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 five phonetic characters to the UCS**
Source: **Richard S. Cook, Jr., and Michael Everson**
Status: **Expert Contribution**
Date: **2001-07-02**

A. Administrative

1. Title
Proposal to add five phonetic characters to the UCS.

2. Requester's name
Richard S. Cook, Jr., and Michael Everson.

3. Requester type
Expert contribution.

4. Submission date
2001-07-02

5. Requester's reference

6a. Completion
This is a complete proposal.

6b. More information to be provided?
No.

B. Technical – General

1a. New script? Name?
No.

1b. Addition of characters to existing block? Name?
Yes. IPA and Latin Extended-B (proposed code points: U+0221, U+0234, U+0235, U+02AE, U+02AF)

2. Number of characters
5.

3. Proposed category
Category A.

4. Proposed level of implementation and rationale
Level 1. Base character with no diacritics.

5a. Character names included in proposal?
Yes.

5b. Character names in accordance with guidelines?
Yes.

5c. Character shapes reviewable?
Yes

6a. Who will provide computerized font?
Michael Everson.

6b. Font currently available?
Yes.

6c. Font format?
TrueType.

7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?
Yes.
7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?
Yes.
8. Does the proposal address other aspects of character data processing?
No.

C. Technical – Justification
1. Contact with the user community?
Yes. The Sino-Tibetan Etymological Dictionary and Thesaurus (STEDT) Project.
2. Information on the user community?
Chinese and Sino-Tibetanist linguists.
3a. The context of use for the proposed characters?
Phonetic transcription.
3b. Reference
4a. Proposed characters in current use?
Yes.
4b. Where?
At least in China and North America.
5a. Characters should be encoded entirely in BMP?
Yes.
5b. Rationale
Keeping them with other Latin characters.
6. Should characters be kept in a continuous range?
No.
7a. Can the characters be considered a presentation form of an existing character or character sequence?
No.
7b. Where?
7c. Reference
8a. Can any of the characters be considered to be similar (in appearance or function) to an existing character?
No.
8b. Where?
8c. Reference
9a. Combining characters or use of composite sequences included?
No.
9b. List of composite sequences and their corresponding glyph images provided?
No.
10. Characters with any special properties such as control function, etc. included?
No.

D. Proposal

At present although [ɕ, ʑ, ɻ, ʵ] used in Sino-Tibetanist linguistics are to be found in the Unicode Standard 3.0, at U+0255, U+0291, U+027F, and U+0285 respectively, the five characters [ɖ, ɳ, ʈ, ʮ, ʯ] are not.

1. Alveolo-palatal consonants. As Pullum and Ladusaw (*Phonetic Symbol Guide*, second edition, 1996, ISBN 0-226-68536-5) report under their entry for [ɕ] (p. 33):

IPA USAGE
Voiceless "alveolo-palatal"median laminal fricative. Articulated further forward than [ç] (true palatal) but not as far forward as [ʃ] (palato-alveolar), and articulated laminally (with the flat blade of the tongue) rather than apically (with the tip of the tongue, as in retroflex [ʂ]).

This character [ɕ] is used to represent the Hanyu Pinyin /x-/ initial of Mainland Standard Chinese (MSC, as in the word *xiàn* 'now'). To this it may be added that, according to Wu Zongji (1992:77) key differences between the [ʃ] as in English and MSC [ɕ] are that, whereas the English sound involves a certain degree of lip-rounding and troughing (grooving) of the tongue, the Chinese sound does not: medial tongue closure is rather tight with [ɕ], and the primary point of frication is rather back in comparison with [ʃ]. Note also that this [ɕ] initial is always found with a following high front vowel, and so is in complementary distribution with MSC's retroflex [ʂ].

What is being distinguished for [ɕ, ʑ] is a place of articulation, such that the following four places are distinguished in the continuum from alveolum to palate:

ALVEOLAR, PALATO-ALVEOLAR, ALVEOLO-PALATAL, PALATAL.

Although IPA [ɕ, ʑ] are apparently marginalized in the "Other Symbols" part of the IPA chart, the "curly-tail" manner of writing in these consonants does, in fact, lend itself to the same logical extension as that seen in the writings of symbols for other places of articulation, e. g. as seen in the IPA symbols for the retroflex series [ɖ, ɳ, ɽ, ʂ, ʈ, ʐ]. Pullum and Ladusaw call this "right tail" a "diacritic for retroflexion" (1996:177; "right tail" seems to us to be a "modification" rather than a "diacritic" per se). By the same reasoning the "curly-tail" mark itself could be viewed as a modification (or "diacritic"). Just as the alveolar [t, d, n,] symbols may receive the "diacritic for retroflexion" to become apico-postalveolar retroflex [ʈ, ɖ, ɳ], so too [ȶ, ȡ, ȵ] are generated by the same principle, after the pattern of [ɕ, ʑ] to indicate laminal alveolo-palatals. This appears, at any rate, to have been the reasoning employed by linguists in their creation and usage of these [ȶ, ȡ, ȵ] curly-tailed symbols. These symbols are employed by linguists working with Chinese and Sino-Tibetan linguistics, in apparent conformity with IPA principles. It is proposed here to encode:

U+0221   [ȶ] LATIN SMALL LETTER T WITH CURL
U+0234   [ȴ] LATIN SMALL LETTER D WITH CURL
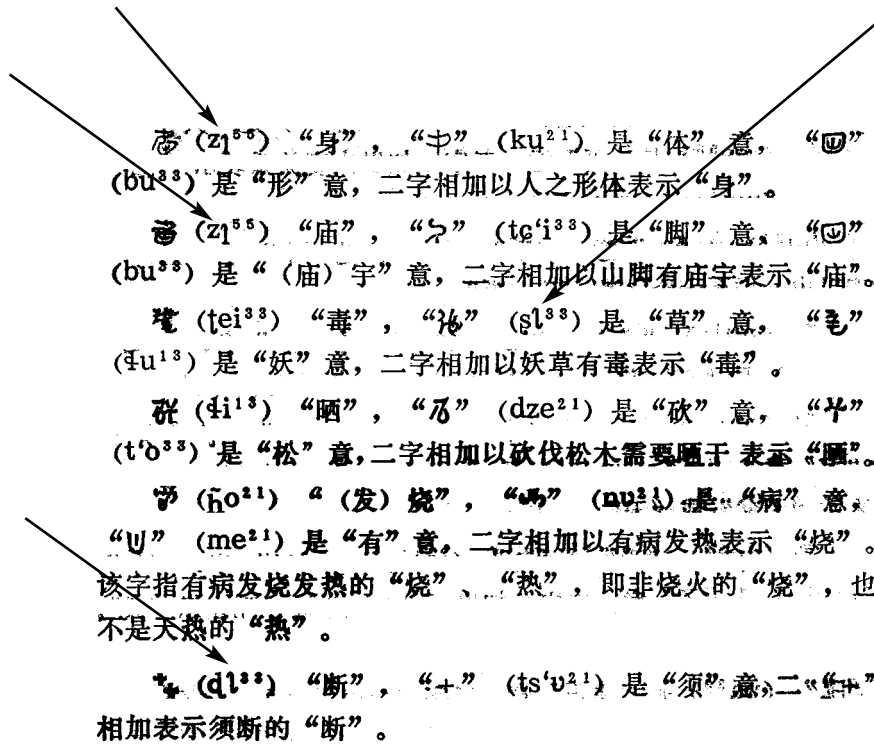U+0235   [ȵ] LATIN SMALL LETTER N WITH CURL

2. Rounded apical vowels. While [ɻ, ʅ] are encoded (U+027F LATIN SMALL LETTER REVERSED R WITH FISHHOOK and U+0285 LATIN SMALL LETTER SQUAT REVERSED ESH), and are noted in Unicode 3.0 for Sinological use, two things may be observed.

1   These two characters [ɻ, ʅ], which represent *unrounded* apical vowels have *rounded* counterparts [ɥ, ɥ] which are as yet unencoded in the UCS. The glyphs are related to, but are not identical with, [ɥ] U+0265 LATIN SMALL LETTER TURNED H. They can be seen on page 5 below in the chart from Wu Zongji (1992), and are also described in Pullum and Ladusaw (1996:81-81). It is proposed here to encode:

U+02AE   [ɥ] LATIN SMALL LETTER TURNED H WITH FISHHOOK
U+02AF   [ɥ] LATIN SMALL LETTER TURNED H WITH FISHHOOK AND TAIL

2   The glyphs for [ɻ, ʅ] shown in the Unicode Standard 3.0 are inconsistent with Sino-Tibetanist use. These letters *should* descend below the baseline, to the depth of the letter <j>. It may be possible that U+027F was derived from U+027E LATIN SMALL LETTER R WITH FISHHOOK – though it can be

noted that Pullum and Ladusaw identify this typographically as a turned small capital ʀ (1996:98) as well as describing it as an *r* with its top left serif removed (1996:161). In any case, in current Sino-Tibetanist use it has grown to the same depth so that all four vowels form a set: [ɿ, ʅ, ʮ, ʯ]. Typographically there are some interesting variations to be found; for instance in *Yi wen lun* (1993) we find that both [ɿ, ʅ] have the same length, though the second one, unusually, sits on the baseline rather than hanging from the x-height line (not to be recommended as this could easily be confused with ʃ U+0283 ʟᴀᴛɪɴ sᴍᴀʟʟ ʟᴇᴛᴛᴇʀ ᴇsʜ).

彦 (zʅ⁵⁵) "身"，"中"（kʮ²¹）是"体"意，"回"（bu³³）是"形"意，二字相加以人之形体表示"身"。

畲 (zʅ⁵⁵) "庙"，"ㄅ"（tɕʻi³³）是"脚"意，"回"（bu³³）是"（庙）宇"意，二字相加以山脚有庙宇表示"庙"。

电 (tɕi³³) "毒"，"池"（sʅ³³）是"草"意，"毛"（ɬu¹³）是"妖"意，二字相加以妖草有毒表示"毒"。

砼 (ɬi¹³) "晒"，"石"（dze²¹）是"砍"意，"牛"（tʻo³³）是"松"意，二字相加以砍伐松木需要晒干表示"晒"。

号 (ɦo²¹) "（发）烧"，"病"（ŋʮ²¹）是"病"意，"山"（me²¹）是"有"意，二字相加以有病发热表示"烧"。该字指有病发烧发热的"烧"、"热"，即非烧火的"烧"，也不是天热的"热"。

牛 (dʑʅ³³) "断"，"十"（tsʻʋ²¹）是"须"意，二字相加表示须断的"断"。

fjʃʃɿɿʮʮʯʮʃsd

zɿ dʑʅ tʻsʅ tʂʅ

附录一　　　　　音　　　综　　　表　　　(吴宗济 WU, 1992)

部位 方法		双唇	唇齿	齿间	舌尖前	舌尖	舌尖后	舌叶(舌尖及面)	舌面前	舌面中	舌根(舌面后)	小舌	喉壁	喉
塞	清 不送气	p			t	t	ʈ		ʨ	c	k	q		ʔ
	清 送气	pʻ			tʻ	tʻ	ʈʻ		ʨʻ	cʻ	kʻ	qʻ		ʔʻ
	浊 不送气	b			d	d	ɖ		ʥ	ɟ	g	ɢ		
	浊 送气	bʻ			dʻ	dʻ	ɖʻ		ʥʻ	ɟʻ	gʻ	ɢʻ		
塞擦	清 不送气	pf	tθ		ts	ts	ʈʂ	tʃ	ʨ					
	清 送气	pfʻ	tθʻ		tsʻ	tsʻ	ʈʂʻ	tʃʻ	ʨʻ					
	浊 不送气	bv	dð		dz	dz	dʐ	dʒ	ʥ					
	浊 送气	bvʻ	dðʻ		dzʻ	dzʻ	dʐʻ	dʒʻ	ʥʻ					
鼻	浊	m	ɱ		n	n	ɳ		ȵ	ɲ	ŋ	ɴ		
滚	浊					r						R		
闪	浊				r	ɾ	ɽ					R		
边	浊				l	l	ɭ			ʎ				
边擦	清				ɬ	t					x			
	浊				ɮ	ɭ				ʎ	ɣ			
擦	清	ɸ	f	θ	s	s	ʂ	ʃ	ɕ	ç	x	χ	ħ	h
	浊	β	v	ð	z	z	ʐ	ʒ	ʑ	ʝ	ɣ	ʁ	ʕ	ɦ
无擦通音及半元音	浊	w ɥ	ʋ		ɹ	ɹ				j(ɥ)	(w)			

元音			
	圆唇元音 (ʯ ʮ ʯ ɥ ʉ)	舌尖元音 前 ɿ ʮ	舌尖元音 后 ʅ ʯ	舌面元音 前 i y　ɪ ʏ　e ø　ɛ œ　æ　a	央 ɨ ʉ　ə　ɵ　ɜ　ɐ	后 ɯ u　ʊ　ɤ o　ʌ ɔ　ʌ　ɑ ɒ
高			
半高	(ø o)		
半低	(œ ɔ)		
低	(ɒ)		

BIBLIOGRAPHIC REFERENCES (SELECTED)

In addition to the above statistics for data in the STEDT *Main Lexicon*, the following references seek to document attested usages of and comments on curly-tail alveolo-palatal place series symbols. Comments on usage of specific symbols are appended to some entries below.

• **CHAO Yuen Ren**　　　趙元任
1928　《現代吳語的研究》<<Xiandai Wuyu de yanjiu>>. [Research on Modern Wu Dialects.] Pei-p'ing: Ch'ing-hua hsueh hsiao yen chiu yuan. 2nd edition. Series: Ch'ing-hua hsueh hsiao yen chiu yuan ts'ung shu ; ti 4 chung. [UCB East Asian 5155.4012]
1934 "The non-uniqueness of phonemic solutions of phonetic systems." Nanking: Extract from v. 4 of the Bulletin of the National research Institute of History and Philology, Academia Sinica. (363-398 p. 28 cm.) [UCB Main Stack P221; .C5]
• **International Phonetic Association (IPA)**
2000 Current Charts Available at <http://www.arts.gla.ac.uk/IPA/ipa.html>. See specifically, <http://www2.arts.gla.ac.uk/IPA/symbols.html> for the Chart of OTHER SYMBOLS.
• **LI Zhenhua**　　　李珍華
1993　《漢字古今音表》，〔美〕李珍華，周長楫編撰。北京：中華書局。<<Han Zi Gu Jin Yin Biao>> (*Historical Chinese Phonologic Tables*). LI Zhenhua and ZHOU Changji. Beijing: Zhonghua Shu Ju, 1993. (Harvard Yenching Library Number: 5120 4414). ISBN 7-101-01198-5/H.103. [8,865 char. readings in 3 historical periods and 7 modern dialects.] This work employs [ȶ, ȡ, ȵ, ɕ, ʑ] in reconstructions of Old Chinese (上古); [ȶ, ȡ, ɕ, ʑ] in transcriptions of Middle Chinese (中古); [ȵ, ɕ, ʑ] for the modern Wu dialect (吳語); [ȵ, ɕ] for the modern Gan dialect (贛語); [ȵ] for the modern Kejia (客家話) dialect. Many of LI and ZHOU's historical forms were also cited in RSC-ECC (Cook 1995).
• **LIANG Min**　　　梁民 (sp?) [Co-authored with ZHANG Junru.]
1996　《侗台語族概論》<<Dong-Tai yuzu gailun>> [Intro to Kam-Tai languages]. Beijing: Zhongguo shehui kexue chubanshe. Cf. pp. 972-973. [UCB EAL TEMP98 3754]
• **MA Xueliang**　　　马学良
1981　《语言学概要》. <<Yuyanxue gailun>> [Elements of Linguistics]. Beijing: Huazhong gongxueyuan chubanshe. Cf. pp. 32,51.
• **PULLUM, Geoffrey K. and William A. LADUSAW.**
1996 *Phonetic Symbol Guide*, Second Edition. Chicago, London: University of Chicago Press.
• **QIAN Nairong**　　　錢乃榮
1992　《當代吳語研究》 <<Dangdai Wuyu yanjiu>> [Contemporary Wu Dialect Studies]. Shanghai: Shanghai jiaoyu chubanshe. 1121 pp. [UCB East Asian PL1931 .C45 1992] Cf. p. 60.
• **Unknown Committee**　　中國大百科全書編輯委員
1988　《中国大百科全书》 <<Zhongguo dabaike quanshu>>. [Great Chinese Encyclopedia]. Beijing: Zhongguo dabaike quanshu chubanshe. Cf. Yuyan-Wenzi vol., p.118 (IPA).
• **WU Zongji**　　　吴宗济
1992　《现代汉语语音概要》，吴宗济 主编。北京: 华语教学出版社。<<Xiandai Hanyu - Yuyin Gaiyao>> [Elements of Modern Chinese Phonetics], 吴宗济 WU Zongji, Ed. Beijing: Sinolingua, 1992. ISBN 7-80052-137-0. A general handbook of Modern Standard (Beijing) Chinese phonetics, this work contains an IPA-style "Phonetic Symbol Table" on page 196 (see attached copy) which includes the [ȶ, ȡ, ȵ, ɕ, ʑ] curly-tail symbols.
• **YU Nae-wing**　　　余迺永 (YU Naiyong)
1993　《新校互註・宋本廣韻》香港：香港中文大學。 <<Xin Jiao Hu Zhu - Song Ben Guang Yun>> [A New Revision of the Sung Edition of the Kuang-yun Rhyming Dictionary]. Hong Kong: Chinese University of Hong Kong. 2nd Ed. (3rd Ed., 2000). 1vol., ~900pp., hardcover, Chinese, indices and English appendices. ISBN: 962-201-413-5. Cf. the table, p. 82.

STEDT *Source Bibliography* ABBREVIATIONS

The Sino-Tibetan Etymological Dictionary and Thesuarus (STEDT) Project *Source Bibliography* abbreviations cited above are as follows (numbers at the end of each entry indicate the total number of records in the STEDT databases from each source):

- AW-TBT = WEIDERT,1987.
- CSL-YIzd = CHEN, 1979.
- DHFRL = DAI, 1991.
- JP-Idu = PULU, 1978.
- LTBA = MATISOFF 1974-.
- RSC-ECC = COOK.
- SHK-Sulung = SUN, 1993.
- SLZO-MLD = SUN 1980.
- TBL = DAI, 1992.
- ZMYYC = SUN 1991.

CHEN Shilin
1979 <<Yihan zidian>> [Yi-Chinese dictionary]. CHEN Shilin, LI Min, et al., eds. 1979. Yi Language Work Unit, People's Committee of Sichuan.
COOK, Richard S. 曲理察
1995 *The Etymology of Chinese* 辰 *Chén*. Monograph volume of LTBA, Vol. 18.2/Fall 1995. Berkeley: University of California. 1vol., 278pp.《 '辰'字的原始義》。藏緬語言學報・卷 18.2。貝克萊市：美國加利福尼亞州大學在貝克萊市語言學部，1995.
DAI Qingxia
1992 <<Zang Mian Yu Zu Yu Yan Ci Hui>> [A Tibeto-Burman lexicon]. DAI Qingxia et al. Beijing: Central Institute of Minorities. {ISBN 7-81001-347-5/H}
1991 <<Zangmianyu shiwuzhong>> [Fifteen Tibeto-Burman languages]. DAI Qingxia et al., eds. Beijing: Yanshan Chubanshe. {DHFRL, authors: DAI, HUANG, FU, RENZENG, LIU.}
MATISOFF, James A.
1974- The Biannual Journal *Linguistics of the Tibeto-Burman Area*. The University of California at Berkeley, James A. MATISOFF, editor. LOC 82-640813, ISSN 0731-3500, OCLC: 4790670. <http://stedt.berkeley.edu/ltba/>
PULU, Jatan.
1978 *Idu phrase book*. Shillong: The Director of Information and Public Relations, Arunachal Pradesh.
SUN Hongkai
1993 Questionnaire (Sulong). Unpublished.
1991 <<Zangmianyu yuyin he cihui>> [Tibeto-Burman phonology and lexicon]. SUN Hongkai et al. Beijing: Chinese Social Sciences Press. {ISBN 7-5004-0797-1/H}
1980 <<Menba, Luoba, Dengren de yuyan>> [The languages of the Menba, Luoba and Deng peoples]. SUN Hongkai, LU Shaozun, ZHANG Jichuan and OUYANG Jueya, eds. Beijing: Social Sciences Press.
WEIDERT, Alfons K.
1987 *Tibeto-Burman tonology: A comparative account*. (Current Issues in Linguistic Theory, Vol. 54.) Amsterdam and Philadelphia: John Benjamins Publishing Co. (5426)