

# Unicode request for extended IPA tone diacritics

Kirk Miller, kirkmiller, gmail.com  
Maksim Fedotov, tequilalime, gmail.com

Valentin Vydrin, vydrine, gmail.com  
2025 August 04

This proposal is for ligatures that combine the IPA tone/pitch diacritics U+030B ◌̂ COMBINING DOUBLE ACUTE ACCENT and U+030F ◌̆ COMBINING DOUBLE GRAVE ACCENT with each other and with simplex acute, macron and grave. This is a productive pattern of extending the IPA tone diacritics.

## Characters

### Compound IPA tone diacritics

- ◌̂̂ 1AF1 COMBINING DOUBLE ACUTE-MACRON.
- ◌̂̆ 1AF2 COMBINING GRAVE-CARON.
- ◌̂̇ 1AF3 COMBINING ACUTE-CIRCUMFLEX.
- ◌̂̈ 1AF4 COMBINING DOUBLE GRAVE-DOUBLE ACUTE.

## Properties

1AF1;COMBINING DOUBLE ACUTE-MACRON;Mn;230;NSM;;;;;N;;;;;  
1AF2;COMBINING GRAVE-CARON;Mn;230;NSM;;;;;N;;;;;  
1AF3;COMBINING ACUTE-CIRCUMFLEX;Mn;230;NSM;;;;;N;;;;;  
1AF4;COMBINING DOUBLE GRAVE-DOUBLE ACUTE;Mn;230;NSM;;;;;N;;;;;

## Deferred

These symbols are productive compound IPA diacritics but are not directly attested.

- ◌̂̂̂ 1... COMBINING MACRON-DOUBLE ACUTE.
- ◌̂̂̆ 1... COMBINING DOUBLE GRAVE-MACRON.
- ◌̂̂̇ 1... COMBINING MACRON-DOUBLE GRAVE.
- ◌̂̂̈ 1... COMBINING CIRCUMFLEX-GRAVE.
- ◌̂̂̉ 1... COMBINING DOUBLE ACUTE-DOUBLE GRAVE.
- ◌̂̂̊ 1... COMBINING TRIPLE GRAVE.

These symbols are an iconic way to disambiguate the ambiguous triple acute and triple grave. However, they are not IPA usage, and attestation is insufficient to request them at this time.

- ◌̂̂̂̂ 1... COMBINING ACUTE PLUS HIGH ACUTE.
- ◌̂̂̂̆ 1... COMBINING HIGH ACUTE PLUS ACUTE.
- ◌̂̂̂̇ 1... COMBINING HIGH GRAVE PLUS GRAVE.
- ◌̂̂̂̈ 1... COMBINING GRAVE PLUS HIGH GRAVE.

# Chart

The proposed character order parallels that of the semantically similar diacritics U+1DC4..1DC7:  
 ̆ (higher rising), ̇ (lower rising), ̈ (lower falling), ̉ (higher falling).

## Combining Diacritical Marks Extended

1AB0

1ADF

	1AB	1AC	1AD	1AE	1AF
0	̆̇	̇̈	̇̉	̇̊	̇̋
1	̆̈	̇̉	̇̊	̇̋	̇̌
2	̆̍	̇̎	̇̏	̇̐	̇̑
3	̆̒	̇̓	̇̔	̇̕	̖̇
4	̗̆	̘̇	̙̇	̇̚	̛̇
5	̜̆	̝̇	̞̇	̟̇	
6	̢̆	̣̇	̤̇	̥̇	
7	̦̆	̧̇	̨̇	̩̇	
8	̬̆	̭̇	̮̇	̯̇	
9	̰̆	̱̇	̲̇	̳̇	
A	̷̆	̸̇	̹̇	̺̇	
B	̻̆	̼̇	̇̽	̇̾	
C	̆̿	̇̾	̇̿	̺̇	
D	̆̽	̇̾	̇̿	̺̇	
E	̆̿	̇̾	̇̿	̺̇	
F	̆̽	̇̾	̇̿	̺̇	

# References

Roger Blench (2025) *Tone Systems in the Bantoid Languages*. Draft.

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

—— (2017) *Индоевропейский Словарь с Ностратическими Этимологиями*, vol I.

Maksim Fedotov (2017) *Гбан язык [Gban]*. In Vydrin, Mazurova, Kibrik & Markus (eds.) *Языки мира: Языки манде [Languages of the world: Mande languages]*, 902–999. Nestor-Historia, St. Petersburg.

Valentin Vydrin (2011) *Déclinaison nominale en dan-gwèètaa (groupe mandé-sud, Côte-d’Ivoire)*. *Faits de langues* 3: 233–258.

—— (2016) *Tonal inflection in Mande languages: the cases of Bamana and Dan-Gwèetaa*. In Palancar & Léonard (eds.) *Tone and Inflection: New facts and new perspectives*, 83–105. De Gruyter Mouton.

—— (2020) *Dan*. In Reiner Vossen & Gerrit Dimmendaal (eds.) *The Oxford Handbook of African Languages*, 451–462. Oxford University Press.

# Background

The modern IPA provides two ways of indicating lexical tone: the combining diacritics <◌́◌̄◌̂◌̃◌̄> for *extra-high*, *high*, *mid*, *low* and *extra-low* pitch, and the Chao tone bars <◌̎◌̍◌̌◌̋◌̊> for *high*, *mid-high*, *mid*, *mid-low* and *low* pitch. The diacritics are inherited from an Africanist tradition, where most languages have relatively simple tone systems with few contour tones, and are thus amenable to a small number of simple diacritics, while the tone letters are inherited from a Sinological convention, where multiple contour tones are typical. However, the Kru and Southern Mande languages of Ivory Coast and Liberia, some of the Jukunoid languages of Nigeria, and scattered languages elsewhere in Africa, break this pattern: they have complex inventories of contour tones, and yet are transcribed using diacritics as in the rest of the continent.

The Chao tone letters combine into ligatures to indicate contour tones. This behavior is handled by the font: only the five component pitches <◌̎◌̍◌̌◌̋◌̊> are encoded in Unicode. The diacritics also combine to indicate contour tones, but in their case the ligatures are encoded separately.

Unicode covers all two-glyph sequences of acute, macron and grave, but not of double acute and grave. The original diacritics were acute, macron and grave, <◌́◌̄◌̂>, for high, mid and low pitch, with extra-high and extra-low <◌́◌̂> later added by iconic doubling. Only compounds of <◌́◌̄◌̂> are listed in the *IPA Handbook*. Compound acute and grave are handled with the caron and circumflex; for the macron, the special ligatures <◌́◌̄◌̂◌̂> are provided, as shown in Table 1:

	H	M	L
H	◌́	◌̄	◌̂
M	◌̎	◌̍	◌̌
L	◌̋	◌̊	◌̉

**Table 1.** The compound tone diacritics found for simple rising and falling pitch in the *IPA Handbook* and IPA charts. All permutations of high <◌́>, mid <◌̄> and low <◌̂> are provided.



## Figures

high-low	ˆ	HL
high-mid	˘	HM
superhigh	˥	SH
superhigh-mid	˥˘	SM

**Figure 1.** Blench (2025:13, table 9). ⟨˥˘⟩ for Mungbam, a Bantoid language of Cameroon. This is typeset as a sequence, but the intent is clear. The author comments (p.c. 2025) that he has the same need for two other languages, Boze and Rigwe, ‘preferably as a single symbol not a composite’; all three accounts are to be published in a new volume on tone in African languages.

There are three modulated tones (all are available on single vowels), they are less frequent than level ones:

- (1) high-falling: *gbê* ‘dog’,
- (2) mid-falling: *dq̄* ‘hunger’,
- (3) extra-high-falling: *zīzī́ dē̄* ‘extremely ancient’.

**Figure 2.** Vydrin (2020:452). Contour tones in Dan, a Mande language of Ivory Coast. Due to font constraints, the extra-high falling tone ⟨˥˘⟩ is approximated as a sequence of diacritics, but that incorrectly suggests that it is an extra-high tone followed by a low floating tone.

(où ils sont appelés « Yacouba »). Le dan-gwèetaa a cinq tonèmes unis et trois tonèmes modulés dont les désignations sont (sur l’exemple de la voyelle *a*) : *á* – extra-haut, *á* – haut, *ā* – moyen, *à* – bas, *à* – extra-bas, *ā̄* – extra-haut-descendant, *â* – haut-descendant, *ā̄* – moyen-descendant. À la différence de la plupart des

**Figure 3.** Vydrin (2011:1). ⟨˥˘⟩ is here substituted with a trema ⟨̄̄⟩, despite the use of proper IPA diacritics for all other tones, such as ⟨˥˘⟩ in the next line. The same substitution is done in Vydrin (2016).

В (высокий)	<i>krá</i> 'быть твердым (о плодах)'	<i>ké</i> 'другой, еще один'	<i>(gbáá</i> 'гбан')	<i>lé</i> 'сота'; 'кастрировать'
Н (низкий)	<i>krà</i> 'поясница'	<i>kè</i> 'друг друга'; IND.NEG	<i>gbà</i> 'давать; поклоняться'	<i>lè</i> 'шип, косточка'
уН (ультранизкий)	<i>krǎ</i> 'сороконожка (вид)'	<i>kě</i> 'гореть'; COND.AFF	<i>gbǎ</i> 'щека', 'разновидность'	<i>lě</i> 'летучая мышь', 'затем'
уН-уВ (восходящий)	<i>krǎ</i> 'задевать, чиркать'	<i>kě</i> COND.NEG	<i>gbǎ</i> 'блоха, вошь'	<i>lě</i> 'дело, история'
уН-В (средне-восходящий)	<i>krǎ</i> 'занавеска из циновки'; 'работа'	<i>kě</i> COND.AFF\PST	<i>gbǎ</i> 'давать'\PFV.PREN	<i>lě</i> 'кастрировать'\PFV.PREN

Две контурные тонемы реализуются как восходящие 1) с ультранизкого до ультра-высокого (*ǎ* [ǎ]) — собственно «восходящий» — и 2) с ультранизкого до высокого (*ǎ* [ǎ]) — «средневосходящий». В принятой графике для них используются диакритика-гачек (*ǎ*) для «восходящего» и трема (*ǎ*) для «средневосходящего».

В настоящем очерке тоны на всех слогах систематически обозначаются только в примерах из собственных материалов автора. В примерах из источников в старой графике (в первую очередь из перевода Нового Завета) часть тонов может быть не обозначена.

**Figure 4.** Fedotov (2017: 906). The spacing diacritics were kerned to simulate IPA tone marking of <ǎ> and <ǎ> for the two rising tones of Gban, a Mande language of Ivory Coast. Again due to font constraints, these are substituted with the generic symbol for rising-tone, <ǎ>, and with *ad hoc* <ǎ> in most of the text, but are transcribed precisely here in the chapter on phonetics. The author will be using these same symbols in an upcoming grammar on Gban.

↗	- high rise tone (as in Wedekind's records of Janjero)
↘	- very low tone
↗	- very high tone

**Figure 5.** Dolgopolsky (2013: 2995). Minimal attestation of the convention of raising one of a pair of accent marks relative to another to indicate a rising or falling tone, in this case high rising <ǎ>. The convention is illustrated here for reference but these diacritics are not requested, pending better evidence of need.

ISO/IEC JTC 1/SC 2/WG 2  
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS  
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646<sup>1</sup>.

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](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 [std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html](http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html).  
See also [std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html](http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html) for latest *Roadmaps*.

**A. Administrative**

1. Title:	<i>Compound IPA tone diacritics</i>
2. Requester's name:	<i>Kirk Miller, Valentin Vydrin, Maksim Fedotov</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>individual</i>
4. Submission date:	<i>2025 August 04</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<i>yes</i>
(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:	<i>yes</i>	
Name of the existing block:	<i>Combining Diacritical Marks Extended</i>	
2. Number of characters in proposal:	<i>4</i>	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary <input checked="" type="checkbox"/>	B.1-Specialized (small collection) <input type="checkbox"/>	B.2-Specialized (large collection) <input type="checkbox"/>
C-Major extinct <input type="checkbox"/>	D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>	
4. Is a repertoire including character names provided?	<i>yes</i>	
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<i>yes</i>	
b. Are the character shapes attached in a legible form suitable for review?	<i>yes</i>	
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Kirk Miller</i>	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>SIL (Gentium Release)</i>	
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<i>yes</i>	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<i>yes</i>	
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)?	<i>yes</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 [www.unicode.org](http://www.unicode.org) for such information on other scripts. Also see Unicode Character Database ([www.unicode.org/reports/tr44/](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.

<sup>1</sup> 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>The authors are members of the user community.</i>  	<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:	  	
4. The context of use for the proposed characters (type of use; common or rare) Reference:	  	<i>phonetic</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	 <i>see illustrations</i>  	<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:	   	<i>no</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	 	<i>no</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:	   	<i>no</i>
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>