ISO/IEC JTC1/SC2/WG2 N2628
Date: 2003-09-29
This is a DOC version of a document to be submitted to WG2.

Title: Proposal to add COMBINING RIGHT DOT ABOVE as required by Taiwanese Holo language Romanization to ISO/IEC 10646

Source: TCA
Status: Member Contribution
Action: For consideration by JTC1/SC2/WG2

A. Administrative

1. Title	Proposal to add COMBINING RIGHT DOT ABOVE as required by Taiwanese Holo language Romanization to ISO/IEC 10646
2. Requester's name	Association for Taiwanese Romanization; Taiwan Ethnic Language Promoting Committee of the Presbyterian Church in Taiwan; Department of Language Education of the National Taitung University; Institute of Taiwan Literature of the National Cheng Kung University, Taiwan.
3. Requester type	Contribution from experts
4. Submission date	2003-09-29
5. Requester's reference	http://69.50.194.230/~lomajic/tavi/index.php?page=Proposal
6a. Completion	This is a complete proposal.
6b. More information to be provided?	No

B. Technical – General

1a. New script? Name?	No
1b. Addition of characters to existing block? Name?	Yes. Combining Diacritical Marks.
2. Number of characters	1
3. Proposed category	Category A.
4. Proposed level of implementation and rationale	Implementation level is 3.

5a. Character names included in proposal?	Yes
5b. Character names in accordance with guidelines?	Yes
5c. Character shapes reviewable?	Yes (see 6b and Appendix)
6a. Who will provide computerized font?	Phahng Taiwanese Workshop; Tan Pek-tiong [Po-Chung Chen]
6b. Font currently available?	Yes (see http://www.phahng.idv.tw/)
6c. Font format?	TrueType
7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?	Yes. An 8-bit font was commonly used. It has de facto use status.
7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?	Yes. See appendix.
8. Does the proposal address other aspects of character data processing?	No

C. Technical – Justification

1. Contact with the user community?	Yes. The Annual World Conference on Taiwanese Languages (TW, US, 1995-2002), the Bible Society (TW), HOTSYS-HAKSYS (US), 5% Taiwanese Translation Project (TW), Taiwanese Writing Forum (TW, US), and users elsewhere (CA, DE, GB, JP).
2. Information on the user community?	Taiwanese Holo, more commonly known as Min Nan ([zh-min-nan]), is used by about 14,345,000 Taiwanese, or 67% of the population in Taiwan. Variants of Peh-oe-ji ("colloquial writing") are attested by a continuous literature dating back to the latter half of the nineteenth century. It remains the only Taiwanese Latin script with complete Bibles published. The script is now taught through the public educational system.
3a. The context of use for the proposed characters?	General, with specialist uses in education, literature, and linguistics. Modern publications often use the Latin characters in conjunction with Han characters.
3b. Reference	See Tiu^{n}(1998).

4a. Proposed characters in current use?	Yes.
4b. Where?	Taiwan and expatriate communities.
5a. Characters should be encoded entirely in BMP?	Yes
5b. Rationale	Latin is a Category A script.
6. Should characters be kept in a continuous range?	Not necessarily.
7a. Can the characters be considered a presentation form of an existing character or character sequence?	No.
7b. Where?	
7c. Reference	See Section E for rationale.
8a. Can any of the characters be considered to be similar (in appearance or function) to an existing character?	No
8b. Where?	
8c. Reference	See Section E for rationale.
9a. Combining characters or use of composite sequences included?	No
9b. List of composite sequences and their corresponding glyph images provided?	No
10. Characters with any special properties such as control function, etc. included?	No

D. SC2/WG2 Administrative

To be completed by SC2/WG2

1. Relevant SC 2/WG 2 document numbers:	
2. Status (list of meeting number and corresponding action or disposition)	
3. Additional contact to user communities, liaison organizations etc.	

4. Assigned category and assigned priority/time frame	
Other Comments	

E. Proposal

User community

The Latin script Peh-oe-ji is used to write the Holo language of Taiwan. Taiwan is an island in the Western Pacific Ocean, north of the Philippines, off the southeastern coast of continental China, about 23°30'N 121°00'E. About 23 million people live in Taiwan. Taiwanese Holo, more commonly known as Min Nan, is used by about 14,345,000 Taiwanese, or 67% of the population in Taiwan. (Ethnologue).

Issues

1. Other Latin scripts for this languages also exist, though far less prevalent, in Taiwan. Most modern publications use a mixture of Han and Latin scripts, with the proportion of each differing from author to author.

2. COMBINING RIGHT DOT ABOVE should be added.

   2.1 COMBINING DOT ABOVE is not a valid alternative to COMBINING RIGHT DOT ABOVE:

      2.1.1 Although the proposed COMBINING RIGHT DOT ABOVE and the existing COMBINING DOT ABOVE have similar appearance and probably the same function indicating an open vowel "o", in no publication has COMBINING DOT ABOVE appeared in the place of COMBINING RIGHT DOT ABOVE. Therefore it is inappropriate to view the proposed character COMBINING RIGHT DOT ABOVE as a presentation form of COMBINING DOT ABOVE.

      2.1.2 Literate Peh-oe-ji reader will not accept COMBINING DOT ABOVE to be a correct (alternative) rendition of COMBINING RIGHT DOT ABOVE. If a new learner writes COMBINING DOT ABOVE for COMBINING RIGHT DOT ABOVE, a responsible teacher will correct the learner.

      2.1.3 No literature ever showed COMBINING DOT ABOVE as the "original" COMBINING RIGHT DOT ABOVE. The dot is not simply moved to the right to make space for tones. Even in the absence of other combining diacritics, it appears as COMBINING RIGHT DOT ABOVE, not COMBINING DOT ABOVE. That is, the position of the COMBINING RIGHT DOT ABOVE is determined relative to the base characters, not other diacritics. In Vietnamese, on the other hand, where two (up class) combining diacritics associate with a base character, some of the diacritics are re-positioned to the left or right of the other diacritics.

      2.1.4 COMBINING DOT ABOVE would be considered as a tonal diacritic, whereas COMBINING RIGHT DOT ABOVE is considered a part of an integral, independent vowel in the alphabet.

      2.1.5 O/o WITH COMBINING RIGHT DOT ABOVE can not be treated as a variation of O/o WITH COMBINING DOT ABOVE since according to the Unicode Standard 4.0

"[t]he base character in a variation sequence is never a combining character nor a decomposable character."

2.2 MIDDLE DOT is not a valid alternative to COMBINING RIGHT DOT ABOVE:

2.2.1 COMBINING RIGHT DOT ABOVE is closer to the vowel (LATIN CAPITAL/SMALL LETTER O) than the tonal diacritic. Using MIDDLE DOT breaks this logic, and creates major difficulty in sorting and other processes.

2.2.2 Although some literature showed MIDDLE DOT as an alternative to COMBINING RIGHT DOT ABOVE, these examples are exceptions created under severe lack of typographic technologies, and should not be used against good typographic practice.

2.2.3 COMBINING RIGHT DOT ABOVE should be a combining diacritical mark, while MIDDLE DOT is a spacing character. When applying a combining diacritical mark (such as COMBINING GRAVE ACCENT, U+0300) to a sequence of LATIN CAPITAL LETTER O + MIDDLE DOT, the diacritical mark is operating on the MIDDLE DOT rather than LATIN CAPITAL LETTER O.

References

- Bible Society. *Sin-ku-iok Seng-keng (Amoy Romanized Bible).* Bible Society, Taipei, Taiwan, 1964. [Cat. No. AR 065 T, first published 1933.]

- Department of Language Education of the National Taitung Teachers College. [N2507: Draft of proposal to add Latin characters required by Latinized Holo language to ISO/IEC 10646], 2002.

- William Campbell, *E-mng-im Sin Ji-tian (Dictionary of the Amoy Vernacular).* Tai-oan Kau-hoe Kong-po-sia, Tainan, Taiwan, 1993-06. [First published 1913-07.]

- Ko Chek Hoan and Tan Pang Tin. Sin-iok (New Testament), the Ko-Tan/Kerymatic colloquial Taiwanese version, 1972-08-31.

- Iau Cheng-to. *Cheng-soan Peh-oe-ji (Concise Colloquial Writing).* Jin-kng, Tainan, Taiwan, 1992.

- Sou Chi-beng (Jemeng Soh). Various versions of *HOTSYS* and *HAKSYS*. OTSYS-HAKSYS, 1994 to 1997. http://www.hotsys-haksys.com/

- Te Khai-su and Michael Everson. [N1593: Proposal to add Latin characters required by Latinized Taiwanese languages to ISO/IEC 10646], 1997.

- Tiu[n] Hak-khiam. Writing in Two Scripts. *Written language and literacy*. 1:225-247, 1998.

Names

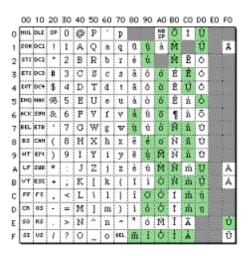## COMBINING RIGHT DOT ABOVE

Ȯ	Ȧ	Ṅ	Ȯ
	ȧ	ṅ	ȯ
	Ė	Ṅ	Ō
	ė	ṅ	ō
	İ	Ṅ	Ȯ
	i	ṅ	ȯ
	Ṁ	Ȯ	Ő
	ṁ	ȯ	ő
	Ḿ	Ơ	Ȯ
	ṁ	ơ	ȯ
	M̄	Ó	Ȯ
	m̄	ó	ȯ
	Ṁ	Ờ	Ú
	ṁ	ờ	ủ

8-bit 2022-compatible code table

8-bit de-facto font code table

[*Note: Due to the limitation of this wiki in processing UTF-8, non-formatting spacing has been inserted within certain syllables to allow reasonable display of affected characters.*]

Appendix: Published examples

Example 1 -- Source : Amoy-English Dictionary p923, 1976, Tâi-tiong : Maryknoll

Phó·-chìo siu-lú-hōe Sisters of Providence (CP) 普照修女會
Phó·-chū hong-chè-hōe Franciscan Sisters of Mercy
　菩慈方濟會
Sa-nî-tek sèng-pó-lók siu-lú-hōe Sisters of St. Paul de
　Chartres 沙爾德聖保祿修女會
Se-tek ki-tok-kàu hūi-bêng bông-jîn hok-lī-hōe Hildesheimer
　Blinden Mission 西德基督教惠明盲人福利會
Sèng-aì-hōe Society of Holy Love 聖愛會
Sèng-bú bû-goân-chōe thoân-kàu siu-lú-hōe Missionary
　Sisters of the Immaculate Conception 聖母無原罪傳教修女會
Sèng-bú bû-liám goân-chōe hong-chè thoân-kàu siu-lú-hōe
　Franciscan Missionary Sisters of the Immaculate Concep-
　tion of the Mother of God 聖母無染原罪方濟傳教修女會
Sèng-bú i-liâu thoân-kàu-hōe Medical Missionaries of Mary
　聖母醫療傳教會

Example 2 -- Source : Source : Bernard L.M. A Dictionary of Southern Min p188, 1973, Hong
　　　　　　　Kong Language Institute

o·-goát-chú N chiah: yellow-throated bunting, *Emberiza*
　elegans elegans 黃眉鵐)
o·-hái-thûn N Zool chiah: black dolphin, *Pseudorca crass-*
　sidus 黑海豚
o·-ham N liáp: a kind of shellfish, 1 *Anomalodiscus*
　squamosus　2 *Anomalocardia producta* 烏蚶
o·-hàn N ê: swarthy big man 黑漢
o·-hek N bé: black croaker (fish), *Atrobucca nibe* 黑䱛
o·-hî N Ich bé/ bóe: Common mullet, *Mugil cephalus* 烏魚
o·-hî-chí N pî: dried roe from the common mullet, *Mugil*
　cephalus 烏魚子
o·-hî-lông N chiah: Chinese cormorant, *Phalacrocorax*
　carbo sinensis 中國鸕鷀
o·-hìa N tè: black roof-tile 黑瓦
o·-hô·-chek N Orn chiah: Chinese stonechat, Japanese sto-
　nechat, Siberian stonechat, *Saxicola torquata* st-*jnegeri*
　黑喉鴝

Example 3 – Source : Tân-lûi, Éng-óan ê k±-hiong p60, 1994, Tâi-pak : Öng-bûn

kā肥ê oˑ-báˑsańg 講：

看he 十 七、 八 歲ê查某

現， 一 領 裙lėk hiah l

聽 著 伊teh 講 啥， 原 在

餅 面ê笑 容， 手 裡thėh

舊 裡chảuh　chảuh 唸：

Example 4 -- Source : Sin-kü-iok ê Sèng-keng , 1916,

SÈNG-KENG BȮK-LIȮK

KŪ IOK

	Chiuⁿ	
Chhòng-sè-kì	50	Thoán-tŏ
Chhut Ai-kip-kì	40	Ngá-ko
Lī-bī-kì	27	I-sài-a
Bîn-sòˑ Kì-liȯk	36	Iâ-lī-bí
Sin-bēng-kì	34	Iâ-lī-bí Ai-ko
Iok-su-a	24	I-se-kiat
Sū-su-kì	21	Tân-i-lí
Lō-tek-kì	4	Hô-se-a
Sat-bó-jíⁿ siōng-koàn	31	Iok-jíⁿ
Sat-bó-jíⁿ hā-koàn	24	A-mô-sū
Liȧt-ông siōng-koàn	22	Gô-pa-tí-a
Liȧt-ông hā-koàn	25	Iok-ná
Lėk-tāi chì-liȯk siōng-koàn	29	Bí-ka
Lėk-tāi chì-liȯk hā-koàn	36	Ná-hông
Í-su-lāh	10	Hap-pa-kok
Nî-hí-bí	13	Se-boan-ngá
Í-su-thiap	10	Hap-kai
Iok-pek-kì	42	Sat-ka-lī-a
Si-phian	150	Má-lāh-kì
Chim-giân	31	

Example 5 – Source : Chiu Tiông-ch…p & Khng Khé-bêng, Tâi-ôan Bän-lâm-öe kàu-thêng p97, 1997, Pîn-tong : An-khó

(一)、分淸 o 韻母和 oʻ 韻母

這兩個單元音韻母都是舌面後的圓唇元音，不同點在於 o 比 oʻ 舌位要高一些，口腔開口度要小一些。國語沒有 o 和 oʻ 的對立，所以這兩個音可以自由變讀。台灣閩南話的 o 和 oʻ 有區別意義的作用，必須讀準這兩個韻母；

窩 o －烏 oʻ　蠔 ô －胡 ôʻ　保 pó －補 póʻ

告 kò －顧 kòʻ　號 hó －雨 hóʻ　韻 só －所 sóʻ

台灣閩南話 oa、oe、oai、oan 等韻母中的 "o" 應是合口呼韻母 "u" 的音位變體，這個 "o" 介乎 u 和 o 音之間。

Example 6 – Source: Âng Hui-chiòng, *Eng-gú Ban-lâm-gú ji-tián* p2, 1995, Tâi-tiong : Maryknoll

ABJURATION sè-choat 誓絕 (Cath.)

　　　　　hoat-sè khì-choat 發誓棄絕

　- of heresy: (Cath.) soan-sè khì-

　　choat ī-toan

ABJURE chju-chōa bô ài ... 誓絕不

ABLE ē-hiáu 會的 ē 能 ū pá-ak 有把握

　- to endure hardships: khek-khóʻ

　　nāi-lô

　Are you - to speak English?

　　Lí ē-hiáu kóng Eng-gí bē?

　Not - to: bē-hiáu

Not - to be on time: bōe-hù

Example 7 – Source : Iâu Chèng-tö, Cheng-sóan Pçh-öe-j,, p67, 1997, Tâi-lâm : Jîn-kng

gō （鵝）	gò （吳）	gō （臥）	gō （誤）
ho （撤）	hoˈ （呼）	hó （好）	hō （虎）
hó （河）	hŏ （狐）	hā （號）	hō （雨）
ko （高）	ko （姑）	kó （果）	kó （古）
kó （筍）	kŏ （糊）	kō （翔）	kō （怙）
kho （科）	khoˈ （顆）	khó （考）	khó （苦）
khó （課）	khŏ （裤）		
ló （老）	lŏ （努）	ló （勞）	ló （力）
lo （猪）	lō （路）		
o （蝸）	o （烏）	ó （果）	ő （挖）
à （與）	ŏ （惡）	ŏ （蠔）	ŏ （湖）
po （褒）	poˈ （埔）	pó （寶）	pó （補）
pó （報）	pò （布）	pŏ （婆）	pō （部）
pō （爆）	pŏ （步）		
pho （坡）	phoˈ （鋪）	phó （皮）	phó （譜）
phó （破）	phŏ （補）	phō （抱）	phō （簿）
so （梭）	soˈ （酥）	só （鎖）	só （所）
só （掃）	sŏ （數）		
to （刀）	toˈ （都）	tó （倒）	tō （賭）
tô （到）	tó （奴）	tō （駝）	tō （圖）
tō （道）	tŏ （肚）		
tho （叫）	thoˈ （偷）	tho （討）	thō （土）
thō （桃）	thŏ （土）	thō （套）	thŏ （兔）

Example 8 – Source : Të° Liông-úi & Siä Siok-koan , Tâi-ôan Hok-kiàn-öe ê gú-im kiat-k± kap
piau-im-hòa p164, Tâi-pak : Hãk-seng

				A:			
1.	所鎖	só	só˙	súo		1.	só˙	só
2.	古果	kú	kó˙	kó		2.	kó˙	kó
3.	導度	tō	tō˙	tū		3.	tō	tō˙
4.	告故	kù	kò˙	kò		4.	kò	kò˙
5.	佈報	pò	pòu	pò˙		5.	pò˙	pò
6.	慕帽	bō	bō˙	bū		6.	bō˙	bō
7.	可苦	khóu	khó˙	khó		7.	khó	khó˙
8.	土桃	thō	thēu	thô˙		8.	thô˙	thô
9.	菊徒	tô	tô˙	thâu		9.	tô	tô˙
10.	五餓	gō	gō˙	gē		10.	gō˙	gō

Example 9 – Source : Siau Pêng-ț,, Tâi-ôan siõk-gú kiâm-sng-ti° (2) p83, 2000, Chiong-hòa : Löa-Kh⁻-jiû ki-kim-höe

ê 心情來聽你這個 ta-ke ê 命令？He 是無可能ê
想，想--著你會 gêng，想--著你會心 kòng 火 tóh
會 bch 生 hiah chē 來拖磨。人講 chē kiáⁿ 餓死 pē，
ta-ke」，怨感！怨感！

➢ 【註解】

chē 新婦：chē sin-pū。

lö 死：『過分操勞』。

Example 10 – Source : Thian-lí täi-hãk, Hiän-täi Bän-lâm-gú sû-tián, 1981, Japan : Thian-lí täi-hãk

ô 壺 〔名〕つぼ，かめ．〔量〕つぼ．
　◇つぼやかめに入れたものを数え
　るとき．1 ～chiú (一～酒) ＝1
　つぼの酒．
ô 芋→ô-á
o-a 烏鴉 〔名〕(鳥) カラス．
ô-á 芋仔 〔名〕 (植) サトイモ (＝
　ô 芋)．
o-àm 烏暗 〔形〕暗い．gōa-bin
　chin～ (外面真～) ＝外はたいへ
　ん暗い．

Example 11 – Source : George Ede, Sam-jü-keng Sin-chöan pçk-öa chù-kái, 1894, Tâi-lâm : Sin-lâu

'siong-kin, siong oán.' Sri-jiân lâng ŭ ben-piat Kun-chú kap Siáu-jîn, iáu-kú in pèⁿ-pèⁿ si ŭ-chöe ê lâng. In tī tú-á chbut-sì ê sí bô sím-mlh hián-chhut koh-iūⁿ. Chit-ê sī siong-kin khah háp-lí ê ì-sù. Aū-lái in ūi in ê só' bák-slp bô saⁿ-tâng, só'-í ū hun in-ê khoán-sit chò Kun-chú Siáu-jîn. Chit-ê sī siong-oán ê ì-sù.

13	苟	kó'	siat-sú, kó'-chhiáⁿ
14	不	pùt	bô
15	教	kàu	kà-sī

Example 12 – Source : Tâi-ôan Hú-siâ° Kàu-höe-pò, 1885, Tâi-lâm

TÂI-OÂN-HÚ-SIÂ
KÀU-HŌE-PÒ.

TE IT TIUⁿ
Kong-sō XI nîⁿ, 6 goéh.

i-oân-hú-siâⁿ ê Kàu-su mñg Kàu-hŏe-lái ê hiaⁿ-tī chí-moái Siōng-tè siúⁿ-sù lín tāi-ke tōa in-tián. án kóe--lái chit-pêng sī in-ūi ài thôan Thian-kok ê tō-l Siōng-tè lái tit-tiòh kiu. Só thôan ê tō-lí lóng sī Sèng-ê ; nāⁿ m̄-sī Sèng-chheh ê tō-lí, goán m̄-káⁿ kóng. Só-í khó'-khñg lín tiòh thák-chheh lái khòaⁿ Sèng-keng, ň