Unicode request for compound tone diacritics

L2/23-188

Kirk Miller, kirkmiller@gmail.com

2023 July 26

This request is for six compound tone diacritics that are based on the vertical line $^{\circ}$ for mid tone.

Transcription systems for languages with four distinctive pitches commonly use a vertical line for the higher mid pitch, thus contrasting \circ *high*, \circ *high-mid*, \circ *low-mid*, \circ *low* (equivalent to \circ 14 \circ in Chao tone letters). This occurs for example in descriptions of Kru and Otomanguean languages, and with the values of \circ and \circ reversed in the General Alphabet of Cameroon Languages. A vertical line may also be used for simple mid tone when the macron is used for vowel length—see for example \circ for a long \circ with mid tone in Figure 6. When transcribing contour tones, this vertical \circ may form ligatures analogous to those that Unicode currently supports for the macron in the Combining Diacritical Marks Supplement. Thus, when the macron is used for vowel length, requested \circ \circ \circ \circ are equivalent to U+1DC4-1DC7 \circ \circ \circ \circ (both sets being 1 \wedge \wedge in Chao tone letters), but \circ \circ \circ also occur as contours distinct from \circ \circ \circ \circ when the macron is used for tone (the two sets being 1 \wedge \wedge 1 and \wedge \wedge 1 in Chao tone letters). In addition, in this case \circ and \circ may form ligatures with each other, for mid-falling \circ and mid-rising \circ (\wedge \wedge in Chao tone letters). I have found such ligatures of the vertical stroke \circ in descriptions of languages from East, Central and West Africa and from North America.

The requested diacritics complete the set of possible two-element ligatures. Ligatures of a larger number of elements are possible, and are theoretically open-ended, though there is a practical limit of what can fit over a single vowel letter. Among ligatures of three tone diacritics, Unicode currently supports only the simplest: $U+1DC8\ \tilde{o}$ and $U+1DC9\ \tilde{o}$. A few others are illustrated in the figures. These will be requested in a follow-up to this proposal.

Chart

Greyed-out cells have already been assigned or (light grey) requested in a parallel proposal.

	0	1	2	3	4	5	6	7	8	9	A	В	C	D	Е	F
Combining Diacritical Marks Extended																
U+1ABx	্	ో	%	៎	៉	្ណ	್ಲ	्	୍	۶	۶	Ö	())	္ပ	ം	្ល
U+1ACx	្ណ	်	ှ	ု	ှ	S	*	ঁ	ţ	;	្ន	″	ঁ	် ်	៍	ř
U+1ADx	័	ৈ	៌	ੰ	៉	ី										

Characters

- 1AD0 COMBINING VERTICAL-LINE-ACUTE.
- ំ 1AD1 COMBINING GRAVE-VERTICAL-LINE.
- 1AD2 COMBINING VERTICAL-LINE-GRAVE.
- 1AD3 COMBINING ACUTE-VERTICAL-LINE.
- ំ 1AD4 COMBINING VERTICAL-LINE-MACRON.
- ី 1AD5 COMBINING MACRON-VERTICAL-LINE.

Properties

```
1AD0;COMBINING VERTICAL-LINE-ACUTE;Mn;230;NSM;;;;N;;;;
1AD1;COMBINING GRAVE-VERTICAL-LINE;Mn;230;NSM;;;;N;;;;
1AD2;COMBINING VERTICAL-LINE-GRAVE;Mn;230;NSM;;;;N;;;;
1AD3;COMBINING ACUTE-VERTICAL-LINE;Mn;230;NSM;;;;N;;;;
1AD4;COMBINING VERTICAL-LINE-MACRON;Mn;230;NSM;;;;N;;;;
1AD5;COMBINING MACRON-VERTICAL-LINE;Mn;230;NSM;;;;N;;;;
```

References

Crazzolara, Pasquale J. (1960) *A study of the Logbara (Ma'di) language: grammar and vocabulary.* London & New York: Oxford Univ. Press.

Dolgopolsky (2013) Indo-European Dictionary with Nostratic Etymologies, vol. I.

Kaye, Jonathan Derek & Hilda Koopman (1982) *Lexique Vata*. In Kaye, Koopman & Sportiche (eds.), *Projet sur les langues kru*, 303–341. Université du Québec à Montréal.

Marchese, Lynell (1983) Listes de lexemes Kru. In *Atlas linguistique Kru: essay de typologie*, 323–405. 3rd edition, Abidjan.

Vogler, Pierre (1988) Le Parler Vata, Strasbourg.

—— (2017) Contribution à la classification des langues kru (Niger-Congo), Strasbourg.

Weltfish, Gene (1937) Caddoan Texts: Pawnee, South Band Dialect. American Ethnological Society, New York.

Figures

Tone Marks

- 33. The tone of a syllable will be shown on vowels (or consonants) by means of the conventional marks given below. The tone of a syllable may be simple (level) or combined (kinetic). (See also § 53.)
 - (a) Marks for simple tones:

```
high tone (h.t.) '; ti cow; éwá beer; ti mvá cow calf mid tone (m.t.) '; dzdó rain; é'da to show; ti mouth low tone (l.t.) '; lè to like; sì with
```

(b) Marks for combined tones (unbroken ascent or descent from one level to another by combination of the simple tones):

```
high-low tone ; tî zá beef; ɔpî nǐ the chief's high-mid tone ; ádríi (or ádrí) orother; átí father mid-high tone ; ɔzɔ or əzɔɔ rain mid-low tone ; tsɔ-tsɔ to strike low-mid tone ; fɛ ne gives low-high tone ; lă they read (past) intermediate tones: see § 53, p. ii.
```

Figure 1. Crazzolara (1960: 7) explaining the diacritics.

INTONATION

11

50. (1) A low-toned verb adds to its low tone a slightly raised tone-tail, a 2/5 tone. Examples:

fè to give,

but: àfa fe to give something

51. (2) A high-toned verbal root, e.g. **Embá**, takes an additional slightly falling tone-tail, a 4/5 tone. Examples:

embá to teach

èri andzi embå he is teaching the children

52. (3) A mid-toned verb sometimes adds a slightly rising tone-tail, from 3/5 to 4/5. Examples:

mvu to drink

éwâ mvů to drink beer

Figure 2. Crazzolara (1960: 11). Use of the diacritics for 'incomplete action' intonation.

- 308. (b) Describing mental state, physical behaviour, manners, &c.:
 - (32) èri a'â kồmú sĩ he was kneeling
 - (33) ma ásíni əvö èmvè (= ètsáandí) be I am heart-broken/sorely grieved
 - (34) อกุอ์ที่กา a'a อ'do อกุอ์ที่โย its claws are like leopard's claws
 - (35) dzó 'alé àndzí; kàni kó èri 'ba, èri avö àlarö the hut is dirty; when they have swept it, it will be clean
 - (36) èmï əvö (= adri) è yere, ma fè rá be quiet, I shall give (you)
 - (37) mvá 'dì èri əvö àfa ni kákārö this boy is ignorant (without knowing things)
 - (38) embá mvíinï əvöző (= mà əvö) mőkêrő he exhorted his son to behave properly
 - (39) èri əvö àvlàvì kəkərö he is/was breathless

Figure 3. Crazzolara (1960: 69). Grammatical examples illustrating three of the requested diacritics.

etú¹-etú = tutu, pipi, vn. to swell; ásínï etú-etú his belly is swollen, raised (cf. next etú).
etú²(cf. tu), vn. to come up, mount; mí etú mávó 'bùá 'dó! come up to me! (say, on a tree); è'bínï etú Énáóá fish swim up the Enyao river.

Figure 4. Crazzolara (1960: 247). A dictionary entry illustrating three of the requested diacritics.

```
    middle tone.

   - falling high-to-mid tone
   - falligh high-to-low tone
   - falling mid-to-low tone
   - rising low-to-high tone
  - rising low-to-mid tone
   - rising mid-to-high tone (mid rise tone)
   - high rise tone (as in Wedekind's records of Janjero)
   - very low tone
  - very high tone
falling mid-to-low \frac{1}{6} = a<sub>4-2</sub>, rising mid-to-high \frac{1}{6} = a<sub>4-6</sub>, etc.
```

Figure 5. Dolgopolsky (2013: 2995) explaining the diacritics.

```
{PH} hāra, Sa {R} ha'rā 'faeces', Sml {ZMO, DSI} hār, Sml N {Abr.} har
'faeces, excrement', Sml har- v. 'defecate', Bn har 'diarrhoea', Rn {PG} har
C: Bi {R} nay- sev. 'spend the night, sleep, rest' ¶ R WBd 187 | Ch:
WCh: Klr {J} na vi. 'lie, sleep', Sha {J} na (pf. nah-i) vi. 'sleep', {IL} na
vi. 'lie (liegen)', ? Dr non 'sleep' | CCh: Mtk {Sb.} ή unε 'lie (liegen)' |
HgB suχú, FlG sεgwi, HgK sakwú / sεχwú, HgG sεgwi v. 'drink' || ECh: Nd
{J} saˈga id. ¶ JI II 111, ChC s.v. 'drink', Stl. ZCh 177 [no. 296], J T 146, J
(pl. žàunànín) id. ! NrBc: My {Sk. in ChC} yáwùn, {ChL} yέwìn, Sir {Sk.}
yöwání, Mbr {Sk.} yáwìn, Jmb yáwàn id. | Bg {J} yuwân, {Csp.} yuwan,
{Sh.} ywān, Grn {Hrn.} yíwán, {Csp.} yōwan id. ∐ CCh: Tr {Nw.} žùwàn
```

Figure 6. Dolgopolsky (2013: 1726, 1966, 2778). The vertical-grave tone mark, some set over a macron indicating vowel length. Note the $\langle \dot{\bar{o}} \rangle$ (blue arrow), where the vertical mid tone contrasts with the horizontal macron. The languages are N. Somali and the Chadic languages Kulere, Sha, Ndam, Siri and Boghom.

```
*? ∇m- 'catch, seize', {JS} *?am 'take' > CCh: Tr {Nw.} om 'catch', ZmB ?ám {J} 'take', {Sa.} 'steal', ?? Msg {Mch.} imã y. 'seize', {Lk.} ime, ime 'fangen, ergreifen, nehmen' || ECh: Ll {Cp., ChC} om v. {Cp.} 'hold', {ChC} 'hold, seize', 'swallowing noisily' | Lame {Sa.} id. gùròk id. || ECh: Mu {J} goró (pl. gòràr), Tmk {Cp.} ger, Kwn {J} kár & kàrí, Ke {Eb.} kur 'neck' (in KwK *g-WCh {Stl.} *?akwi 'he-goat' ({AD} 'goat') > Hs àkuyà 'he-goat', àkwíyà 'goat' | Kfr {Nt.} ok id. || Ron {J}: Df ʔāh, Btr āχ, Klr ʔāh id. || Krkr {Lk, J} sb. {Kr.}: HgNk yɛ, HgF nègyò, HgG ùyè, HgK ?yɛ(-ndɛ), FlK yfy, FlG
```

Figure 7. Dolgopolsky (2013: 130, 728, 833, 1600). The vertical-acute tone mark. The languages are Lele, Tumak, Daffo, Butura and Fali of Kiria, all Chadic.

```
bβ 'soul' (transcribed in Gk [Horapollon] as βαί) (× N *beʔy∇ 'body, self', q.v. ffd.) ¶¶ Cf. also Ch: Ke {Eb.} fu:li 'blasen, wehen', Kir {ChL} 'small' || CCh: Mln {ChL} kyέŋ, Bdm {Lk.} gená, Lame {ChL} kwánàyằτ id. || ECh: Kwn {J} končé, Jg {J} koyan, Kjr kani id. ¶ ChC s.v. 'small',
```

Figure 8. Dolgopolsky (2013: 1157). The acute-vertical tone mark. The languages are Kera and Jegu (Mogum), both Chadic.

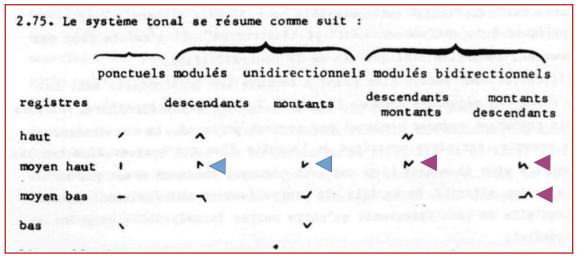


Figure 9. Vogler (1988: 52). The diacritics marked in blue are requested. The more complex ligatures marked in purple will be requested separately.

Figure 10. Vogler (1988: 421). List of phonemes.

```
(40) w : bassa kpá ; klao kpā' ; guéré kpå ; grebo klá' ; tepo klá ; glio-oubi kala ; e : bété kwá 'os'
```

Figure 11. Vogler (2017: 13), with the compound tone $\langle \ \ \ \ \rangle$ set as a sequence $\langle \ \ \ \ \rangle$ due to a lack of font support. The problem with this solution is that normally a tone mark such as $\langle \ \ \rangle$ without a consonant or vowel letter to carry it will be interpreted as a *floating tone*, a very different phenomenon. (Not that the same *ad hoc* solution is used for the diacritic $\ \ \ \$ on the second word in the set, $kp\tilde{a}$.)

```
SIA -
                                    : V "gagner, avoir le dessus"
                                    : N sg (-kli) "coquille d'escargot"
sıākuklu
kopopá
                                           : N sg (-p5) "albinos" (
ko sa 🤜
                                           : V "choisir"
kosa (-se, -sw)
                                           : Ad sg (-s), -si) "sec"
ku jlo
                                                 : V "se relever
ku klalè
                                                 : V "attraper"
kukl A
                                                 : N sg (-kl) "
kDk la
                                                 : N sg (k7k11)
kuk lúku
                                                 : Ph "cocorico"
kakla pia
                                                 : V "être mal.
                                         --- et subordinatif "alors que
ıla 🕙
                                   : N cl ( | lai ) "sommeil"
ıla .
                                      "s'amuser de façon répréhensible"
ila yla
                                   : V "rêver" ("déchirer le sommeil")
```

Figure 12. Vogler (1988: 490, 464, 460, 453). Contrasts between ⟨⋄⟩ (blue) and ⟨⋄⟩ (yellow). A more complex diacritic is marked in purple.

```
mininyici (minyici) : N sg (-ci) "narine"

de gla : V "réfléchir"
```

Figure 13. Vogler (1988: 472, 434). Vertical-acute $\langle " \rangle$ (blue) contrasts with macronacute $\langle " \rangle$ (yellow).

```
107 na la na kulú dūdūku si na la na kulú dūdūku Ne baisse pas ton front,
```

Figure 14. Vogler (1988: 419). A contrast in running text.

asi	ЬE	masi'ngUa	MACHINE
Udyma IV	i. A	m Ú dým Č	POT (POMMADE)
en	N A	mini.	NEZ
yenu	40	nyeni	MIRROIR

Figure 15. Kaye & Koopman (1982). A manual-typewriter substitution for $\langle \mathring{\circ} \rangle$. An analogous substitution is made for $\langle \mathring{\circ} \rangle$, $\langle \mathring{b} \rangle$ is used for δ , etc. (The initial letter of each line is clipped in this photocopy.)

These four distinctions can occur with normal, short, or double length vowels. Double length vowels also occur with combinations of the above tones, the shift from one tone to the other often giving the impression of a gliding repetition of the vowel sound, rather than of a continuous voicing:

- $\triangleright \hat{a} \cdot \text{ high to middle high}$
- ▶ a normal to middle high
- $ightharpoonup \dot{a} \cdot \text{ middle high to normal}$
 - \vec{a} · normal to high
- - \hat{a} · normal to low

Figure 16. Weltfish (1937: 3). In addition to $\langle " \rangle$ as in Dolgopolsky (yellow), we have the vertical-macron ligatures $\langle " \rangle$ (red).

4 Publications, American Ethnological Society Vol. XVII

I. TEXTS BY EFFIE BLANE

pi·tahawira·t^a Band, tsa·stawirahi·ka Woman-ceremonial-gift-horse (in the Pipe Dance)

e tiratúkspari ekuxruksutska tiráwa hat kare sika pa kis "- In my travels he (God) wanted God, Heaven for him not to be poor

Figure 17. Weltfish (1937: 4). $\langle \overline{\ } \ \overline{\ } \rangle$ in addition to simple $\langle \dot{\ } \rangle$.

nawahi weretka pa kis "See, now, I am poor 'It

tipi raski
'It is a boy.'

Figure 18. Weltfish (1937: 5). The vertical diacritic combined with the macron, $\langle \overline{\ } \ \overline{\ } \rangle$, and with the acute, $\langle \overline{\ } \ \overline{\ } \rangle$. Combinations with the grave do not occur in these texts.

Figure 19. Marchese (1983: 237). A typewriter substitute for vertical and macron (the two orders do not contrast and so did not need to be distinguished).

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 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 std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html. See also std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest Roadmaps.

A. Administrative

1. Title: Additional p	honetic click letters					
2. Requester's name:	Kirk Miller					
3. Requester type (Member body/Liaison/Individual contribution):						
4. Submission date:	2023 July 26					
5. Requester's reference (if applicable):						
6. Choose one of the following:						
This is a complete proposal:	<u>yes</u>					
(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):	<u>no</u>					
Proposed name of script:						
b. The proposal is for addition of character(s) to an existing l						
Name of the existing block:	Combining Diacritical Marks Extended					
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 na	aming guidelines" in Annex L of					
P&P document?	yes					
b. Are the character shapes attached in a legible form suitable	e for review? <u>yes</u>					
5. Fonts related:						
a. Who will provide the appropriate computerized font to the	e Project Editor of 10646 for publishing the standard?					
Kirk Mil						
b. Identify the party granting a license for use of the font by						
SIL (Gentium	Release)					
6. References:						
a. Are references (to other character sets, dictionaries, descr	iptive texts etc.) provided?					
b. Are published examples of use (such as samples from news	spapers, magazines, or other					
sources) of proposed characters attached?	yes					
7. Special encoding issues:						
Does the proposal address other aspects of character data pr						
presentation, sorting, searching, indexing, transliteration et	c. (if yes please enclose information)? <u>no</u>					
8. Additional Information:						
Submitters are invited to provide any additional information abou	t 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 www.unicode.org for such information on other scripts. Also see Unicode Character Database (www.unicode.org/reports/tr44/) and associated Unicode Technical Reports for information needed for consideration by the						
(www.unicode.org/reports/tr44/) and associated Unicode Technic						

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.)?	<u>yes</u>
If YES, with whom? Author is a member of the user 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? Reference:	
4. The context of use for the proposed characters (type of use; common or rare)	transcription
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?	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)?	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?	no
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
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?	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 characters?	no
If YES, are the equivalent corresponding unified ideographic characters identified?	
If YES, reference:	