

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

TITLE: A PROPOSAL TO ADD DESCRIPTION FOR BOPOMOFO USAGE TO FOUR COMBINING CHARACTERS, AND TO ENCODE TWO BOPOMOFO COMBINING TONE MARKS
SOURCE: TCA, BOBBY TUNG, SELENA WEI
STATUS: INDIVIDUAL CONTRIBUTION
ACTION: FOR CONSIDERATION BY JTC1/SC2/WG2 AND UTC
DATA: 2019-06-14

1. Introduction

This is a proposal for two purposes. One is to add description for Bopomofo usage to four Combining characters. Another purpose is to encode two Bopomofo Combining tone marks

(, ).

2. Issues

(1) Add descriptions to four characters for Bopomofo Combining tone marks' usage

Tone marks' of Bopomofo - a phonetic system for Chinese - had been encoded into code points in 'Combining Diacritical ' block. We'd like to add descriptions to four characters for Bopomofo Combining tone marks' usage as follows as boldface:

codepoint	name / descriptions	glyph
U+0301	COMBINING ACUTE ACCENT = stress mark = Greek oxia, tonos = Mandarin Chinese second tone → 0027 apostrophe → 00B4 acute accent → 02B9 modifier letter prime → 02CA modifier letter acute accent → 0384 greek tonos	
U+030C	COMBINING CARON = hacek, V above	

	= Mandarin Chinese third tone → 02C7 caron	
U+0300	COMBINING GRAVE ACCENT = Greek varia = Mandarin Chinese fourth tone → 0060 grave accent → 02CB caron	
U+0307	COMBINING DOT ABOVE derivative (Newtonian notation) = Min-Nan Chinese eighth tone • IPA(withdrawn in 1976): palatalization → 02D0 dot above	

(2) *Encode two Bopomofo Combining tone marks for Dialect*

Two Spacing tone marks used for Dialect had been encoded as U+02EB (f) and U+02EA (L). But there are no correspond Combining marks now. To complete it, we want to encode two Bopomofo Combining tone marks for Dialect as follows:

name	glyph
COMBINING YANG DEPARTING TONE MARK	
COMBINING YIN DEPARTING TONE MARK	

Bopomofo system has its own layout rule for print and web published by Ministry of Education, Taiwan. Not only text books for elementary school , educational materials should follow the rule including digital content as web page and ebooks.

To fulfill the requirement, we consider to use OpenType feature to adjust tone marks' position. The mechanisms should work with web technology and applications that support latest OpenType layout features. And we also like to extend the coverage not only for Mandarin but also Dialects like Hakka(客家語) and Hokkein(閩南語, Min-nan). But it is not semantically right to use current characters with OpenType features. It will be better to use characters in 'Combining Diacritical' block. That's why we want to encode two Bopomofo Combining tone marks for Dialect.

3. *Unicode Properties*

COMBINING YANG DEPARTING TONE MARK;Lo;0;L;;;;N;;;;;

COMBINING YIN DEPARTING TONE MARK;Lo;0;L;;;;N;;;;;

4. Figures

Fig 1. Sample of Bopomofono tone marks' usage in vertical writing mode
(吳昭新,《新編台灣閩南語用字彙編》, 2013.09, ISBN 978-9574308002(pdf))



ㄗ

- 華音不用這個符號
- 台音『會^ㄗ曉^ㄍ』的『會^ㄗ』

ㄗ

- 華音不用這個符號
- 台音『紅^ㄉ嬰^ㄗ仔^ㄩ』的『嬰^ㄗ』

115

ㄨ

- 華音不用這個符號
- 台音『暗^ㄨ時^ㄣ』的『暗^ㄨ』

ㄨ

- 華音不用這個符號
- 台音用『掩^ㄨ嘴^ㄗ』的『掩^ㄨ』

117

Fig 2. Sample of Bopomofo tone marks' usage in horizontal writing mode
(遠流出版社《臺灣語語音入門》, 2001.01, ISBN 9573242354)



董忠司◎著

字調的舉例

下面的聲調舉例請聽每一個詞的末字，請跟著 CD 發音，同時體會聲音的高低、長短：

7-1 第一調 1 (¹) 次高平調 (高平調)

千秋	く ㄋ	く ㄨ	chian ¹ chiu ¹
鞦韆	く ㄨ	く ㄋ	chiu ¹ chian ¹
觀光	ㄍ ㄨ	ㄍ ㄨ	kuan ¹ kong ¹
公關	ㄍ ㄨ	ㄍ ㄨ	kong ¹ kuan ¹

7-2 第二調 2 (²) 高降調

手股	く ㄨ、	ㄍ ㄨ、	chiu ² koo ²
鼓手	ㄍ ㄨ、	く ㄨ、	koo ² chiu ²
馬屎	ㄨ ㄨ、	ㄨ ㄨ、	be ² sai ²
駛馬	ㄨ ㄨ、	ㄨ ㄨ、	sai ² be ²

7-3 第三調 3 (³) 低降調 或 低平調

愛意	ㄨ ㄨ	ㄨ ㄨ	ai ³ i ³
意愛	ㄨ ㄨ	ㄨ ㄨ	i ³ ai ³
政見	ㄨ ㄨ	ㄍ ㄨ	cing ³ kian ³
見證	ㄍ ㄨ	ㄨ ㄨ	kian ³ cing ³

7-4 第四調 4 (ˋ) 中降短調

法國	ㄉㄨㄛˋ ㄩㄢˋ	ㄎㄨㄛˋ	huat ⁴ kok ⁴
國法	ㄎㄨㄛˋ	ㄉㄨㄛˋ ㄩㄢˋ	kok ⁴ huat ⁴
出脫	ㄘㄨㄛˋ	ㄘㄨㄛˋ ㄩㄢˋ	chut ⁴ thuat ⁴
脫出	ㄘㄨㄛˋ ㄩㄢˋ	ㄘㄨㄛˋ	thuat ⁴ chut ⁴

7-5 第五調 5 (ˊ) 中升調

魚鱸	ㄩˊ ㄌㄩˊ	ㄎㄩˊ	hi ⁵ ki ⁵
旗魚	ㄎㄩˊ	ㄩˊ ㄌㄩˊ	ki ⁵ hi ⁵
原油	ㄍㄨㄢˊ	ㄩˊ ㄍㄨㄢˊ	guan ⁵ iu ⁵
猶原	ㄩˊ ㄍㄨㄢˊ	ㄍㄨㄢˊ	iu ⁵ guan ⁵

7-6 第七調 7 (ˊ) 中平調

戶櫂(門檻)	ㄕㄨˊ ㄉㄨˊ	ㄉㄨˊ ㄌㄨˊ	hoo ⁷ ting ⁷
訂戶	ㄉㄨˊ ㄌㄨˊ	ㄕㄨˊ ㄉㄨˊ	ting ⁷ hoo ⁷
會議	ㄕㄨˊ ㄉㄨˊ	ㄍㄨˊ ㄉㄨˊ	hue ⁷ gi ⁷
議會	ㄍㄨˊ ㄉㄨˊ	ㄕㄨˊ ㄉㄨˊ	gi ⁷ hue ⁷

7-7 第八調 8 (ˊ) 高短調

十六	ㄉㄨˊ ㄌㄨˊ	ㄕㄨˊ ㄉㄨˊ	cap ⁸ lak ⁸
----	---------	---------	-----------------------------------

Fig 3. Sample of Bopomofo tone marks layout adjustment in HTML ruby structure

Rendered display

Structure data

U+0301 : `<ruby>條<rt>ㄊㄞˊ ㄅㄠˋ</rt></ruby>`

U+030C : `<ruby>火<rt>ㄏㄨㄛˋ ㄉㄜˊ</rt></ruby>`

U+0300 : `<ruby>到<rt>ㄉㄠˋ ㄅㄠˋ</rt></ruby>`

U+0307 : `<ruby>食<rt>ㄕㄞˊ ㄩˋ ㄖˊ</rt></ruby>`

`<ruby>第<rt>ㄉㄞˋ ㄙㄜˊ</rt></ruby>`

`<ruby>厝<rt>ㄉㄞˋ ㄇㄞˊ</rt></ruby>`

4. Reference

The Manual of Phonetic Symbols of Mandarin Chinese (國語注音符號手冊)

https://language.moe.gov.tw/001/Upload/files/site_content/M0001/juyin/index.html

**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 Add description for Bopomofo usage to four Combining characters, AND to encode two Bopomofo Combining tone marks**
2. Requester's name: *TCA, Bobby Tung, Selena Wei*
3. Requester type (Member body/Liaison/Individual contribution): *Individual contribution*
4. Submission date: *2019-06-14*
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: *Combining Diacritical Marks Extended*
2. Number of characters in proposal: *2*
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? *But Ko*
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): *But Ko (buttaiwan@gmail.com)*
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)? 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 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? If YES explain	<i>NO</i>
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:	<i>YES</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	<i>NO</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>rare</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>YES</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:	
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<i>YES</i>
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:	<i>NO</i>
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:	<i>NO</i>
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? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>NO</i>
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:	<i>NO</i>
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)	<i>NO</i>
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	<i>NO</i>

(Detach the following pages from the filled summary form before submitting).

Information accompanying submissions

The process of deciding which characters should be included in the repertoire of the standard by WG 2 depends on the availability of accurate and comprehensive information about any proposed additions. WG 2, at its San Francisco meeting 26, designed a form (template) that will assist the submitters in gathering and providing the relevant information, and will assist WG 2 in making more informed decisions.

This form is part of the [WG 2 Principles and Procedures document](http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html) and is available on line at <http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

A duly completed proposal summary form must accompany each new submission. The latest version of this form must be used in submissions. Such a form will assist WG 2 to better evaluate the proposal, and progress the proposal towards a speedier acceptance and inclusion in the standard. Submitters are also requested to ensure that a proposed character does not already exist in the standard.

Submitters are encouraged to visit the "[Where is my Character](#)" page on the Unicode web site for more information on checking if their proposed character or script is already encoded in the standard, or a similar proposal has already been made by someone else. There are also several electronic discussion lists maintained by the Unicode consortium that one could use to discuss with other experts internationally on various subjects related to the standard. Submitters are also encouraged to familiarize themselves with ISO/IEC TR15285 – Character Glyph Model. The latest version of ISO/IEC 10646 as well as ISO/IEC TR15285 are available on line under 'Freely Available Standards' at: <http://standards.iso.org/ittf/PubliclyAvailableStandards/>.

National bodies should take note of the guidelines in section *C.4 Some guidelines on proposing new material as ballot comments* in the [P&P document](#).

In addition to text extracted from the standard in *Annex L: Character-naming guidelines* in the [P&P document](#), the following definitions from the standard are also referenced in the proposal summary form:

Clause 4.12 Combining character:

A member of an identified subset of the coded character set of ISO/IEC 10646 intended for combination with the preceding non-combining graphic character, or with a sequence of combining characters preceded by a non-combining character (see also 4.14).

NOTE – ISO/IEC 10646 specifies several subset collections, which include combining characters.

Clause 4.14 Composite sequence:

A sequence of graphic characters consisting of a non-combining character followed by one or more combining characters (see also 4.12).

NOTE 1 – A graphic symbol for a composite sequence generally consists of the combination of the graphic symbols of each character in the sequence.

NOTE 2 – A composite sequence is not a character and therefore is not a member of the repertoire of ISO/IEC 10646.

Submitter's responsibilities

The national body or liaison organization (or any other organization or an individual) proposing new character(s) or a new script shall provide:

1. Proposed category for the script or character(s), character name(s), and description of usage.
2. Justification for the category and name(s).
3. A representative glyph(s) image on paper:
If the proposed glyph image is similar to a glyph image of a previously encoded ISO/IEC 10646 character, then additional justification for encoding the new character shall be provided.
Note: Any proposal that suggests that one or more of such variant forms is actually a distinct character requiring separate encoding should provide detailed, printed evidence that there is actual, contrastive use of the variant form(s). It is insufficient for a proposal to claim a requirement to encode as characters in the Standard, glyphic forms which happen to occur in another character encoding that did not follow the Character-Glyph Model that guides the choice of appropriate characters for encoding in ISO/IEC 10646.
Note: WG 2 has resolved in Resolution [M38.12](#) not to add any more Arabic presentation forms to the standard and suggests users to employ appropriate input methods, rendering and font technologies to meet the user requirements.
4. Mappings to accepted sources, for example, other standards, dictionaries, accessible published materials.
5. Font resource:
A font must be provided to the Editor promptly after the characters have been approved by WG2 for use in printing the charts. If a font is not provided, the Editor cannot include the glyphs in the charts and, as a result, the repertoire of characters corresponding to these glyphs will not be included in draft amendments. Fonts must be in one of the following formats (in preferential order): OpenType, TrueType, Postscript Type 1.
Fonts submitted must have no license restrictions that prevent embedding into PDF documents. Because of synchronization between ISO/IEC 10646 and The Unicode Standard, any grant of license must cover use in publishing both standards and related documents (see resolution [M45.30](#)). For technical reasons, the editors must be able under that license to freely modify or replace glyph outlines in their copies of the fonts. The submitter should be prepared to provide fonts of suitable quality and license conditions, unless equivalent fonts are already available to the editors.
6. List of all the parties consulted:
Towards ensuring that all significant stakeholders have been consulted, WG 2 requests that submitters solicit feedback from other groups who use the proposed characters, and list these contacts. Submitters are encouraged to provide the email id-s of the submitters as well as other experts who have been consulted to facilitate any clarification queries.
7. Equivalent glyph images:
If the submission intends using composite sequences of proposed or existing combining and non-combining characters, a list consisting of each composite sequence and its corresponding glyph image shall be provided to better understand the intended use.
8. Compatibility equivalents:
If the submission includes compatibility ideographic characters, identify (per resolution [M45.29](#)):
 - the source, which contains two distinct code positions that correspond to a single unified CJK Ideograph character of ISO/IEC 10646
 - the ISO 10646 unified CJK Ideograph
 - the code position in the source for the unified CJK ideograph
 - the code position in the source for the proposed compatibility ideograph
9. Properties that may affect the BiDi processing
Any BiDirectional algorithm related properties associated with the characters should be spelled out (see UAX#9 - <http://www.unicode.org/reports/tr9/>).
10. Any additional information that will assist in correct understanding of the different characteristics and linguistic processing of the proposed character(s) or script.
11. If any of the proposed characters are suitable as *syntax* characters please take note of guideline D.2.5 (on page 24 and of Unicode Standard Annex 31 at <http://www.unicode.org/reports/tr31/>).
12. If the proposal is for a new script, identify the standardized characters that are commonly used directly in the script, or proposed to be unified with the characters of the script, in particular standardized characters allocated in different blocks. Examples include punctuation marks and combining marks. Such information will assist in assigning properties for characters shared across multiple scripts or in identifying character repertoires needed to support particular languages.
13. If you are aware of already standardized characters that are visually close to any of the proposed characters, you are invited to list them in the proposal. This will assist in the analysis of the script for 'visually confusables', towards providing additional guidance on use of the standard from a security perspective (see UTR#36 - Unicode Security Considerations – at <http://www.unicode.org/reports/tr36/>).