

Title: Preliminary Proposal to Encode the Pau Cin Hau Alphabet in ISO/IEC 10646
Source: Script Encoding Initiative (SEI)
Author: Anshuman Pandey (pandey@umich.edu)
Status: Liaison Contribution
Action: For consideration by UTC and WG2
Date: 2010-10-27

1 Introduction

This is a proposal to encode the Pau Cin Hau Alphabet in the Universal Character Set. The information presented here supersedes that given in the following documents:

- N3781 L2/10-080 “Preliminary Proposal to Encode the Pau Cin Hau Script in ISO/IEC 10646”
- N3784R L2/10-092R “Defining Properties for Tone Marks of the Pau Cin Hau Script”
- N3865R L2/10-073R1 “Allocating the Pau Cin Hau Scripts in the Unicode Roadmap”

The script described here is identical to that described in N3781 L2/10-080. The difference is the name assigned to the script: ‘Pau Cin Hau’ is now ‘Pau Cin Hau Alphabet’. The present proposal replaces N3781 and N3784R and extends the information provided in those documents. It provides a brief background of the script, describes its structure, and presents details on the writing system to assist in implementation. The ISO/IEC proposal summary form is enclosed.

A logographic script related to the Pau Cin Hau Alphabet is mentioned in this proposal. More information about it is available in “Introducing the Logographic Script of Pau Cin Hau” (N3961 L2/10-438).

2 Background

The Pau Cin Hau Alphabet is an ecclesiastical script of the Laipian religious tradition, which emerged in the Chin Hills region of present-day Myanmar at the turn of the 20th century. The script is named after Pau Cin Hau (1859–1948), a ‘Sukte’ (Tedim) Chin, who founded the Laipian tradition and developed the script in order to convey his teachings. In an account given by J. J. Bennison in the 1931 *Census of India* report for Burma, Pau Cin Hau stated that the characters of his script were revealed to him in a dream in 1902.¹

The script was designed to represent Tedim [ctd], a language of the northern branch of the Kuki-Chin group of the Tibeto-Burman family, which is spoken in Chin State, Myanmar. It is known traditionally as *pau cin hau lai* “script of Pau Cin Hau” and *tual lai* “local script”. The Tedim word *lai* means “writing” and Pau Cin Hau himself is referred to as *laipianpa* “script creator” (and *laipian* is the ‘script-based religion’).² The name is also romanized as ‘Pau Chin Hau’ and is spelled in one source as ‘Bow-chinhow’.³

There are two distinct writing systems associated with Pau Cin Hau and the Laipian tradition. One is a logographic script and the other is an alphabet. Both are attested in written and printed primary and secondary sources. The script originally devised by Pau Cin Hau in 1902 is said to have been reformed twice. The

¹ Bennison 1933: 217.

² Pau: 11.

³ American Bible Society 1938: 82.

logographic script likely represents the first reform. It is said to consist of 1,050 characters, a number that is based on a traditional count of the characters used in a Laipian recitation text (see Figure 6). The final reform took place in 1931, at which time the logographic script was simplified.⁴ The result was the 57-character alphabet, which is proposed here for encoding.

The alphabet of Pau Cin Hau was used for writing and printing Laipian and Christian literature. Primers of the script were printed for the purpose of teaching it (see Figure 3). Books of Laipian ritual songs were written and it is believed that some of these were also printed.⁵ In 1931, the Baptist and Foreign Bible Society printed the “Sermon on the Mount” from the book of St. Matthew in the ‘Kamhow’ (now known as Tedim) dialect in the alphabet (see Figure 4). At least one metal font of the alphabet was developed in order to print the materials of the Baptist and Foreign Bible Society. The script charts shown in Figure 2 and Figure 3 appear to be printed using different fonts, but the source of the other font is unknown.

Some proposed characters resemble those found in Burmese, Latin, and other writing systems. These occurrences are coincidental since neither script has a genetic relationship with any other writing system. The similarities in glyph shapes may be a result of the influence of major scripts used in and around Chin State.

The use of the Pau Cin Hau alphabet diminished with the decline of the Laipian tradition and the rise of Christian missionary activity. Baptist missionaries introduced Latin-based scripts for languages of the Chin Hills and in many cases developed the first written forms for these languages. A Latin-based script for Tedim was introduced by Joseph. H. Cope, an American Baptist missionary.⁶ This Latin alphabet is the regular script for Tedim and has replaced Pau Cin Hau. Although practice of Laipian and usage of the Pau Cin Hau script have declined, both still survive to an extent. It is reported that both the logo-syllabic and alphabetic systems are still in limited use by the Laipian community.⁷ Information on the size of the user community was unavailable to Bennison in 1931;⁸ the same is true at present. In addition to the use of the Pau Cin Hau Alphabet by Laipian practitioners, the script enjoys scholarly attention, as is evidenced by articles published by members of Chin-speaking communities in several websites.

3 Script Details

3.1 Script Name

The name of the script is ‘Pau Cin Hau Alphabet.’ A native name is yet to be identified. It is unclear if the Tedim name *pau cin hau lai* refers to the logographic or alphabetic script, or to both. In the preliminary proposal (N3781 L2/10-080), the proposed script was simply called ‘Pau Cin Hau’. This name was assigned before research was conducted upon the logographic script. In the document requesting allocations for the script in the Unicode roadmaps (N3865R L2/10-073R1), the names ‘Pau Cin Hau A’ and ‘Pau Cin Hau B’ were suggested as temporary names for the alphabetic and logographic scripts, respectively.

The names for the two scripts are based upon functional classifications. The alphabetic script proposed here, which was called ‘Pau Cin Hau A’, is now aptly named as ‘Pau Cin Hau Alphabet’. The script named ‘Pau Cin Hau B’ has been assigned the tentative name ‘Pau Cin Hau Logographs’. Determining suitable names for these scripts has been identified as an issue in Section 5.

⁴ Pau: 10.

⁵ Banks 1967: 46.

⁶ Banks 1967: 59.

⁷ Button 2009: 23 fn. 20.

⁸ Bennison 1933: 217.

3.2 Character Repertoire

The Pau Cin Hau Alphabet has a repertoire of 57 characters, which consists of 21 consonant letters, 7 vowel letters, 9 coda letters, and 20 tone marks. The script is allocated to the Supplementary Multilingual Plane (SMP) at the range U+11840..1187F. A preliminary code chart and nameslist is provided in Figure 1.

3.3 Character Glyphs

The glyphs of the Pau Cin Hau font used in the code chart are based upon forms shown in script charts. The font was designed by the present author. Text typeset in the font is shown in Figure 5.

3.4 Character Names

Characters are named according to those given in the script charts. The only exceptions are the names of ɣ_{AA} and ɛ_{AAI} , which are shown in the charts as ‘á’ and ‘ái’, respectively. The doubling of ‘A’ is a normalization of the acute accent.

The names given in the charts may present some confusion. Names for onset consonant letters are written with an ‘A’, which suggests an inherent-vowel feature that does not reflect the actual phonological properties of the letters. The naming practice resembles character names found in Brahmi-based scripts, eg. PA, KA, LA, etc., which may be an influence from Burmese or Indic scripts. Additionally, the names for coda consonants are also misleading. Codas in Tedim can be only voiceless, liquid, or nasal consonants.⁹ Therefore, names such as AB, AG, AD that are given in the charts contradict basic phonological properties of the letters; ie. ‘B’, ‘G’, and ‘D’ represent voiced sounds, whose voiceless analogues are ‘P’, ‘K’, and ‘T’. Also, the use of Latin ‘A’ in the names of coda consonants suggests an initial vocalic property that does not inhere in the letter. The use of ‘A’ in the names for coda may be a prosthetic element to assist in pronunciation of these consonants in isolation.

Names for the tone marks have been assigned serially based upon the order in which they appear in the charts. The script was intended for writing languages other than Tedim. However, other northern Chin languages have correlations between tone contour and tone category that differ from Tedim. Therefore, the semantic value of the tone marks is dependent upon the language being represented. A generic approach to the names of tone marks supports the intended broader use of the script. Names for the tone marks based upon the tonal system of Tedim are given in Section 5.

4 Writing System

4.1 Structure

Pau Cin Hau is an alphabetic script that is written from left to right. Consonants, vowels, and tone marks are written as independent characters. The script was designed to represent the Tedim language. The Tedim syllable canon may be described as $(C_1)V_1(V_2)(C_2)T$.¹⁰ The presence of consonant letters, such as FA and CHA, that represent phonemes not used in Tedim suggest that the script was also intended for writing other Chin languages, such as Mizo.

The onset (C_1) is represented by one of the following consonants.¹¹

⁹ Button 2009: 30.

¹⁰ Thang 2001: 33.

¹¹ Button 2009: 34.

ᄁ	PA	/p/	ᄃ	NGA	/ŋ/	ᄅ	TA	/t/
ᄇ	KA	/k/	ᄉ	HA	/h/	ᄋ	HTA	/t ^h /
ᄍ	LA	/l/	ᄏ	GA	/g/	ᄑ	NA	/n/
ᄓ	MA	/m/	ᄕ	XA	/x/	ᄗ	HPA	/p ^h /
ᄕ	DA	/d/	ᄛ	HSA	/s/	ᄙ	RA	/r/
ᄝ	YA	/j/	ᄟ	BA	/b/	ᄡ	FA	/f/
ᄟ	VA	/v/	ᄡ	TGA	/j/	ᄣ	CHA	/c ^h /

The nucleus (V_1) is represented using the following vowels. The letters UA and IA are diphthongs,¹² which are coded as $V_1(V_2)$.

ᄥ	AA	/a/	ᄦ	AW	/o/	ᄧ	IA	/ia/
ᄨ	AAI	/e/	ᄩ	U	/u/			
ᄪ	I	/i/	ᄫ	UA	/ua/			

The coda (C_2) is represented using one of the following consonants.¹³ The letters AU and AI are glides.¹⁴

ᄬ	AB	/p/	ᄭ	AM	/m/	ᄮ	AU	/w/
ᄮ	AG	/k/	ᄯ	AN	/n/	ᄰ	ANG	/ŋ/
ᄰ	AD	/t/	ᄱ	AL	/l/	ᄲ	AI	/j/

The tone (T) is represented using one of the 20 tone marks (see Section 4.2):

ᄴ	PAU CIN HAU TONE MARK-1	ᄴ	PAU CIN HAU TONE MARK-11
ᄶ	PAU CIN HAU TONE MARK-2	ᄶ	PAU CIN HAU TONE MARK-12
ᄸ	PAU CIN HAU TONE MARK-3	ᄸ	PAU CIN HAU TONE MARK-13
ᄺ	PAU CIN HAU TONE MARK-4	ᄺ	PAU CIN HAU TONE MARK-14
ᄼ	PAU CIN HAU TONE MARK-5	ᄼ	PAU CIN HAU TONE MARK-15
ᄾ	PAU CIN HAU TONE MARK-6	ᄾ	PAU CIN HAU TONE MARK-16
ᄿ	PAU CIN HAU TONE MARK-7	ᄿ	PAU CIN HAU TONE MARK-17
ᄽ	PAU CIN HAU TONE MARK-8	ᄽ	PAU CIN HAU TONE MARK-18
ᄽ	PAU CIN HAU TONE MARK-9	ᄽ	PAU CIN HAU TONE MARK-19
ᄽ	PAU CIN HAU TONE MARK-10	ᄽ	PAU CIN HAU TONE MARK-20

4.2 Tone Marks

The tone marks reportedly represent the complete tonal system of Tedim. Charts of the script simply show the tone marks arranged in a matrix of six columns and four rows, as shown below, but do not explain the logic behind the arrangement or specify the values of the marks:

ᄴ	ᄶ	ᄸ	ᄺ	ᄼ	ᄾ
ᄿ	ᄽ		ᄽ	ᄽ	
	ᄴ	ᄶ	ᄸ	ᄺ	ᄼ
ᄼ	ᄾ	ᄿ	ᄽ	ᄽ	ᄽ

¹² Button 2009: 26.

¹³ Button 2009: 30.

¹⁴ Button 2009: 27.

Christopher Button, a specialist of northern Chin linguistics, has shed much light upon the properties and arrangement of the tonal signs.¹⁵ He suggested that the signs are used for marking vowel length, tone, and glottal stop and that particular signs are used as punctuation. Thus, the term ‘tone mark’ does not fully describe the function of the characters.

At the highest level, the arrangement can be divided into two sets of three columns. The first set of three columns contains the ‘basic’ marks. Each column in the second set corresponds by position to a column in the first set and represents the ‘sentence-final’ form of the respective ‘basic’ mark. Not all ‘basic’ and ‘sentence-final’ marks have corresponding forms. Of the total 20 ‘tonal signs’, the 15 characters in columns (1, 4) and (2, 5) represent tone marks and the 5 characters in (3, 6) represent glottal stop.

The 15 tone marks may be grouped by column according to their vocalic length and positional features and by row according to tone category as follows:

	BASIC		SENTENCE FINAL	
	LONG VOWEL	SHORT VOWEL	LONG VOWEL	SHORT VOWEL
TONE II		·		·
SANDHI TONE	┆	┆·	┆┆	┆┆·
TONE I		·	↗	↗·
TONE III	┆	┆·	┆┆	┆┆·

Each mark in the ‘basic’ class has two features: tone (basic or sandhi) and vowel length. Those of the ‘sentence-final’ possess the same two features, but have a structural property as well. They are contextual variants of the corresponding ‘basic’ marks. When a ‘basic’ mark appears at the end of a sentence, it is replaced with its corresponding ‘sentence-final’ form.

The 5 marks used for indicating glottal stop may be classified as shown below. There are regular and sandhi-variant forms. These marks are used for writing codas ending in a glottal stop.

	BASIC	SENTENCE FINAL
	GLOTTAL STOP	GLOTTAL STOP
SANDHI VARIANT	┆·	┆┆·
SANDHI VARIANT II (?)	·	
REGULAR	┆·	┆┆·

The placement of glottal stop marks along with tone marks in the traditional arrangement is peculiar. According to Button, the glottal stop has no actual correlation with tones in Tedim phonology. He states that connected speech allows fluctuations in pronunciation that are functionally matched with the tones orthographically, but with which there is no real phonological association. The regular glottal stop mark ʔ is paired

¹⁵ Button, February 2010, personal correspondence.

with tone III orthographically (row 4 in the traditional arrangement) and the sandhi-variant form 𑜀 is paired with tone II orthographically (row 1). The sandhi-variant form 𑜁, which is orthographically paired with tone I (row 3), does not have real representation in Tedim, as there is only one possible sandhi variant of words with glottal stops.

The ‘sentence-final’ marks are a peculiar innovation. Their glyph shapes hint at their function as punctuation marks. They are written by doubling the elemental shape of the ‘basic’ mark, eg. 𑜀 → 𑜀𑜀, 𑜁 → 𑜁𑜁, etc. This typology may be influenced by the orthography of Burmese punctuation, in which 𑜀 U+104A MYANMAR SIGN LITTLE SECTION is doubled to produce 𑜀 U+104B MYANMAR SIGN SECTION. The exceptions to this composition rule are the shapes of marks in the tone I category, for which the ‘sentence-final’ marks are produced by introducing doubles of graphical elements not found in the ‘basic’ mark, eg. 𑜀 → 𑜀𑜀 and 𑜁 → 𑜁𑜁. Alternately, the horizontal doubled lines may represent the absence of an orthographic symbol in non-sentence final position.

While the Pau Cin Hau Alphabet may have been designed for writing Tedim, it appears that the tone-mark repertoire of the script was intended for the comprehensive representation of the northern Chin languages. For example, as mentioned above, not all tone marks have actual representation in Tedim. There is only one possible sandhi variant for the ‘basic’ glottal stop in the language, but three marks for the ‘basic’ glottal stop in the script. It is likely that the mark 𑜁 was included for writing other Chin languages that possess a sandhi variant of glottal stop that is orthographically paired with tone I (row 1).

The contours of the basic tone categories of the northern Chin languages vary by language, as shown below:¹⁶

	MIZO	ZAHAU	THADO	ZO	TEDIM	SIZANG
TONE I	𑜀	𑜀	𑜀	𑜀	𑜀	𑜀
TONE IIA	𑜀	𑜀	𑜀	𑜀	𑜀	𑜀
TONE IIB	𑜀	𑜀				
TONE III	𑜀	𑜀	𑜀	𑜀	𑜀	𑜀

Based upon the differences in the correlation between tone contour and tone category, the semantic value of tone marks in the Pau Cin Hau Alphabet is dependent upon the language being represented.

4.3 Digits

Script-specific digits are unattested in the Pau Cin Hau Alphabet. Latin-based digits are used in the text excerpt shown in Figure 4.

4.4 Punctuation

Sentence boundaries are indicated through the use of certain tone marks. Word boundaries are indicated through spaces. Other script-specific punctuation is unattested. Latin punctuation, such as periods, appear in the text excerpt in Figure 4.

¹⁶ Adapted from Button (2009: 37).

4.5 Collation

The sort order for the script is being determined. See Section 5.

4.6 Line Breaking

Tone marks: No breaking may occur between a tone mark and the character that precedes it. All tone marks have the linebreaking property ‘AL’.

4.7 Character Properties

Properties for characters of the Pau Cin Hau Alphabet in the Unicode Character Database format:

```

11840;PAU CIN HAU LETTER PA;Lo;0;L;;;;N;;;;;
11841;PAU CIN HAU LETTER KA;Lo;0;L;;;;N;;;;;
11842;PAU CIN HAU LETTER LA;Lo;0;L;;;;N;;;;;
11843;PAU CIN HAU LETTER MA;Lo;0;L;;;;N;;;;;
11844;PAU CIN HAU LETTER DA;Lo;0;L;;;;N;;;;;
11845;PAU CIN HAU LETTER YA;Lo;0;L;;;;N;;;;;
11846;PAU CIN HAU LETTER VA;Lo;0;L;;;;N;;;;;
11847;PAU CIN HAU LETTER NGA;Lo;0;L;;;;N;;;;;
11848;PAU CIN HAU LETTER HA;Lo;0;L;;;;N;;;;;
11849;PAU CIN HAU LETTER GA;Lo;0;L;;;;N;;;;;
1184A;PAU CIN HAU LETTER XA;Lo;0;L;;;;N;;;;;
1184B;PAU CIN HAU LETTER HSA;Lo;0;L;;;;N;;;;;
1184C;PAU CIN HAU LETTER BA;Lo;0;L;;;;N;;;;;
1184D;PAU CIN HAU LETTER TGA;Lo;0;L;;;;N;;;;;
1184E;PAU CIN HAU LETTER TA;Lo;0;L;;;;N;;;;;
1184F;PAU CIN HAU LETTER HTA;Lo;0;L;;;;N;;;;;
11850;PAU CIN HAU LETTER NA;Lo;0;L;;;;N;;;;;
11851;PAU CIN HAU LETTER HPA;Lo;0;L;;;;N;;;;;
11852;PAU CIN HAU LETTER RA;Lo;0;L;;;;N;;;;;
11853;PAU CIN HAU LETTER FA;Lo;0;L;;;;N;;;;;
11854;PAU CIN HAU LETTER CHA;Lo;0;L;;;;N;;;;;
11855;PAU CIN HAU LETTER AA;Lo;0;L;;;;N;;;;;
11856;PAU CIN HAU LETTER AAI;Lo;0;L;;;;N;;;;;
11857;PAU CIN HAU LETTER I;Lo;0;L;;;;N;;;;;
11858;PAU CIN HAU LETTER AW;Lo;0;L;;;;N;;;;;
11859;PAU CIN HAU LETTER U;Lo;0;L;;;;N;;;;;
1185A;PAU CIN HAU LETTER UA;Lo;0;L;;;;N;;;;;
1185B;PAU CIN HAU LETTER IA;Lo;0;L;;;;N;;;;;
1185C;PAU CIN HAU LETTER AB;Lo;0;L;;;;N;;;;;
1185D;PAU CIN HAU LETTER AG;Lo;0;L;;;;N;;;;;
1185E;PAU CIN HAU LETTER AD;Lo;0;L;;;;N;;;;;
1185F;PAU CIN HAU LETTER AM;Lo;0;L;;;;N;;;;;
11860;PAU CIN HAU LETTER AN;Lo;0;L;;;;N;;;;;
11861;PAU CIN HAU LETTER AL;Lo;0;L;;;;N;;;;;
11862;PAU CIN HAU LETTER AU;Lo;0;L;;;;N;;;;;
11863;PAU CIN HAU LETTER ANG;Lo;0;L;;;;N;;;;;
11864;PAU CIN HAU LETTER AI;Lo;0;L;;;;N;;;;;
11865;PAU CIN HAU TONE MARK-1;Lo;0;L;;;;N;;;;;
11866;PAU CIN HAU TONE MARK-2;Lo;0;L;;;;N;;;;;
11867;PAU CIN HAU TONE MARK-3;Lo;0;L;;;;N;;;;;
11868;PAU CIN HAU TONE MARK-4;Lo;0;L;;;;N;;;;;
11869;PAU CIN HAU TONE MARK-5;Lo;0;L;;;;N;;;;;
1186A;PAU CIN HAU TONE MARK-6;Lo;0;L;;;;N;;;;;
1186B;PAU CIN HAU TONE MARK-7;Lo;0;L;;;;N;;;;;
1186C;PAU CIN HAU TONE MARK-8;Lo;0;L;;;;N;;;;;

```

```

1186D;PAU CIN HAU TONE MARK-9;Lo;0;L;;;;;N;;;;;
1186E;PAU CIN HAU TONE MARK-10;Lo;0;L;;;;;N;;;;;
1186F;PAU CIN HAU TONE MARK-11;Lo;0;L;;;;;N;;;;;
11870;PAU CIN HAU TONE MARK-12;Lo;0;L;;;;;N;;;;;
11871;PAU CIN HAU TONE MARK-13;Lo;0;L;;;;;N;;;;;
11872;PAU CIN HAU TONE MARK-14;Lo;0;L;;;;;N;;;;;
11873;PAU CIN HAU TONE MARK-15;Lo;0;L;;;;;N;;;;;
11874;PAU CIN HAU TONE MARK-16;Lo;0;L;;;;;N;;;;;
11875;PAU CIN HAU TONE MARK-17;Lo;0;L;;;;;N;;;;;
11876;PAU CIN HAU TONE MARK-18;Lo;0;L;;;;;N;;;;;
11877;PAU CIN HAU TONE MARK-19;Lo;0;L;;;;;N;;;;;
11878;PAU CIN HAU TONE MARK-20;Lo;0;L;;;;;N;;;;;

```

4.8 Confusables

Several Pau Cin Hau characters are ‘confusable’ with those of Latin:

```

PAU CIN HAU LETTER LA      ; 0043 LATIN CAPITAL LETTER C
PAU CIN HAU LETTER DA      ; 0042 LATIN CAPITAL LETTER B
PAU CIN HAU LETTER YA      ; 0055 LATIN CAPITAL LETTER U
PAU CIN HAU LETTER TA      ; 004C LATIN CAPITAL LETTER L
PAU CIN HAU LETTER HPA     ; 005A LATIN CAPITAL LETTER Z
PAU CIN HAU LETTER AW      ; 0041 LATIN CAPITAL LETTER A
PAU CIN HAU LETTER AG      ; 0054 LATIN CAPITAL LETTER T
PAU CIN HAU LETTER AM      ; 0056 LATIN CAPITAL LETTER V
PAU CIN HAU LETTER AI      ; 004F LATIN CAPITAL LETTER O
PAU CIN HAU TONE MARK-11   ; 00B7 MIDDLE DOT
PAU CIN HAU TONE MARK-12   ; 003A COLON

```

Other ‘confusable’ glyphs resemble those of the Myanmar script:

```

PAU CIN HAU LETTER KA      ; 1005 MYANMAR LETTER CA
PAU CIN HAU LETTER LA      ; 1004 MYANMAR LETTER NGA
PAU CIN HAU LETTER MA      ; 1013 MYANMAR LETTER DHA
PAU CIN HAU LETTER HA      ; 1001 MYANMAR LETTER KHA
PAU CIN HAU LETTER HSA     ; 1054 MYANMAR LETTER VOCALIC L
PAU CIN HAU LETTER HTA     ; 1027 MYANMAR LETTER E
PAU CIN HAU LETTER AB      ; 1016 MYANMAR LETTER PHA
PAU CIN HAU LETTER AI      ; 101D MYANMAR LETTER WA
PAU CIN HAU TONE MARK-1    ; 104A MYANMAR SIGN LITTLE SECTION
PAU CIN HAU TONE MARK-4    ; 104B MYANMAR SIGN SECTION

```

5 Outstanding Issues

Script Name The indigenous names for the two scripts have not yet been identified. Names have been assigned based upon the structure of the writing systems: ‘Pau Cin Hau Alphabet’ and ‘Pau Cin Hau Logographs’. Are such names acceptable? Are there better alternatives?

Collation What is the preferred collation order for the script? Is the encoding order suitable as the sort order? Should collation follow that of the Latin-based script for Tedim?

Names for Tone Marks The value of tone marks is language dependent (see Section 4.2). For this reason, generic names have been assigned. Alternate names based upon the Tedim tonal system are given below. Are such generic names preferred to Tedim-specific names:

l	PAU CIN HAU TONE-2 LONG
l̥	PAU CIN HAU TONE-2 SHORT
l̥ˀ	PAU CIN HAU TONE-2 GLOTTAL STOP
ll	PAU CIN HAU TONE-2 LONG FINAL
ll̥	PAU CIN HAU TONE-2 SHORT FINAL
ll̥ˀ	PAU CIN HAU TONE-2 GLOTTAL STOP FINAL
l̥ʰ	PAU CIN HAU SANDHI TONE LONG
l̥ʰ̥	PAU CIN HAU SANDHI TONE SHORT
l̥ʰ̥ˀ	PAU CIN HAU SANDHI TONE LONG FINAL
l̥ʰ̥̥	PAU CIN HAU SANDHI TONE SHORT FINAL
·	PAU CIN HAU TONE-1 SHORT
ˀ	PAU CIN HAU TONE-1 GLOTTAL STOP
⇒	PAU CIN HAU TONE-1 LONG FINAL
⇒ˀ	PAU CIN HAU TONE-1 GLOTTAL STOP FINAL
l̥	PAU CIN HAU TONE-3 LONG
l̥̥	PAU CIN HAU TONE-3 SHORT
l̥̥ˀ	PAU CIN HAU TONE-3 GLOTTAL STOP
ll̥̥	PAU CIN HAU TONE-3 LONG FINAL
ll̥̥̥	PAU CIN HAU TONE-3 SHORT FINAL
ll̥̥̥ˀ	PAU CIN HAU TONE-3 GLOTTAL STOP FINAL

6 References

- American Bible Society. 1938. *The Book of a Thousand Tongues: Being Some Account of the Translation and Publication of All or Part of The Holy Scriptures Into More Than a Thousand Languages and Dialects With Over 1100 Examples from the Text*. Edited by Eric M. North. New York and London: Harper & Brothers.
- Banks, E. Pendleton. 1967. "Pau Cin Hau: A Case of Religious Innovation among the Northern Chin." In *American Historical Anthropology: Essays in Honor of Leslie Spier*. London and Amsterdam: Southern Illinois University Press.
- Bennison, J. J. 1933. "The Pau Cin Hau Movement in the Chin Hills." *Census of India, 1931*, vol. XI, Burma, pt. 1, ch. XI. pp. 217–218.
- Button, Christopher T. J. 2009. "A Reconstruction of Proto Northern Chin in Old Burmese and Old Chinese Perspective". Ph.D. dissertation. School of Oriental and African Studies, University of London. <http://victoria.linguistlist.org/~lapolla/downloads/Button-Proto-Northern-Chin.pdf>
- Go, Khup Za. 2008. *Zo Chronicles: A Documentary Study of History and Culture of the Kuki-Chin-Lushai Tribe*. New Delhi: Mittal Publications.
- Pandey, Anshuman. 2010a. "Preliminary Proposal to Encode the Pau Cin Hau Script in ISO/IEC 10646" (N3781 L2/10-080), February 28, 2010. <http://std.dkuug.dk/JTC1/SC2/WG2/docs/n3781.pdf>
- . 2010b. "Defining Properties for Tone Marks of the Pau Cin Hau Script" (N3784R L2/10-092R), May 22, 2010. <http://std.dkuug.dk/JTC1/SC2/WG2/docs/n3784.pdf>
- . 2010c. "Allocating the Pau Cin Hau Scripts in the Unicode Roadmap" (N3865R L2/10-073R1), July 26, 2010. <http://std.dkuug.dk/JTC1/SC2/WG2/docs/n3865.pdf>

- . 2010d. “Introducing the Logographic Script of Pau Cin Hau” (N3961 L2/10-438), October 27, 2010. <http://std.dkuug.dk/JTC1/SC2/WG2/docs/n3961.pdf>
- Pau, Pum Khan. “Religious conversions reconsidered: Christian missions and indigenous movement among the Zo people in colonial Chin Hills.” <http://www.scribd.com/doc/25568052/Religious-Conversion-Reconsidered>
- “Pu Pau Cin Hau Lai.” Zomi Online Library, Item #200, <http://zomilibrary.com/main/items/show/200>
- Thang, Khoi Lam. 2001. “A Phonological Reconstruction of Proto-Chin”. M.A. thesis. Payap University, Chiang Mai, Thailand. http://ic.payap.ac.th/graduate/linguistics/theses/Khoi_Lam_Thang_Thesis.pdf

7 Acknowledgments

I am indebted to Christopher Button for his willingness to share his knowledge of the northern Chin language family. My understanding of the Pau Cin Hau tone marks owes a great deal to his observations and analysis of the correspondences between the tone marks and tone system of Tedim. He patiently answered my questions and overlooked my ignorance of Chin linguistics. I also thank him for providing me with materials written in the Pau Cin Hau logographic script.

This project was made possible in part by a grant from the United States National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at the University of California, Berkeley). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.

	1184	1185	1186	1187
0	Ⓔ 11840	Ⓒ 11850	Ⓔ 11860	⋮ 11870
1	Ⓔ 11841	Ⓔ 11851	Ⓔ 11861	Ⓔ 11871
2	Ⓒ 11842	Ⓔ 11852	Ⓔ 11862	Ⓔ 11872
3	Ⓔ 11843	Ⓔ 11853	Ⓔ 11863	Ⓔ 11873
4	Ⓔ 11844	Ⓔ 11854	Ⓔ 11864	Ⓔ 11874
5	Ⓔ 11845	Ⓔ 11855	Ⓔ 11865	Ⓔ 11875
6	Ⓔ 11846	Ⓔ 11856	Ⓔ 11866	Ⓔ 11876
7	Ⓔ 11847	Ⓔ 11857	Ⓔ 11867	Ⓔ 11877
8	Ⓔ 11848	Ⓔ 11858	Ⓔ 11868	Ⓔ 11878
9	Ⓔ 11849	Ⓔ 11859	Ⓔ 11869	
A	Ⓔ 1184A	Ⓔ 1185A	Ⓔ 1186A	
B	Ⓔ 1184B	Ⓔ 1185B	Ⓔ 1186B	
C	Ⓔ 1184C	Ⓔ 1185C	Ⓔ 1186C	
D	Ⓔ 1184D	Ⓔ 1185D	Ⓔ 1186D	
E	Ⓔ 1184E	Ⓔ 1185E	Ⓔ 1186E	
F	Ⓔ 1184F	Ⓔ 1185F	Ⓔ 1186F	

Consonants

- 11840 Ⓔ PAU CIN HAU LETTER PA
- 11841 Ⓔ PAU CIN HAU LETTER KA
- 11842 Ⓒ PAU CIN HAU LETTER LA
- 11843 Ⓔ PAU CIN HAU LETTER MA
- 11844 Ⓔ PAU CIN HAU LETTER DA
- 11845 Ⓔ PAU CIN HAU LETTER YA
- 11846 Ⓔ PAU CIN HAU LETTER VA
- 11847 Ⓔ PAU CIN HAU LETTER NGA
- 11848 Ⓔ PAU CIN HAU LETTER HA
- 11849 Ⓔ PAU CIN HAU LETTER GA
- 1184A Ⓔ PAU CIN HAU LETTER XA
- 1184B Ⓔ PAU CIN HAU LETTER HSA
- 1184C Ⓔ PAU CIN HAU LETTER BA
- 1184D Ⓔ PAU CIN HAU LETTER TGA
- 1184E Ⓔ PAU CIN HAU LETTER TA
- 1184F Ⓔ PAU CIN HAU LETTER HTA
- 11850 Ⓒ PAU CIN HAU LETTER NA
- 11851 Ⓔ PAU CIN HAU LETTER HPA
- 11852 Ⓔ PAU CIN HAU LETTER RA
- 11853 Ⓔ PAU CIN HAU LETTER FA
- 11854 Ⓔ PAU CIN HAU LETTER CHA

Vowels

- 11855 Ⓔ PAU CIN HAU LETTER AA
- 11856 Ⓔ PAU CIN HAU LETTER AAI
- 11857 Ⓔ PAU CIN HAU LETTER I
- 11858 Ⓔ PAU CIN HAU LETTER AW
- 11859 Ⓔ PAU CIN HAU LETTER U
- 1185A Ⓔ PAU CIN HAU LETTER UA
- 1185B Ⓔ PAU CIN HAU LETTER IA

Final consonants

- 1185C Ⓔ PAU CIN HAU LETTER AB
- 1185D Ⓔ PAU CIN HAU LETTER AG
- 1185E Ⓔ PAU CIN HAU LETTER AD
- 1185F Ⓔ PAU CIN HAU LETTER AM
- 11860 Ⓔ PAU CIN HAU LETTER AN
- 11861 Ⓔ PAU CIN HAU LETTER AL
- 11862 Ⓔ PAU CIN HAU LETTER AU
- 11863 Ⓔ PAU CIN HAU LETTER ANG
- 11864 Ⓔ PAU CIN HAU LETTER AI

Tone marks

- 11865 Ⓔ PAU CIN HAU TONE MARK-1
- 11866 Ⓔ PAU CIN HAU TONE MARK-2
- 11867 Ⓔ PAU CIN HAU TONE MARK-3
- 11868 Ⓔ PAU CIN HAU TONE MARK-4
- 11869 Ⓔ PAU CIN HAU TONE MARK-5
- 1186A Ⓔ PAU CIN HAU TONE MARK-6
- 1186B Ⓔ PAU CIN HAU TONE MARK-7
- 1186C Ⓔ PAU CIN HAU TONE MARK-8
- 1186D Ⓔ PAU CIN HAU TONE MARK-9
- 1186E Ⓔ PAU CIN HAU TONE MARK-10
- 1186F Ⓔ PAU CIN HAU TONE MARK-11
- 11870 Ⓔ PAU CIN HAU TONE MARK-12
- 11871 Ⓔ PAU CIN HAU TONE MARK-13
- 11872 Ⓔ PAU CIN HAU TONE MARK-14
- 11873 Ⓔ PAU CIN HAU TONE MARK-15
- 11874 Ⓔ PAU CIN HAU TONE MARK-16
- 11875 Ⓔ PAU CIN HAU TONE MARK-17
- 11876 Ⓔ PAU CIN HAU TONE MARK-18
- 11877 Ⓔ PAU CIN HAU TONE MARK-19
- 11878 Ⓔ PAU CIN HAU TONE MARK-20

Figure 1: Proposed code chart and nameslist for the Pau Cin Hau Alphabet.

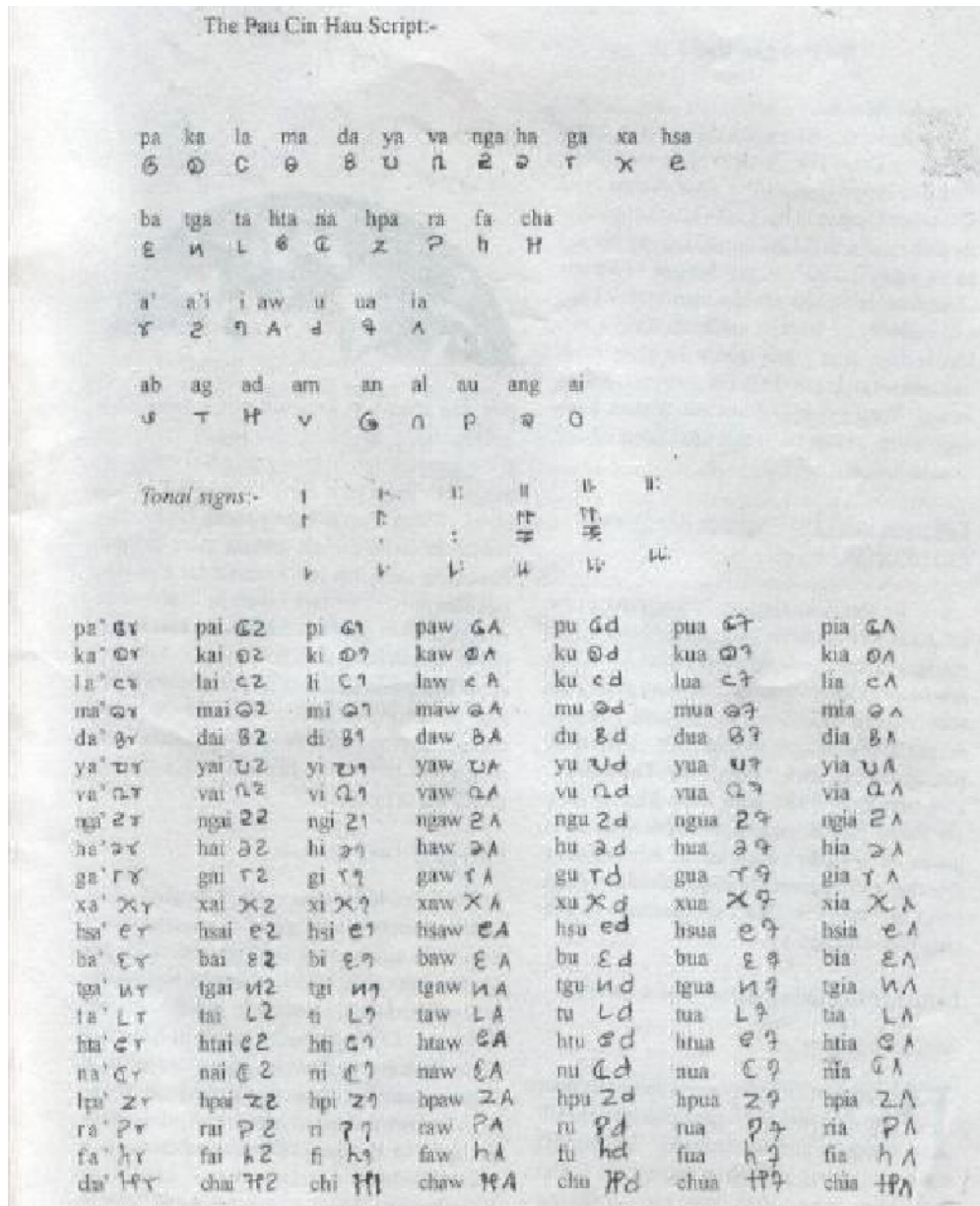


Figure 3: Characters of the Pau Cin Hau Alphabet (from “Pu Pau Cin Hau Lai”: 4).

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹**

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: **Preliminary Proposal to Encode the Pau Cin Hau Alphabet in ISO/IEC 10646**
2. Requester's name: *Script Encoding Initiative (SEI) / Anshuman Pandey <pandey@umich.edu>*
3. Requester type (Member body/Liaison/Individual contribution): *Liaison contribution*
4. Submission date: *2010-10-27*
5. Requester's reference (if applicable):
6. Choose one of the following:
This is a complete proposal: _____
(or) More information will be provided later: *Yes*

B. Technical – General

1. Choose one of the following:
 - a. This proposal is for a new script (set of characters): *Yes*
Proposed name of script: *Pau Cin Hau Alphabet*
 - b. The proposal is for addition of character(s) to an existing block:
Name of the existing block: _____
2. Number of characters in proposal: *57*
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. 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*
5. Fonts related:
 - a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard? *Anshuman Pandey, author of the proposal*
 - b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): *Anshuman Pandey <pandey@umich.edu>*
6. 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*
7. 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, see proposal for additional details*

8. 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.

¹ Form number: N3702-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain	No
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? If YES, available relevant documents:	Yes <i>Dr. Christopher Button <chris.button@hotmail.com></i> <i>See citations in text of proposal</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	Yes <i>See text of proposal</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	Rare <i>See text of proposal</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	Yes <i>Limited use by Laipian community in Chin State, Myanmar</i>
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:	No
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	Yes
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:	No
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:	No
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:	No
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:	No
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)	No
13. Does the proposal contain any Ideographic compatibility character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? If YES, reference:	No