Unicode request for modifier voiceless implosive letters

Kirk Miller, kirkmiller@gmail.com   2024 June 15

This proposal, officially supported by the International Phonetic Association after evaluation by the IPA Alphbets, Charts and Fonts Committee (Nicolaides 2024), follows on L2/20-252 and L2/20-253, which requested modifier (superscript) variants of nearly all letters of the modern IPA alphabet that were as yet unsupported by Unicode. Those two proposals included a few retired IPA letters, but consideration of historical IPA modifiers was largely deferred until the IPA could more fully discuss the desired scope of such encoding through the newly established Alphabet, Charts and Fonts Committee. This proposal is a result of that consultation.

The Kiel Convention of 1989 introduced a series of letters for voiceless implosives, ⟨ƥ ƭ ƈ ƙ ʠ⟩, with retroflex ⟨ʈ⟩ implicit. Voiceless implosives are quite rare, however, and /ƙ/ is not known to be phonemic in any language (though it does occur allophonically in Mayan languages and paralinguistically in English), and in 1993 the letters were retired. Voiceless implosives are now usually transcribed into IPA with a devoicing ring: ⟨ɓ̥ ɗ̥ ᶑ̥ ʄ̊ ɠ̊ ʛ̥⟩. This solution is not entirely satisfactory, however, because the airstream mechanism differs: voiceless [ƥ ƭ ʈ ƈ ƙ ʠ] are purely glottalic ingressive sounds, whereas voicing requires [ɓ ɗ ᶑ ʄ ɠ ʛ] to be simultaneously glottalic ingressive and pulmonic egressive (Laver 1994: 173). [ɓ̥] is therefore not precisely the same sound as [ƥ]. Because of this, some linguists continue to use the dedicated voiceless letters ƥ ƭ ƈ ƙ ʠ. Indeed, the ICPLA is considering resurrecting ƥ ƭ ʈ ƈ ƙ ʠ for ExtIPA (Martin Ball, p.c. 2024), so modifier variants may prove useful there as well. Regardless, to provide full Unicode coverage of the IPA, modifier versions of historical symbols should be supported if those symbols are attested in the recent literature (Nicolaides 2024, cited in Figure 1). We therefore request the six modifier voiceless implosive letters, ⟨ᶂ ᶋ ᶘ ᶜ ᶄ ᶐ⟩, to parallel the existing voiced series ⟨ᶔ ᶑ ᶚ ᶡ ᶢ ᶃ⟩.

As for what these characters might be used for, there are possibilities for both phonetic modifier and non-modifier use. Figure 2 illustrates superscript non-modifier segments in the transcription of speech pathology. This convention indicates that the segments didn't meet their target articulation, and may be applied to any consonant or vowel, though such research involving languages with implosives is not yet attested. Modifier use is less obvious, but one of the letters of the doubly articulated implosive [ƙɓ] of Central Igbo might be superscripted rather than linked with a tie bar, producing ⟨ᶄɓ⟩ or ⟨ƙᵝ⟩, parallel to superscript conventions for other doubly articulated consonants (such as ⟨ᶢb⟩ for [g͡b] and ⟨ᵏp⟩ for [k͡p] – see Figure 3 and Figure 4), clicks (such as ⟨�q‡⟩ for [q͡ǂ]), affricates (such as ⟨tˢ⟩ for [t͡ʃ]), and diphthongs (such as ⟨oᵛ⟩ for [oʊ̯]).

Characters

Modifier voiceless implosive letters

ꞔ   1DFFA MODIFIER LETTER SMALL C WITH HOOK.

ƙ   1DFFB MODIFIER LETTER SMALL K WITH HOOK.

ƥ   1DFFC MODIFIER LETTER SMALL P WITH HOOK.

ʠ   1DFFD MODIFIER LETTER SMALL Q WITH HOOK.

ƭ   1DFFE MODIFIER LETTER SMALL T WITH HOOK.

ƭ   1DFFF MODIFIER LETTER SMALL T WITH HOOK AND RETROFLEX HOOK.

Properties

```
1DFFA;MODIFIER LETTER SMALL C WITH HOOK;Lm;0;L;<super> 0188;;;;N;;;;;
1DFFB;MODIFIER LETTER SMALL K WITH HOOK;Lm;0;L;<super> 0199;;;;N;;;;;
1DFFC;MODIFIER LETTER SMALL P WITH HOOK;Lm;0;L;<super> 01A5;;;;N;;;;;
1DFFD;MODIFIER LETTER SMALL Q WITH HOOK;Lm;0;L;<super> 02A0;;;;N;;;;;
1DFFE;MODIFIER LETTER SMALL T WITH HOOK;Lm;0;L;<super> 01AD;;;;N;;;;;
1DFFF;MODIFIER LETTER SMALL T WITH HOOK AND RETROFLEX HOOK;Lm;0;L;<super> 1DF09;;;;N;;;;;
```

DoNotEmit data

For historical reasons, the letter with retroflex hook is not canonically equivalent to the base letter plus a retroflex hook diacritic. It should thus be listed in DoNotEmit.txt.

```
1DFFE 0322; 1DFFF; Precomposed_Form # MODIFIER LETTER SMALL T WITH HOOK, COMBINING
     RETROFLEX HOOK BELOW;  MODIFIER LETTER SMALL T WITH HOOK AND RETROFLEX HOOK
```

References

Thomas & Elizabeth Brewster (1976) *Language acquisition made practical: Field methods for language learners.* Lingua House, Colorado Springs.

John Laver (1994) *Principles of Phonetics.* Cambridge University Press.

Katerina Nicolaides (2024) 'Unicode support for historical and para-IPA letters.' Letter submitted to the Unicode Technical Committee, 01 January 2024. L2/24-049.

Jill Perry (2000) Phonological/phonetic assessment of an English-speaking adult with dysarthria. Masters thesis, Memorial University of Newfoundland.

Kenneth Pike (1947) *Phonemics.* University of Michigan Press.

Chart

Now that the block for modifier phonetic letters, Latin Extended-F, is full, we propose placing any additional modifier Latin letters at the end of Latin Extended-G, to help keep them separate from baseline letters. Greyed out cells are assigned (medium grey) or proposed elsewhere (light grey).

Latin Extended-G

1DF00   1DFFF

	1DF0	1DF1	1DF2	1DF3	1DF4	1DF5	1DF6	1DF7	1DF8	1DF9	1DFA	1DFB	1DFC	1DFD	1DFE	1DFF
0	ffŋ	ʞ	ʣ	ɢ												
1	ȶ	ⱶ	ʥ	ɣ̷												
2	ɔ	ʣ̷	ƚ	ħ												
3	ⱦ	ʄ	ʧ	ɸ̷												
4	ⱡ	ŋ̷	tθ	q̷												
5	ꬸ	ɟ	ˀd	ʁ												
6	ꭕ	ſ	ɫ	ʁ̷												
7	ɑ	ʧ̷	ˀn	ʀ												
8	ɺ	ʒ̷	ˀr	ʦ												
9	ɭ	ʤ̷	ˀs	ʏ												
A	ǃ	ɨ	ˀt	ʃ												ɗ
B	ʃ	ʠ	ʤ̷													ƙ
C	ʄ	ʧ̷	ʧ̷													ƥ
D	ꝛ	ç	ɖ													ɗ
E	ꝝ	s̨	ʣ̷													ƭ
F	ɕ	ꬴ	ǥ̃													ƭ

3

Figures

Modifier versions of:

- ◆ the pre-Kiel click letters ⟨ʇ ʗ ʖ ʄ ʘ ⟩.
- ◆ the voiceless implosive letters ⟨ɓ̥ ƭ ʄ̊ ɗ̥ ƙ ɗ̥ ⟩.
- ◆ the Sinological extensions to the IPA ⟨ȶ ȡ ȵ ᶅ ᴀ ᴇ ω ㄗ ㄗ ㆠ ㆡ ⟩.
- ◆ the common palatal-hook letters ⟨ɖ ɳ ʂ ʐ⟩ (⟨ȶ ǀ⟩ are already encoded).

Superscript modifiers: Superscripting is a crucial dimension of IPA representation that requires hard encoding for preservation of the underlying data structure. In the opinion of the IPA Council, reported in Ashby (2020), modifier versions of "historical letter-like symbols should be supported if [the historical symbols] are attested in the recent literature." As was done in L2/20-252 and 253, accidental gaps in the literature need to be filled, as it would be detrimental to the practical use of the alphabet if Unicode were to formalize such gaps. In the unanimous opinion of the Alphabet, Charts and Fonts Committee, the set of letters listed in the "Consultation" section above meets these criteria.

Figure 1. Nicolaides (2024: 2–3) Relevant portions of L2/24-049, the formal request for these characters from the IPA.

Figure 2. Perry (2000: 91). IPA superscripts are not used only as modifiers of other letters. They are also used for epenthetic sounds and for weak or incomplete articulation, for example in transcriptions of speech acquisition and pathology. Here is an example of the superscript affricate ligature ⟨ᵈʒ⟩ for a pronunciation of [ʤ] that didn't meet its target. The transcription is ⟨ðəbæᵈʒəʷᵈʒʌmd⟩; the target phrase is *the badger jumped*. The literature is heavily centered on English, but the same convention could be applied to a language with voiceless implosives.

time — a **double stop.** It is voiceless, so it is a combination of p and k released simultaneously. It can be written [ᵏp]. As

Figure 3. Brewster & Brewster (1976: 275). ⟨ᵏp⟩ for [k͜p].

Exercise 1a. Pronounce [aᵏpa]. Make the closure at the velum come first, then release the lips and velum simultaneously.

1b. Repeat, releasing the lip closure first.

1c. Practice the sound in further sequences such as [ᵏpa], [aᵏp], [aᵏpa].

Figure 4. Pike (1947: 34). The use of a modifier for doubly articulated stops has a long history.

<div style="border:1px solid black">

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*.

</div>

A. Administrative

1. **Title:** *Modifier voiceless implosive letters*
2. Requester's name: *Kirk Miller*
3. Requester type (Member body/Liaison/Individual contribution): *individual*
4. Submission date: *2024 June 15*
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: *Latin Extended-G*
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 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?
      Kirk Miller
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
      SIL (Gentium Release)
6. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *no*
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources)
      of proposed characters attached? *no*
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 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
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? *The International Phonetic Organization* _ _ _ _ _ _

    If YES, available relevant documents: *(see letter of support)* _ _ _ _ _ _

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) *phonetic*

    Reference: _

5. Are the proposed characters in current use by the user community? *potentially*

    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: _