L2/13-157

Title:
A proposal to encode the Akarmatrik music notation symbols in UCS

Author:
Chandan Misra

Submission Date:
7-26-2013

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
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:** *Proposal to encode Akarmatrik notation symbols in UCS*
2. Requester's name: *Chandan Misra*
3. Requester type (Member body/Liaison/Individual contribution): *Individual contribution*
4. Submission date: *5-20-2013*
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: *Yes*
   (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: *Yes*
      Name of the existing block: *Bengali*
2. Number of characters in proposal: *6*
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?
   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?
   Chandan Misra
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
   Chandan Misra
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*

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 Unicode Character Database (http://www.unicode.org/reports/tr44/) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

[1] Form number: N4102-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, 2011-03, 2012-01)

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.)? *No*
   If YES, with whom?
   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*
   Reference: *Proposal document*
4. The context of use for the proposed characters (type of use; common or rare) *common*
   Reference: *Proposal document*
5. Are the proposed characters in current use by the user community? *Yes*
   If YES, where? Reference: *Worldwide, http://tagoreweb.in, http://www.nltr.org, published books*
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: *Proposal document*
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? *No*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:
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, or could be confused with, an existing character? *Yes*
    If YES, is a rationale for its inclusion provided? *Yes*
    If YES, reference: *Proposal 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:
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
    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 characters? *No*
    If YES, are the equivalent corresponding unified ideographic characters identified?
    If YES, reference:

Proposal to encode Akarmatrik music notation symbols in the UCS

Rabindrasangeet is arguably the most popular form of music for about 200 millions Bengalis around the world. Native speakers of the language inhabit Bangladesh, the states of West Bengal, Tripura and Jharkhand in India; sizable number of Bengalis are also found in countries such as the UK and the USA. Rabindrasangeet refers the body of about 2200 songs written by *Rabindranath Tagore* (1861-1941) mostly as standalone pieces but also for his twenty seven dance dramas and operas. The music is scored in the *Akarmatrik* form of transcription which arranges the notes along with the lyrics for the vocal performance, including repetitions, variations etc. Any instrumentation or other arrangement is not specified. The scoring system is mainly meant to aid the vocal performer.

This *Akarmatrik* system was developed in 1905 by Rabindranath's elder brother *Jyotirindranath Tagore* which fused the tools of western staff notation with features of Indian classical music system into a unique and very effective and compact scoring system for the vocalist. Between 1889 and 1905 he made various experimentations for a proper technique for writing the musical arrangement and his first arrangement was published in 1897. This *Akarmatrik* system for the music is in practice world-wide in a standard manner by a huge user group and its usefulness can be increased if they were encoded in a compact way, that is, as character codes.

This proposal proposes 6 code points for Akarmatrik music notation symbols in the existing Bengali block (Range: 0980 - 09FF) in BMP. The table *Complete Outline of Akarmatrik Music Notation Symbols* below shows all 51 characters required to represent notation symbols in Unicode.

Table 1: Complete Outline of Akarmatrik Music Notation Symbols

Sl. No.	Name	Appearance	Unicode
1.	Akarmatrik Suddha Shadaj	স	U+09B8
2.	Akarmatrik Suddha Rishabha	র	U+09B0
3.	Akarmatrik Suddha Gandhar	গ	U+0997
4.	Akarmatrik Suddha Madhyam	ম	U+09AE
5.	Akarmatrik Suddha Pancham	প	U+09AA
6.	Akarmatrik Suddha Dhaivat	ধ	U+09A7
7.	Akarmatrik Suddha Nishad	ন	U+09A8
8.	Akarmatrik Komal Rishabha	ঋ	U+098B
9.	Akarmatrik Komal Gandhar	জ্ঞ	Not in Unicode
10.	Akarmatrik Kari Madhyam	ক্ষ	Not in Unicode
11.	Akarmatrik Komal Dhaivat	দ	U+09A6
12.	Akarmatrik Komal Nishad	ণ	U+09A3
13.	Akarmatrik Mandra Saptak Symbol	়	U+0316
14.	Akarmatrik Taar Saptak Symbol	◌́	U+0954 or U+0301
15.	Akarmatrik Ekmatra Symbol	া	U+09BE
16.	Akarmatrik Ardhamatra Symbol	ঃ	U+0983
17.	Akarmatrik Sikimatra Symbol	◌	Not in Unicode
18.	Meend	◌͡◌	Not in Unicode
19.	Akarmatrik Starting Gumpha Bandhani Symbol	{	U+007B
20.	Akarmatrik Ending Gumpha Bandhani Symbol	}	U+007D
21. and 22.	Akarmatrik Swara Cancellation Symbols While Repetition	(and)	U+0028 and U+0029
23.	Akarmatrik Avasan Symbol	‖	Not in Unicode
24.	Akarmatrik Khali or Phaank Symbol	০	U+09E6
25. to 33.	Akarmatrik Taalanka or Atikomal or Anukomal Symbols	১২৩৪৫৬৭৮৯	U+09E7 - U+09EF
34. to 45.	Akarmatrik Sparsha Swaras	Same as symbols 1 – 12 but used as superscripts	
46.	Akarmatrik Yugal Danda Symbol	Ⅱ	U+2161
47. and 48.	Akarmatrik Melody change Symbol	[and]	U+005B and U+005D
49.	Akarmatrik Danda Symbol	I	U+2160
50.	Akarmatrik Taala Bibhaga Symbol	|	Not in Unicode
51.	Akarmatrik Hyphen Symbol	-	U+2010

There are altogether 51 symbols in the above table. Among these 51 symbols, 41 symbols are already defined in the Unicode chart. Now there are certain symbols like 9 and 10 which were not in the Unicode but they are compound characters of Bengali language. So, these two notes and their grace notes (total 4 symbols) need not to be encoded. So, in 51 symbols we have only 6 symbols which need to be encoded.

The justification of each symbol (marked 'not in Unicode' in above table) to encode it in UCS is given below –

- Symbol 9 and 10 are not in Unicode. But they are written using three symbols which are in Unicode. Symbol 9 can be written as U+0999 + U+09CD + U+0997. Similarly symbol 10 can be written as U+09B9 + U+09CD + U+09AE. Hence these symbols need not to be encoded in Unicode.
- Symbol 17 or *Akarmatrik Sikimatra symbol* is similar to Bengali digit zero. But the use of this symbol is quite different. Digit zero is used to denote a phaank or khali (symbol 24) which is always the first beat of a particular Bibhaga. On the other hand Sikimatra symbol which is standalone by nature, most of the time it is associated with a note or Swara to play it for one quarter of the duration of a whole note. Using same Unicode code point for two different purposes is complex and erroneous. Moreover Sikimatra symbol is slightly smaller than its counterpart as shown in **Figure 1**. So, one code point is needed for it.
- Though Akarmatrik Meend symbol or symbol 18 is depicted as a single glyph but in places the width of this symbol varies as depicted in the **Figure 2** where some span across two notes while some span across three notes. That is why it is divided into three glyphs for application purpose like **Figure 3** so that it can be stretched as per our need. Each glyph can be considered as a new symbol here. So, three code points are needed for them.
- Symbol 23 or Akarmatrik Avasan symbol (**Figure 4**) is not in Unicode. It is similar to U+1D101. But complication will arise if developers want to create western and Indic music sheet in a single application. The same is true for any web page displaying any Western score along with its Indic counterpart. In that case only one code point is insufficient for rendering both the symbols depicting different meaning. So, one code point is needed for it.
- Akarmatrik Taala Bibhaga symbol (**Figure 5**) looks similar to the symbol Devanagari Danda (U+0964). But careful observation reveals that Akarmatrik Taala Bibhaga symbol is taller than U+0964. One probable reason for making it little bit taller than usual character size is that it divides two matras of a Taala. Similarities between symbols among different notation systems in India also support this fact. For example in Bhatkhande notation system Taala Bibhaga symbol stretches from top to bottom of a page as in **Figure 6**. So, one code point is needed for it.
- Symbol 46 or Akarmatrik Yugal Danda Symbol and symbol 49 or Akarmatrik Danda Symbol are similar to Roman numeral U+2161 and U+2160 respectively as shown in **Figure 7**. But it is recommended in Unicode technical note #19 to avoid the use of characters identified as Roman numerals. So, two code points are needed for the two characters.

Table 2: Code points and Akarmatrik music notation symbols

Hex	Name
09D0	AKARMATRIK SIKIMATRA SYMBOL
09D1	AKARMATRIK MEEND SYMBOL START
09D2	AKARMATRIK MEEND SYMBOL CONTINUE
09D3	AKARMATRIK MEEND SYMBOL END
09D4	AKARMATRIK AVASAN SYMBOL
09D5	AKARMATRIK TAALA BIBHAGA SYMBOL
09D6	AKARMATRIK DANDA SYMBOL
09D8	AKARMATRIK YUGAL DANDA SYMBOL

Appendix

Figure1: Digit zero (squared) and Sikimatra symbol (circled) [1]

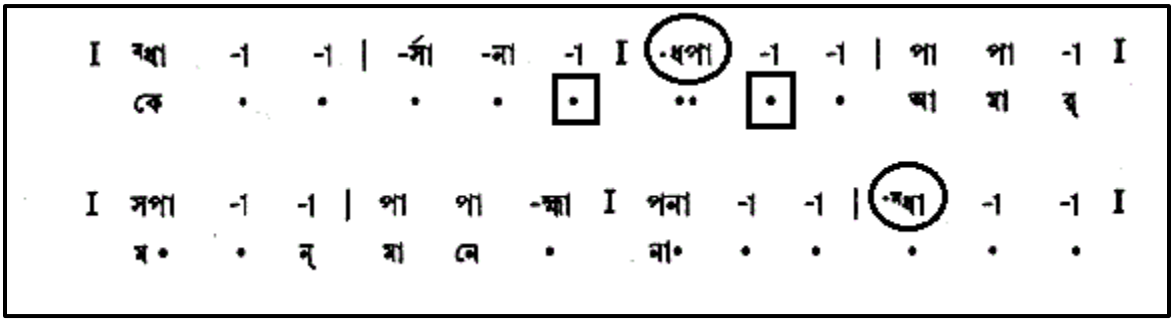


Figure 2: Example of Akarmatrik Meend Symbol [1]

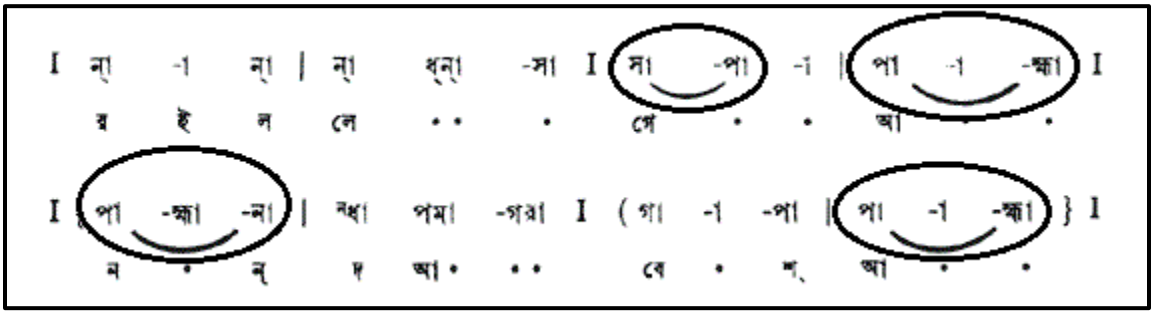


Figure 3: Three parts of Akarmatrik Meend Symbol

Figure 4: Example of Akarmatrik Avasan Symbol [1]

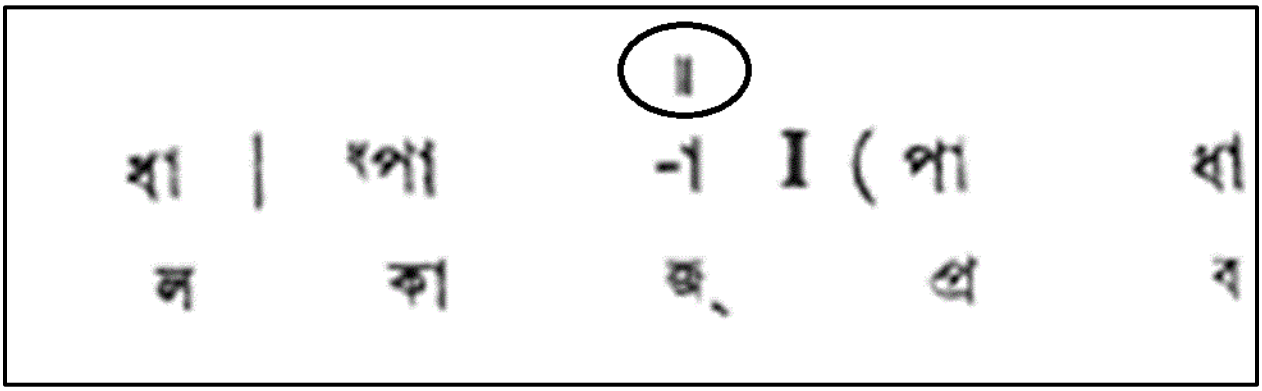


Figure 5: Example of Akarmatrik Taala Bibhaga Symbol [1]

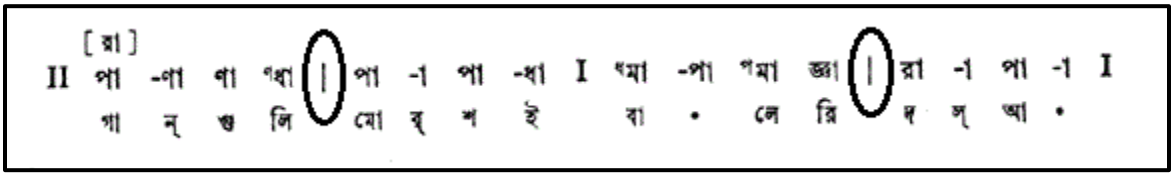


Figure 6: Taala Bibhaga Symbol in Bhatkhande notation system [2]

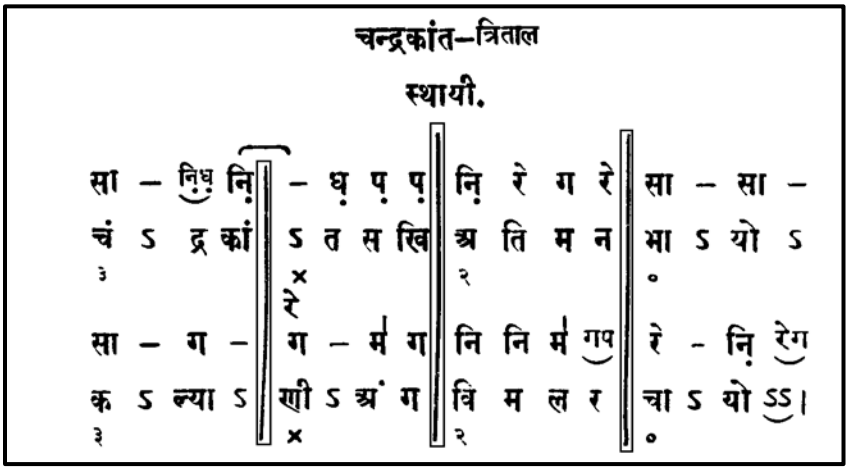
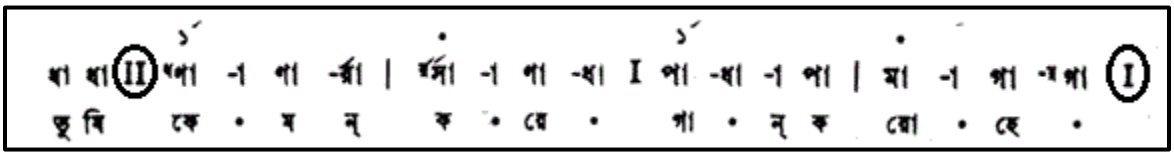


Figure 7: Akarmatrik Danda and Yugal Danda symbol [1]



Bibliography:

[1] Scan copy of music sheet of Tagore songs published by Visva-bharti University. Online repository can be found in http://tagoreweb.in

[2] Scan copy of music sheet of Pt. Vishnu Narayan Bhatkhande archived in Digital Library of India, http://www.dli.ernet.in/, accessed on 25.7.2013.