Doc Type:	**Working Group Document**
Title:	**Proposal to Disunify Khamti Style Letters from Myanmar**
Source:	**Martin Hosken, Stephen Morey**
Status:	**Individual contribution**
Action:	**For consideration by UTC and WG2**
Date:	**2014-04-23**

Executive Summary This proposal is to disunify the Khamti and Aiton and Phake style Myanmar consonants (those with the dots) into their own letters as part of the Myanmar script block, and thus create a new Myanmar Extended block to hold them in. The effect is to add 16 characters. In addition the representative glyphs for Khamti specific letters in the Myanmar Extended block, are adjusted to have the Khamti style.

Introduction There are a number of stylistic variants to the use of the Myanmar script across the different languages of Burma. One significant style, which we will call the Khamti style[1], adds dot terminals to characters. Thus:

Burmese Style   Khamti Style

က   က

Most Khamti, Aiton and Phake users living in Burma are also fluent in Burmese and Shan, and use those languages, as well as their own language, on a computer. In a plain text context (such as is most commonly used, including Facebook, SMS, email) where these languages are being used, the Burmese style of characters gets used exclusively. This is because it makes even less sense to view Burmese using Khamti style characters than to view Khamti using the dotless Burmese style. This has the effect of users rarely seeing their language written in an appropriate style. There is no option to select an appropriate font since the same codepoints are being used for Burmese as for Khamti and so, in a plain text context, there is no way to see the two styles. The communities, therefore, have a real concern that a significant aspect of their cultural heritage, tied up in their script, will be lost. They therefore request that the characters that have a Khamti style be disunified from their Burmese equivalents, so that in multi-lingual plain text, the contrast may be conserved.

This language mix has been evident textually for a long time. Disunifying these characters will facilitate the better handling of texts in plain text both for analysis and rendition.

In addition to adding dots to various consonants, the Khamti style also fills in the various dot type characters found in the Myanmar block. This dot filling is also a style that is used in other languages, including Burmese. The danger, therefore, is that introducing new codepoints for these other forms will cause confusion for users of other languages. There would be a temptation to use the codepoint that represents the presentation that a user desires over using an appropriate styling. In addition, in a plain text context, inappropriate styling of the dot type characters is much more acceptable to inappropriate styling of the base character consonants. Rendering Burmese with filled dots or Khamti with unfilled dots, while perhaps not ideal, would be accepted much more than rendering the consonants inappropriately. For analysis purposes, only disunifying the consonants is sufficient for language identification in plain text..

---

1   For brevity we use the term 'Khamti style' to cover Khamti, Aiton and Phake styles. There is some difference, but in general they are the same. Likewise 'Burmese' for 'Burmese and Shan'

The impact on existing text of such a disunification is minimal. There is, as yet, little text in these languages, using Unicode. Although, there is sufficient interest in using Unicode, that this situation will soon change. Existing text, using the Burmese codepoints, will continue to render as before using language specific styling. The only situation is where old text needs to interact with new text in terms of searching. Such long term texts will need to be transcoded with simple search and replace. The greater concern is how long the transition will take before implementation can effectively occur.

Proposal The disunification creates 16 new characters thus requiring a new block allocation of 1 column. Since this is not a new script and the existing script is all in the BMP, it is proposed that the new columns come from the BMP. One option for finding such space is to place the block immediately preceding the existing Myanmar block, either as a new block or as a change to the existing block structure. While it is odd starting a major block not on a XX00 boundary, the BMP is rather full. This proposal will be written in terms of such a new block, but the block can be moved.

 The new characters inherit their properties from the existing characters and will have compatibility decompositions to those characters.

```
0FF0;MYANMAR LETTER KHAMTI KA;Lo;0;L;<compat> 1000;;;;N;;;;;
0FF1;MYANMAR LETTER KHAMTI KHA;Lo;0;L;<compat> 1002;;;;N;;;;;
0FF2;MYANMAR LETTER KHAMTI NGA;Lo;0;L;<compat> 1004;;;;N;;;;;
0FF3;MYANMAR LETTER KHAMTI TA;L0;0;L;<compat> 1010;;;;N;;;;;
0FF4;MYANMAR LETTER AITON THA;L0,0,L;<compat> 1011;;;;N;;;;;
0FF5;MYANMAR LETTER KHAMTI PA;Lo;0;L;<compat> 1015;;;;N;;;;;
0FF6;MYANMAR LETTER KHAMTI MA;Lo;0;L;<compat> 1019;;;;N;;;;;
0FF7;MYANMAR LETTER KHAMTI YA;Lo;0;L;<compat> 101A;;;;N;;;;;
0FF8;MYANMAR LETTER KHAMTI LA;Lo;0;L;<compat> 101C;;;;N;;;;;
0FF9;MYANMAR LETTER AITON WA;Lo;0;L;<compat> 101D;;;;N;;;;;
0FFA;MYANMAR LETTER KHAMTI A;Lo;0;L;<compat> 1022;;;;N;;;;;
0FFB;MYANMAR VOWEL SIGN AITON E;Mc;0;L;<compat> 1031;;;;N;;;;;
0FFC;MYANMAR LETTER KHAMTI SHAN KA;Lo;0;L;<compat> 1075;;;;N;;;;;
0FFD;MYANMAR LETTER KHAMTI SHAN CA;Lo;0;L;<compat> 1078;;;;N;;;;;
0FFE;MYANMAR LETTER AITON SHAN NYA;Lo;0;L;<compat> 107A;;;;N;;;;;
0FFF;MYANMAR LETTER KHAMTI SHAN THA;Lo;0;L;<compat> 1080;;;;N;;;;;
```

There are 4 characters in the list which are not used by Khamti, where Khamti uses the undotted Burmese form, while Aiton and Phake use dots in their characters. The intent is that the Khamti would use only the codes required for Khamti, while the Aiton and Phake would use all of them. Also the Khamti style their characters slightly differently from Aiton and Phake. There is no intent to provide a plain text contrast between these languages.

The E Vowel sign is proposed for disunification on the grounds of keeping an appropriate style with the consonants.

In addition to the disunified characters, we propose updates to the representative glyphs of characters only used in Khamti style. The characters involved are: AA61, AA62, AA63, AA64, AA65, AA66, AA6B, AA6C, AA6F. While this does not necessarily cover all the styling changes for Aiton and Phake, showing the Khamti style indicates which glyphs would be changed for a typical font. The Aiton and Phake variants can be implemented without having to follow the representative glyphs slavishly.

If this proposal is accepted, it is requested that the proposal be fast tracked due to the block change. The characters interact with others in other Myanmar blocks and therefore it is important that the block be allocated and implementations be updated to support the block as part of the Myanmar script otherwise text runs will be broken within a text.

	0FF
0	ဟ
1	ၡ
2	ၢ
3	ၣ
4	ၤ
5	ၥ
6	ၦ
7	ၧ
8	ၨ
9	ၩ
A	ၪ
B	ၫ◌
C	ၬ
D	ၭ
E	ၮ
F	ၯ

	AA6	AA7
0	ကာ	⌢
1	ကာ	ကာ
2	ဟ	ကာ
3	ဃာ	၃
4	ဖာ	ဖျ
5	ဖာ	ဖျာ
6	၆ာ	ကျာ
7	၀၀	ဖျ
8	၆ၓ	ဖ
9	၆၀	၄
A	၀ၓ	၆ဩ
B	၅	◌ေ
C	၈	◌
D	၇	◌
E	၆	၄ျ
F	၆၀	၄ျ

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[2]
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:**	*Khamti Variants*
2. Requester's name:	*Martin Hosken*
3. Requester type (Member body/Liaison/Individual contribution):	*Individual contribution*
4. Submission date:	*30/Apr/2014*
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	*X*
(or) More information will be provided later:	

B. Technical – General

1. Choose one of the following:
    a. This proposal is for a new script (set of characters):
        Proposed name of script:
    b. The proposal is for addition of character(s) to an existing block: *X*
        Name of the existing block: *Myanmar*

2. Number of characters in proposal: *16*

3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary *X*	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. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? *SIL*
    If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

6. References:
    a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *no*
    b. Are published examples of use (such as samples from newspapers, magazines, or other sources)
    of proposed characters attached? *no*

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)? *no*

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.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? — *no*
   If YES explain

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.)? — *yes*
   If YES, with whom? — *Stephen Morey, Khamti community*
   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? — *no*
   Reference:

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

5. Are the proposed characters in current use by the user community? — yes
   If YES, where? Reference:

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely
   in the BMP? — *yes*
   If YES, is a rationale provided? — *yes*
   If YES, reference: — *addition to existing BMP*

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? — *no*

8. Can any of the proposed characters be considered a presentation form of an existing
   character or character sequence? — *yes*
   If YES, is a rationale for its inclusion provided? — *yes*
   If YES, reference: — *This document*

9. Can any of the proposed characters be encoded using a composed character sequence of either
   existing characters or other proposed characters? — *no*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:

10. Can any of the proposed character(s) be considered to be similar (in appearance or function)
    to an existing character? — *yes*
    If YES, is a rationale for its inclusion provided? — *yes*
    If YES, reference: — *this document*

11. Does the proposal include use of combining characters and/or use of composite sequences? — *no*
    If YES, is a rationale for such use provided?
    If YES, reference: — *no*
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? — *no*
    If YES, reference:

12. Does the proposal contain characters with any special properties such as
    control function or similar semantics? — *no*
    If YES, describe in detail (include attachment if necessary)

13. Does the proposal contain any Ideographic compatibility character(s)? — *no*
    If YES, is the equivalent corresponding unified ideographic character(s) identified?
    If YES, reference: