pama - rwira.

tşangadzira - pumbudzira.

tşene - shanangurwa - 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).



24 I Zino pamazuva ake Nebu-  
 kadhezari mambo weBabironi  
 akasika, Jehoyakimi akaitwa  
 muranda wake makore matatu; ipapo  
 akamuramba akamumukira. z Jehova  
 akatuma kwaari hondo dzavaKardea,  
 nehondo dzavaSiria, nehondo dzava-  
 Moabi, nehondo dzavana vaAmoni,  
 kuzorwa naye; akadzituma kuzoparadza  
 Juda, sezakarebga neshoko raJehova,  
 raakataura nemiromo yavaranda vake  
 vaprofitu. 3 Zirokwazo izi zakaitirwa

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. *Va/sikana va/viri va/kuru va/kaenda kumusha kwa/vo,*  
 (lit. Girls two big they-have-gone to-village their;  
 i.e. The two big girls have gone to their village).

*Zi/garo zi/kuru zi/viri zi/ri-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).

more frequent in Tsonga, where they are used with retroflex consonants, and labio-alveolar and prepalatal fricatives. The various types of orthography are discussed at the commencement of each chapter V–IX. In contradistinction to the use of diacritics, according to the Lepsius system, Shona uses several phonetic symbols of the system as advocated by the International Phonetic Association, viz. *b*, *d*, *ʒ*, *v*, *ɣ*. These are not all I.P.A. symbols, but have been devised on the principle of ‘one sound one symbol’.

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

## F. References

- Doke, C.M. 1954. *Handbook of African languages: The Southern Bantu Languages*. London: Oxford University Press for International African Institute.
- Floyd, Rick, comp. 1981. *A Manual for Articulatory Phonetics*. Dallas, TX: Summer Institute of Linguistics.
- Fortune, G. 1957. *Elements of Shona (Zezuru Dialect)*. London: Longmans, Green and Co., Ltd.
- Pike, Kenneth L. 1947. *Phonemics: A technique for Reducing Languages to Writing*. Ann Arbor, MI: The University of Michigan Press.
- Shona language. 1949, 1999 Impression. *Bhaibheri: muNdimi yeUnion Shona* (The Bible in Shona). British and Foreign Bible Society.
- Ladefoged, Peter and Ian Maddieson. 1996. *The Sounds of the World's Languages*. Blackwell Publishers. p. 310.
- Smalley, William A. 1989. *Manual of Articulatory Phonetics*. Revised Edition. Lanham, MD: University Press of America, Inc.