Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

1. Introduction. The Sora Sompeng script is used to write the Sora language. Sora is a member of the Munda family of languages, which, together with the Mon-Khmer languages, makes up Austro-Asiatic. The Sora people live between the Oriya- and Telugu-speaking populations in what is now the Orissa-Andhra border area. Sora Sompeng was devised on 1936-06-18 by Mangei Gomango, son-in-law of charismatic community leader Malia Gomango, on the inspiration of a vision of the 24 letters he had. The story is retold in translation in a document by Khageswar Mahapatra written in 1978–1979.

> The Savara people used to worship "Daru-Bramha" (Wooden Image of God). But when Lalitha (The Savara Princes) falling in love with a Bramhin revealed to him the secrets of the worship, the Brahmin won over the Lord. The Lord vanished and no more responded to the prayers of the Savara people. Then the Savara said, "Wonderful! You don't listen to me; you don't respond to me. But you listen to the Brahmin, who is an outsider! All right, I say, from today I will offer you only blood and liquor. When you come back again to us and teach us to be wise and good, we will worship you then only with due sanctity." The Savara people took a resolution like that and since then they were sacrificing cocks, goats, buffaloes, and human beings too at the rituals. They worship, pay obeisance, and then pour down blood and liquor on the deity's head. However, after many days, at the end of the Age of Kali, the Lord come back and said, "Now I have come to you, not as Daru-Bramha, but as Akshara-Bramha (Alphabetic Image of God). You worship me in this form. I will be visible on the *matter bnom vijnan* 'hills'. The Savara then went and saw. The twenty-four letters appeared in his vision. Then a shrine was built on that site and the worship of the Akshara-Bramha commenced from that day.

The script was promulgated as part of a comprehensive cultural program, and was offered as an improvement over IPA-based scripts used by linguists and missionaries, and the Telugu and Oriya scripts used by Hindus. Sora Sompeng is used in religious contexts, and is published in a variety of printed materials, tracts, almanacs, invitation cards, and similar ephemera. Its use is somewhat analogous to the use made of the Cherokee script.

2. Structure. Sora Sompeng is of the Brahmic type: the consonant letters contain an inherent vowel. Consonant clusters are written not with conjunct characters, and there is no visible vowel killer to show the deletion of the inherent vowel; the reader must determine the presence or absence of the inherent schwa word by word. The character repertoire does not match the phonemic repertoire of Sora very well; LETTER I is used for both [i] and [ɨ], and LETTER O is used for both [o] and [ɔ], for instance. The glottal stop is written with LETTER HAH, and the sequence of LETTER RAH and LETTER DAH is used to write retroflex [ɽ]. There is also an additional "auxiliary" LETTER MAE used to transcribe foreign sounds. Mahapatra says:

The auxiliary letter imaᴇ' is used for transcription of sounds foreign to Sora; e.g. Oriya /a, t, n, d, l, s/ are transcribed as /ox, tx, nx, dx, lx, sx/. But Oriya /s/ and aspirite stops are to be transcribed by names [*leg.* means] of adding Sora /j/ after its /s/ and non- aspirate stops.

This description is not very precise, but it is a matter of orthography, not of character identity.

3. Collating order. The letters get their names from 24 gods in the Sora pantheon, e.g. *s* for Sundaṅ, *t* for Tənod; no rationale is apparent for the relative ordering, but collation order is as in the code chart. Mahapatra gives the following names (the spellings may be slightly suspect and are under review):

⌐	saʔa	Sundaṅ
ॐ	taʔa	Tṅod
ৰ	baʔa	Babu
ৎ	caʔa	Candi
৳	daʔa	Daṅki
ॐ	gaʔa	Gada
ৎ	maʔa	Mundaxra
ৰ	ṅaʔa	Ṅal
৶	laʔa	Ləbo
ॐ	naʔa	Naa
ৰ	vaʔa	Varox
ॐ	paʔa	Patta
ॐ	yaʔa	Yuyu
ৰ	raʔa	Rṅgi
৲	haʔa	Haro
ॐ	kaʔa	Ketuṅ
৳	jaʔa	Janṅlo
ৰ	ñaʔa	Ñarana
৶	aʔa	Aṅge
৳	eʔe	Elda
৳	iʔi	Ida
ॻ	uʔu	Uyuṅ
৭	oʔo	Onal
৶	ɛʔɛ	Ɛmaṅda
৲	maɛ	Almraɛ

4. Character names. Consonant letter names are derived by adding [aʔa] (written *ah*) to the consonant. (Sora free-standing nouns must have two syllables, and this is often achieved by breaking the vowel of a word into two with a glottal stop. The glottal stop appears only between like vowels, so it's sufficient just to write *ah*, *eh*, etc.)

5. Punctuation. Sora Sompeng has no script-specific punctuation. Punctuation common to Latin, Telugu, and Oriya may be used. Mahapatra indicates that the following international characters are in use: FULL STOP, COMMA, HYPHEN, EXCLAMATION MARK, PLUS SIGN, EQUALS SIGN, MULTIPLICATION SIGN, SEMICOLON, LEFT PARENTHESIS, RIGHT PARENTHESIS.

6. Linebreaking. Letters and digits behave as in Latin and other alphabetic scripts.

7. Unicode Character Properties.

```
110D0;SORA SOMPENG LETTER SAH;Lo;0;L;;;;;N;;;;;
110D1;SORA SOMPENG LETTER TAH;Lo;0;L;;;;;N;;;;;
110D2;SORA SOMPENG LETTER BAH;Lo;0;L;;;;;N;;;;;
110D3;SORA SOMPENG LETTER CAH;Lo;0;L;;;;;N;;;;;
110D4;SORA SOMPENG LETTER DAH;Lo;0;L;;;;;N;;;;;
110D5;SORA SOMPENG LETTER GAH;Lo;0;L;;;;;N;;;;;
110D6;SORA SOMPENG LETTER MAH;Lo;0;L;;;;;N;;;;;
110D7;SORA SOMPENG LETTER NGAH;Lo;0;L;;;;;N;;;;;
110D8;SORA SOMPENG LETTER LAH;Lo;0;L;;;;;N;;;;;
110D9;SORA SOMPENG LETTER NAH;Lo;0;L;;;;;N;;;;;
110DA;SORA SOMPENG LETTER VAH;Lo;0;L;;;;;N;;;;;
110DB;SORA SOMPENG LETTER PAH;Lo;0;L;;;;;N;;;;;
110DC;SORA SOMPENG LETTER YAH;Lo;0;L;;;;;N;;;;;
110DD;SORA SOMPENG LETTER RAH;Lo;0;L;;;;;N;;;;;
110DE;SORA SOMPENG LETTER HAH;Lo;0;L;;;;;N;;;;;
110DF;SORA SOMPENG LETTER KAH;Lo;0;L;;;;;N;;;;;
110E0;SORA SOMPENG LETTER JAH;Lo;0;L;;;;;N;;;;;
110E1;SORA SOMPENG LETTER NYAH;Lo;0;L;;;;;N;;;;;
110E2;SORA SOMPENG LETTER AH;Lo;0;L;;;;;N;;;;;
110E3;SORA SOMPENG LETTER EEH;Lo;0;L;;;;;N;;;;;
110E4;SORA SOMPENG LETTER IH;Lo;0;L;;;;;N;;;;;
110E5;SORA SOMPENG LETTER UH;Lo;0;L;;;;;N;;;;;
110E6;SORA SOMPENG LETTER OH;Lo;0;L;;;;;N;;;;;
110E7;SORA SOMPENG LETTER EH;Lo;0;L;;;;;N;;;;;
110E8;SORA SOMPENG LETTER MAE;Lo;0;L;;;;;N;;;;;
110F0;SORA SOMPENG DIGIT ZERO;Nd;0;L;;0;0;0;N;;;;;
110F1;SORA SOMPENG DIGIT ONE;Nd;0;L;;1;1;1;N;;;;;
110F2;SORA SOMPENG DIGIT TWO;Nd;0;L;;2;2;2;N;;;;;
110F3;SORA SOMPENG DIGIT THREE;Nd;0;L;;3;3;3;N;;;;;
110F4;SORA SOMPENG DIGIT FOUR;Nd;0;L;;4;4;4;N;;;;;
110F5;SORA SOMPENG DIGIT FIVE;Nd;0;L;;5;5;5;N;;;;;
110F6;SORA SOMPENG DIGIT SIX;Nd;0;L;;6;6;6;N;;;;;
110F7;SORA SOMPENG DIGIT SEVEN;Nd;0;L;;7;7;7;N;;;;;
110F8;SORA SOMPENG DIGIT EIGHT;Nd;0;L;;8;8;8;N;;;;;
110F9;SORA SOMPENG DIGIT NINE;Nd;0;L;;9;9;9;N;;;;;
```

8. Bibliography.

Mahapatra, Khagehwar. 1978–1979. "'Soraɴ Sompeɴ': A Sora Script". Unpublished conference paper. Delhi, Mysore.

Zide, Norman. 1996. "Scripts for Munda languages", in Peter T. Daniels and William Bright, eds. *The world's writing systems*. New York; Oxford: Oxford University Press. ISBN 0-19-507993-0

Zide, Norman. 1999. "Three Munda scripts", in *Linguistics of the Tibeto-Burman Area*. Vol. 22.2—Fall 1999.

Figure 1. Sora Sompeng alphabet and digits. The text at the top reads ᱈᱉᱇᱃᱆᱄᱈ ᱉᱈᱊ *sompeṅ tub*.

A. Administrative

1. Title

Proposal for encoding the Sora Sompeng script in the UCS

2. Requester's name

UC Berkeley Script Encoding Initiative (Universal Scripts Project)

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

Liaison contribution.

4. Submission date

2009-06-08

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

No.

6b. More information will be provided later

Yes.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

Yes.

1b. Proposed name of script

Sora Sompeng.

1c. The proposal is for addition of character(s) to an existing block

No.

1d. Name of the existing block

2. Number of characters in proposal

35.

3. Proposed category (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)

Category A.

4a. Is a repertoire including character names provided?

Yes.

4b. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?

Yes.

4c. Are the character shapes attached in a legible form suitable for review?

Yes.

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson.

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, Fontographer.

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

7. 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.

8. 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 Unicode Character Database http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

See above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

Yes, in N1957 and N3410.

2a. 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.

2b. If YES, with whom?

David Stampe and Piers Vitebsky..

2c. If YES, available relevant documents

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

People living at the Orissa-Andhra border in India.

4a. The context of use for the proposed characters (type of use; common or rare)

Restricted to religious use.

4b. Reference

5a. Are the proposed characters in current use by the user community?

Yes.

5b. If YES, where?

In India.

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

No.

6b. If YES, is a rationale provided?

6c. If YES, reference

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

Yes.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

No.

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

No.

11b. If YES, is a rationale for such use provided?

11c. If YES, reference

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

11e. If YES, reference

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

No.

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?

	110D	110E	110F
0	110D0	110E0	110F0
1	110D1	110E1	110F1
2	110D2	110E2	110F2
3	110D3	110E3	110F3
4	110D4	110E4	110F4
5	110D5	110E5	110F5
6	110D6	110E6	110F6
7	110D7	110E7	110F7
8	110D8	110E8	110F8
9	110D9		110F9
A	110DA		
B	110DB		
C	110DC		
D	110DD		
E	110DE		
F	110DF		

Consonants

110D0   SORA SOMPENG LETTER SAH
110D1   SORA SOMPENG LETTER TAH
110D2   SORA SOMPENG LETTER BAH
110D3   SORA SOMPENG LETTER CAH
110D4   SORA SOMPENG LETTER DAH
110D5   SORA SOMPENG LETTER GAH
110D6   SORA SOMPENG LETTER MAH
110D7   SORA SOMPENG LETTER NGAH
110D8   SORA SOMPENG LETTER LAH
110D9   SORA SOMPENG LETTER NAH
110DA   SORA SOMPENG LETTER VAH
110DB   SORA SOMPENG LETTER PAH
110DC   SORA SOMPENG LETTER YAH
110DD   SORA SOMPENG LETTER RAH
110DE   SORA SOMPENG LETTER HAH
110DF   SORA SOMPENG LETTER KAH
110E0   SORA SOMPENG LETTER JAH
110E1   SORA SOMPENG LETTER NYAH

Vowels

110E2   SORA SOMPENG LETTER AH
110E3   SORA SOMPENG LETTER EEH
110E4   SORA SOMPENG LETTER IH
110E5   SORA SOMPENG LETTER UH
110E6   SORA SOMPENG LETTER OH
110E7   SORA SOMPENG LETTER EH

Other letter

110E8   SORA SOMPENG LETTER MAE

Digits

110F0   SORA SOMPENG DIGIT ZERO
110F1   SORA SOMPENG DIGIT ONE
110F2   SORA SOMPENG DIGIT TWO
110F3   SORA SOMPENG DIGIT THREE
110F4   SORA SOMPENG DIGIT FOUR
110F5   SORA SOMPENG DIGIT FIVE
110F6   SORA SOMPENG DIGIT SIX
110F7   SORA SOMPENG DIGIT SEVEN
110F8   SORA SOMPENG DIGIT EIGHT
110F9   SORA SOMPENG DIGIT NINE