

Proposal to Encode Three Balinese Punctuation Marks

Rikza F. Sh.
rikza.f.sh@gmail.com

Aditya Bayu Perdana
jimeildotkomm@gmail.com

January 2022

1. INTRODUCTION

The island of Bali has a long literary history, with extensive literature in Sanskrit, Kawi (Old Javanese), and Balinese language traditionally written with the Balinese script. While today most everyday writing in Bali uses the Latin alphabet, the Balinese script is still taught in the local curriculum and used on certain occasions. The script is currently well-supported in Unicode, and nearly all traditional texts can be accurately encoded. However, in the process of researching some older documents, we have discovered a number of punctuation marks currently not encoded in Unicode. The current Balinese block still has four vacant slots, three of which can be filled with these marks.

2. REQUEST

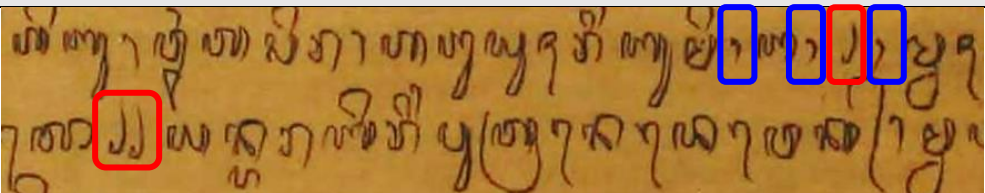
This proposal requests the addition of three new characters in the Balinese block with the following names and code points:

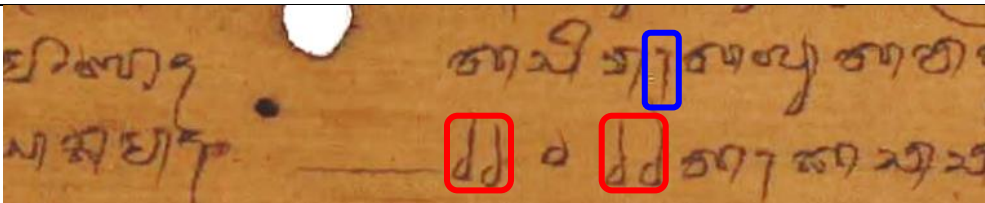
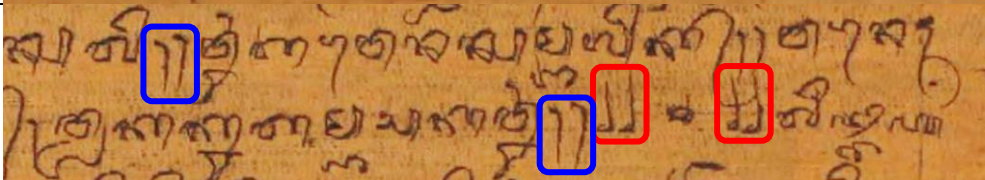
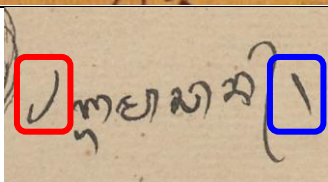
U+1B4E	↘	BALINESE INVERTED CARIK SIKI
U+1B4F	↘↘	BALINESE INVERTED CARIK PAREREN
U+1B7F	↪	BALINESE PANTI BAWAK

3. ATTESTATION

Inverted Carik Siki and Carik Pareren

Inverted versions of CARIK SIKI (U+1B5E) and CARIK PAREREN (U+1B5F) can be found contrastively with their regular counterpart. So far, we have only encountered these characters in handwritten documents. Examples can be seen below enclosed in red square (characters in blue square are their upright counterpart):

Source	Image
Kramaning Alaki Arabi , f. 2A	

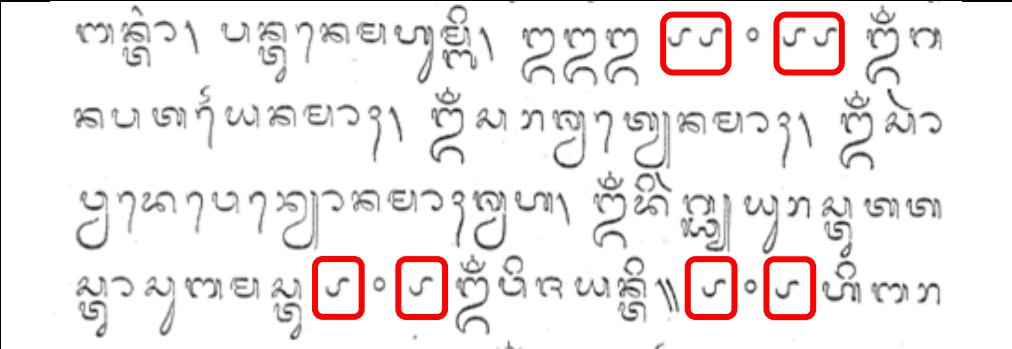
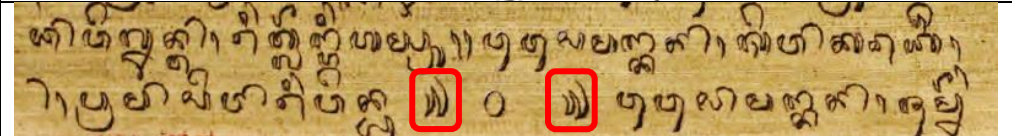
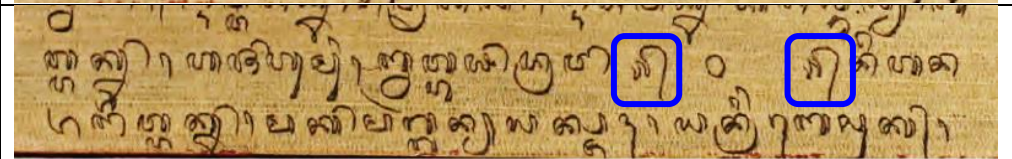
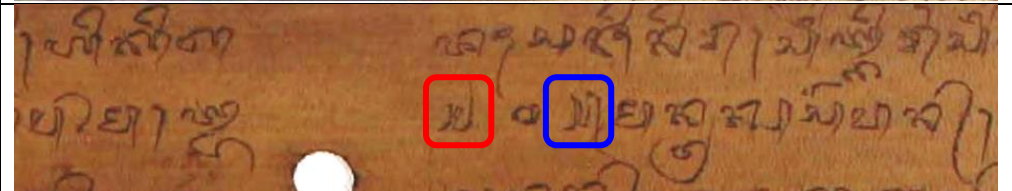
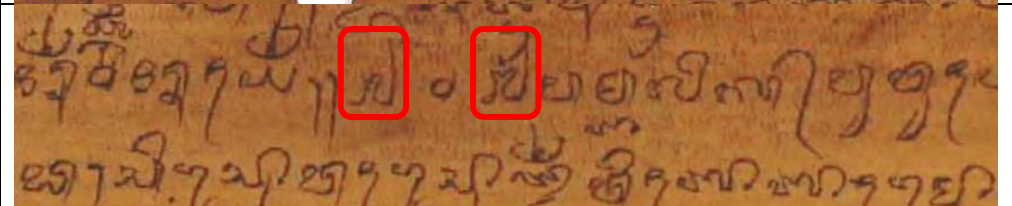
Krahak , f. 11A	
Krahak , f. 18B	
Painting of Geguritan Gunawati, Or. 3390: 185	

According to one contemporary scribe, these characters seemed to be used to indicate passages with some degree of importance or a finer division within the text. Scribe might opted to use these characters when other punctuations are already reserved for other purpose. Like their upright counterpart, both inverted CARIK can also be used in sequence with PADA WINDU (U+1B5C).

Panti Bawak

A truncated version of PANTI (U+1B5A) can be found used contrastively with its regular counterpart. This character can be encountered in handwritten and printed documents. Examples can be seen below enclosed in red square (characters in blue square is its regular counterpart):

Source	Image
70 Weda's, Catalogue No. 1406 in Brandes (1901)	

Kakawin Bharatayuddha Lontar 4, Catalogue No. 205 in Brandes (1901)	
Pawintenan Sastra Ring Sarira , f. 9B	
Pawintenan Sastra Ring Sarira , f. 10A	
Krahah , f. 7b	
Krahah , f. 11a	

This character may serve the same purpose as the aforementioned inverted CARIK in providing finer division of text under the discretion of the scribe. We named this character PANTI BAWAK, *bawak* being Balinese for “short” or “abbreviated”, in contrast to the already encoded PANTI LANTANG (U+1B7D), *lantang* being Balinese for “long”.

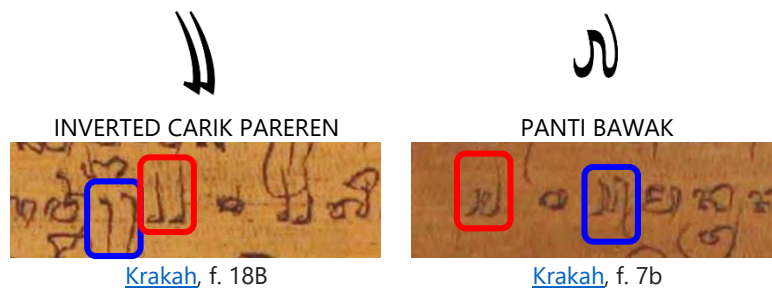
4. CHARACTER DATA

Glyph Confusability

In typographic representation, the proposed characters can be made by modifying existing characters into distinct glyphs (here using Vimala font as example):

 U+1B5E	 INVERTED CARIK SIKI	 U+1B5F	 INVERTED CARIK PAREREN	 U+1B5A	 PANTI BAWAK
---	---	---	--	---	--

However, manuscript samples show that some glyphs may be confusable with one another. The glyph of INVERTED CARIK PAREREN may be confusable with HALF PANTI as so:



Looking at the regular counterpart of these characters (namely CARIK PAREREN and PANTI) might help establish which is which, but the difference might not be obvious at first glance.

The glyph of INVERTED CARIK SIKI may be confusable with MUSICAL SYMBOL DUNG (U+1B63) in isolation. But as the latter is only used in musical notation, the difference can be more readily discerned from context:



Character Data

In UnicodeData.txt format:

```
1B4E;BALINESE INVERTED CARIK SIKI;Po;0;L;;;;N;;;;;
1B4F;BALINESE INVERTED CARIK PAREREN;Po;0;L;;;;N;;;;;
1B7F;BALINESE PANTI BAWAK;Po;0;L;;;;N;;;;;
```

5. BIBLIOGRAPHY

Brandes, Dr. J., 1901. *Beschrijving der Javaansche, Balineesche en Sasaksche Handschriften*. Batavia: Landsdrukkerij.

Suasta, I B Made. 1996. *Modernisasi dan Pelestarian Perkembangan Metode dan Teknik Penulisan Aksara Bali*. Jakarta: Departemen Pendidikan dan Kebudayaan.

**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://std.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://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title:	Proposal to Encode Three Balinese Punctuation Marks		
2. Requester's name:	<i>Rikza F. Sh. And Aditya Bayu Perdana</i>		
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>		
4. Submission date:			
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):			NO
Proposed name of script:			
b. The proposal is for addition of character(s) to an existing block:			YES
Name of the existing block:	<i>Balinese</i>		
2. Number of characters in proposal:			3
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?			
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?			
5. Fonts related:			
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Aditya Bayu Perdana</i>		
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Aditya Bayu Perdana, OFL</i>		
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		
<i>See proposal</i>			

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: N4502-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.)?	YES
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: <i>See proposal</i>	
4. The context of use for the proposed characters (type of use; common or rare)	common
Reference: <i>Used in many traditional Balinese manuscripts and their modern transcriptions</i>	
5. Are the proposed characters in current use by the user community?	YES
If YES, where? Reference: <i>Modern copies and transcriptions of traditional manuscripts</i>	
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	NO
If YES, is a rationale provided?	
If YES, reference:	
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	
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?	
If YES, reference: <i>See proposal</i>	
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:	