

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

**Doc Type:** Working Group Document  
**Title:** Proposal to add phonetic click characters to the UCS  
**Source:** Michael Everson  
**Status:** Individual Contribution  
**Date:** 2004-06-10

This document requests a number of phonetic characters used to represent clicks and other sounds used in transcriptions of the Khoisian language  $\text{ǀxũ}$ : (Kxoe) to be added to the UCS and contains the proposal summary form.

## A. Administrative

### 1. Title

Proposal to add phonetic click characters to the UCS.

### 2. Requester's name

Michael Everson.

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

Individual contribution.

### 4. Submission date

2004-06-10

### 5. Requester's reference (if applicable)

#### 6. Choose one of the following:

##### 6a. This is a complete proposal

Yes.

##### 6b. More information will be provided later

No.

## B. Technical – General

### 1. Choose one of the following:

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

No.

#### Proposed name of script

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

Yes.

#### 1b. Name of the existing block

Latin Extended-B.

### 2. Number of characters in proposal

15

### 3. Proposed category (see section II, Character Categories)

Category B.1.

#### 4a. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000)

Level 1.

#### 4b. Is a rationale provided for the choice?

Yes.

#### 4c. If YES, reference

Spacing characters.

#### 5a. Is a repertoire including character names provided?

Yes.

#### 5b. If YES, are the names in accordance with the naming guidelines in Annex L of ISO/IEC 10646-1: 2000?

Yes.

**5c. Are the character shapes attached in a legible form suitable for review?**

Yes.

**6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?**

Michael Everson. TrueType.

**6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:**

Michael Everson. Fontographer.

**7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?**

No.

**7b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?**

Yes.

**8. 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, see below.

**9. 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.**

The characters have the same properties as other Latin phonetic characters.

## **C. Technical – Justification**

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

No.

**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.)?**

No. The characters were printed in *Bantu Studies* in 1925; the characters have historical use.

**2b. If YES, with whom?**

**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?**

Yes.

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

Used in phonetic representation of Bantu languages by linguists.

**4b. Reference**

See examples below.

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

Probably not. IPA and practical orthographies are more generally used at the present time.

**5b. If YES, where?**

See examples below.

**6a. After giving due considerations to the principles in Principles and Procedures document (a WG 2 standing document) must the proposed characters be entirely in the BMP?**

Yes.

**6b. If YES, is a rationale provided?**

Yes.

**6c. If YES, reference**

Keep with other phonetic characters.

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

Not necessarily.

**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?**

Yes.

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

Yes.

**10c. If YES, reference**

Two of the characters look like arrows, but differ in function and meaning from arrows which are used in phonetic contexts. ↓ VOICELESS ALVEOLAR CLICK and ↑ VOICED ALVEOLAR CLICK are consonant letters which differ from ↓ DOWNWARDS ARROW and ↑ UPWARDS ARROW, which mean ingressive airflow and egressive airflow respectively. ɿ MODIFIER LETTER INVERTED GLOTTAL STOP looks like ʅ LATIN LETTER INVERTED GLOTTAL STOP, but is used alongside ʔ modifier letter glottal stop and not ʔ LATIN LETTER GLOTTAL STOP in Doke's orthography. See figure 6.

**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**

**12a. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?**

No.

**12b. If YES, reference**

**13a. Does the proposal contain characters with any special properties such as control function or similar semantics?**

No.

**13b. If YES, describe in detail (include attachment if necessary)**

**14a. Does the proposal contain any Ideographic compatibility character(s)?**

No.

**14b. If YES, is the equivalent corresponding unified ideographic character(s) identified?**

**14c. If YES, reference**













## D. Proposal

Professor Clement Martyn Doke, who was Senior Lecturer in the Department of Bantu Studies, University of the Witwatersrand, Johannesburg, wrote widely on African linguistics and orthography. He published "An Outline of the Phonetics of the Language of the ǀǀ: Bushmen of North-West Kalahari" in *Bantu Studies* in December 1925. The phonetic transcription he devised and used throughout contains a number of characters which are not yet encoded. Three of these characters represent pre-palatal stops (between alveolar and palatal stops), and twelve of them represent clicks of various kinds. The repertoire requested is given here with proposed code positions:

t̪	023A LATIN SMALL LETTER PRE-PALATAL T
----	---------------------------------------

d̪	023B LATIN SMALL LETTER PRE-PALATAL D
----	---------------------------------------

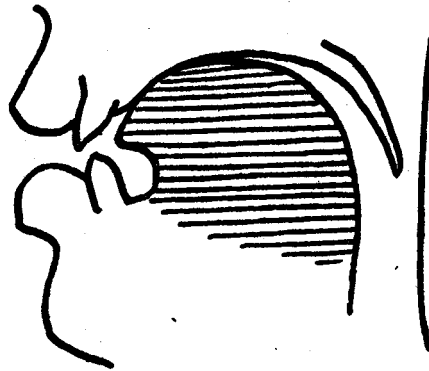
ɲ	023C LATIN SMALL LETTER PRE-PALATAL N → latin small letter n with left hook
---	--------------------------------------------------------------------------------

	023D LATIN SMALL LETTER N WITH RIGHT LOOP nasal dental click
	023E LATIN LETTER VOICELESS ALVEOLAR CLICK → 2193 downwards arrow (ingressive airflow)
	023F LATIN LETTER VOICED ALVEOLAR CLICK → 2191 upwards arrow (egressive airflow)
	0240 LATIN SMALL LETTER ENG WITH LEFT LOOP • nasal alveolar click
	0241 LATIN LETTER SCRIPT CAPITAL Q • voiced palatoalveolar click → 1D4AC mathematical script capital q
	0242 LATIN SMALL LETTER N WITH LEFT LOOP • nasal palatoalveolar click
	0243 LATIN LETTER VOICELESS RETROFLEX CLICK → 22D4 pitchfork (proper intersection)
	0244 LATIN LETTER VOICED RETROFLEX CLICK
	0245 LATIN SMALL LETTER N WITH LEFT HOOK AND RIGHT LOOP • nasal retroflex click
	0246 MODIFIER LETTER INVERTED GLOTTAL STOP → 0296 latin letter inverted glottal stop • voiceless lateral click
	0247 LATIN LETTER OWL • voiced lateral click
	0248 LATIN SMALL LETTER N WITH LEFT LOOP AND RIGHT LOOP • nasal lateral click

## 2. The Nasals

çhũ: employs four nasal emitted consonants, **m**, **n**, **p** and **ŋ**. **m**, **n**, and **ŋ** are, as far as can be ascertained, pronounced as in English, **ŋ** being the nasal sound in the English words *singing*, *ping-pong* (**sɪŋŋ**, **pɪŋpɔŋ**), or the nasal found often before **k** and **g** in English, as in *ink*, *angle* (**ɪŋk**, **æŋgɪ**).

→ **p** in çhũ: needs a little more explanation than the other sounds. It is not identical with the Romance palatal nasal, represented by the symbol **ɲ**, and found in such words as French *règne* (**rɛɲ**) or Italian *kampapa*, in the pronunciation of which sound the tongue-tip is usually kept down behind the lower teeth, and the back of the tongue brought forward to touch the hard palate near its apex. **p** in çhũ: is as **p** in Zulu, and Suto, a pre-palatal sound in which the "flat" of the tongue or middle of the tongue is brought somewhat forward to touch the division between the alveolus and the hard palate. This sound then is intermediate in position between **n** and **p**; and the tongue position is represented in the following chart :



→ It is noted that the tip of the tongue is not against the alveolus as for **n**, nor is it so far forward as to be able to rest behind the lower teeth. With the exception of the case of *maka:pa* (tobacco), all the instances I have recorded of the use of this nasal are before the pre-palatal affricates **tʃʰ** and **dʒ**; and, in these cases, **p** is lengthened to form a syllable by itself, e.g. **ptʃʰŋ** (to weep).

All the nasals in çhũ: are very commonly used with syllabic value, e.g. :

**sɲ** (to see).

**fɲ:** (liver).

**ŋ** (yes).

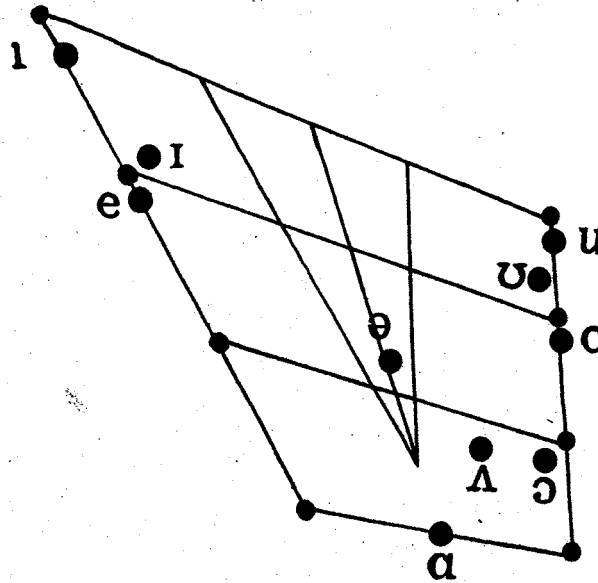
**ya:m** (jaw).

Syllabic nasals, like vowels, may form both short and long syllables, and are subject to their own individual tones.

**Figure 1.** Text describing the three çhũ: pre-palatal letters **t**, **d**, and **p**, noting the distinction between **[n]** and **[ɲ]** and **[p]**

## I. VOWELS

The following is a chart showing approximately the position of the *chũ*: vowels in comparison with the position of the eight cardinal vowels, as now used by the International Phonetic Association.<sup>1</sup>



## II. EMITTED CONSONANTS

*Chart of the Emitted Consonants of chũ:*

	Bi-Labial.	Alveolar.	Retroflex.	Pre-Palatal.	Velar.	Glottal.
Explosive .	(p) b	t th d		[t] [d]	k kh g	ʔ
Nasal . . .	m	n		ɲ	ŋ	
Fricative .	v	s z z̄s		ʃ ʒ	x	h ɦ
Flapped . .			ɭ			
Affricate .		tsʔ		tʃʔ dʒ	kxʔ	
Semi-vowel	w			ɟ		

The *chũ*: language employs twenty-four elementary emitted consonant sounds, to represent which the following symbols are used :— p, b, m, v, w, t, d, n, s, z, l, t, d, p, ʃ, ʒ, ɟ, k, g, ŋ, x, ʔ, h and ɦ.

[Note.—It is doubtful whether ɲ is a genuine *chũ*: sound, I therefore enclose it in round brackets ; t and d never being used except in affricate combination, I enclose them in square brackets.]

Figure 2. Text describing the vowels and consonants of *chũ*:

### III. REVERSED CONSONANTS OR CLICKS

čhū: is what is sometimes called a “ five-click language ”, that is to say it employs five positional types of click. Each of these five types is used in three main ways: (1) unvoiced, (2) voiced, and (3) nasal.

Chart of čhū: Clicks :

	Dental.	Alveolar.	Palato-Alveolar.	Retroflex.	Lateral.
Unvoiced .	ɿ	ɥ	ɿ	ʍ	ɸ
Voiced . .	ʎ	ɿ	ʑ	ɸ	ɸ
Nasal . . .	ɿ	ɿ	ɿ	ɿ	ɿ

General Observations :

Clicks are reversed sounds, in the production of which a somewhat complicated mechanism has to be set in operation. There are always two points of tongue articulation, between which a space of rarefaction is created, the “ smack ” of the click being made on the release of one of these points. One of these points of articulation is invariably velar, i.e. the back of the tongue is raised to touch the soft palate. The second point of articulation varies with the different types of click.

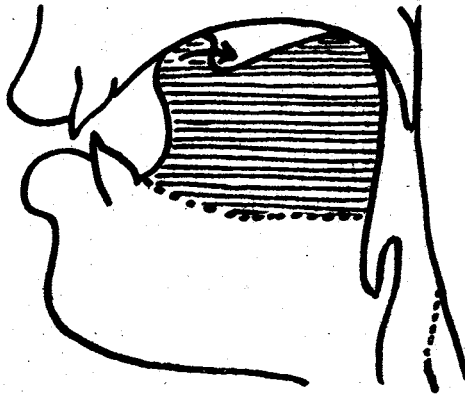
In the formation of the click the front of the tongue rises into position first, the back of the tongue rising immediately after; the front or side (in the case of the lateral click) of the tongue is the first to be released, the back of the tongue coming away from the velum perceptibly after it, thus completing the click.

In čhū: each click may be pronounced unvoiced (i.e. without concurrent vibration of the vocal chords), and these unvoiced varieties of the clicks may immediately precede the accompanying vowel, or may be aspirated (i.e. followed immediately by a rush of air causing throat friction), or they may be accompanied by the glottal stop. In this last case, during the closure of the glottis, the velar point of articulation is released silently, and no “ k-sound ” is heard. When the unvoiced click is *immediately* followed by the vowel, a “ k-sound ” (the result of the release of the back of the tongue from the velum) is distinctly heard. In čhū: an aspirate is very often found after the glottal stop accompanying a click. Unvoiced clicks also occur followed by the velar fricative and the ejective velar affricate.

The voiced clicks are pronounced with the same mechanical preparation as the unvoiced, but with the addition of concurrent vibration of the vocal chords. It is a mistake to consider these forms to be accompanied by a “ g ”. What is heard is the voiced “ g-sound ” resultant on the release of the back point of articulation. In čhū: voiced clicks are often followed by the glottal stop, but the pause necessary for this provides the insertion of the neutral vowel, and I consider it better to use this consistently, writing ɸəʔa instead of ɸʔa.

In the case of the nasal clicks, nasalization accompanies the click throughout. There is no question of any nasal consonant accompanying the unvoiced click; the whole click is nasalized. For this reason I employ separate symbols. Nasal clicks in čhū: may become syllabic, e.g. ɿ: (take).

Figure 3. Text describing the čhū: clicks and their articulation in detail. Of the 15 characters listed, only three (ɿ, ʎ, and ɿ) have been encoded.

(4) *The Retroflex Clicks.*

Tongue-tip curled back, and underside of tip pressed against hard palate ; back of tongue raised to touch velum, and sides of tongue touching upper side gums to enclose space of rarefaction. The tongue-tip is released backwards, sliding along the palate, causing a harsh-sounding click which is not "instantaneous". This click is foreign to Nama Hottentot, and has been miscalled a "lateral". The nasal form resembles the nasal click found in Suto, and represented in that language by *nǃ*. The symbols I employ are as follows : unvoiced,  $\psi$  ; voiced,  $\psi$  ; and nasal,  $\eta$ .

Examples :  $\psi\acute{a}$ : (to run away).

$\psi^?á$  (rib).

$\psi h\acute{a}$  (meat).

$\psi x\acute{u}$  (to cut).

$\acute{h}o$ : (man).

$\eta\phi$ : (calabash).

$\acute{\eta}$ : (to sigh).

$\eta^?a$  (big).

Figure 4. Text describing the  $\psi\acute{h}$ : retroflex clicks and their articulation in detail.



The following, then, is the order of the forty-nine *chũ*: letters:—

1. i	26. ʒ
2. ɪ	27. j
3. e	28. k
4. ə	29. g
5. a	30. ŋ
6. ʌ	31. x
7. ɔ	32. ʔ
8. o	33. h
9. u	34. ɦ
10. u	35. ɿ
11. p	36. ɣ
12. b	37. ŋ
13. m	38. ↓
14. v	39. ↑
15. w	40. ɔ̣
16. t	41. ɕ
17. d	42. ɖ
18. n	43. ɓ
19. s	44. ɸ
20. z	45. ʈ
21. l	46. ɺ
22. ʈ	47. ɣ
23. ɖ	48. ʌ
24. ɓ	49. ɺ
25. ɸ	

*Note.*—The tones of words in the following vocabulary, which are marked with an asterisk (\*), have not been ascertained.

*chũ*: English Vocabulary

u = to go.

atá = a person.

q: = thou.

qle:je = almost.

aw\* = to give.

q:hn̩ = thou (form used in isolation).

aw mí ku ʈũ ʃn̩: = give me  
water to drink.

owa:la = to scatter about.

aw mí ku maka:pq ɣʰr = give  
me tobacco to smoke.

oko = and.

aw mí ku ɸhd naʰm̩: = give me  
meat to eat.

okxʰw̩: = to say, to speak,  
word.

oʈeʰi = to scare away.

**Figure 5.** Text showing the repertoire of the *chũ*: letters, in the idiosyncratic sorting order (based on place of articulation) which Doke uses in his glossary. Of the click letters shown (numbers 35–49) the IPA representations would be *approximately* (!) as follows: *Dental*: ɿ [k̠], ɣ [g̠], ŋ [ŋ̠]; *Alveolar*: ↓ [k̠], ↑ [g̠], ɔ̣ [ŋ̠]; *Palatoalveolar*: ɕ [k̠], ɖ [g̠], ɓ [ŋ̠]; *Retroflex*: ʈ [k̠], ɸ [g̠], ɺ [ŋ̠]; *Lateral*: ɣ [k̠], ʌ [g̠], ɺ [ŋ̠]. IPA does not represent the retroflex clicks, which are extremely rare; Pullum and Ladusaw note that Desmond T. Cole used a triple solidus /// (“The Bushman languages”, *Encyclopaedia Britannica* 1966) so I have used the triple vertical in the transcription above. The Unicode name for !U+01C3 appears to be infelicitous though the term is often met with. The IPA gives ɔ bilabial, | dental, ! (post)alveolar, † palatoalveolar, and || alveolar lateral clicks; the *chũ*: retroflex does not seem to be represented in the IPA.

sāi = thin.	s <sup>2</sup> ha: = to fill (pipe).
sāi ɿ <sup>2</sup> h = to press down.	s <sup>2</sup> ha: maka:ɿq = to put tobacco in pipe.
sao* = ratel.	s <sup>2</sup> hā = skin cap, to cover head.
samo = to be morose.	s <sup>2</sup> hāi = to wrench, to pull.
saroq = to dress.	s <sup>2</sup> hwā: = to be green. ←
sāwa = evil.	s <sup>2</sup> hwā = shade, shadow.
sarj = a scratch.	sa <sup>2</sup> ɿ* = Countess beetle.
sarjɿ = again.	sa <sup>2</sup> ɿ = to cover with bushes.
sarjɿm = to cut firewood, to strike with hand.	sa <sup>2</sup> u = arrow haft.
soua* = upper arm.	sa <sup>2</sup> u = walking stick.
sorɿdm = to loosen, to open.	sa <sup>2</sup> o = apron of egg-shell beads.
sosoge* = reply to Bushman greeting.	sa <sup>2</sup> u = to bark (of dog).
sum = leopard (also ɿum).	ɿ: = to take, to receive.
sɿ:ɿɿ = pimple, boil, etc.	ɿ̂: = to tie knot (see also ŝɿ̂).
su:sa = to smell something.	ɿ̂a: = to let loose, to free prisoner.
su:lu = green boomslang.	ɿ̂a: ɿɿ̂ = leave me alone ! ←
ŝɿ̂ = to tie knot, to bind, to in- span.	ɿ̂aɿ = men.
ŝɿ̂ gumɿ = inspan the oxen.	ɿ̂a <sup>2</sup> * = bow.
s <sup>2</sup> a: = (1) to pain, (2) to be full (of food).	ɿ̂a:ɿ = to make fire with fire- stick.
s <sup>2</sup> d: = to like, to love.	ɿ̂a:ɿ ɿ <sup>2</sup> hulu dɿm = to bore hole.
s <sup>2</sup> āi = grass.	ɿ̂a:ɿa = now (syllable repeated for emphasis).
s <sup>2</sup> āu = gift, present.	ɿ̂a:ɿa:ɿe: = just now, presently.
s <sup>2</sup> ama = to buy.	ɿ̂a:ɿja = whirlwind.
s <sup>2</sup> āɿ <sup>2</sup> ā = walking stick.	ɿ̂a:ɿ = to roast.
s <sup>2</sup> d:ɿa = tortoise.	ɿ̂a:ɿha: = to throw.
s <sup>2</sup> hr = to smoke.	

## ADDENDA

I. A short Vocabulary comparing some words of the Otjimagare type of ɿhū: with the preceding spoken around Neitsas.

English.	Otjimagare.	Neitsas-Nurugas.
arm	ɿhā	ɿ <sup>2</sup> hā:
arrow	sāu	sa <sup>2</sup> u
bow	ɿ̂a <sup>2</sup>	ɿ̂a <sup>2</sup> ←
bow-string	ts <sup>2</sup> a <sup>2</sup> ɿ	s <sup>2</sup> ɿ <sup>2</sup> o
bring	na:	na:

Figure 6. Text from Doke's ɿhū: vocabulary sorted according to the scheme shown in Figure 3. Note the word s<sup>2</sup>ama 'to buy', with two modifier letters, not \*ɿama. Of typographic interest is the placement of the tone diacritics with the double inverted breve in s<sup>2</sup>hwā: 'to be green', ɿ̂a<sup>2</sup> 'men', and ɿ̂a<sup>2</sup> 'bow'. UTC advice as to the correct encoding of these sequences would be welcome.